



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

Table of Contents

List of Tables	v
List of Figures.....	vi
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	18
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.....	30
a) Pearson's correlation coefficient and Spearman's rank-order correlation coefficient	30

b) Top-Down correlation coefficient	32
3.3 Journal similarity matrix	33
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 with hierarchical clustering.....	47
5. Programming code in Python	50
References.....	53
Appendix A.....	55
Appendix B.....	66
Appendix C.....	77

List of Tables

Table 1: Publication statistics of Chemistry research groups (2002-2008)	4
Table 2. Publication statistics of Chemistry panel members	4
Table 3. Euclidean distances between barycenter of Chemistry individual research groups, panel members, research groups together and panel using WoS SCs VOS map.....	17
Table 4. Euclidean distances between SAPVs of Chemistry individual groups, panel members, research groups together and panel in WoS SCs similarity matrix.	23
Table 5. WCS value of Chemistry individual research groups, panel members, research groups together and panel using WoS SCs similarity matrix.....	27
Table 6. WCD value between Chemistry 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 Chemistry individual research groups and panel members.....	28
Table 8. Euclidean distances between barycenter of Chemistry individual research groups, panel members, research groups together and panel using the journal VOS map.....	39
Table 9. Euclidean distances between SAPV of Chemistry individual research groups, panel members, research groups together and panel using the journal similarity matrix.....	44
Table 10. WCS value of the Chemistry groups, panel members, panel and research groups together using the journal similarity matrix.	46
Table 11. WCD value of the Chemistry groups, panel members, panel and research groups together using the journal similarity matrix.	46

List of Figures

Figure 1. Excerpt of Chemistry research groups and panel members_WoS SCs.xlsx file	5
Figure 2. Excerpt of the Chemistry panel and research groups together_WoS SCs.xlsx file	6
Figure 3. Excerpt of the Chemistry 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 Chemistry 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. CHEM-B research group's publication overlay map in WoS SCs	13
Figure 12. Barycenter coordinates of the Chemistry 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 Chemistry individual research groups, panel members, research groups together and panel using WoS SCs VOS map	17
Figure 14. Barycenter overlay map of Chemistry individual research groups, panel members (PM), research groups together and panel in WoS SCs.....	18
Figure 15. Barycenter map of Chemistry individual research groups, panel members (PM), research groups together and panel in WoS SCs (zoomed).....	19
Figure 16. Barycenter overlay map of Chemistry 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 Chemistry 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 Chemistry individual research groups, panel members, research groups together and panel together using WoS SCs similarity matrix.....	22
Figure 20. Excerpt of CHEM-B.csv file	24
Figure 21. Location of the SAPV of CHEM-B in the WoS SCs similarity matrix.....	25

Figure 22. Excerpt of WCS value matrix of the Chemistry individual research groups, panel members, research groups together and panel using WoS SCs similarity matrix	26
Figure 23. Scatter plots of the of correlation between barycenter, SAPV and WCS methods in WoS SCs in Chemistry department.....	28
Figure 24. Excerpt of Chemistry research groups and panel_journal title.xlsx file	29
Figure 25. Excerpt of the Chemistry panel and research groups together_journals title.xlsx file	30
Figure 26. Excerpt of the Chemistry 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 Chemistry department	31
Figure 28. Excerpt of the journal VOS map data.....	34
Figure 29. Journal overlay map of the CHEM-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 Chemistry 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 Chemistry groups, panel members, research groups together and panel using the journal VOS map	38
Figure 34. Barycenter overlay map of Chemistry panel, panel members (PM), research groups and research groups together.....	40
Figure 35. Barycenter overlay map of Chemistry panel, panel members (PM), research groups and research groups together (zoomed)	40
Figure 36. Barycenter overlay map of Chemistry panel, panel members (PM), research groups and research groups together with their confidence regions	41
Figure 37. Excerpt of SAPV of the Chemistry 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 Chemistry 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 Chemistry 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	48

Figure 41. Heat map with hierarchical clustering based on correlation coefficient between six approaches in the Chemistry department	49
Figure 42. WoS SCs overlay map of CHEM-A research group's publications	55
Figure 43. WoS SCs overlay map of CHEM-B research group's publications.....	55
Figure 44. WoS SCs overlay map of CHEM-C research group's publications.....	56
Figure 45. WoS SCs overlay map of CHEM-D research group's publications	56
Figure 46. WoS SCs overlay map of CHEM-E research group's publications.....	57
Figure 47. WoS SCs overlay map of CHEM-F research group's publications	57
Figure 48. WoS SCs overlay map of CHEM-G research group's publications	58
Figure 49. WoS SCs overlay map of CHEM-H research group's publications	58
Figure 50. WoS SCs overlay map of CHEM-I research group's publications	59
Figure 51. WoS SCs overlay map of CHEM-J research group's publications.....	59
Figure 52. WoS SCs overlay map of CHEM-K research group's publications	60
Figure 53. WoS SCs overlay map of CHEM-L research group's publications.....	60
Figure 54. WoS SCs overlay map of Chemistry groups' publications	61
Figure 55. WoS SCs overlay map of PM1's publications	61
Figure 56. WoS SCs overlay map of PM2's publications	62
Figure 57. WoS SCs overlay map of PM3's publications	62
Figure 58. WoS SCs overlay map of PM4's publications	63
Figure 59. WoS SCs overlay map of PM5's publications	63
Figure 60. WoS SCs overlay map of PM6's publications	64
Figure 61. WoS SCs overlay map of PM7's publications	64
Figure 62. WoS SCs overlay map of panel's publications	65
Figure 63. SAPV of the CHEM-A research group's publications in WoS SCs similarity matrix	66
Figure 64. SAPV of the CHEM-B research group's publications in WoS SCs similarity matrix.....	66
Figure 65. SAPV of the CHEM-C research group's publications in WoS SCs similarity matrix.....	67
Figure 66. SAPV of the CHEM-D research group's publications in WoS SCs similarity matrix	67

Figure 67. SAPV of the CHEM-E research group's publications in WoS SCs similarity matrix	68
Figure 68. SAPV of the CHEM-F research group's publications in WoS SCs similarity matrix	68
Figure 69. SAPV of the CHEM-G research group's publications in WoS SCs similarity matrix	69
Figure 70. SAPV of the CHEM-H research group's publications in WoS SCs similarity matrix	69
Figure 71. SAPV of the CHEM-I research group's publications in WoS SCs similarity matrix	70
Figure 72. SAPV of the CHEM-J research group's publications in WoS SCs similarity matrix	70
Figure 73. SAPV of the CHEM-K research group's publications in WoS SCs similarity matrix	71
Figure 74. SAPV of the CHEM-L research group's publications in WoS SCs similarity matrix	71
Figure 75. SAPV of the Chemistry research group's publications in WoS SCs similarity matrix	72
Figure 76. SAPV of the PM1's publications in WoS SCs similarity matrix	72
Figure 77. SAPV of the PM2's publications in WoS SCs similarity matrix	73
Figure 78. SAPV of the PM3's publications in WoS SCs similarity matrix	73
Figure 79. SAPV of the PM4's publications in WoS SCs similarity matrix	74
Figure 80. SAPV of the PM5's publications in WoS SCs similarity matrix	74
Figure 81. SAPV of the PM6's publications in WoS SCs similarity matrix	75
Figure 82. SAPV of the PM7's publications in WoS SCs similarity matrix	75
Figure 83. SAPV of the panel publications in WoS SCs similarity matrix	76
Figure 84. Journal overlay map of CHEM-A research group's publications	77
Figure 85. Journal overlay map of CHEM-B research group's publications	78
Figure 86. Journal overlay map of CHEM-C research group's publications	78
Figure 87. Journal overlay map of CHEM-D research group's publications	79
Figure 88. Journal overlay map of CHEM-E research group's publications	79
Figure 89. Journal overlay map of CHEM-F research group's publications	80
Figure 90. Journal overlay map of CHEM-G research group's publications	80
Figure 91. Journal overlay map of CHEM-H research group's publications	81
Figure 92. Journal overlay map of CHEM-I research group's publications	81

Figure 93. Journal overlay map of CHEM-J research group's publications.....	82
Figure 94. Journal overlay map of CHEM-K research group's publications	82
Figure 95. Journal overlay map of CHEM-L research group's publications.....	83
Figure 96. Journal overlay map of Chemistry research groups' publications.....	83
Figure 97. Journal overlay map of PM1's publications	84
Figure 98. Journal overlay map of PM2's publications	84
Figure 99. Journal overlay map of PM3's publications	85
Figure 100. Journal overlay map of PM4's publications	85
Figure 101. Journal overlay map of PM5's publications	86
Figure 102. Journal overlay map of PM6's publications	86
Figure 103. Journal overlay map of PM7's publications	87
Figure 104. Journal overlay map of the panel's publications	87

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., Chemistry) are to be evaluated by the same panel and the language of communication is English. Following the Dutch Standard Evaluation Protocol

(VSNU, 2003; VSNU, KNAW, & NWO, 2014), the peer panels assess the quality, the productivity, the relevance and the viability of each research group.

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

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

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

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

panel members and panel (all panel members together) in WoS SCs and journals in the appendix.

2 Cognitive distance based on Web of Science subject categories

2.1 Data collection process

We collect data from the 2009 assessment of the 12 research groups of the Department of Chemistry, 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 CHEM-A, CHEM-B etc.

a) Research groups data retrieval

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

Table 1 lists the publication profile of the Chemistry research groups during the eight years preceding their evaluation. The Chemistry research groups generated 920 publications in 300 journals. Members of two research groups co-authored 43 publications. In total, their publications are distributed over 94 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’.

Table 1: Publication statistics of Chemistry research groups (2002-2008)

Group code	Number of Publications	Number of Journals	Number of WoS SCs
CHEM-A	129	47	27
CHEM-B	65	24	17
CHEM-C	156	52	26
CHEM-D	32	17	13
CHEM-E	70	39	23
CHEM-F	21	17	8
CHEM-G	161	47	42
CHEM-H	62	33	28
CHEM-I	51	24	19
CHEM-J	27	11	15
CHEM-K	97	66	48
CHEM-L	92	42	24
All groups	920	300	94

b) Panel members data retrieval

The Chemistry panel was composed of 7 panel members (including the chair). We have obtained the names and curricula vitae of the panel members from the ADOC. We replace the original name of each panel member with a code name: PM1, PM2 etc. We perform an advanced search for each panel member in WoS through checking the SCIE and SSCI. All the publications of the individual panel members up to the year of assessment (2009) were taken into account. We use the ‘Analyze Results’ option in the WoS, and rank the record by WoS SCs with the minimum set to 1. We save the resulting list as ‘analyze.txt’ and subsequently save a copy of the file named ‘[PM code]_WoS SCs.txt’ for example, ‘PM1_WoS SCs.txt’.

Table 2. Publication statistics of Chemistry panel members

Panel code	Number of Publications	Number of Journals	Number of WoS SCs
PM1	694	72	11
PM2	221	60	25
PM3	152	42	16
PM4	254	55	34
PM5	206	58	34
PM6	113	32	14
PM7	512	69	12
Panel	2150	248	66

Table 2 lists the publication profile of the Chemistry panel members. The combined publication output of the Chemistry panel members consists of 2150 publications two of

which are co-authored publications between two panel members. The number of publications per panel member ranges from 113 to 694. In total, these publications appeared in 248 different journals and are assigned to 66 different WoS SCs.

	A	B	C
1	Web of Science Categories	records	% of 920
2	CHEMISTRY PHYSICAL	198	21.522
3	CHEMISTRY ANALYTICAL	194	21.087
4	SPECTROSCOPY	164	17.826
5	PHYSICS ATOMIC MOLECULAR CHEMICAL	100	10.87
6	PHYSICS APPLIED	77	8.37
7	MATERIALS SCIENCE MULTIDISCIPLINARY	75	8.152
8	ENVIRONMENTAL SCIENCES	59	6.413
9	BIOCHEMICAL RESEARCH METHODS	43	4.674
10	PHYSICS CONDENSED MATTER	43	4.674
11	CHEMISTRY MULTIDISCIPLINARY	39	4.239
12	CHEMISTRY ORGANIC	36	3.913
13	CHEMISTRY APPLIED	29	3.152
14	BIOCHEMISTRY MOLECULAR BIOLOGY	28	3.043
15	INSTRUMENTS INSTRUMENTATION	27	2.935

GroupsTogether CHEM-A CHEM-B CHEM-C CHEM-D CHEM-E CHEM-F CHEM-G CHEM-H

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

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’.

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 (CHEM-A, CHEM-B, CHEM-C, etc.), the panel members’ code names, (PM1, PM2, PM3, etc.), Panel together and Groups together. The Excel file is saved as ‘Chemistry 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 'Chemistry panel and research groups together_WoS SCs.xlsx' (Figure 2) and export data from 'Panel_WoS SCs.txt' and 'Groups together_WoS SCs.txt' in two different sheets.

	A	B	C
1	Web of Science Categories	records	% of 920
2	CHEMISTRY PHYSICAL	198	21.522
3	CHEMISTRY ANALYTICAL	194	21.087
4	SPECTROSCOPY	164	17.826
5	PHYSICS ATOMIC MOLECULAR CHEMICAL	100	10.87
6	PHYSICS APPLIED	77	8.37
7	MATERIALS SCIENCE MULTIDISCIPLINARY	75	8.152
8	ENVIRONMENTAL SCIENCES	59	6.413
9	BIOCHEMICAL RESEARCH METHODS	43	4.674
10	PHYSICS CONDENSED MATTER	43	4.674
11	CHEMISTRY MULTIDISCIPLINARY	39	4.239
12	CHEMISTRY ORGANIC	36	3.913
13	CHEMISTRY APPLIED	29	3.152
14	BIOCHEMISTRY MOLECULAR BIOLOGY	28	3.043
15	INSTRUMENTS INSTRUMENTATION	27	2.935
16	PHYSICS FLUIDS PLASMAS	27	2.935
17	COMPUTER SCIENCE ARTIFICIAL INTELLIG	25	2.717
18	MICROSCOPY	25	2.717
19	METEOROLOGY ATMOSPHERIC SCIENCES	23	2.5
20	NANOSCIENCE NANOTECHNOLOGY	20	2.174
21	MATERIALS SCIENCE COATINGS FILMS	19	2.065
22	CRYSTALLOGRAPHY	18	1.957
23	NUCLEAR SCIENCE TECHNOLOGY	18	1.957
24	CHEMISTRY INORGANIC NUCLEAR	17	1.848
25	OPTICS	17	1.848
26	PHYSICS MATHEMATICAL	17	1.848
27	RADIOLOGY NUCLEAR MEDICINE MEDICA	17	1.848
28	ELECTROCHEMISTRY	16	1.739
29	ENGINEERING ELECTRICAL ELECTRONIC	15	1.63
30	ENGINEERING ENVIRONMENTAL	15	1.63
31	COMPUTER SCIENCE THEORY METHODS	14	1.522
32	ENGINEERING CHEMICAL	13	1.413

	A	B	C
1	Web of Science Categories	records	% of 2150
2	CHEMISTRY INORGANIC NUCLEAR	798	37.116
3	CHEMISTRY ORGANIC	458	21.302
4	CHEMISTRY ANALYTICAL	350	16.279
5	CHEMISTRY MULTIDISCIPLINARY	324	15.07
6	CHEMISTRY PHYSICAL	177	8.233
7	PHYSICS ATOMIC MOLECULAR CHEMICAL	115	5.349
8	ENVIRONMENTAL SCIENCES	93	4.326
9	SPECTROSCOPY	63	2.93
10	BIOCHEMICAL RESEARCH METHODS	57	2.651
11	ENGINEERING CHEMICAL	41	1.907
12	GEOCHEMISTRY GEOPHYSICS	39	1.814
13	ENGINEERING ENVIRONMENTAL	36	1.674
14	CRYSTALLOGRAPHY	33	1.535
15	METEOROLOGY ATMOSPHERIC SCIENCES	33	1.535
16	NEUROSCIENCES	32	1.488
17	CHEMISTRY APPLIED	31	1.442
18	BIOCHEMISTRY MOLECULAR BIOLOGY	25	1.163
19	ELECTROCHEMISTRY	24	1.116
20	PHYSICS APPLIED	23	1.07
21	MATERIALS SCIENCE MULTIDISCIPLINARY	20	0.93
22	MINERALOGY	18	0.837
23	MULTIDISCIPLINARY SCIENCES	14	0.651
24	MATERIALS SCIENCE COATINGS FILMS	13	0.605
25	WATER RESOURCES	13	0.605
26	INSTRUMENTS INSTRUMENTATION	10	0.465
27	PHYSICS CONDENSED MATTER	9	0.419
28	CHEMISTRY MEDICINAL	7	0.326
29	IMMUNOLOGY	7	0.326
30	BIOTECHNOLOGY APPLIED MICROBIOLOG	6	0.279
31	FOOD SCIENCE TECHNOLOGY	6	0.279
32	GEOLOGY	6	0.279

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

This produces a new Excel file called 'Chemistry panel and groups together_WoS SCs-joined.xlsx' (Figure 3). To calculate the correlation, the value zero was kept on the corresponding WoS SCs in which either the panel or the groups had no publications (but not both). Using the data from the file, we calculate correlation coefficient in SPSS (Statistical Package for the Social Sciences) and find value ($r = 0.39$, $p = 0.37$).

	A	B	C	D
1		Web of Science Categories	records_x	records_y
2	0	CHEMISTRY PHYSICAL	198	177
3	1	CHEMISTRY ANALYTICAL	194	350
4	2	SPECTROSCOPY	164	63
5	3	PHYSICS ATOMIC MOLECULAR CHEMICAL	100	115
6	4	PHYSICS APPLIED	77	23
7	5	MATERIALS SCIENCE MULTIDISCIPLINARY	75	20
8	6	ENVIRONMENTAL SCIENCES	59	93
9	7	BIOCHEMICAL RESEARCH METHODS	43	57
10	8	PHYSICS CONDENSED MATTER	43	9
11	9	CHEMISTRY MULTIDISCIPLINARY	39	324
12	10	CHEMISTRY ORGANIC	36	458
13	11	CHEMISTRY APPLIED	29	31
14	12	BIOCHEMISTRY MOLECULAR BIOLOGY	28	25
15	13	INSTRUMENTS INSTRUMENTATION	27	10
16	14	PHYSICS FLUIDS PLASMAS	27	0

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

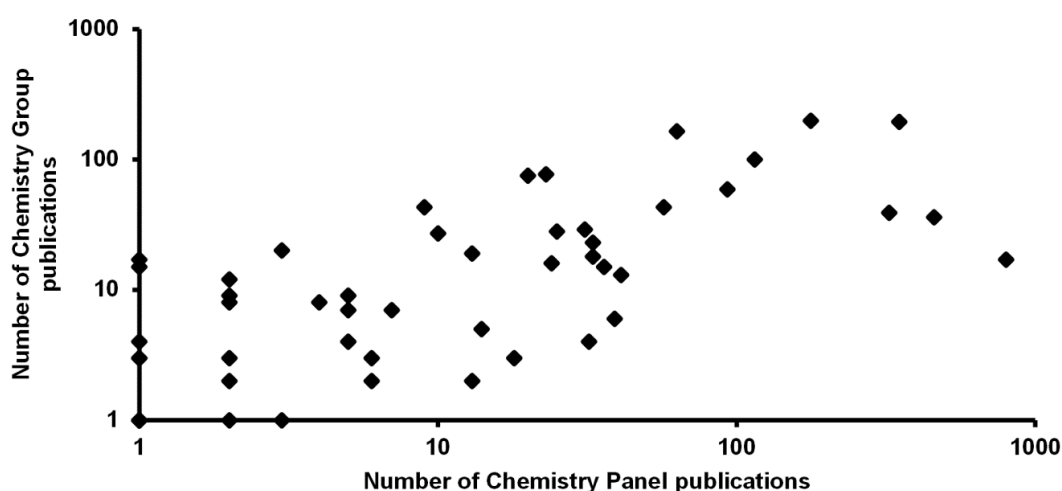


Figure 4. Log-log plot of the number of publications (log-log scale) per WoS SC for the panel (vertical axis) and research groups together (horizontal axis) of the Chemistry department

Figure 4 shows a log-log plot of the number of publications per WoS SCs for the Chemistry panel and research groups together.

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

The outcome shows that the top-down correlation between Chemistry research groups together and the panel's profile in the WoS SCs is moderate (0.54).

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

2.3 Web of Science subject categories similarity matrix

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

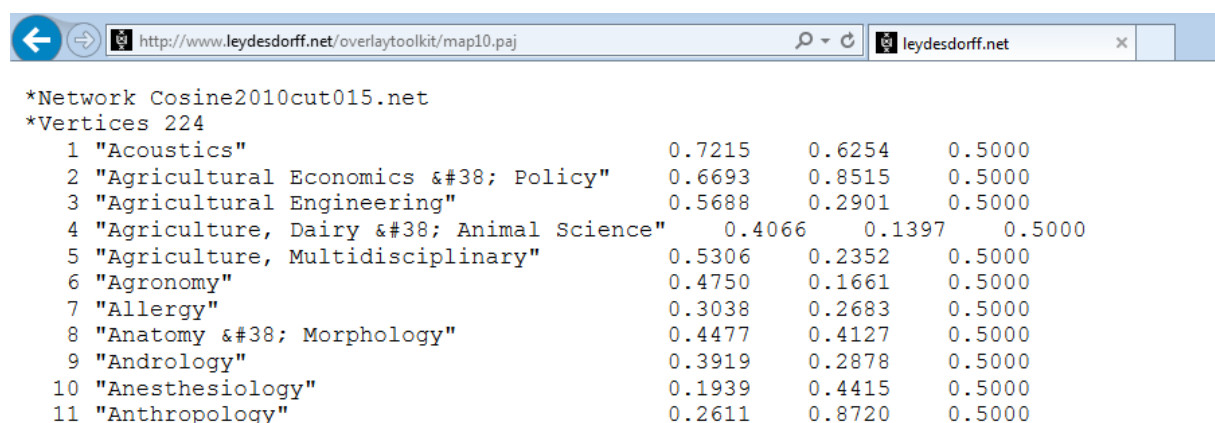


Figure 5. Excerpt of the map10.paj file

We download the 'map10.paj' (Figure 5) file and open the file in Pajek (available at <http://mrvar.fdv.uni-lj.si/pajek>) and save the network as 'map10.net'. The information in the network file can be visualized. The subfield of bibliometric mapping is dedicated to the

visualization, clustering and interpretation of similarity matrices or networks like the one we use. Many different algorithms or layout techniques have been developed for this purpose. We have used two layout techniques:

- i) Kamada-Kawai (Kamada & Kawai, 1989) is a spring-based layout algorithm for networks, which is implemented in, among others, Pajek (de Nooy, Mrvar, & Batagelj, 2012). Kamada-Kawai is the algorithm used by Rafols et al., (2010).
- ii) VOS (van Eck & Waltman, 2007) stands for ‘visualization of similarities’ and is a variant of multidimensional scaling (Borg & Groenen, 2005; van Eck, Waltman, Dekker, & van den Berg, 2010). It is implemented in VOSviewer and in recent versions of Pajek.

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

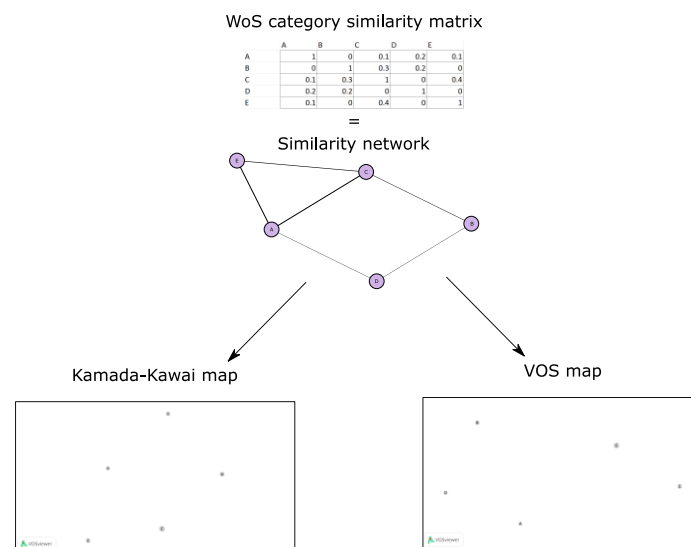


Figure 6. Transformation of WoS SCs similarity matrix to Kamada-Kawai map and VOS map

We run VOSviewer (<http://www.vosviewer.com>) and click on ‘Create’ from the action tab. It offers to create a map based on a network. We select this option and in the next step through

Pajek tab, we choose the ‘map10.net’ file and click on the next button. It prompts us to choose whether we want to use the coordinates that are in the file or want to calculate new ones (Figure 7).

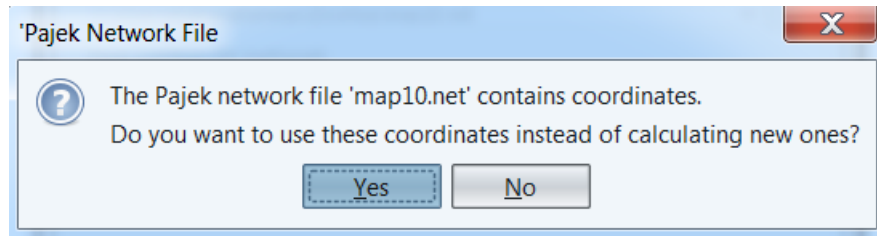


Figure 7. VOSviewer message before choosing Kamada Kawai map or VOS map data

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

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

2.4 Web of Science subject categories overlay map creation

During data collection (see section 2.1, the resulting files are downloaded using the default name ‘analyze.txt’.

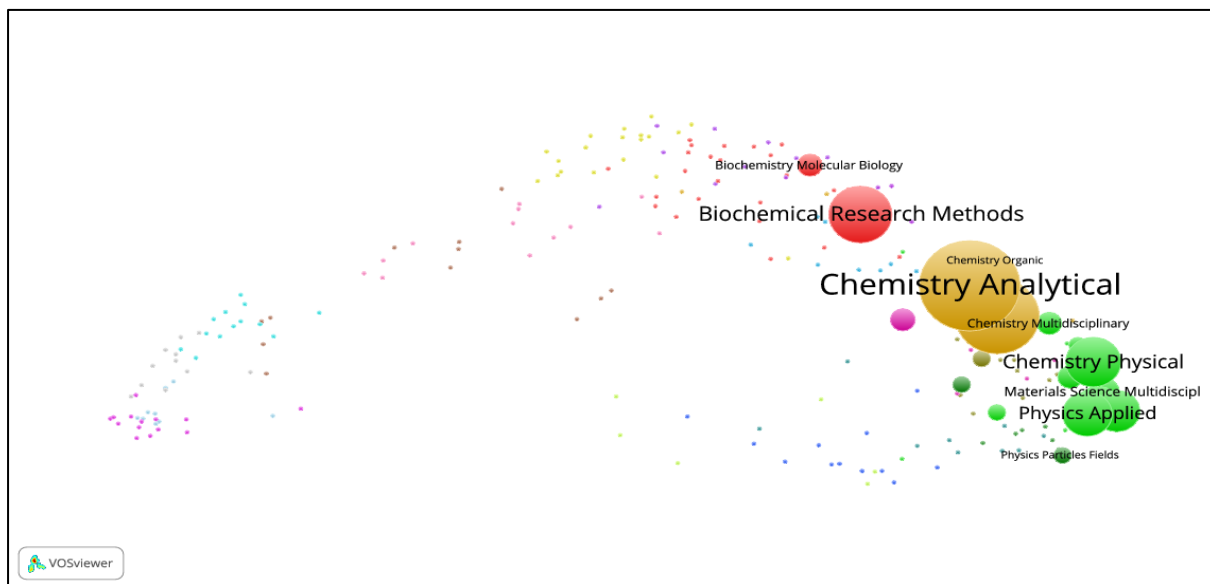


Figure 11. CHEM-B research group's publication overlay map in WoS SCs

We download the ‘WC10.exe’ program from <http://www.leydesdorff.net/overlay> toolkit. This file ‘analyze.txt’ transformed by the mini-program ‘WC10.exe’ to ‘WC10.vec’ for upload into Pajek as a vector, and generate files like ‘vos4.csv’, ‘vos6.csv’, and ‘vos19.csv’ for use in VOSviewer (with 4, 6 or 19 base colors for the clusters, respectively). We keep ‘analyze.txt’ and ‘WC10.exe’ in a folder and run the exe file. The program ‘WC10.exe’ generates three map files: ‘vos4.csv’, ‘vos6.csv’, and ‘vos19.csv’. We open the ‘vos19.csv’ in VOSviewer. For example, Figure 11 shows CHEM-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).

2.5 Bootstrapping and confidence intervals

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

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

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

By generating a large amount of independent bootstrap samples (in our case 1000) and each time calculating the bootstrap replication, we can approximate the variability within the data set. Since we have a two-sample problem (distance between two entities; Efron & Tibshirani, 1998, Ch. 8), we calculate the distances between pairs of bootstrap replications, from which we obtain a CI using a bootstrap percentile approach (Efron & Tibshirani, 1998, Ch. 13). In

the case of WCS, we generate 1000 independent bootstrap sample for both entities and calculate the similarity between them using formula 7. A more detailed explanation and implementation of our method is available on Github (<http://nbviewer.jupyter.org/gist/rafguns/6fa3460677741e356538337003692389> and <http://nbviewer.jupyter.org/gist/rafguns/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 (‘Chemistry 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" "Chemistry research groups
and panel_WoS SCs.xlsx"
```

This program calculates the barycenter and generates an output file ‘Chemistry research groups and panel_WoS SCs-barycenter.xlsx’. Figure 12 shows the barycenter coordinates of

the Chemistry individual research groups, panel members, research groups together and panel.

	A	B	C
1		x	y
2	Groups	0.771959	-0.08276
3	CHEM-A	0.872563	-0.14536
4	CHEM-B	0.897422	-0.10331
5	CHEM-C	0.873185	-0.18104
6	CHEM-D	0.778243	0.086195
7	CHEM-E	0.600217	0.127043
8	CHEM-F	0.548838	0.183573
9	CHEM-G	0.740755	-0.04122
10	CHEM-H	0.68064	-0.04813
11	CHEM-I	0.794036	-0.02874
12	CHEM-J	0.619974	-0.13582
13	CHEM-K	0.513167	-0.12709
14	CHEM-L	0.944611	-0.09699
15	Panel	0.845875	0.003736
16	PM 1	0.913505	0.025651
17	PM 2	0.644473	0.057483
18	PM 3	0.791262	-0.02594
19	PM 4	0.701055	0.007732
20	PM 5	0.875586	-0.09763
21	PM 6	0.771366	0.126061
22	PM 7	0.966086	-0.02981

Figure 12. Barycenter coordinates of the Chemistry 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).

	A	B	C	D	E	F	G	H	I	J	K	L
1		CHEM-A	CHEM-B	CHEM-C	CHEM-D	CHEM-E	CHEM-F	CHEM-G	CHEM-H	CHEM-I	CHEM-J	CHEM-K
2	CHEM-A	0	0.043769	0.056809	0.287558	0.423313	0.43367	0.207821	0.141852	0.155409	0.345482	0.453434
3	CHEM-B	0.043769	0	0.100354	0.264635	0.4108	0.42062	0.205129	0.147654	0.142317	0.368478	0.466345
4	CHEM-C	0.056809	0.100354	0	0.317934	0.438428	0.449317	0.217216	0.146588	0.180864	0.315128	0.43511
5	CHEM-D	0.287558	0.264635	0.317934	0	0.166575	0.173575	0.135166	0.188168	0.137269	0.339183	0.346915
6	CHEM-E	0.423313	0.4108	0.438428	0.166575	0	0.011459	0.221482	0.292147	0.269169	0.32144	0.257992
7	CHEM-F	0.43367	0.42062	0.449317	0.173575	0.011459	0	0.232498	0.303132	0.279245	0.331646	0.264929
8	CHEM-G	0.207821	0.205129	0.217216	0.135166	0.221482	0.232498	0	0.070696	0.071851	0.223055	0.277072
9	CHEM-H	0.141852	0.147654	0.146588	0.188168	0.292147	0.303132	0.070696	0	0.063727	0.233193	0.319116
10	CHEM-I	0.155409	0.142317	0.180864	0.137269	0.269169	0.279245	0.071851	0.063727	0	0.283409	0.348602
11	CHEM-J	0.345482	0.368478	0.315128	0.339183	0.32144	0.331646	0.223055	0.233193	0.283409	0	0.145338
12	CHEM-K	0.453434	0.466345	0.43511	0.346915	0.257992	0.264929	0.277072	0.319116	0.348602	0.145338	0
13	CHEM-L	0.04666	0.00298	0.103288	0.263974	0.410851	0.420628	0.206022	0.149194	0.142625	0.370736	0.468028
14	Groups	0.156386	0.160406	0.161635	0.177341	0.277292	0.288308	0.055811	0.015118	0.061126	0.226996	0.307563
15	PM1	0.166941	0.128518	0.217009	0.165164	0.329288	0.337214	0.178594	0.165302	0.110572	0.393724	0.454459
16	PM2	0.350347	0.341645	0.361958	0.128654	0.078919	0.090333	0.144792	0.215372	0.199387	0.258753	0.227823
17	PM3	0.170966	0.160543	0.19217	0.129071	0.252453	0.262739	0.052661	0.061288	0.019729	0.268627	0.32971
18	PM4	0.268886	0.262207	0.280288	0.107538	0.158483	0.16956	0.063088	0.133782	0.121104	0.231846	0.250013
19	PM5	0.05634	0.055078	0.09091	0.232441	0.367237	0.377531	0.154113	0.092972	0.099074	0.315064	0.4113
20	PM6	0.301596	0.275668	0.335204	0.026506	0.175046	0.180653	0.160555	0.210406	0.156006	0.365551	0.369998
21	PM7	0.11609	0.072386	0.172296	0.234953	0.394896	0.403552	0.215856	0.177635	0.144285	0.410159	0.490999
22	Panel	0.16587	0.141322	0.202439	0.123329	0.275226	0.284316	0.107905	0.106532	0.044115	0.326106	0.38367

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

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

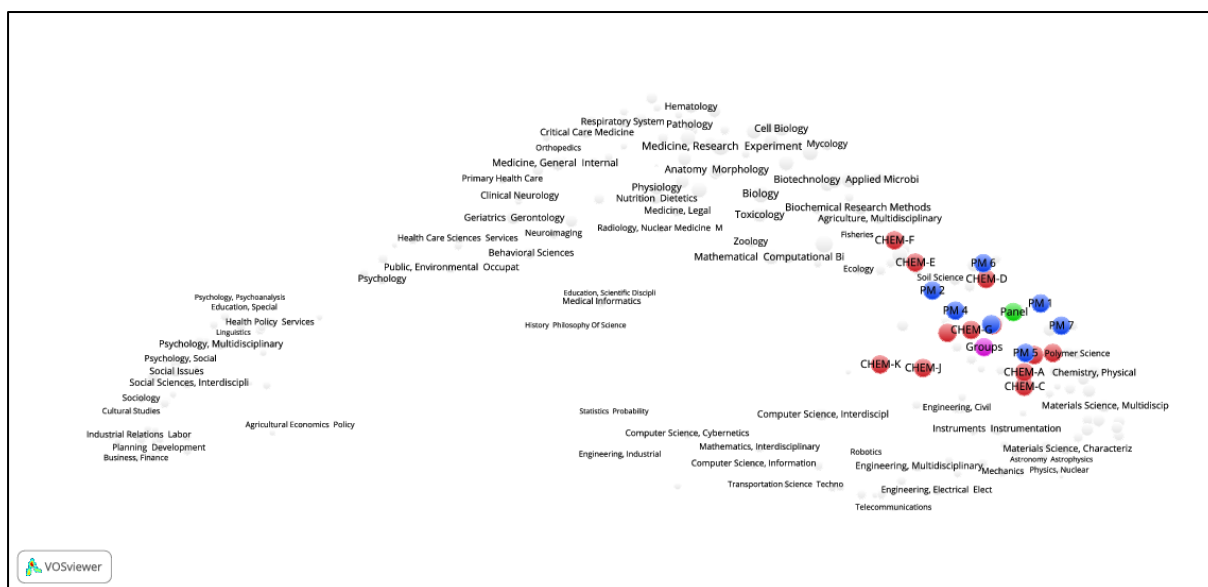
	Group	CHEM-A	CHEM-B	CHEM-C	CHEM-D	CHEM-E	CHEM-F	CHEM-G	CHEM-H	CHEM-I	CHEM-J	CHEM-K	CHEM-L
Panel	0.105	0.166	0.141	0.202	0.123	0.275	0.284	0.108	0.107	0.044	0.326	0.384	0.141
PM 1	0.168	0.167	0.129	0.217	0.165	0.329	0.337	0.179	0.165	0.111	0.394	0.454	0.127
PM 2	0.200	0.350	0.342	0.362	0.129	<u>0.079</u>	<u>0.090</u>	0.145	0.215	0.199	0.259	<u>0.228</u>	0.342
PM 3	0.054	0.171	0.161	0.192	0.129	0.252	0.263	<u>0.053</u>	<u>0.061</u>	<u>0.020</u>	0.269	0.330	0.161
PM 4	0.119	0.269	0.262	0.280	0.108	0.158	0.170	0.063	0.134	0.121	<u>0.232</u>	0.250	0.263
PM 5	0.106	<u>0.056</u>	<u>0.055</u>	<u>0.091</u>	0.232	0.367	0.378	0.154	0.093	0.099	0.315	0.411	<u>0.057</u>
PM 6	0.200	0.302	0.276	0.335	<u>0.027</u>	0.175	0.181	0.161	0.210	0.156	0.366	0.370	0.275
PM 7	0.186	0.116	0.072	0.172	0.235	0.395	0.404	0.216	0.178	0.144	0.410	0.491	0.070

For each research group we determined the panel member at the shortest distance. Average of shortest distance is 0.087 (SD 0.070). The number in the row of this panel member is indicated in bold and underlined. Distances whose confidence intervals overlap with that of the shortest distance are in bold (same column).

From the matrix of Euclidean distances, 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.

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

c) Barycenter overlay map



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 Chemistry 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 Chemistry 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).

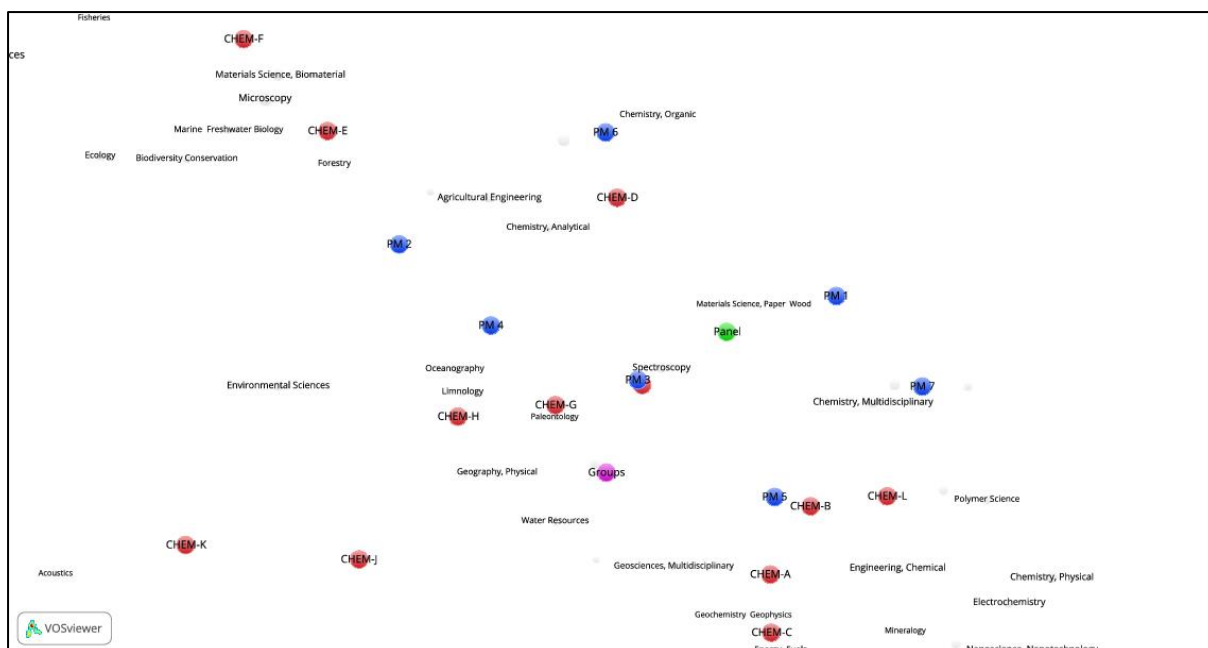


Figure 15. Barycenter map of Chemistry individual research groups, panel members (PM), research groups together and panel in WoS SCs (zoomed)

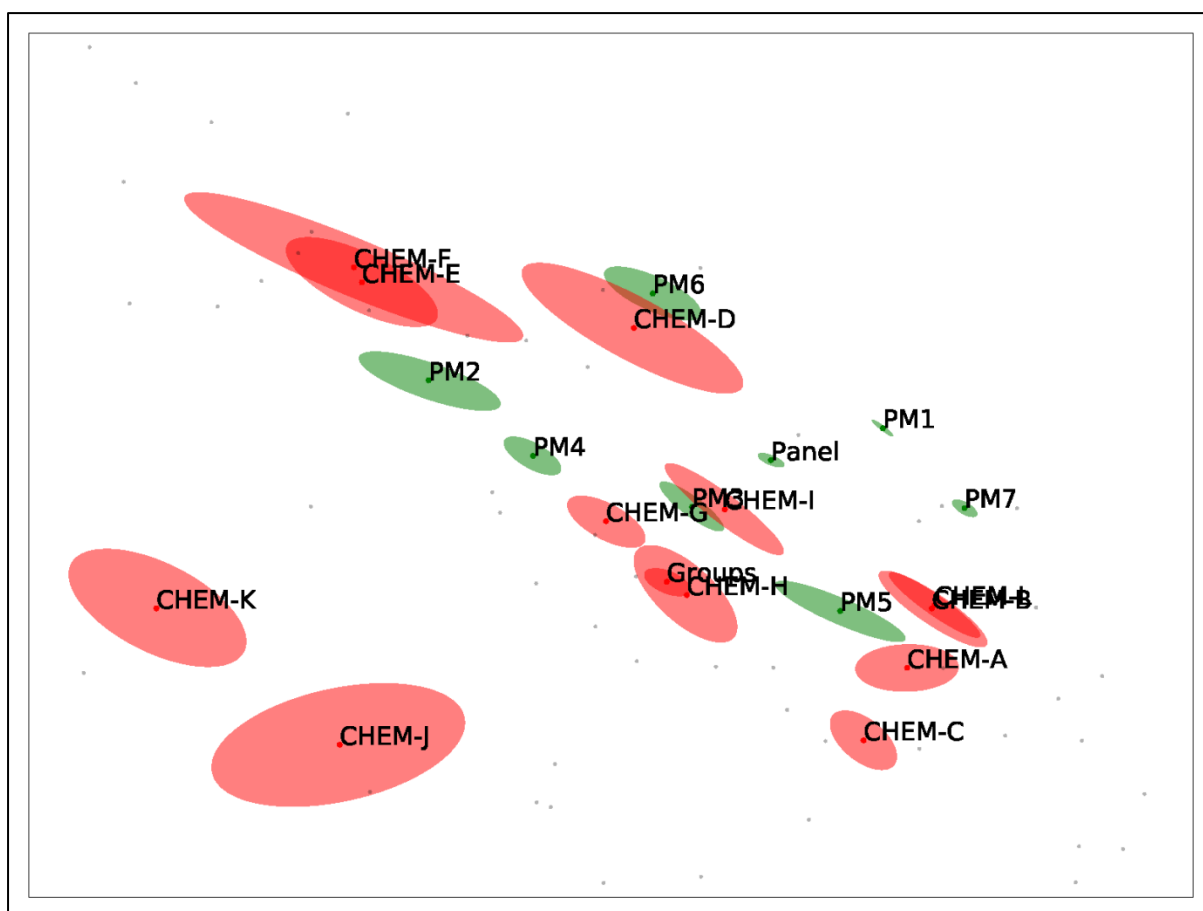


Figure 16. Barycenter overlay map of Chemistry panel, panel members (PM), research groups and research groups together (groups) with their confidence regions

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

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

2.7 Similarity-adapted publication vector method

a) Similarity-adapted publication vector calculation

A similarity-adapted publication vector (SAPV) is determined as the vector $C = (C_1, C_2, \dots, C_N)$, 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 L_1 -norm of the unnormalized vector.

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

Figure 17. Excerpt of WoS SCs similarity matrix

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

A python script ‘sa-vector-categories.py’ is used that takes as input the WoS SCs similarity matrix (Figure 17) and the number of publications of Chemistry 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).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1		Acoustics	al Econom	tural Engir	Dairy Anire, Multidi	Agronomy	Allergy	my Morpl	Andrology	iesthesiolon	thropolog	area Studie	omy Astro	
2	CHEM-A	0.000614	0.000495	0.004711	0.000147	0.001039	0.000227	0.00011	0.000593	0.000285	0	0	0	0
3	CHEM-B	0	0	0.006523	0.00018	0.002771	0.000705	0.000182	0.001831	0.000788	0.000134	0	0	0.000397
4	CHEM-C	0.001462	4.82E-05	0.00418	0.000227	0.003793	0.000278	0	0.002056	0.000535	0	0	0	0.001665
5	CHEM-D	0	0	0.007684	0.000484	0.009978	0.003325	0.000728	0.004744	0.00295	0.000375	0	0	0
6	CHEM-E	8.02E-05	6.18E-05	0.011349	0.002914	0.010884	0.005064	0.001898	0.011936	0.007022	0.000342	0	0	0
7	CHEM-F	0	0	0.011966	0.002311	0.01308	0.002761	0.001769	0.008934	0.005959	0.001998	0	0	0
8	CHEM-G	0.000457	0	0.015285	0.000396	0.010396	0.004335	0.000412	0.003047	0.000945	8.17E-05	0	0	9.86E-05
9	CHEM-H	0.001723	0	0.009486	0.001197	0.009406	0.00242	0.000115	0.00427	0.001108	0.000189	0	0	0.00211
10	CHEM-I	0.000167	0	0.009361	0.001145	0.008612	0.001582	0.000144	0.004854	0.002179	0	0	0	0.000332
11	CHEM-J	0.004849	0.00247	0.0064	0	0.005689	0.000562	0	0.001414	0	0	0	0	0.001293
12	CHEM-K	0.003997	0	0.002613	0.000427	0.002318	0.000555	0.000298	0.011009	0.004966	0.002907	0	0	0.000424
13	CHEM-L	0	0	0.008702	0.002063	0.00601	0.00209	0.000264	0.002479	0.001046	4.04E-05	0	0	0
14	Groups together	0.001038	0.000138	0.007952	0.000925	0.006105	0.00194	0.000412	0.004446	0.002027	0.000442	0	0	0.000515
15	PM1	0	0	0.003231	0.000267	0.004568	0.000294	3.54E-05	0.000368	0.000167	0	0	0	0
16	PM2	0.000229	0	0.009033	0.001589	0.007846	0.001708	0.000935	0.008756	0.004562	0.002529	0	0	0
17	PM3	0	0	0.009307	0.000897	0.009526	0.001102	0.000481	0.005273	0.001861	0.000313	0	0	0
18	PM4	0.000209	0	0.021497	0.001583	0.014451	0.006118	0.000746	0.002384	0.001528	2.3E-05	0	0	0
19	PM5	8.45E-05	0	0.005234	0.000412	0.001446	0.000338	0.001673	0.001823	0.001138	0.000778	0	0	0.000149
20	PM6	0.000273	0	0.006414	0.002114	0.012382	0.002586	0.000612	0.004133	0.002176	0.000208	0	0	0
21	PM7	5.4E-05	0	0.003872	0.000102	0.00156	0.000369	0	0.000276	0.000187	0	0	0	0
22	Panel together	7.52E-05	0	0.00796	0.000779	0.006492	0.001618	0.000496	0.00248	0.001291	0.000442	0	0	1.73E-05

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

We run the program as:

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

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

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:

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

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

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1		CHEM-A	CHEM-B	CHEM-C	CHEM-D	CHEM-E	CHEM-F	CHEM-G	CHEM-H	CHEM-I	CHEM-J	CHEM-K	CHEM-L	Groups together	PM1
2	CHEM-A		0.014998	0.036699	0.076944	0.109042	0.11001	0.084073	0.060364	0.056628	0.084636	0.10227	0.025248	0.04578	0.080863
3	CHEM-B	0.014998		0.038953	0.069387	0.10048	0.100916	0.077475	0.052709	0.047344	0.085	0.102564	0.024272	0.04064	0.078831
4	CHEM-C	0.036699	0.038953		0.090981	0.105161	0.107166	0.081496	0.051104	0.053526	0.073307	0.091927	0.042308	0.043103	0.107949
5	CHEM-D	0.076944	0.069387	0.090981		0.066268	0.062813	0.075669	0.067336	0.056723	0.095516	0.099073	0.06997	0.059635	0.061192
6	CHEM-E	0.109042	0.10048	0.105161	0.066268		0.024731	0.062986	0.065859	0.063441	0.088568	0.07662	0.096599	0.067651	0.123627
7	CHEM-F	0.11001	0.100916	0.107166	0.062813	0.024731		0.068838	0.068589	0.060918	0.092408	0.085253	0.097057	0.070784	0.119146
8	CHEM-G	0.084073	0.077475	0.081496	0.075669	0.062986	0.068838		0.040963	0.051071	0.070748	0.084015	0.071403	0.04741	0.116027
9	CHEM-H	0.060364	0.052709	0.051104	0.067336	0.065859	0.068589	0.040963		0.025512	0.055881	0.076894	0.053958	0.024334	0.103811
10	CHEM-I	0.056628	0.047344	0.053526	0.056723	0.063441	0.060918	0.051071	0.025512		0.069028	0.084831	0.047673	0.026403	0.093209
11	CHEM-J	0.084636	0.085	0.073307	0.095516	0.088568	0.092408	0.070748	0.055881	0.069028		0.059455	0.084326	0.057709	0.128617
12	CHEM-K	0.10227	0.102564	0.091927	0.099073	0.07662	0.085253	0.084015	0.076894	0.084831	0.059455		0.095776	0.068996	0.141
13	CHEM-L	0.025248	0.024272	0.042308	0.06997	0.096599	0.097057	0.071403	0.053958	0.047673	0.084326	0.095776		0.036609	0.0851
14	Groups together	0.04578	0.04064	0.043103	0.059635	0.067651	0.070784	0.04741	0.024334	0.026403	0.057709	0.068996	0.036609		0.094494
15	PM1	0.080863	0.078831	0.107949	0.061192	0.123627	0.119146	0.116027	0.103811	0.093209	0.128617	0.141	0.0851	0.094494	
16	PM2	0.082228	0.074046	0.079192	0.053899	0.036145	0.032324	0.055113	0.045944	0.036367	0.075474	0.070987	0.069912	0.043385	0.104809
17	PM3	0.082444	0.074229	0.080284	0.065673	0.057486	0.057867	0.040215	0.039826	0.042156	0.075109	0.085747	0.072632	0.047813	0.107131
18	PM4	0.105577	0.098646	0.10441	0.085072	0.064098	0.070007	0.026582	0.063137	0.070731	0.085469	0.094209	0.090769	0.068722	0.129032
19	PM5	0.015178	0.012873	0.034376	0.073936	0.100273	0.101968	0.076978	0.052822	0.050329	0.082346	0.09621	0.02364	0.038411	0.085946
20	PM6	0.09312	0.086982	0.110597	0.025124	0.084589	0.080188	0.095947	0.08971	0.079626	0.113496	0.11621	0.08813	0.082128	0.053954
21	PM7	0.068439	0.068083	0.09722	0.071506	0.128101	0.124544	0.112946	0.098902	0.089391	0.12457	0.139799	0.075349	0.089254	0.024789
22	Panel together	0.053766	0.047198	0.073222	0.032619	0.080339	0.077948	0.066576	0.055822	0.04639	0.088506	0.10076	0.049286	0.046851	0.051464

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

	Groups	CHEM-A	CHEM-B	CHEM-C	CHEM-D	CHEM-E	CHEM-F	CHEM-G	CHEM-H	CHEM-I	CHEM-J	CHEM-K	CHEM-L
Panel	0.047	0.054	0.047	0.073	0.033	0.080	0.078	0.067	0.056	0.046	0.089	0.101	0.049
PM1	0.094	0.081	0.079	0.108	0.061	0.124	0.119	0.116	0.104	0.093	0.129	0.141	0.085
PM2	0.043	0.082	0.074	0.079	0.054	<u>0.036</u>	<u>0.032</u>	0.055	0.046	<u>0.036</u>	0.075	<u>0.071</u>	0.070
PM3	0.048	0.082	0.074	0.08	0.066	0.057	0.058	0.040	<u>0.040</u>	0.042	<u>0.075</u>	0.086	0.073
PM4	0.069	0.106	0.099	0.104	0.085	0.064	0.070	<u>0.027</u>	0.063	0.071	0.085	0.094	0.091
PM5	0.038	<u>0.015</u>	<u>0.013</u>	<u>0.034</u>	0.074	0.100	0.102	0.077	0.053	0.050	0.082	0.096	<u>0.024</u>
PM6	0.082	0.093	0.087	0.111	<u>0.025</u>	0.085	0.080	0.096	0.090	0.080	0.113	0.116	0.088
PM7	0.089	0.068	0.068	0.097	0.072	0.128	0.125	0.113	0.099	0.089	0.125	0.140	0.075

For each research group we determined the panel member at the shortest distance. Average of shortest distance is 0.036 (SD 0.019). 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 CHEM-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 ‘CHEM-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 "CHEM-B_similarity matrix.xlsx" Sheet1
```

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

A	B	C	D	E	F
id	label	x	y	weight	cluster
166	Acoustics	-0.679	-0.1244	4.596	5
167	Computer Science, Artificial Intelligence	-0.6212	-0.2486	10.374	5
168	Critical Care Medicine	0.2449	0.1876	15.2022	1
169	Electrochemistry	-0.9707	-0.0975	15.8479	3
170	Chemistry, Applied	-0.6888	0.0381	21.3935	3
171	CHEM-B	-0.9019	-0.0913	20	3
172	Spectroscopy	-0.8294	-0.0648	16.774	3
173	Materials Science, Textiles	-0.932	-0.0652	13.0056	3
174	Rheumatology	0.2067	0.2069	6.6262	1
175	Linguistics	1.2002	-0.0252	2.124	2
176	Psychology	0.8384	0.0345	16.1112	2
177	Imaging Science Photographic Technology	-0.8268	-0.0523	10.55	4
178	Transportation	1.2204	-0.0986	8.53	2
179	Psychology, Multidisciplinary	1.194	-0.0406	15.642	2
180	Physics, Nuclear	-1.0957	-0.2548	4.546	3
181	Substance Abuse	0.965	0.0215	11.1081	2
182	Water Resources	-0.8956	0.0529	11.735	4

Figure 20. Excerpt of CHEM-B.csv file

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

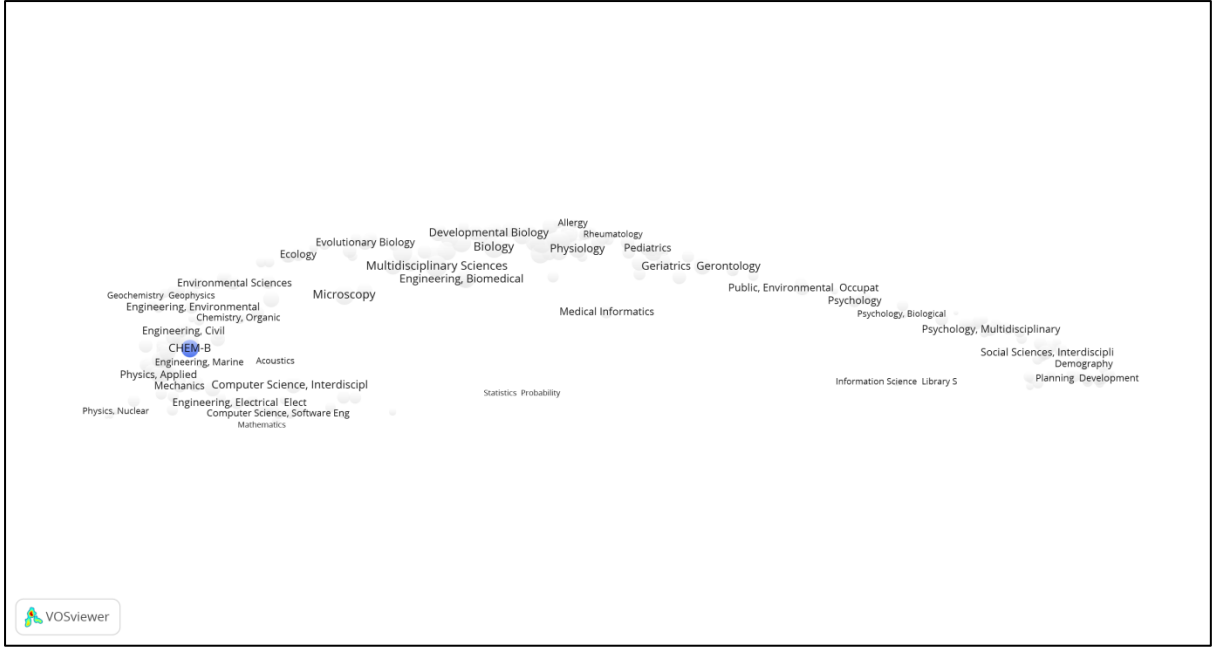


Figure 21. Location of the SAPV of CHEM-B in the WoS SCs similarity matrix

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:

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

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 (‘`Chemistry research groups and panel_WoS SCs.xlsx`’, see Figure 1), and calculates the weighted cosine similarity between all entities.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		CHEM-A	CHEM-B	CHEM-C	CHEM-D	CHEM-E	CHEM-F	CHEM-G	CHEM-H	CHEM-I	CHEM-J	CHEM-K	CHEM-L
2	CHEM-A	1	0.982187	0.847677	0.723419	0.571821	0.478582	0.605332	0.732093	0.760878	0.588363	0.540567	0.953971
3	CHEM-B	0.982187	1	0.839411	0.726605	0.645866	0.545493	0.661534	0.786989	0.806496	0.602171	0.494112	0.937794
4	CHEM-C	0.847677	0.839411	1	0.521896	0.602663	0.545308	0.678772	0.847588	0.846625	0.70738	0.609624	0.831461
5	CHEM-D	0.723419	0.726605	0.521896	1	0.683398	0.648156	0.542204	0.573729	0.647193	0.428559	0.358141	0.713267
6	CHEM-E	0.571821	0.645866	0.602663	0.683398	1	0.922139	0.832458	0.830423	0.87718	0.671146	0.465463	0.598489
7	CHEM-F	0.478582	0.545493	0.545308	0.648156	0.922139	1	0.795824	0.767794	0.869016	0.641315	0.386938	0.531089
8	CHEM-G	0.605332	0.661534	0.678772	0.542204	0.832458	0.795824	1	0.892046	0.886994	0.736063	0.419041	0.654594
9	CHEM-H	0.732093	0.786989	0.847588	0.573729	0.830423	0.767794	0.892046	1	0.944852	0.83936	0.526218	0.718718
10	CHEM-I	0.760878	0.806496	0.846625	0.647193	0.87718	0.869016	0.886994	0.944852	1	0.777271	0.505392	0.778148
11	CHEM-J	0.588363	0.602171	0.70738	0.428559	0.671146	0.641315	0.736063	0.83936	0.777271	1	0.645927	0.535623
12	CHEM-K	0.540567	0.494112	0.609624	0.358141	0.465463	0.386938	0.419041	0.526218	0.505392	0.645927	1	0.541216
13	CHEM-L	0.953971	0.937794	0.831461	0.713267	0.598489	0.531089	0.654594	0.718718	0.778148	0.535623	0.541216	1
14	Groups together	0.909811	0.924307	0.914322	0.71691	0.803164	0.732339	0.839523	0.919434	0.939212	0.772159	0.643656	0.912846
15	PM1	0.709127	0.66713	0.444769	0.921842	0.469094	0.449248	0.395478	0.439929	0.50718	0.32345	0.273289	0.661074
16	PM2	0.669697	0.713058	0.725942	0.674571	0.914099	0.945098	0.836889	0.846748	0.947317	0.703266	0.526572	0.712985
17	PM3	0.594176	0.655356	0.672531	0.569279	0.838593	0.831134	0.865932	0.880408	0.894059	0.711036	0.402606	0.603752
18	PM4	0.458664	0.517483	0.504265	0.483841	0.780965	0.776784	0.951019	0.758188	0.769409	0.626373	0.315195	0.548758
19	PM5	0.982693	0.990455	0.841568	0.669192	0.580631	0.475445	0.613543	0.747302	0.757836	0.57288	0.512372	0.93342
20	PM6	0.613188	0.600151	0.377413	0.973038	0.545148	0.519024	0.390607	0.410412	0.483971	0.294064	0.279892	0.602593
21	PM7	0.758308	0.71277	0.502893	0.850408	0.460341	0.438778	0.439836	0.49421	0.550242	0.373334	0.29022	0.699569
22	Panel together	0.827546	0.815165	0.635087	0.933756	0.687837	0.663057	0.663624	0.686769	0.74953	0.532121	0.394613	0.803754

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

We run the program as:

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

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

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

	Groups	CHEM-A	CHEM-B	CHEM-C	CHEM-D	CHEM-E	CHEM-F	CHEM-G	CHEM-H	CHEM-I	CHEM-J	CHEM-K	CHEM-L
Panel	0.816	0.828	0.815	0.635	0.934	0.688	0.663	0.664	0.687	0.750	0.532	0.395	0.804
PM1	0.607	0.709	0.667	0.445	0.922	0.469	0.449	0.395	0.440	0.507	0.323	0.273	0.661
PM2	0.871	0.670	0.713	0.726	0.675	0.914	0.945	0.837	0.847	0.947	0.703	0.527	0.713
PM3	0.806	0.594	0.655	0.673	0.569	0.839	0.831	0.866	0.880	0.894	0.711	0.403	0.604
PM4	0.709	0.459	0.517	0.504	0.484	0.781	0.777	0.951	0.758	0.769	0.626	0.315	0.549
PM5	0.901	0.983	0.990	0.842	0.669	0.581	0.475	0.614	0.747	0.758	0.573	0.512	0.933
PM6	0.570	0.613	0.600	0.377	0.973	0.545	0.519	0.391	0.410	0.484	0.294	0.280	0.603
PM7	0.650	0.758	0.713	0.503	0.850	0.460	0.439	0.440	0.494	0.550	0.373	0.290	0.700

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

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

	Groups	CHEM-A	CHEM-B	CHEM-C	CHEM-D	CHEM-E	CHEM-F	CHEM-G	CHEM-H	CHEM-I	CHEM-J	CHEM-K	CHEM-L
Panel	0.184	0.172	0.185	0.365	0.066	0.312	0.337	0.336	0.313	0.250	0.468	0.605	0.196
PM1	0.393	0.291	0.333	0.555	0.078	0.531	0.551	0.605	0.56	0.493	0.677	0.727	0.339
PM2	0.129	0.330	0.287	0.274	0.325	0.086	0.055	0.163	0.153	0.053	0.297	0.473	0.287
PM3	0.194	0.406	0.345	0.327	0.431	0.161	0.169	0.134	0.12	0.106	0.289	0.597	0.396
PM4	0.291	0.541	0.483	0.496	0.516	0.219	0.223	0.049	0.242	0.231	0.374	0.685	0.451
PM5	0.099	0.017	0.010	0.158	0.331	0.419	0.525	0.386	0.253	0.242	0.427	0.488	0.067
PM6	0.430	0.387	0.400	0.623	0.027	0.455	0.481	0.609	0.59	0.516	0.706	0.720	0.397
PM7	0.350	0.242	0.287	0.497	0.150	0.540	0.561	0.56	0.506	0.45	0.627	0.710	0.300

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

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

In Table 7, the upper triangle refers to Pearson's correlations while the lower triangle refers to Spearman correlation values. Table 7 and Figure 23 show there is moderate correlation

between barycenter and WCS methods, while the correlation is strong between SAPV and WCS methods.

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

	Pearson		
Spearman	Barycenter	SAPV	WCS
Barycenter	1	0.75	0.64
SAPV	0.72	1	0.92
WCS	0.62	0.92	1

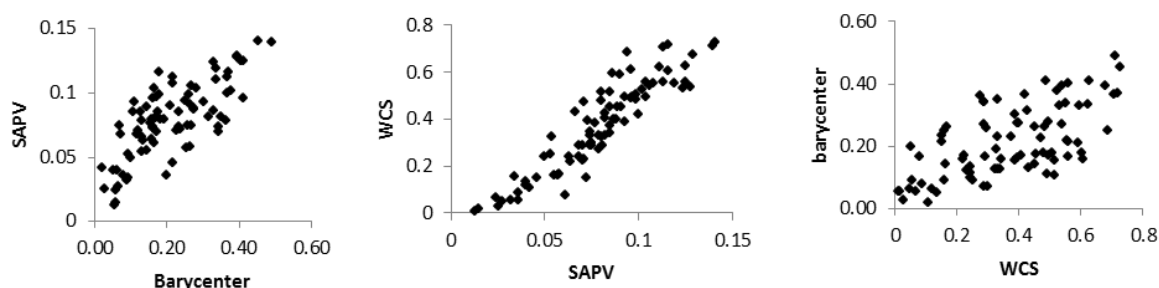


Figure 23. Scatter plots of the of correlation between barycenter, SAPV and WCS methods in WoS SCs in Chemistry department

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 ‘CHEM-B_journals title.txt’. For panel members we rename to ‘[Panel member code]_ journals title.txt’, for example ‘PM2_journals title.txt’.

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

	A	B	C
1	Source Titles	records	% of 920
2	SPECTROCHIMICA ACTA PART B ATOMIC SPECTROSCOPY	37	4.022
3	JOURNAL OF PHYSICAL CHEMISTRY A	37	4.022
4	JOURNAL OF ANALYTICAL ATOMIC SPECTROMETRY	35	3.804
5	X RAY SPECTROMETRY	27	2.935
6	ANALYTICAL CHEMISTRY	26	2.826
7	JOURNAL OF APPLIED PHYSICS	24	2.609
8	RAPID COMMUNICATIONS IN MASS SPECTROMETRY	23	2.5
9	JOURNAL OF PHYSICAL CHEMISTRY B	20	2.174
10	ULTRAMICROSCOPY	17	1.848
11	ANALYTICA CHIMICA ACTA	17	1.848
12	TETRAHEDRON	15	1.63
13	MICROPOROUS AND MESOPOROUS MATERIALS	15	1.63
14	JOURNAL OF MOLECULAR STRUCTURE	15	1.63
15	PHYSICAL REVIEW E	14	1.522

Figure 24. Excerpt of Chemistry 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 (CHEM-A, CHEM-B, CHEM-C, etc.), the panel members’ code names, (PM1, PM2, PM3, etc.), Panel members together and Groups together. The Excel file is saved as ‘Chemistry 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 'Chemistry panel and groups together_ journals title.xlsx' and export data from 'Panel_ journals titles. txt' and 'Groups together_ journals title.txt' in two different sheets (Figure 25).

	A	B	C
1	Source Titles	records	% of 920
2	SPECTROCHIMICA ACTA PART B ATOMIC SP	37	4.022
3	JOURNAL OF PHYSICAL CHEMISTRY A	37	4.022
4	JOURNAL OF ANALYTICAL ATOMIC SPECTRO	35	3.804
5	X RAY SPECTROMETRY	27	2.935
6	ANALYTICAL CHEMISTRY	26	2.826
7	JOURNAL OF APPLIED PHYSICS	24	2.609
8	RAPID COMMUNICATIONS IN MASS SPECTR	23	2.5
9	JOURNAL OF PHYSICAL CHEMISTRY B	20	2.174
10	ULTRAMICROSCOPY	17	1.848
11	ANALYTICA CHIMICA ACTA	17	1.848
12	TETRAHEDRON	15	1.63
13	MICROPOROUS AND MESOPOROUS MATER	15	1.63
14	JOURNAL OF MOLECULAR STRUCTURE	15	1.63
15	PHYSICAL REVIEW E	14	1.522
16	ATMOSPHERIC ENVIRONMENT	14	1.522
17	NUCLEAR INSTRUMENTS METHODS IN PHY	13	1.413
18	MICROCHIMICA ACTA	11	1.196
19	JOURNAL OF CHROMATOGRAPHY A	11	1.196
20	APPLIED SURFACE SCIENCE	11	1.196
21	ACTA CRYSTALLOGRAPHICA SECTION E STR	11	1.196
22	PHYSICAL CHEMISTRY CHEMICAL PHYSICS	10	1.087
23	LECTURE NOTES IN COMPUTER SCIENCE	10	1.087
24	CHEMICAL PHYSICS LETTERS	10	1.087
25	SPECTROCHIMICA ACTA PART A MOLECULA	9	0.978
26	JOURNAL OF CHEMICAL PHYSICS	9	0.978
27	TALANTA	7	0.761
28	PHYSICS AND CHEMISTRY OF LIQUIDS	7	0.761
29	JOURNAL OF MOLECULAR STRUCTURE THE	7	0.761
30	JOURNAL OF ATMOSPHERIC CHEMISTRY	7	0.761
31	INTERNATIONAL JOURNAL OF QUANTUM C	7	0.761
32	ENVIRONMENTAL SCIENCE TECHNOLOGY	7	0.761
	GroupsTogether	PanelTogether	

	A	B	C
1	Source Titles	records	% of 2150
2	INORGANIC CHEMISTRY	213	9.879
3	ORGANOMETALLICS	174	8.071
4	JOURNAL OF ORGANOMETALLIC CHEMIST	106	4.917
5	ANALYTICAL CHEMISTRY	94	4.36
6	JOURNAL OF THE AMERICAN CHEMICAL SC	86	3.989
7	INORGANICA CHIMICA ACTA	79	3.664
8	JOURNAL OF THE CHEMICAL SOCIETY DAL	67	3.108
9	FRESENIUS JOURNAL OF ANALYTICAL CHEM	55	2.551
10	EUROPEAN JOURNAL OF INORGANIC CHEM	44	2.041
11	JOURNAL OF CHEMICAL PHYSICS	40	1.855
12	TETRAHEDRON LETTERS	38	1.763
13	JOURNAL OF ANALYTICAL ATOMIC SPECTR	36	1.67
14	DALTON TRANSACTIONS	36	1.67
15	JOURNAL OF PHYSICAL CHEMISTRY A	32	1.484
16	JOURNAL OF ORGANIC CHEMISTRY	30	1.391
17	CHEMISTRY A EUROPEAN JOURNAL	28	1.299
18	ABSTRACTS OF PAPERS OF THE AMERICAN	28	1.299
19	JOURNAL OF THE CHEMICAL SOCIETY CHEM	27	1.252
20	RECUEIL DES TRAVAUX CHIMIQUES DES PA	24	1.113
21	ANALYTICAL AND BIOANALYTICAL CHEMIS	21	0.974
22	JOURNAL OF AEROSOL SCIENCE	20	0.928
23	ENVIRONMENTAL SCIENCE TECHNOLOGY	19	0.881
24	CHEMICAL COMMUNICATIONS	19	0.881
25	ACTA CRYSTALLOGRAPHICA SECTION C CRY	19	0.881
26	ORGANIC LETTERS	18	0.835
27	CHEMICAL PHYSICS LETTERS	17	0.788
28	ANGEWANDTE CHEMIE INTERNATIONAL E	17	0.788
29	ADVANCED SYNTHESIS CATALYSIS	17	0.788
30	EUROPEAN JOURNAL OF ORGANIC CHEMIS	16	0.742
31	TETRAHEDRON	15	0.696
32	JOURNAL OF PHYSICAL CHEMISTRY	14	0.649
	GroupsTogether	PanelTogether	

Figure 25. Excerpt of the Chemistry panel and research groups together_ journals title.xlsx file

	A	B	C	D
1		Source Titles	records_x	records_y
2	0	SPECTROCHIMICA ACTA PART B ATOMIC SP	37	13
3	1	JOURNAL OF PHYSICAL CHEMISTRY A	37	32
4	2	JOURNAL OF ANALYTICAL ATOMIC SPECTRO	35	36
5	3	X RAY SPECTROMETRY	27	0
6	4	ANALYTICAL CHEMISTRY	26	94
7	5	JOURNAL OF APPLIED PHYSICS	24	5
8	6	RAPID COMMUNICATIONS IN MASS SPECTR	23	3
9	7	JOURNAL OF PHYSICAL CHEMISTRY B	20	5
10	8	ULTRAMICROSCOPY	17	0
11	9	ANALYTICA CHIMICA ACTA	17	14
12	10	TETRAHEDRON	15	15
13	11	MICROPOROUS AND MESOPOROUS MATER	15	0
14	12	JOURNAL OF MOLECULAR STRUCTURE	15	13
15	13	PHYSICAL REVIEW E	14	0
16	14	ATMOSPHERIC ENVIRONMENT	14	5
17	15	NUCLEAR INSTRUMENTS METHODS IN PHY	13	0
18	16	MICROCHIMICA ACTA	11	0
19	17	JOURNAL OF CHROMATOGRAPHY A	11	12
20	18	APPLIED SURFACE SCIENCE	11	6
21	19	ACTA CRYSTALLOGRAPHICA SECTION E STRI	11	9
22	20	PHYSICAL CHEMISTRY CHEMICAL PHYSICS	10	11

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

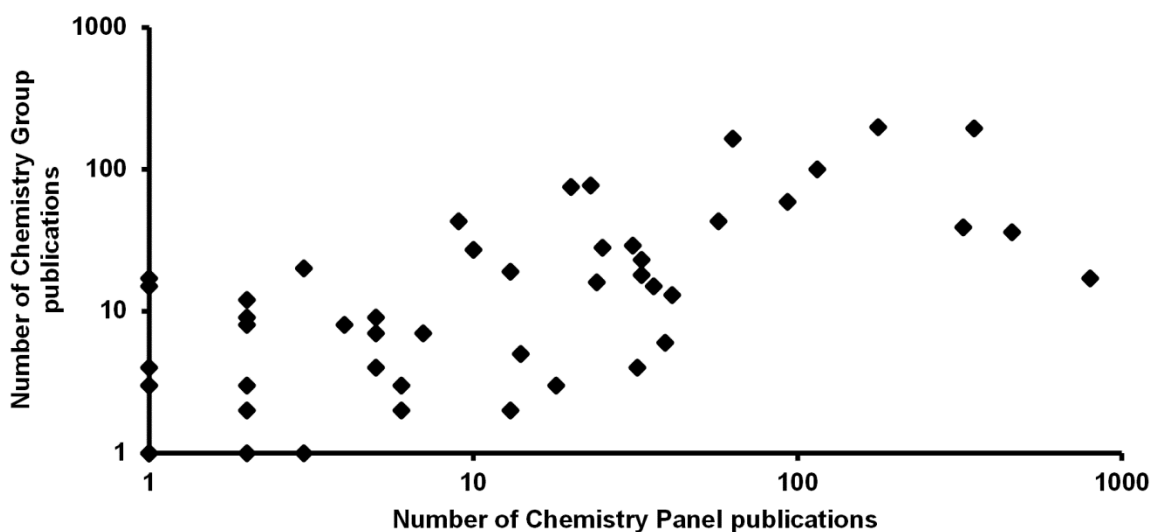


Figure 27. Log-log plot of the number of publications (log-log scale) per journals for the panel (horizontal axis) and research groups together (vertical axis) of the Chemistry department

We reuse the Python script ‘join-sheets.py’ (see section 2.2) to take the data of the two sheets and join it into one. We run the program as:

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

It produces a new Excel file named ‘Chemistry panel and research groups together_journals title-joined.xlsx’ (Figure 26). To calculate correlation, the value zero was kept on the corresponding journals in which either the panel or the groups had no publications (but not both). We calculate correlation coefficient using SPSS and obtain value ($r = 0.16$, $\rho = -0.36$). A log-log plot of the number of publications per journal for the Chemistry panel and research groups together is shown in Figure 27.

b) Top-Down correlation coefficient

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

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

The outcome shows that the top-down correlation between Chemistry research groups together and the panel based on the journals in which they publish is low (0.14). We conclude that both Spearman’s and top-down correlation suggest that there is low correlation between the journal publication portfolios of panel and research groups together.

In our opinion, the correlations are an insufficient measure in this case, as the similarity of journals is not taken into account here. This is reminiscent of the way diversity is sometimes studied using only the dimensions of variety and balance. As discussed by Stirling (2007), the

additional dimension of disparity – the opposite concept of similarity – is needed to provide a complete picture. Likewise, a comparison of publication profiles based on journals that does not consider journal similarity might yield distorted results.

3.3 Journal similarity matrix

Journal similarity data were received as a NET file (file name cosine.net) from Loet Leydesdorff in the context of the joint paper (Rahman, Guns, Leydesdorff, & Engels, 2016). While we did not construct this similarity matrix ourselves, we briefly outline the main steps that were taken to create it. The data was harvested from Clarivate Analytics's (formerly Thomson Reuters') Journal Citation Reports (JCR) of the Science and Social Science Editions 2011. An aggregated journal-journal citation matrix of 10,675 journals¹ was constructed with a grand total of 35,295,459 citations over the entire matrix, which was subsequently normalized in the citing direction. The similarities between journals are calculated using the cosine similarity between their citing distributions respectively (see Leydesdorff, Rafols, & Chen (2013) for details). The resulting journal similarity matrix can be considered as an adjacency matrix, and thus is equivalent to a weighted network where similar journals are linked and link weights increase with similarity strength.

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

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

This way, we store the adjacency matrix in HDF5 (Hierarchical Data Format version 5), which was found to be the most efficient way of storing the data in terms of speed and memory requirements. From http://www.leydesdorff.net/journals11/citing_all.txt, we download the journal VOS map and save it to a file named 'Journal_VOS_map.xlsx' (Figure 28).

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

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

Figure 28. Excerpt of the journal VOS map data

3.4 Journal overlay map creation

During data collection, the resulting files were downloaded using the default name ‘analyze.txt’ (see section 3.1).

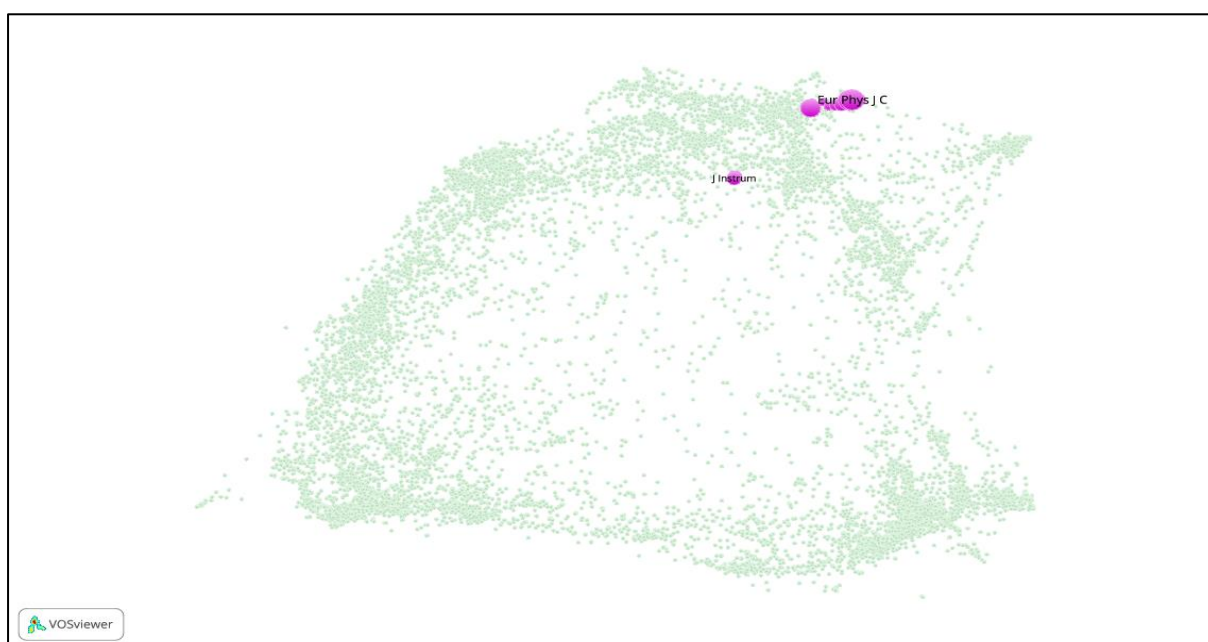


Figure 29. Journal overlay map of the CHEM-B research group

We downloaded the ‘Analyze.exe’ program, as well as the file ‘citing.dbf’ from <http://www.leydesdorff.net/journals11>. For each entity (Individual research groups, panel members, research groups together and panel), we save the corresponding ‘analyze.txt’ file in the concerned folder and run the program ‘Analyze.exe’. ‘Analyze.exe’ reads ‘analyze.txt’, and generates an output file ‘citing.txt’. We open the latter in VOSviewer to obtain an overlay map. For example, Figure 29 shows the journal overlay map of the CHEM-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 ‘Chemistry research groups and panel_journals title.xlsx’ (Figure 24) as input. We run the program as follows

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

Figure 31. Excerpt of journal name change.xlsx file

	A	B	C
1		x	y
2	CHEM-A	0.067802	0.772591
3	CHEM-B	0.054251	0.772454
4	CHEM-C	0.193197	0.66659
5	CHEM-D	-0.16272	0.73104
6	CHEM-E	-0.30254	0.485175
7	CHEM-F	-0.22018	0.560578
8	CHEM-G	-0.00767	0.536403
9	CHEM-H	0.046428	0.557191
10	CHEM-I	-0.05375	0.629241
11	CHEM-J	-0.00532	0.526184
12	CHEM-K	0.036307	0.338472
13	CHEM-L	0.018509	0.707481
14	Groups	0.015412	0.613007
15	PM1	-0.03938	0.81528
16	PM2	-0.21366	0.469588
17	PM3	-0.07934	0.620878
18	PM4	-0.12069	0.557891
19	PM5	0.069165	0.663625
20	PM6	-0.10473	0.772805
21	PM7	-0.02309	0.802566
22	Panel	-0.0576	0.718767

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

We keep the ‘Chemistry 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 "Journal_VOS_map.xlsx" "Chemistry research groups
and panel journals title.xlsx"
```

This program calculates the barycenter coordinates of Chemistry individual research groups, panel members, groups and panel in the journals VOS map in an output file ‘Chemistry 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.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		CHEM-A	CHEM-B	CHEM-C	CHEM-D	CHEM-E	CHEM-F	CHEM-G	CHEM-H	CHEM-I	CHEM-J	CHEM-K	CHEM-L
2	CHEM-A	0	0.013552	0.164195	0.234234	0.468784	0.357606	0.247953	0.216458	0.187948	0.257029	0.43526	0.081665
3	CHEM-B	0.013552	0	0.17468	0.220885	0.458068	0.346702	0.244037	0.215405	0.179372	0.253373	0.434353	0.074155
4	CHEM-C	0.164195	0.17468	0	0.361702	0.527885	0.426752	0.239364	0.183055	0.249756	0.243155	0.363697	0.17941
5	CHEM-D	0.234234	0.220885	0.361702	0	0.282841	0.179886	0.248845	0.271964	0.149119	0.258338	0.440137	0.182751
6	CHEM-E	0.468784	0.458068	0.527885	0.282841	0	0.111662	0.299285	0.356317	0.287487	0.300028	0.369238	0.390499
7	CHEM-F	0.357606	0.346702	0.426752	0.179886	0.111662	0	0.213881	0.266627	0.180035	0.217589	0.339287	0.280271
8	CHEM-G	0.247953	0.244037	0.239364	0.248845	0.299285	0.213881	0	0.057952	0.103646	0.010484	0.202757	0.173069
9	CHEM-H	0.216458	0.215405	0.183055	0.271964	0.356317	0.266627	0.057952	0	0.123397	0.06033	0.218953	0.152861
10	CHEM-I	0.187948	0.179372	0.249756	0.149119	0.287487	0.180035	0.103646	0.123397	0	0.113868	0.304396	0.106504
11	CHEM-J	0.257029	0.253373	0.243155	0.258338	0.300028	0.217589	0.010484	0.06033	0.113868	0	0.192273	0.182857
12	CHEM-K	0.43526	0.434353	0.363697	0.440137	0.369238	0.339287	0.202757	0.218953	0.304396	0.192273	0	0.369438
13	CHEM-L	0.081665	0.074155	0.17941	0.182751	0.390499	0.280271	0.173069	0.152861	0.106504	0.182857	0.369438	0
14	GroupsTogether	0.167964	0.164109	0.185684	0.213686	0.342683	0.241353	0.080005	0.063854	0.071043	0.089265	0.275329	0.094525
15	PM1	0.115369	0.102958	0.276043	0.149362	0.422163	0.312349	0.280674	0.271979	0.186594	0.291095	0.482778	0.122358
16	PM2	0.413559	0.404355	0.452041	0.266368	0.090233	0.091223	0.216556	0.274443	0.225963	0.215885	0.282267	0.332408
17	PM3	0.211346	0.202042	0.276341	0.138158	0.261215	0.153206	0.110782	0.140971	0.026918	0.120188	0.305167	0.130667
18	PM4	0.285699	0.27684	0.332172	0.178177	0.195849	0.099528	0.115044	0.167116	0.097832	0.11964	0.2698	0.204334
19	PM5	0.108975	0.109846	0.124067	0.241482	0.412318	0.307145	0.148622	0.108835	0.127634	0.156329	0.326809	0.067002
20	PM6	0.172529	0.158977	0.31629	0.071465	0.349084	0.241597	0.25555	0.263318	0.152345	0.265899	0.456657	0.139478
21	PM7	0.095704	0.082992	0.255475	0.156884	0.422882	0.312095	0.266609	0.255031	0.176017	0.276952	0.467879	0.103785
22	PanelTogether	0.136464	0.124066	0.256165	0.105832	0.338467	0.226839	0.189076	0.192167	0.089609	0.199552	0.391718	0.07694

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

We reuse the formula 4 and determine the Euclidean distance between barycenters. We note that the python script ‘journals-categories.py’ executes both formula 3 and 4.

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

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

	Groups	CHEM-A	CHEM-B	CHEM-C	CHEM-D	CHEM-E	CHEM-F	CHEM-G	CHEM-H	CHEM-I	CHEM-J	CHEM-K	CHEM-L
Panel	0.129	0.136	0.124	0.256	0.106	0.338	0.227	0.189	0.192	0.09	0.200	0.392	0.077
PM1	0.210	0.115	0.103	0.276	0.149	0.422	0.312	0.281	0.272	0.187	0.291	0.483	0.122
PM2	0.270	0.414	0.404	0.452	0.266	<u>0.090</u>	<u>0.091</u>	0.217	0.274	0.226	0.216	0.282	0.332
PM3	0.095	0.211	0.202	0.276	0.138	0.261	0.153	<u>0.111</u>	0.141	<u>0.027</u>	0.120	0.305	0.131
PM4	0.147	0.286	0.277	0.332	0.178	0.196	0.100	0.115	0.167	0.098	<u>0.120</u>	<u>0.270</u>	0.204
PM5	0.074	0.109	0.110	<u>0.124</u>	0.241	0.412	0.307	0.149	<u>0.109</u>	0.128	0.156	0.327	<u>0.067</u>
PM6	0.200	0.173	0.159	0.316	<u>0.071</u>	0.349	0.242	0.256	0.263	0.152	0.266	0.457	0.139
PM7	0.193	<u>0.096</u>	<u>0.083</u>	0.255	0.157	0.423	0.312	0.267	0.255	0.176	0.277	0.468	0.104

For each research group we determined the panel member at the shortest distance. Average of shortest distance is 0.104 (SD 0.058). The number in the row of this panel member is indicated in bold and underlined. Distances whose confidence intervals overlap with that of the shortest distance are in bold (same column).

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

c) Barycenter overlay map

We take the Journal level_VOS map (Figure 28) file and manually add the Chemistry individual groups, panel members, research groups together and panel's coordinates (Figure 32) after the 10,673 journals title. We fill up the 'weight' column with 20 (we can put other numbers too) to highlight the size of the bubble of the added entities. In the 'cluster' column, we assign 1 to all the 10,673 journals, 2 to the research groups together, 3 to all research groups, 4 to the panel, and 5 to individual panel members. We save the map file as 'Barycenter map of Chemistry 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 Chemistry department and include the confidence regions of the respective barycenter of panel, panel members (PM), research groups and research groups together using the journal VOS map (Figure 36). The bootstrap replications of barycenters are also used to add a 95% confidence region for each barycenter to the maps. In particular, we stress that overlapping confidence regions as seen in Figure 36

(figure with overlapping regions in it) does not correspond to overlap between CIs for distances. For detail process about confidence regions see section 2.6c.

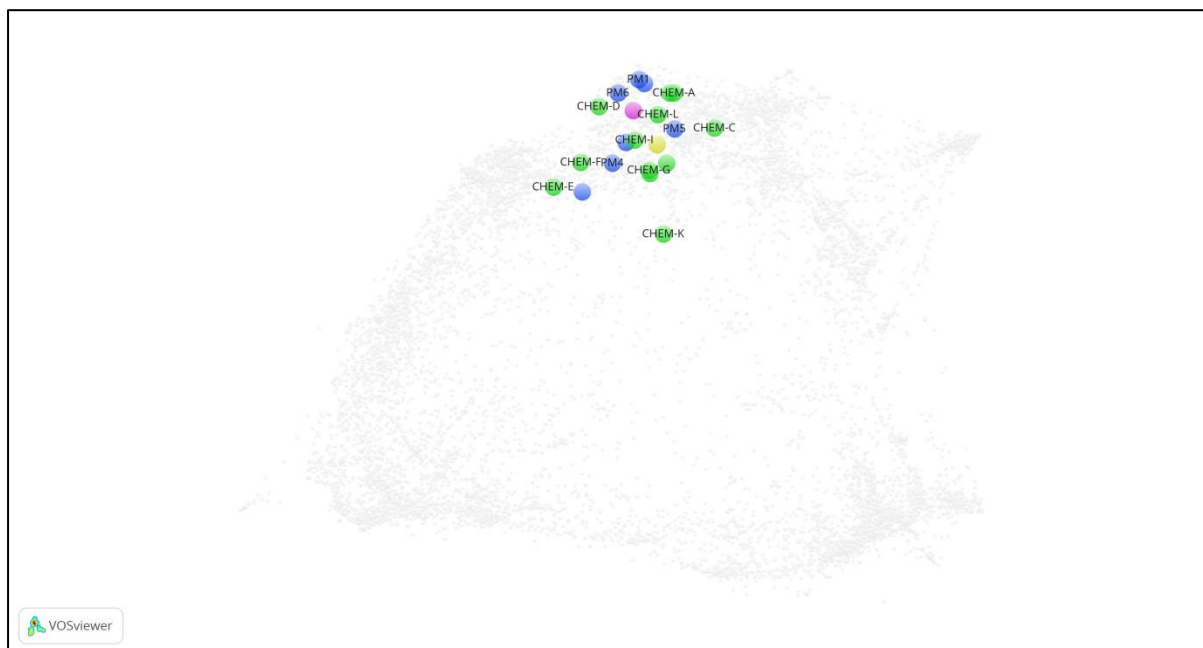


Figure 34. Barycenter overlay map of Chemistry panel, panel members (PM), research groups and research groups together

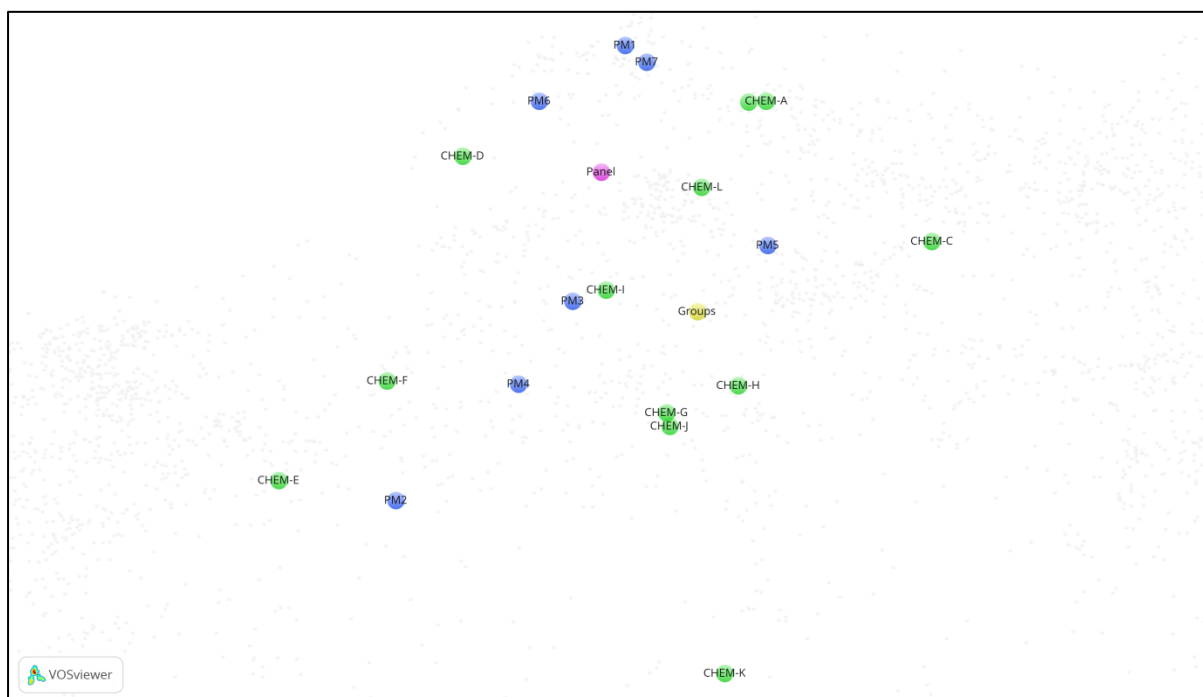


Figure 35. Barycenter overlay map of Chemistry panel, panel members (PM), research groups and research groups together (zoomed)

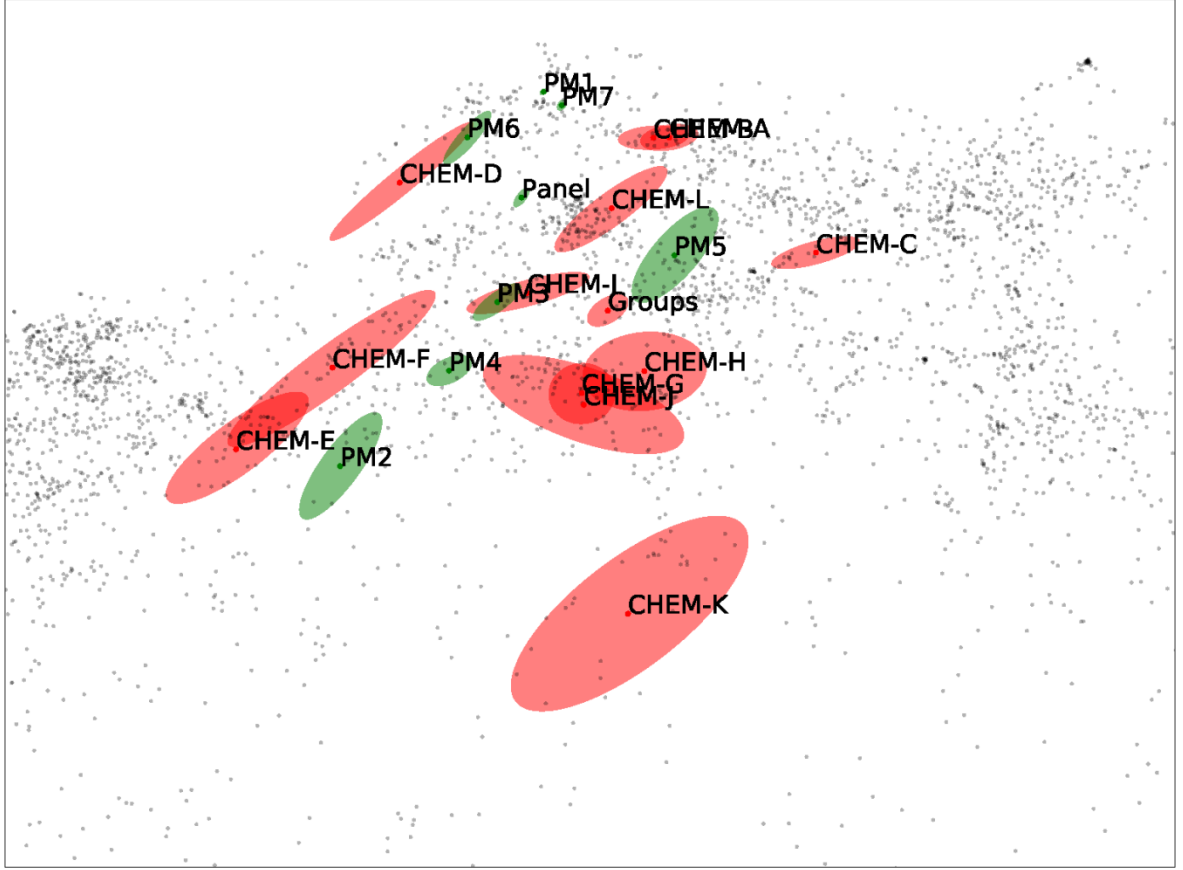


Figure 36. Barycenter overlay map of Chemistry panel, panel members (PM), research groups and research groups together with their confidence regions

3.6 Similarity-adapted publication vector method

a) Similarity-adapted publication vector calculation

Recall formula 5. A similarity-adapted publication vector is determined as the vector $C = (C_1, C_2, \dots, C_N)$, where:

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

Here, s_{kj} denotes the similarity value between the k -th and the j -th journal, and m_j is the number of publications in journal j . The numerator of the formula is equal to the k -th element of $S * M$, the multiplication of the similarity matrix S and the column matrix of publications $M = (m_j)_j$. The denominator is the L_1 -norm of the unnormalized vector.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		40R A QUJHN JOURI	PG BULLE	PS JOURN	PHARMS	TCC REVIEM	INAL IMAT	SCHEN	SND APPLIEM	ERGENCE	CEMIC MED	MIC PEDI	
2	CHEM-A	5.23E-07	1.27E-05	2.55E-05	0.000131	6.92E-05	0.000136	1.17E-06	5.86E-07	1.44E-05	2.45E-06	6.48E-06	2.48E-06
3	CHEM-B	1.31E-07	1.35E-05	2.49E-05	0.000142	7.66E-05	0.000151	1.04E-06	5.85E-07	1.39E-05	2.43E-06	6.76E-06	2.09E-06
4	CHEM-C	4.4E-07	1.76E-05	3.56E-05	0.00011	4.06E-05	0.000203	1.97E-06	1.93E-06	2.43E-05	3.73E-06	8.69E-06	3.84E-06
5	CHEM-D	0	1.21E-05	1.41E-05	0.000243	8.57E-05	0.000122	4.14E-06	5.24E-07	6.35E-06	5.3E-06	8.39E-06	3.93E-06
6	CHEM-E	1.62E-07	3.26E-05	3.35E-05	0.000289	5.41E-05	0.000125	5.68E-06	9.7E-07	1.61E-05	8.52E-06	1.46E-05	8.78E-06
7	CHEM-F	3.01E-07	2.71E-05	2.2E-05	0.000475	0.000303	0.000272	1.29E-05	1.05E-06	9.15E-06	1.34E-05	1.62E-05	1.01E-05
8	CHEM-G	2.54E-06	6.46E-05	6.79E-05	0.000183	8.39E-05	0.000338	9.29E-06	5.93E-07	7.05E-06	1E-05	1.57E-05	1.25E-05
9	CHEM-H	6.23E-07	3.79E-05	8.08E-05	0.000178	5.53E-05	0.000288	5.63E-06	9.86E-07	1.33E-05	7.49E-06	1.44E-05	7.89E-06
10	CHEM-I	1.24E-07	2.33E-05	4.31E-05	0.000237	0.000102	0.000333	3.28E-06	5.49E-07	8.81E-06	5.54E-06	9.56E-06	5.54E-06
11	CHEM-J	1.21E-05	2.71E-05	2.89E-05	0.000264	0.000143	0.000387	9.37E-06	8.72E-07	1.33E-05	8.98E-06	1.4E-05	8.28E-06
12	CHEM-K	1.12E-05	3.01E-05	2.97E-05	0.000143	2.53E-05	8.69E-05	0.000127	7.61E-06	2.84E-05	2.01E-05	2.26E-05	1.37E-05
13	CHEM-L	2.82E-07	1.8E-05	2.45E-05	0.00015	7.32E-05	0.000166	4.55E-06	3.95E-07	1.25E-05	4.61E-06	1E-05	3.88E-06
14	GroupsTogether	1.74E-06	2.7E-05	3.7E-05	0.000178	6.99E-05	0.000194	1.48E-05	1.46E-06	1.55E-05	6.96E-06	1.18E-05	6.55E-06
15	PM1	1.23E-08	9.94E-06	1.35E-05	0.000144	5.93E-05	0.00014	1.3E-06	2.28E-07	5.63E-06	1.83E-06	5.03E-06	1.52E-06
16	PM2	7.55E-08	2.51E-05	2.62E-05	0.000277	6.98E-05	0.000168	8.09E-06	1.88E-06	1.46E-05	1.14E-05	1.85E-05	9.93E-06
17	PM3	1.1E-07	3.15E-05	0.000208	0.000203	5.81E-05	0.000263	3.82E-06	6.2E-07	9.46E-06	6.8E-06	1.47E-05	6.99E-06
18	PM4	1.48E-06	4.88E-05	3.54E-05	0.000272	0.000114	0.000257	7.35E-06	1.09E-06	8.69E-06	1.06E-05	1.45E-05	1.11E-05
19	PM5	2.29E-07	2.68E-05	3.39E-05	0.000133	6.89E-05	0.000112	9.32E-06	7.69E-07	1.68E-05	1.69E-05	2.1E-05	3.27E-05
20	PM6	1.06E-07	1.04E-05	1.19E-05	0.000198	6.93E-05	0.000128	7.33E-06	2.62E-07	6.61E-06	2.99E-06	5.95E-06	2.2E-06
21	PM7	1.02E-07	1.27E-05	1.73E-05	0.000139	6.06E-05	0.000139	2.4E-06	3.14E-07	7.18E-06	2.08E-06	5.67E-06	1.75E-06
22	PanelTogether	2.53E-07	2.06E-05	3.38E-05	0.000182	6.92E-05	0.000162	4.58E-06	6.53E-07	9.03E-06	6.26E-06	1.05E-05	7.62E-06

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

A Python script ‘sim_adapted_pub_vectors_journals.py’ (the calculation of SAPVs is carried out by the sa_vector function, see section 5) is used that takes as input the similarity matrix and the weights (number of publications) per journals (‘Chemistry 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 ‘Chemistry research groups and panel_journals title.xlsx’ in a folder. We run the program as:

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

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

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1		CHEM-A	CHEM-B	CHEM-C	CHEM-D	CHEM-E	CHEM-F	CHEM-G	CHEM-H	CHEM-I	CHEM-J	CHEM-K	CHEM-L	Groups
2	CHEM-A	0	0.001679	0.016863	0.015738	0.018471	0.018395	0.018578	0.016591	0.01555	0.018716	0.018709	0.008041	0.010351
3	CHEM-B	0.001679	0	0.017528	0.015151	0.018044	0.017765	0.018246	0.01657	0.015276	0.018358	0.018918	0.008066	0.010322
4	CHEM-C	0.016863	0.017528	0	0.027503	0.019191	0.01969	0.017366	0.010989	0.013815	0.016051	0.012748	0.015032	0.011549
5	CHEM-D	0.015738	0.015151	0.027503	0	0.021869	0.020771	0.023597	0.02426	0.021817	0.024042	0.026139	0.016974	0.018942
6	CHEM-E	0.018471	0.018044	0.019191	0.021869	0	0.009706	0.01259	0.011378	0.012121	0.014192	0.01122	0.013264	0.010503
7	CHEM-F	0.018395	0.017765	0.01969	0.020771	0.009706	0	0.009658	0.011508	0.007706	0.008806	0.016015	0.014322	0.011031
8	CHEM-G	0.018578	0.018246	0.017366	0.023597	0.01259	0.009658	0	0.008129	0.007978	0.008851	0.015395	0.014602	0.010399
9	CHEM-H	0.016591	0.01657	0.010989	0.02426	0.011378	0.011508	0.008129	0	0.007365	0.008427	0.010106	0.012771	0.007095
10	CHEM-I	0.01555	0.015276	0.013815	0.021817	0.012121	0.007706	0.007978	0.007365	0	0.007006	0.014153	0.011726	0.007477
11	CHEM-J	0.018716	0.018358	0.016051	0.024042	0.014192	0.008806	0.008851	0.008427	0.007006	0	0.016068	0.0159	0.011117
12	CHEM-K	0.018709	0.018918	0.012748	0.026139	0.01122	0.016015	0.015395	0.010106	0.014153	0.016068	0	0.014097	0.010524
13	CHEM-L	0.008041	0.008066	0.015032	0.016974	0.013264	0.014322	0.014602	0.012771	0.011726	0.0159	0.014097	0	0.006334
14	GroupsTogether	0.010351	0.010322	0.011549	0.018942	0.010503	0.011031	0.010399	0.007095	0.007477	0.011117	0.010524	0.006334	0
15	PM1	0.015417	0.015248	0.029331	0.010592	0.026715	0.025432	0.027073	0.027634	0.024921	0.027742	0.029753	0.018468	0.021805
16	PM2	0.017108	0.016722	0.01733	0.021506	0.004653	0.008131	0.011423	0.009966	0.009577	0.012297	0.010494	0.011713	0.008742
17	PM3	0.018528	0.018268	0.016805	0.023682	0.009302	0.010538	0.008679	0.007566	0.009601	0.011678	0.012112	0.013977	0.00989
18	PM4	0.019117	0.018578	0.019216	0.022618	0.010502	0.005722	0.005395	0.009922	0.007526	0.008537	0.015964	0.014874	0.011028
19	PM5	0.005507	0.005777	0.013461	0.01919	0.016033	0.017167	0.016875	0.013428	0.013912	0.016929	0.014749	0.007543	0.007869
20	PM6	0.01713	0.016685	0.029676	0.00493	0.025117	0.024087	0.026403	0.027135	0.024566	0.027056	0.02896	0.018866	0.021597
21	PM7	0.011565	0.011474	0.025695	0.011775	0.02352	0.022452	0.023755	0.023979	0.021472	0.02449	0.026191	0.01476	0.018059
22	PanelTogether	0.008403	0.007959	0.020421	0.010597	0.016229	0.015244	0.016793	0.016949	0.014509	0.017749	0.019586	0.008845	0.011043

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

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 Chemistry individual research groups, panel members, research groups together and panel (Figure 38), we extract Table 9 containing only the distances between the research groups and research groups together on the one hand and the panel and panel members on the other, for the convenience of analysis.

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

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

	Groups	CHEM-A	CHEM-B	CHEM-C	CHEM-D	CHEM-E	CHEM-F	CHEM-G	CHEM-H	CHEM-I	CHEM-J	CHEM-K	CHEM-L
Panel	0.011	0.008	0.008	0.02	0.011	0.016	0.015	0.017	0.017	0.015	0.018	0.020	0.009
PM1	0.022	0.015	0.015	0.029	0.011	0.027	0.025	0.027	0.028	0.025	0.028	0.030	0.018
PM2	0.009	0.017	0.017	0.017	0.022	<u>0.005</u>	0.008	0.011	0.010	0.010	0.012	<u>0.010</u>	0.012
PM3	0.010	0.019	0.018	0.017	0.024	0.009	0.011	0.009	<u>0.008</u>	0.010	0.012	0.012	0.014
PM4	0.011	0.019	0.019	0.019	0.023	0.011	<u>0.006</u>	<u>0.005</u>	0.010	<u>0.008</u>	<u>0.009</u>	0.016	0.015
PM5	0.008	<u>0.006</u>	<u>0.006</u>	<u>0.013</u>	0.019	0.016	0.017	0.017	0.013	0.014	0.017	0.015	<u>0.008</u>
PM6	0.022	0.017	0.017	0.030	<u>0.005</u>	0.025	0.024	0.026	0.027	0.025	0.027	0.029	0.019
PM7	0.018	0.012	0.011	0.026	0.012	0.024	0.022	0.024	0.024	0.021	0.024	0.026	0.015

For each research group we determine the panel member at the highest similarity. Average of shortest distance is 0.007 (SD 0.002). 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).

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:

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

The numerator is nothing but the matrix multiplication: $(M^k)^t * S * R^m$, 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: $(M^k)^t * S * M^k$ and $(R^m)^t * S * R^m$. 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 ‘`Chemistry 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 (‘`Chemistry research groups and panel_journals title.xlsx`’, Figure 24), and calculates SAPVs for all entities, as well as the pairwise distances between them. We run the program as:

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

This program calculates the WCS value in an output file as ‘`Chemistry research groups and panel _journals title-cosine.xlsx`’ (Figure 39).

	A	B	C	D	E	F	G	H	I	J	K	L
1		CHEM-A	CHEM-B	CHEM-C	CHEM-D	CHEM-E	CHEM-F	CHEM-G	CHEM-H	CHEM-I	CHEM-J	CHEM-K
2	CHEM-A	1	0.967911	0.392257	0.540703	0.393971	0.270767	0.251355	0.266399	0.324413	0.175454	0.251822
3	CHEM-B	0.967911	1	0.362388	0.557634	0.426366	0.298174	0.27138	0.272252	0.344063	0.184795	0.227855
4	CHEM-C	0.392257	0.362388	1	0.123043	0.34957	0.334856	0.487083	0.699829	0.659265	0.406718	0.522601
5	CHEM-D	0.540703	0.557634	0.123043	1	0.405358	0.305201	0.183163	0.134594	0.230817	0.13104	0.082936
6	CHEM-E	0.393971	0.426366	0.34957	0.405358	1	0.680471	0.565995	0.489691	0.730449	0.41342	0.268693
7	CHEM-F	0.270767	0.298174	0.334856	0.305201	0.680471	1	0.712962	0.517309	0.780361	0.616682	0.175274
8	CHEM-G	0.251355	0.27138	0.487083	0.183163	0.565995	0.712962	1	0.808558	0.777021	0.695775	0.224241
9	CHEM-H	0.266399	0.272252	0.699829	0.134594	0.489691	0.517309	0.808558	1	0.751357	0.771858	0.360482
10	CHEM-I	0.324413	0.344063	0.659265	0.230817	0.730449	0.780361	0.777021	0.751357	1	0.593699	0.278444
11	CHEM-J	0.175454	0.184795	0.406718	0.13104	0.41342	0.616682	0.695775	0.771858	0.593699	1	0.211747
12	CHEM-K	0.251822	0.227855	0.522601	0.082936	0.268693	0.175274	0.224241	0.360482	0.278444	0.211747	1
13	CHEM-L	0.751098	0.739012	0.400598	0.457612	0.43007	0.322038	0.317804	0.282105	0.396984	0.177939	0.292408
14	GroupsTogether	0.76454	0.761645	0.74219	0.484638	0.687679	0.630917	0.714582	0.729693	0.779748	0.551924	0.481513
15	PM1	0.669612	0.633057	0.143976	0.730365	0.336014	0.243914	0.178914	0.127173	0.220436	0.098569	0.091407
16	PM2	0.372576	0.397381	0.431747	0.315305	0.830047	0.804971	0.648762	0.54592	0.795203	0.513388	0.308185
17	PM3	0.234576	0.247236	0.549041	0.175532	0.558004	0.58486	0.726229	0.774355	0.793208	0.481178	0.243124
18	PM4	0.24915	0.280874	0.369998	0.230788	0.72419	0.876935	0.850219	0.612607	0.805445	0.593613	0.183856
19	PM5	0.940646	0.944299	0.452333	0.392626	0.371115	0.229375	0.230562	0.285202	0.31268	0.165755	0.295116
20	PM6	0.533898	0.53847	0.110121	0.957006	0.357575	0.256791	0.164756	0.110834	0.198278	0.098904	0.076084
21	PM7	0.727295	0.68867	0.186101	0.599116	0.345324	0.260604	0.211867	0.161915	0.252798	0.119728	0.117067
22	PanelTogether	0.768746	0.746963	0.303983	0.716916	0.535197	0.460528	0.399316	0.325297	0.454968	0.260449	0.183987

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

From the calculated WCS value matrix of Chemistry 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 Chemistry groups, panel members, panel and research groups together using the journal similarity matrix

	Groups	CHEM-A	CHEM-B	CHEM-C	CHEM-D	CHEM-E	CHEM-F	CHEM-G	CHEM-H	CHEM-I	CHEM-J	CHEM-K	CHEM-L
Panel	0.728	0.769	0.747	0.304	0.717	0.535	0.461	0.399	0.325	0.455	0.260	0.184	0.682
PM1	0.520	0.670	0.633	0.144	0.730	0.336	0.244	0.179	0.127	0.220	0.099	0.091	0.580
PM2	0.719	0.373	0.397	0.432	0.315	0.830	0.805	0.649	0.546	0.795	0.513	0.308	0.470
PM3	0.649	0.235	0.247	0.549	0.176	0.558	0.585	0.726	0.774	0.793	0.481	0.243	0.284
PM4	0.664	0.249	0.281	0.370	0.231	0.724	0.877	0.850	0.613	0.805	0.594	0.184	0.319
PM5	0.741	0.941	0.944	0.452	0.393	0.371	0.229	0.231	0.285	0.313	0.166	0.295	0.702
PM6	0.465	0.534	0.538	0.110	0.957	0.358	0.257	0.165	0.111	0.198	0.099	0.076	0.495
PM7	0.563	0.727	0.689	0.186	0.599	0.345	0.261	0.212	0.162	0.253	0.12	0.117	0.614

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

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

	Groups	CHEM-A	CHEM-B	CHEM-C	CHEM-D	CHEM-E	CHEM-F	CHEM-G	CHEM-H	CHEM-I	CHEM-J	CHEM-K	CHEM-L
Panel	0.272	0.231	0.253	0.696	0.283	0.465	0.539	0.601	0.675	0.545	0.740	0.816	0.318
PM1	0.480	0.330	0.367	0.856	0.270	0.664	0.756	0.821	0.873	0.780	0.901	0.909	0.420
PM2	0.281	0.627	0.603	0.568	0.685	0.170	0.195	0.351	0.454	0.205	0.487	0.692	0.530
PM3	0.351	0.765	0.753	0.451	0.824	0.442	0.415	0.274	0.226	0.207	0.519	0.757	0.716
PM4	0.336	0.751	0.719	0.630	0.769	0.276	0.123	0.150	0.387	0.195	0.406	0.816	0.681
PM5	0.259	0.059	0.056	0.548	0.607	0.629	0.771	0.769	0.715	0.687	0.834	0.705	0.298
PM6	0.535	0.466	0.462	0.890	0.043	0.642	0.743	0.835	0.889	0.802	0.901	0.924	0.505
PM7	0.437	0.273	0.311	0.814	0.401	0.655	0.739	0.788	0.838	0.747	0.880	0.883	0.386

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

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

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

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

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

WoS SCs

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

Journals

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

We calculate Spearman's rank-order correlation coefficient between each pair of the six approaches. More specifically, we determine the correlation using the distances between barycenters and between SAPVs, and dissimilarity of individual research groups and panel members using $1 - \text{WCS}$. For the sake of simplicity, the results are shown under the WCS method. We create an MS Excel file (Figure 40) containing:

- i) Euclidean distances between barycenters of the Chemistry individual research groups and panel members at the level of WoS SCs,
- ii) Euclidean distances between barycenters of the Chemistry individual research groups and panel members at the level of journals,

- iii) Euclidean distances between SAPVs of the Chemistry individual research groups and panel members at the level of WoS SCs,
- iv) Euclidean distances between SAPVs of the Chemistry individual research groups and panel members at the level of journals,
- v) WCS value of Chemistry individual research groups and panel members at the level of WoS SCs,
- vi) WCS value of Chemistry individual research groups and panel members at the level of journals.

	A	B	C	D	E	F
1	Barycenter_ CHEM_WoS SCs	Barycenter_ CHEM_Journals	SAPV_ CHEM_WoS SCs	SAPV_ CHEM_Journals	WCS_ CHEM_WoS SCs	WCS_ CHEM_Journals
2	0.167	0.115	0.081	0.015	0.291	0.330
3	0.290	0.103	0.079	0.015	0.333	0.367
4	0.217	0.276	0.108	0.029	0.555	0.856
5	0.165	0.149	0.061	0.011	0.078	0.270
6	0.329	0.422	0.124	0.027	0.531	0.664
7	0.337	0.312	0.119	0.025	0.551	0.756
8	0.179	0.281	0.116	0.027	0.605	0.821
9	0.165	0.272	0.104	0.028	0.560	0.873
10	0.111	0.187	0.093	0.025	0.493	0.780
11	0.394	0.291	0.129	0.028	0.677	0.901
12	0.454	0.483	0.141	0.030	0.727	0.909
13	0.127	0.122	0.085	0.018	0.339	0.420
14	0.350	0.414	0.082	0.017	0.330	0.627
15	0.342	0.404	0.074	0.017	0.287	0.603
16	0.362	0.452	0.079	0.017	0.274	0.568
17	0.129	0.266	0.054	0.022	0.325	0.685
18	0.079	0.090	0.036	0.005	0.086	0.170
19	0.090	0.091	0.032	0.008	0.055	0.195
20	0.145	0.217	0.055	0.011	0.163	0.351

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

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

The heat map with hierarchical clustering (Figure 41) shows that correlations between the two levels of aggregation based on barycenter ($\rho = 0.81$), SAPV ($\rho = 0.83$) and WCS ($\rho = 0.90$) 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.

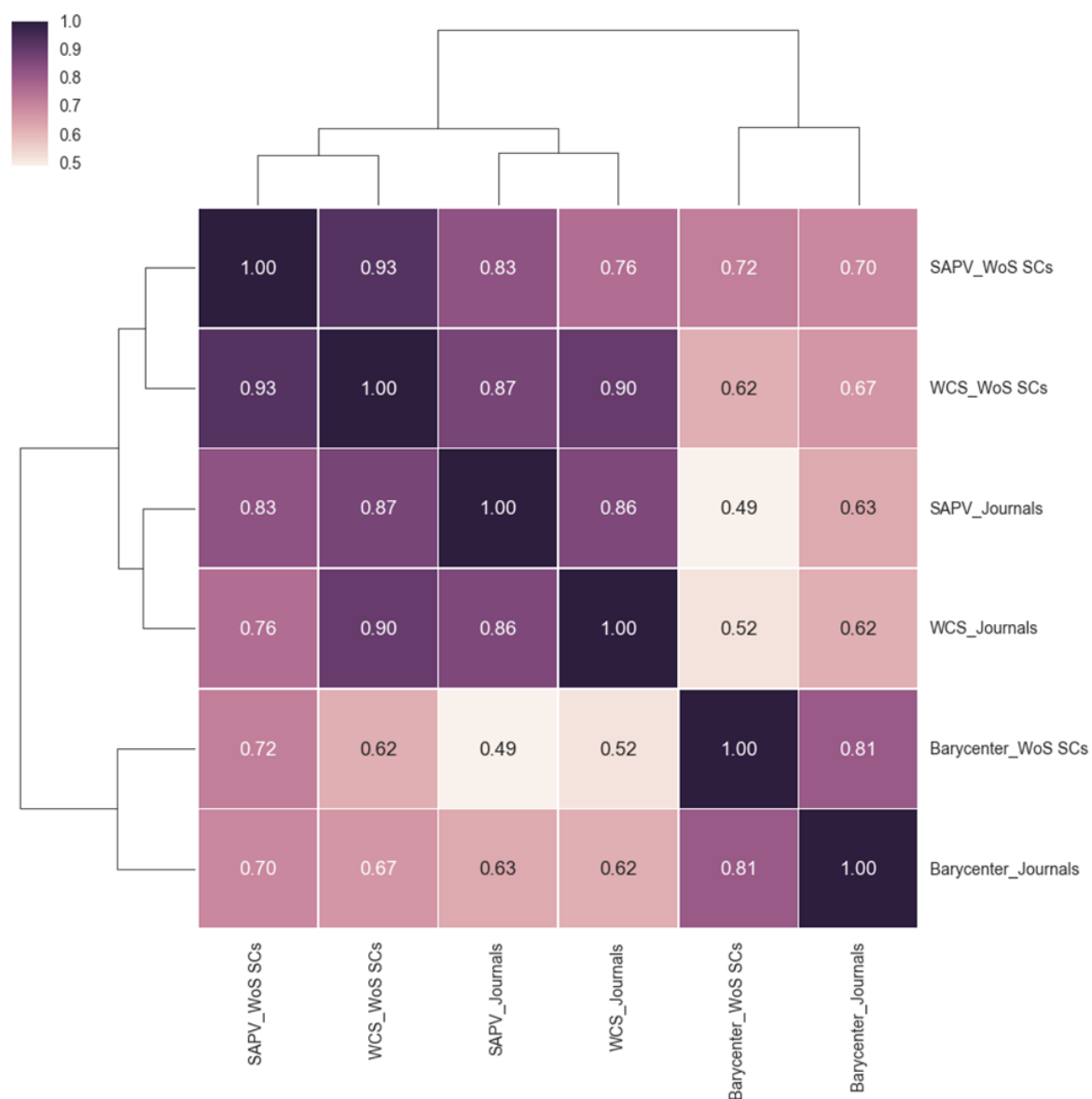


Figure 41. Heat map with hierarchical clustering based on correlation coefficient between six approaches in the Chemistry department

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*. Boca Raton, Fla: 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

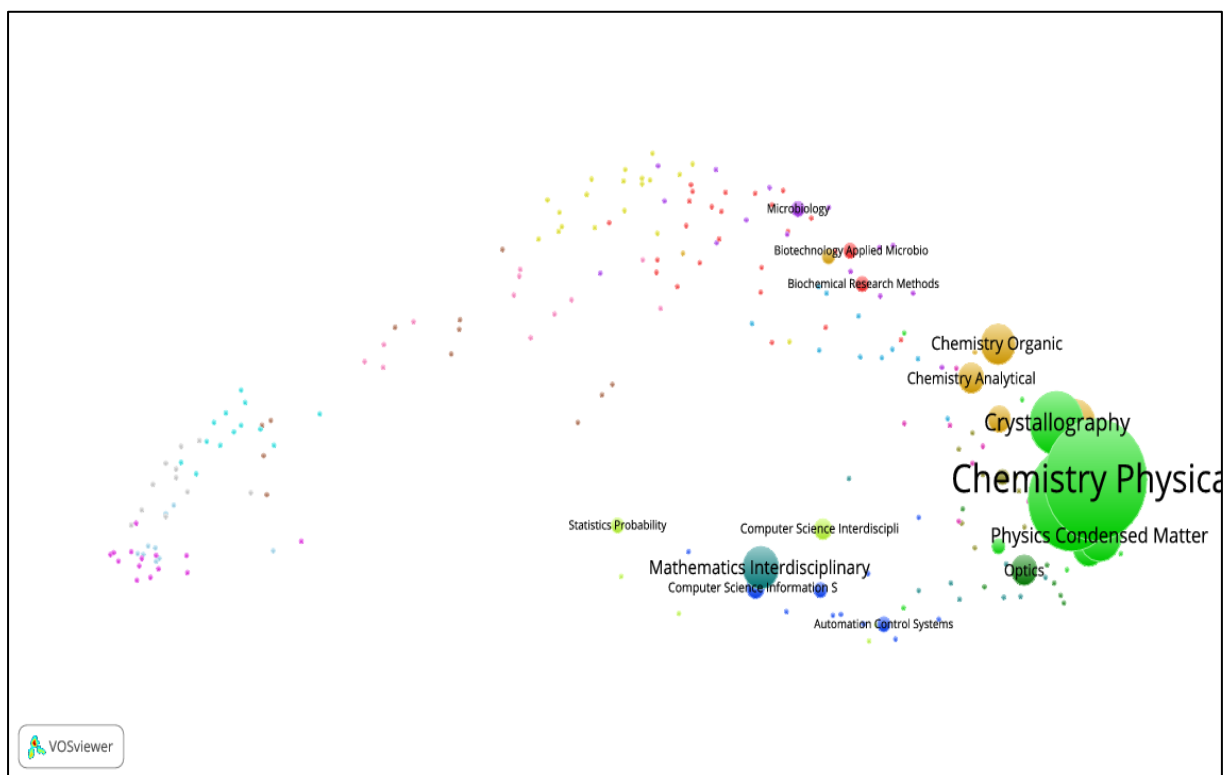


Figure 42. WoS SCs overlay map of CHEM-A research group's publications

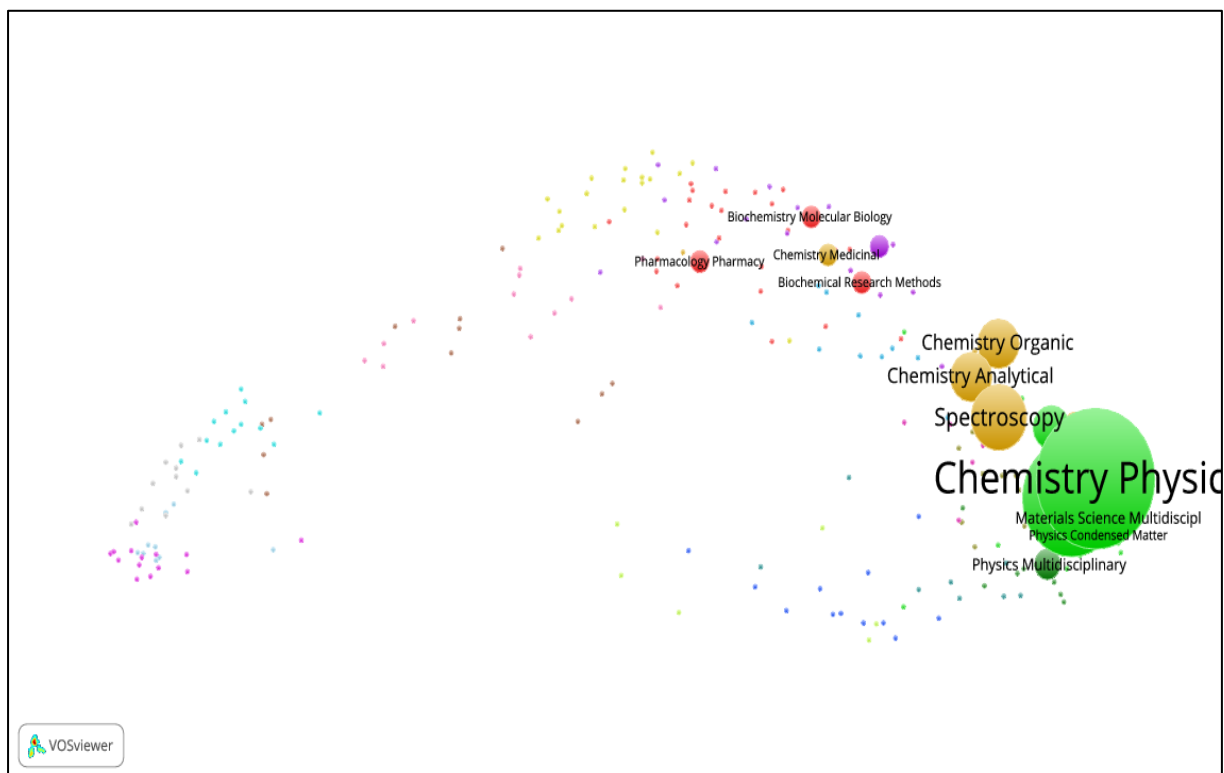


Figure 43. WoS SCs overlay map of CHEM-B research group's publications

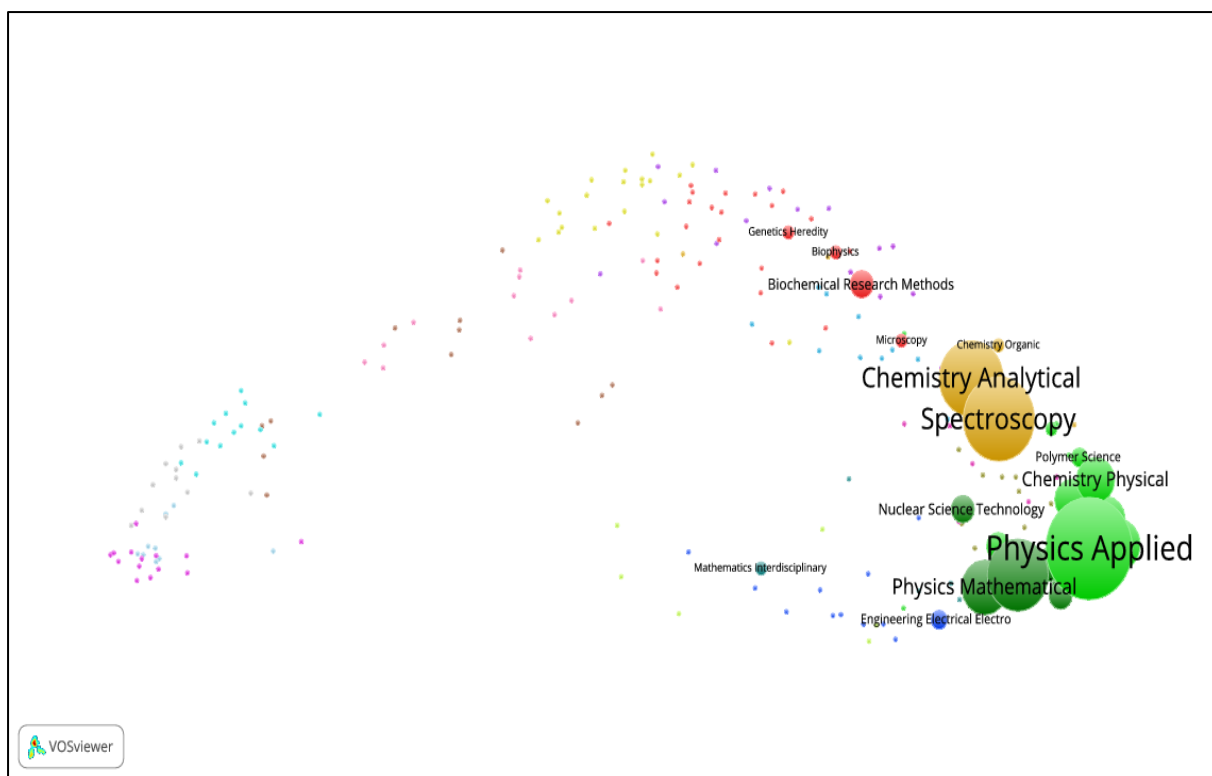


Figure 44. WoS SCs overlay map of CHEM-C research group's publications

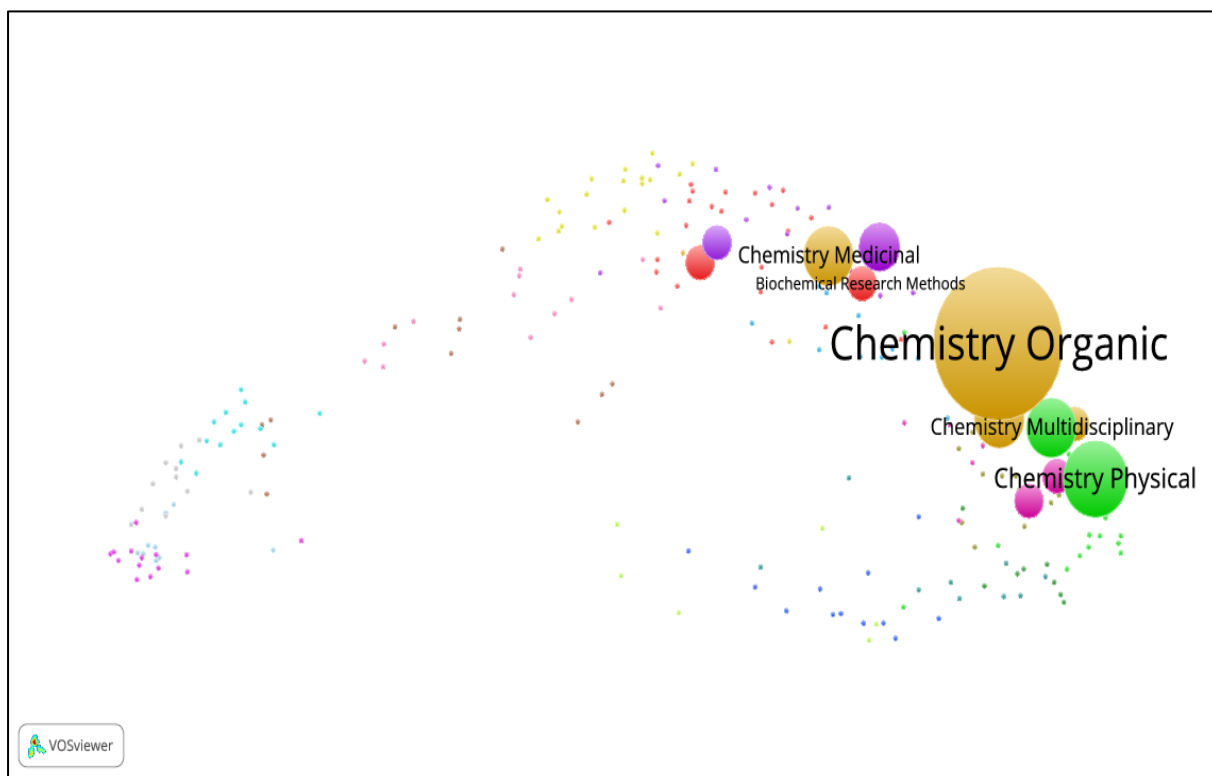


Figure 45. WoS SCs overlay map of CHEM-D research group's publications

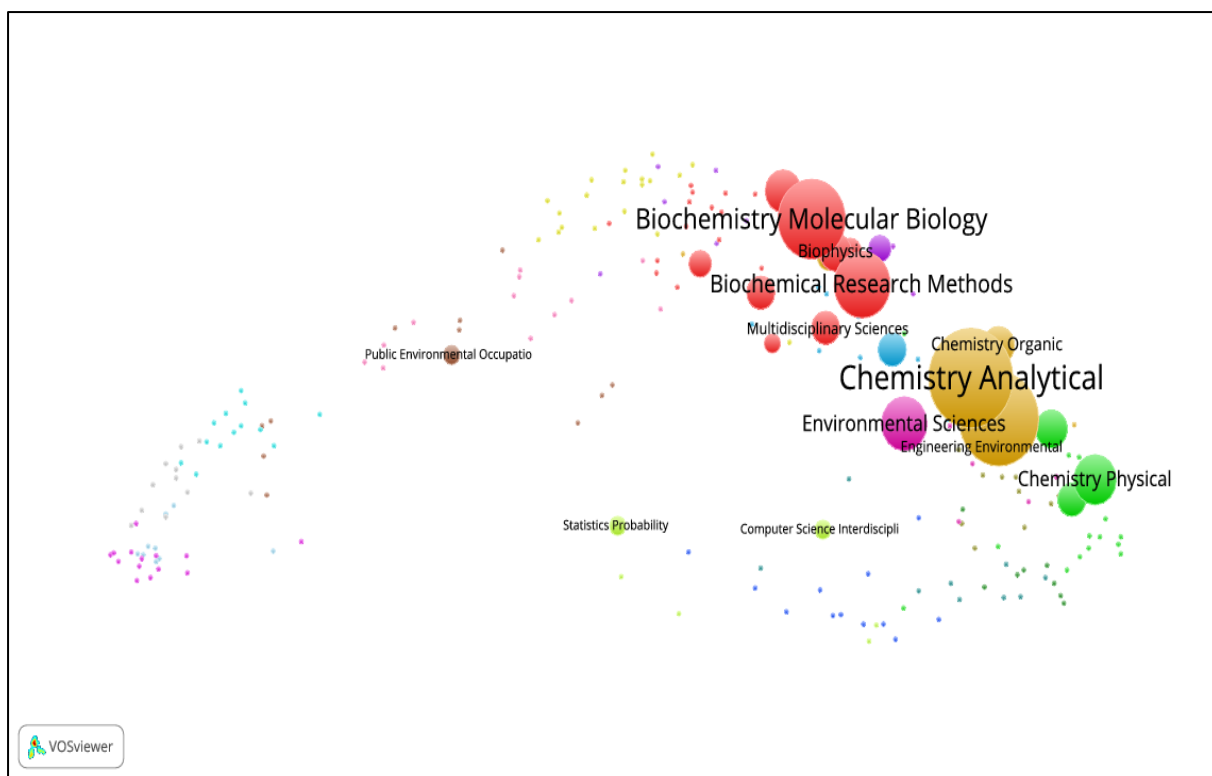


Figure 46. WoS SCs overlay map of CHEM-E research group's publications

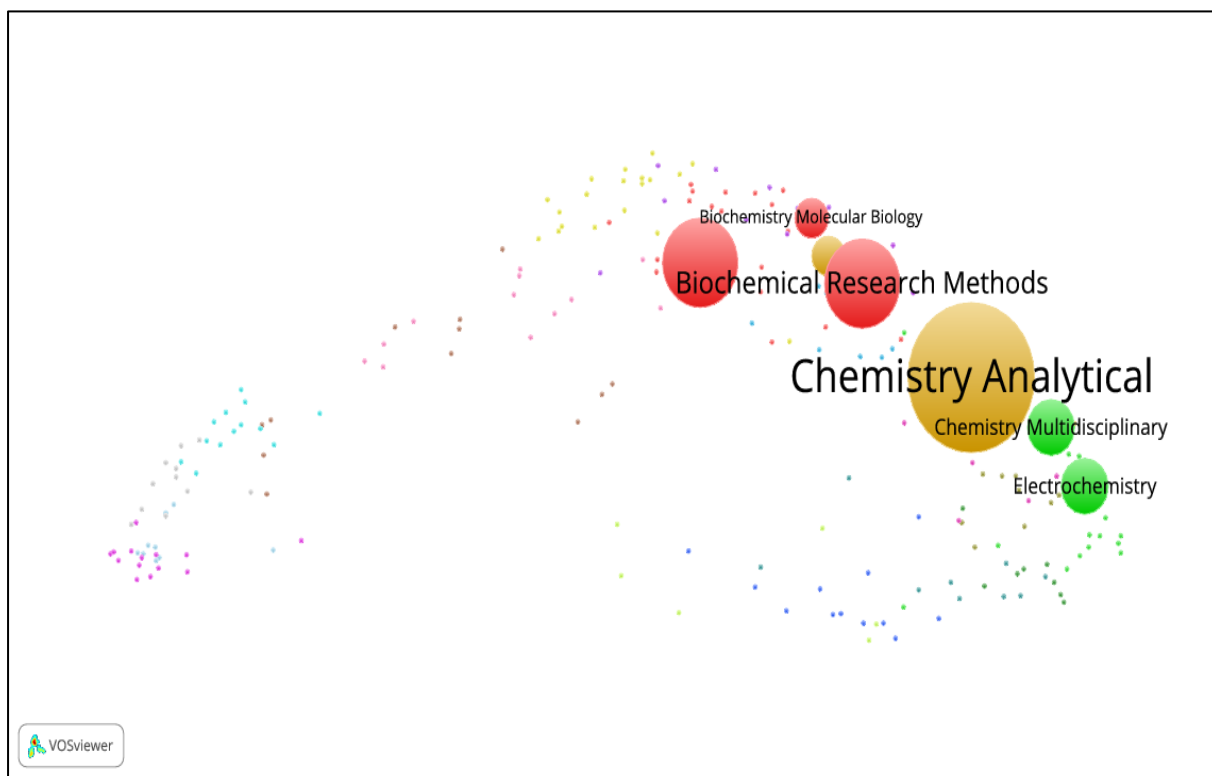


Figure 47. WoS SCs overlay map of CHEM-F research group's publications

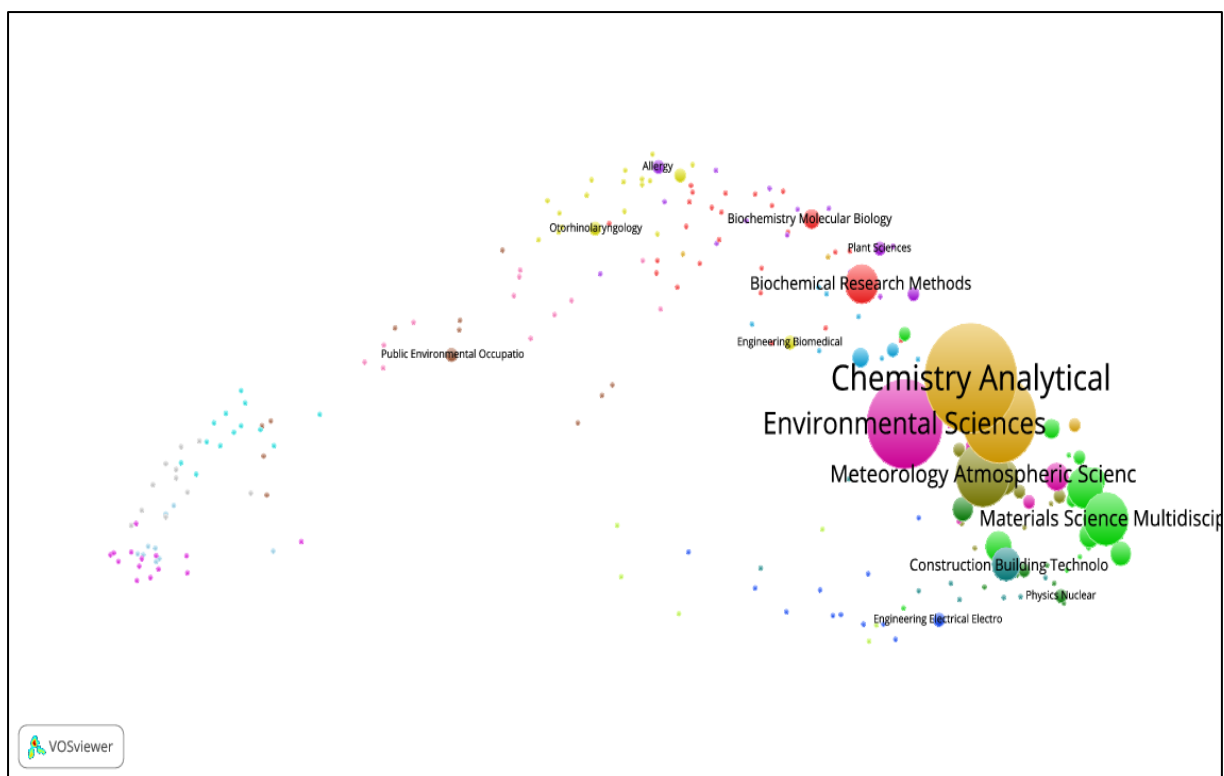


Figure 48. WoS SCs overlay map of CHEM-G research group's publications

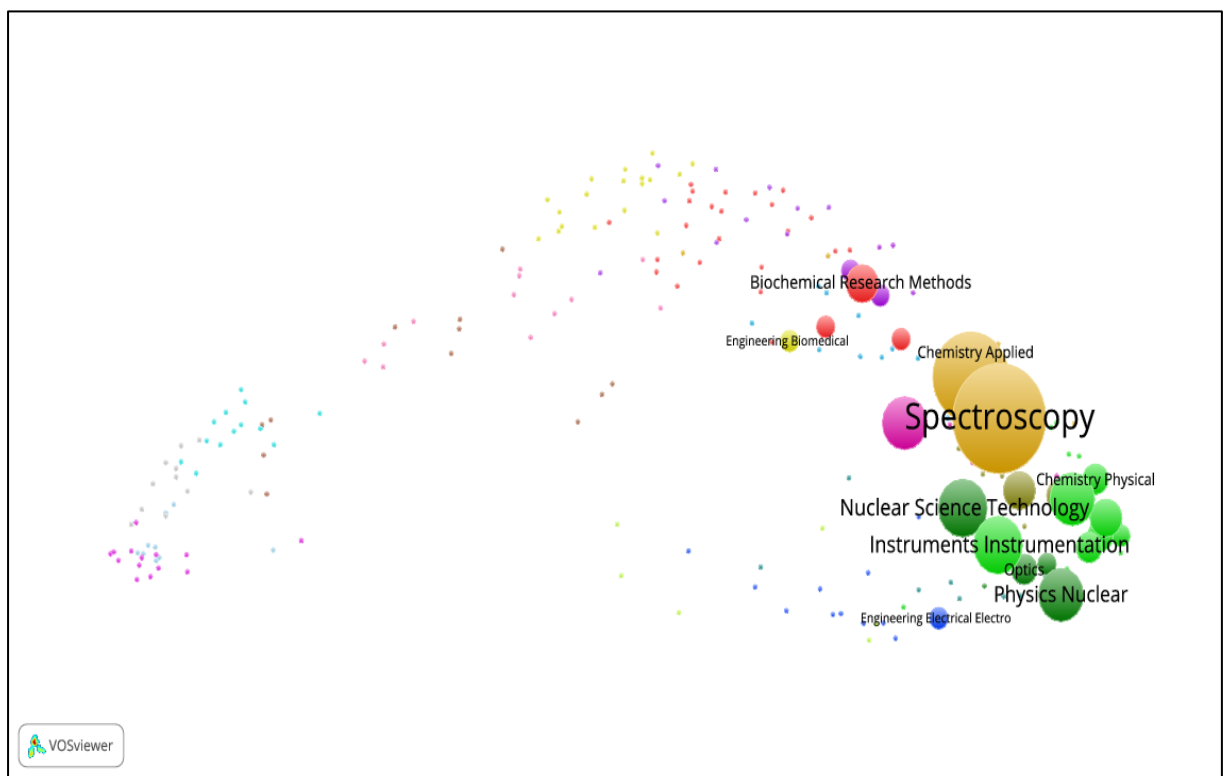


Figure 49. WoS SCs overlay map of CHEM-H research group's publications

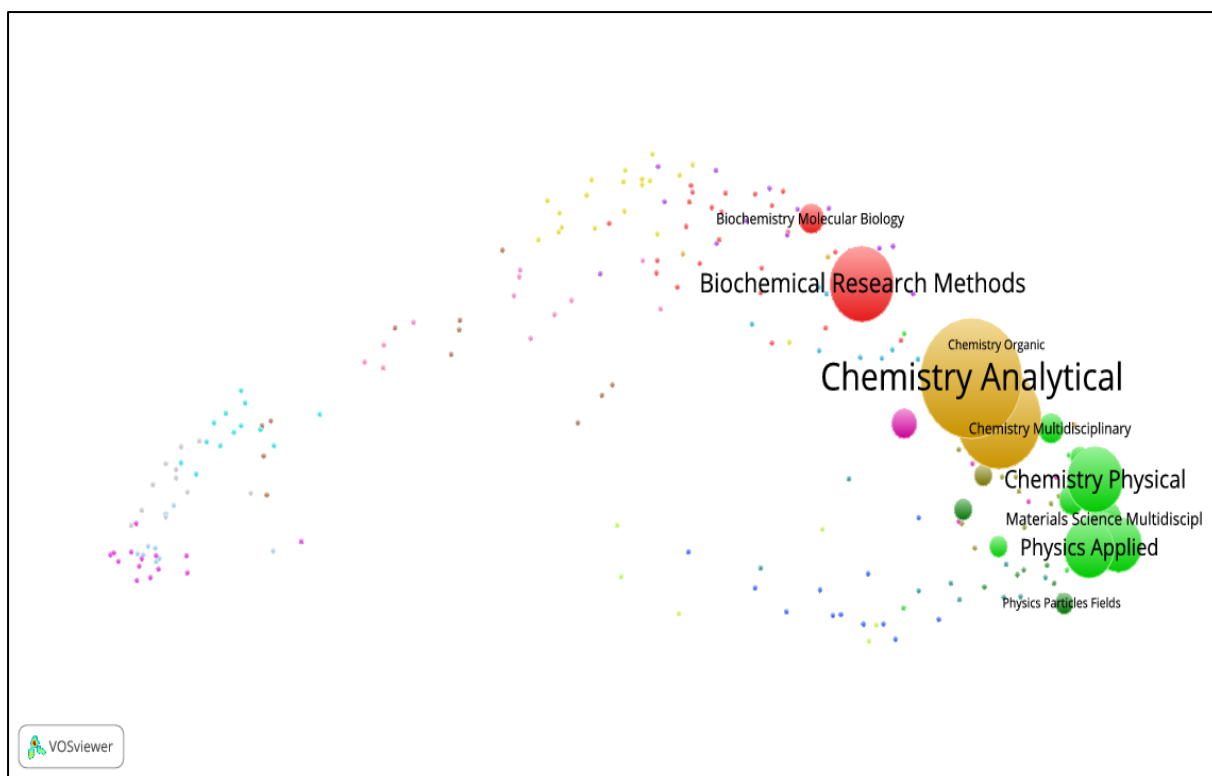


Figure 50. WoS SCs overlay map of CHEM-I research group's publications

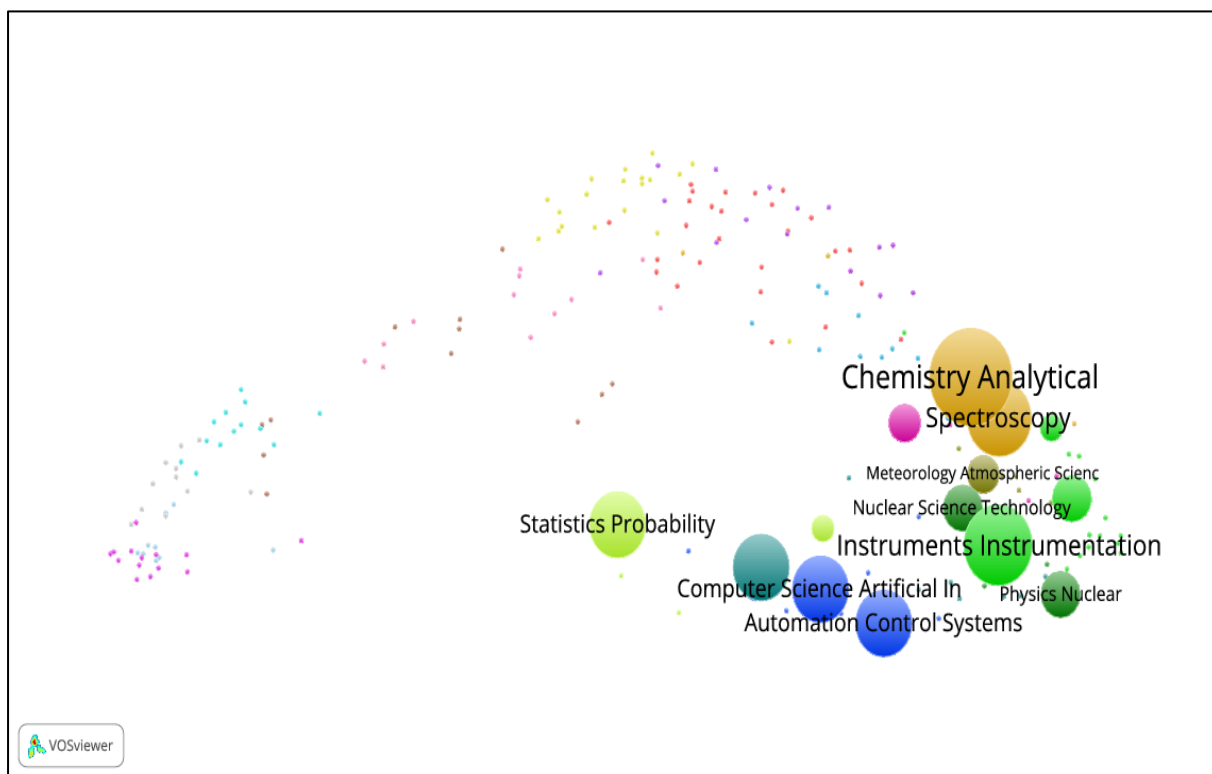


Figure 51. WoS SCs overlay map of CHEM-J research group's publications

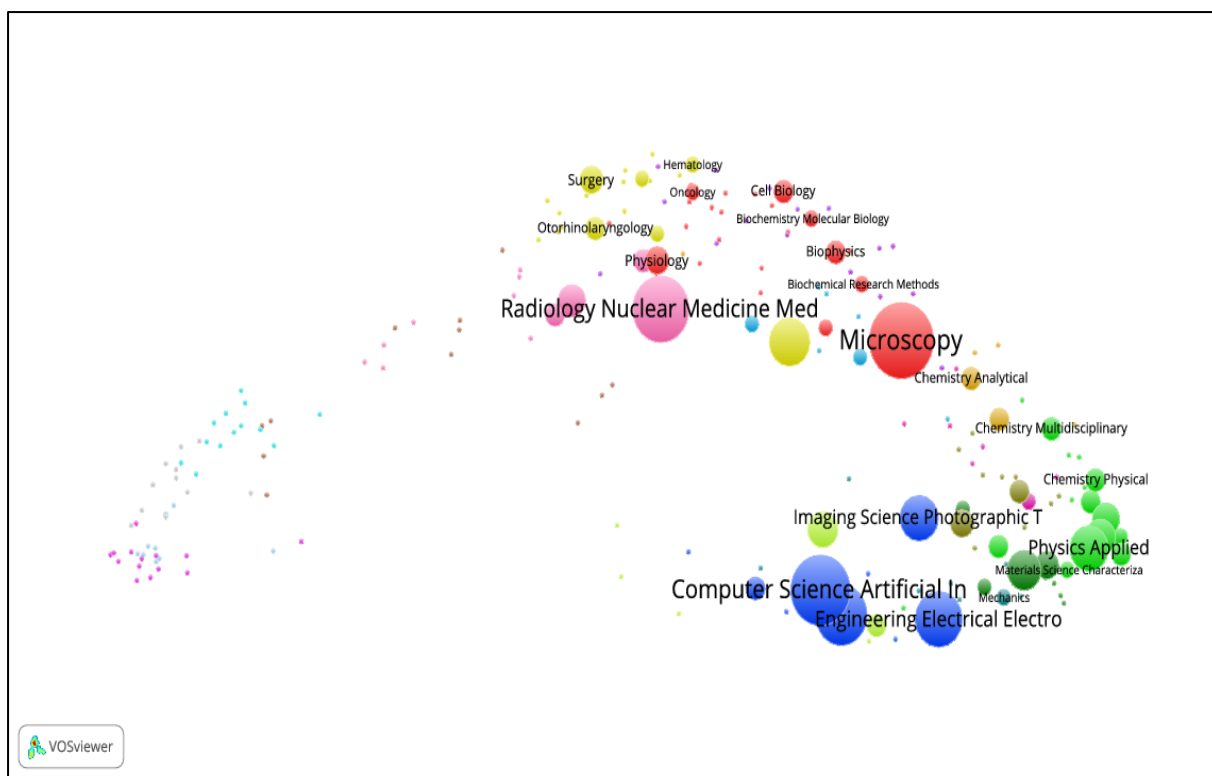


Figure 52. WoS SCs overlay map of CHEM-K research group's publications

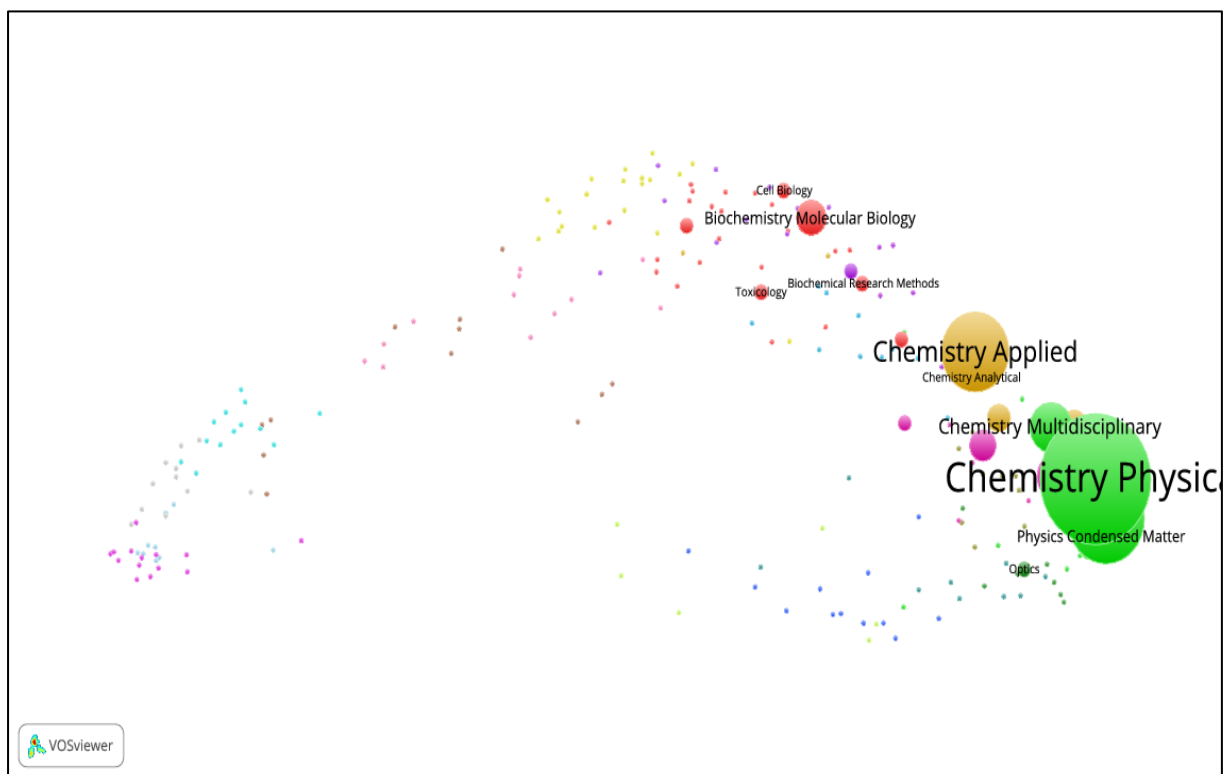


Figure 53. WoS SCs overlay map of CHEM-L research group's publications

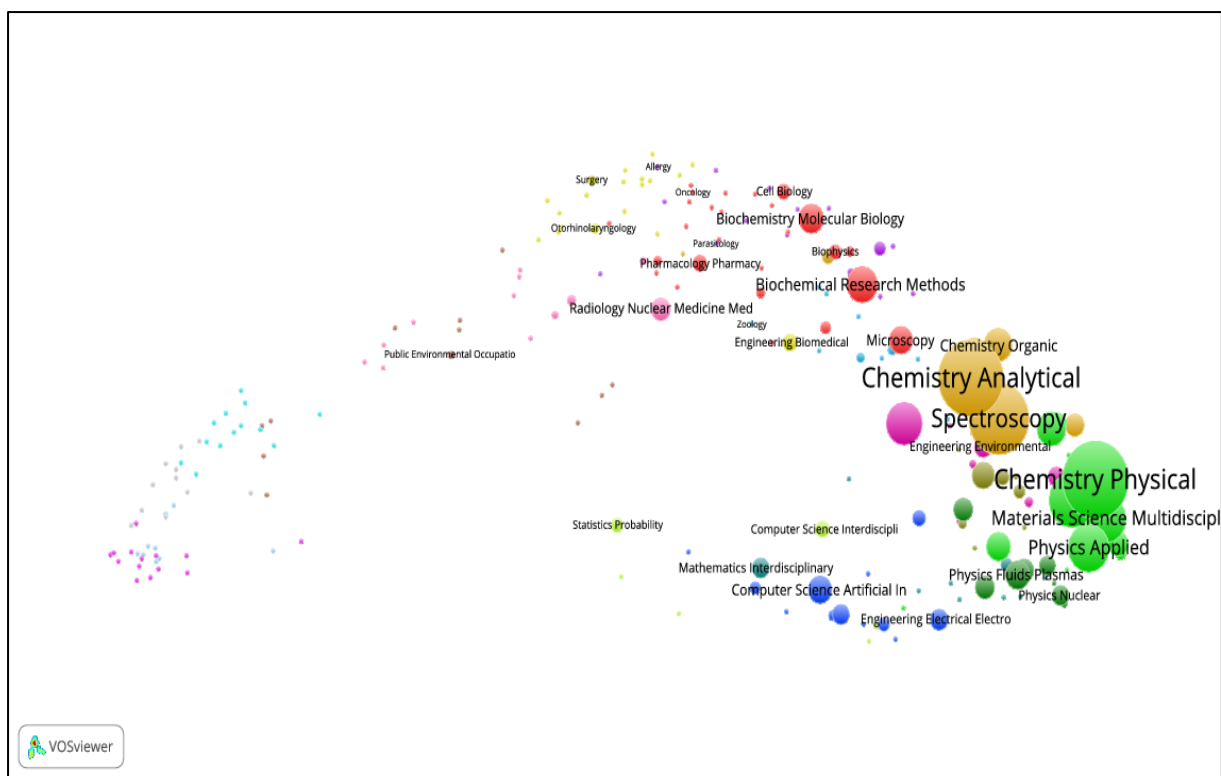


Figure 54. WoS SCs overlay map of Chemistry groups' publications

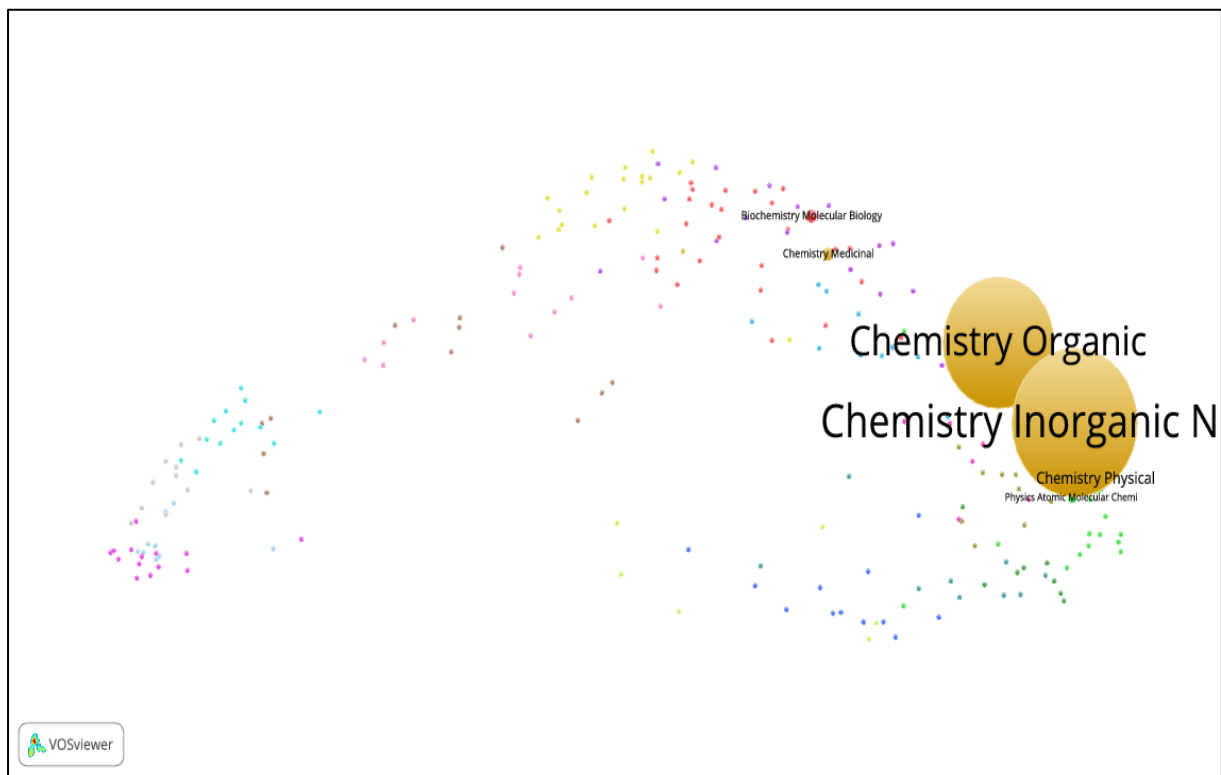


Figure 55. WoS SCs overlay map of PM1's publications

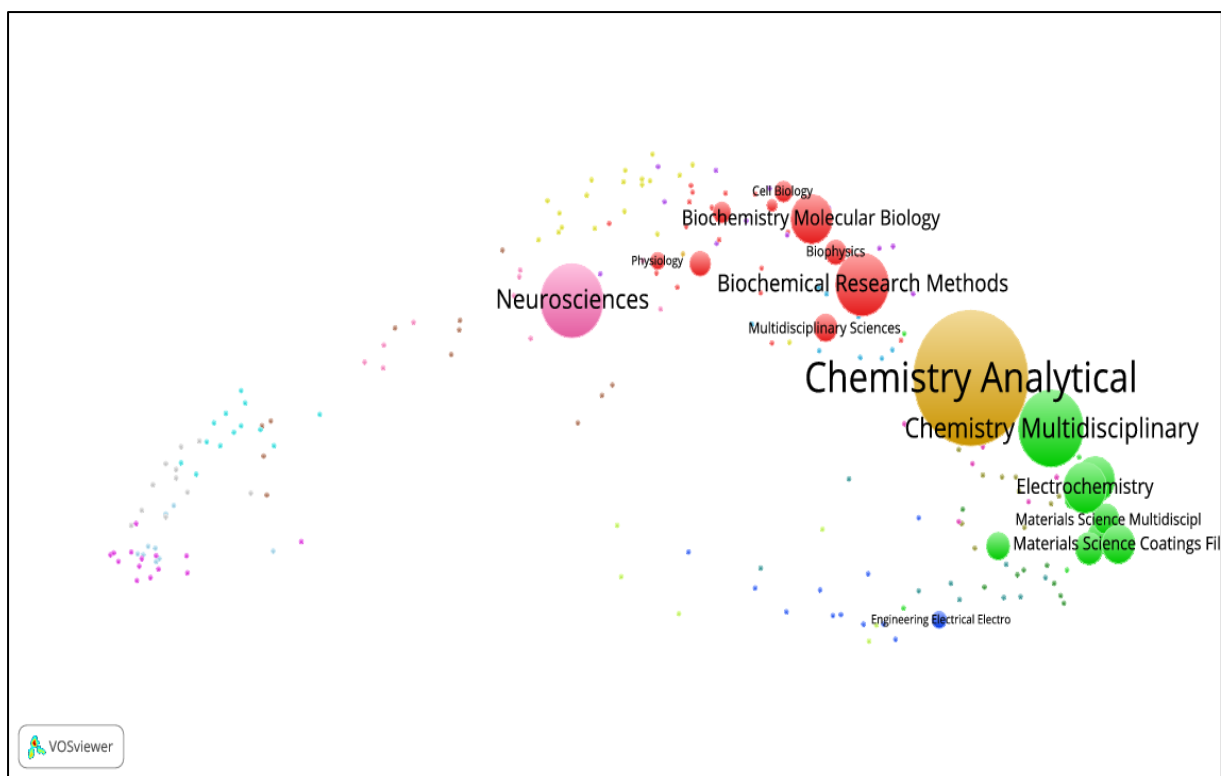


Figure 56. WoS SCs overlay map of PM2's publications

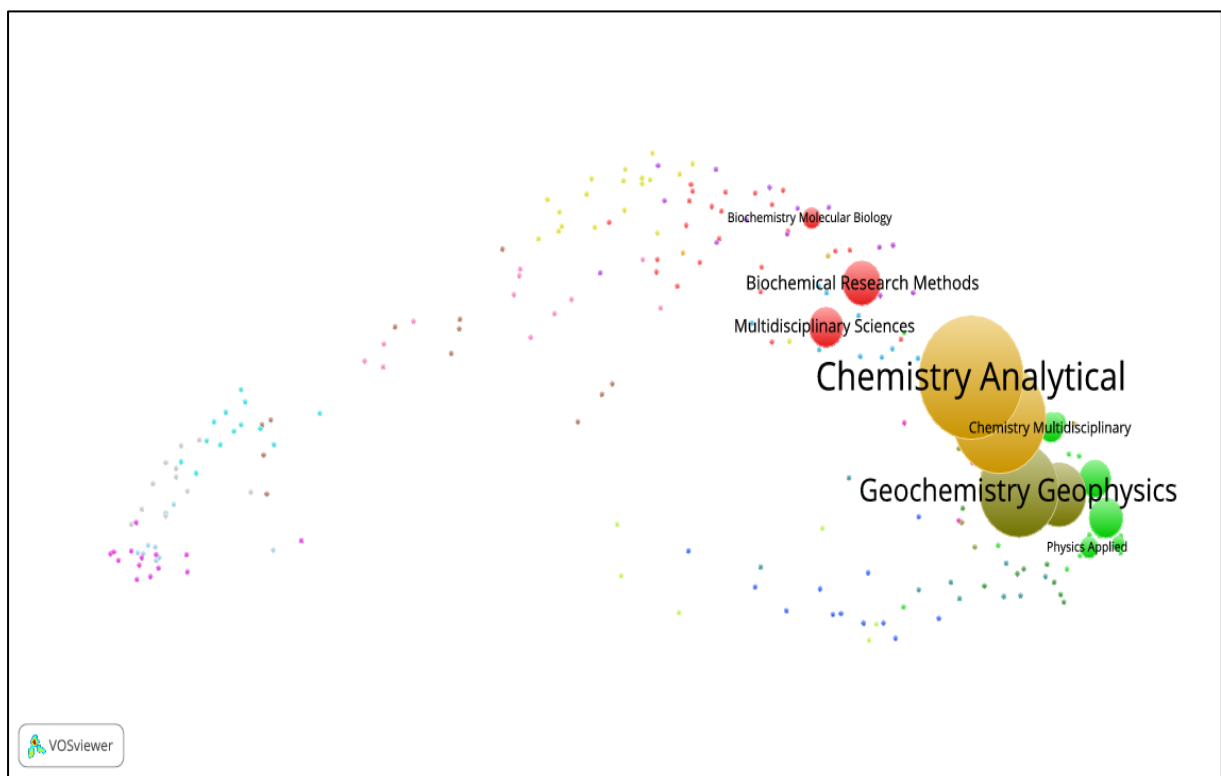


Figure 57. WoS SCs overlay map of PM3's publications

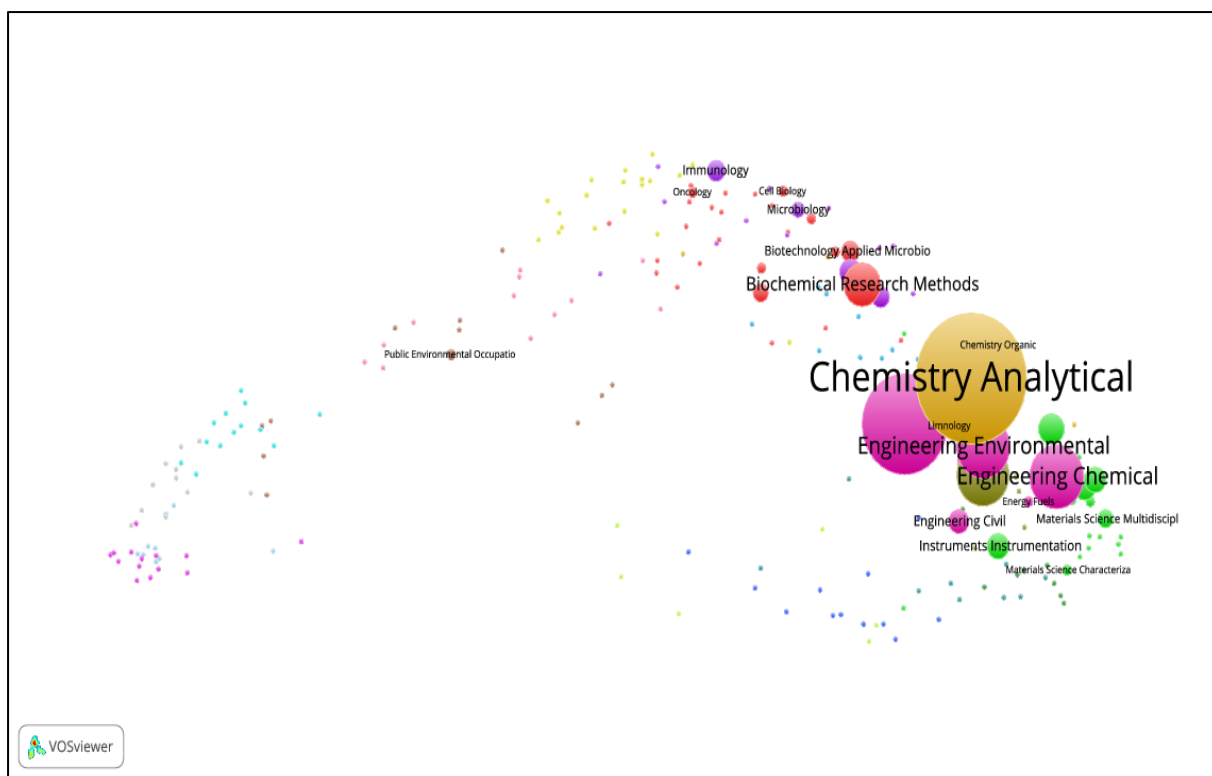


Figure 58. WoS SCs overlay map of PM4's publications

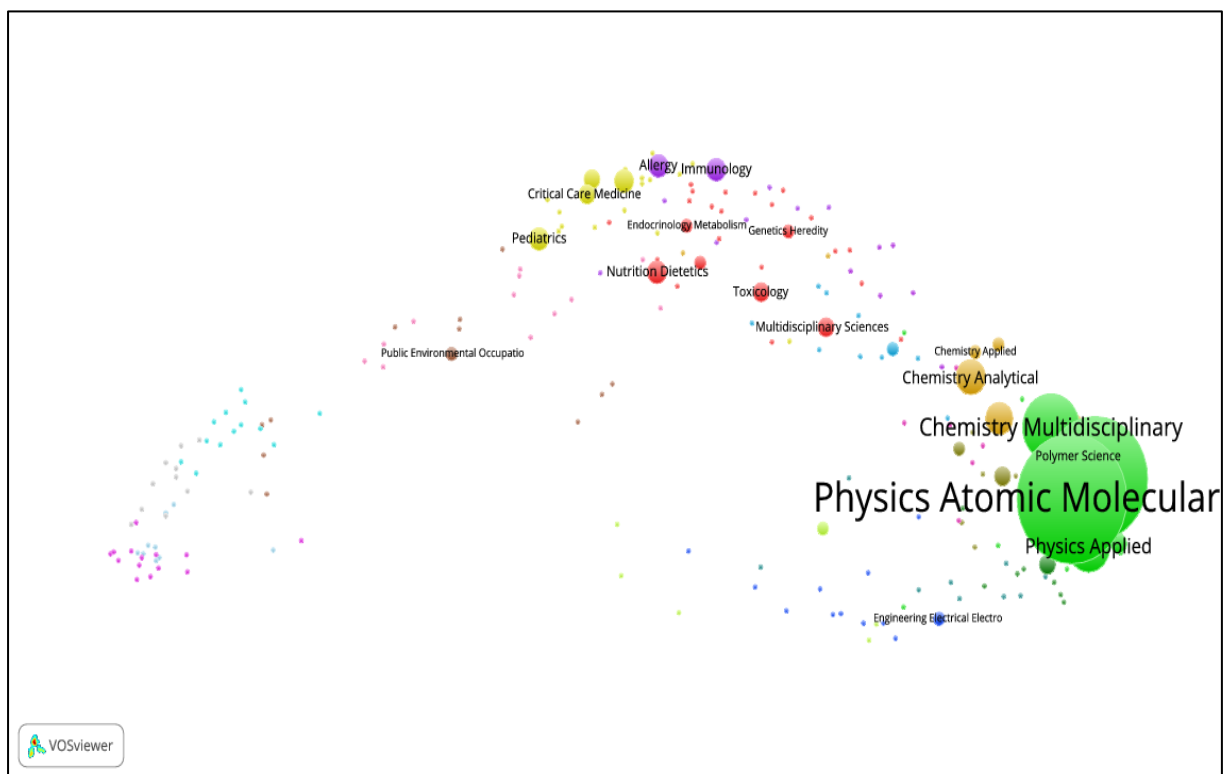


Figure 59. WoS SCs overlay map of PM5's publications

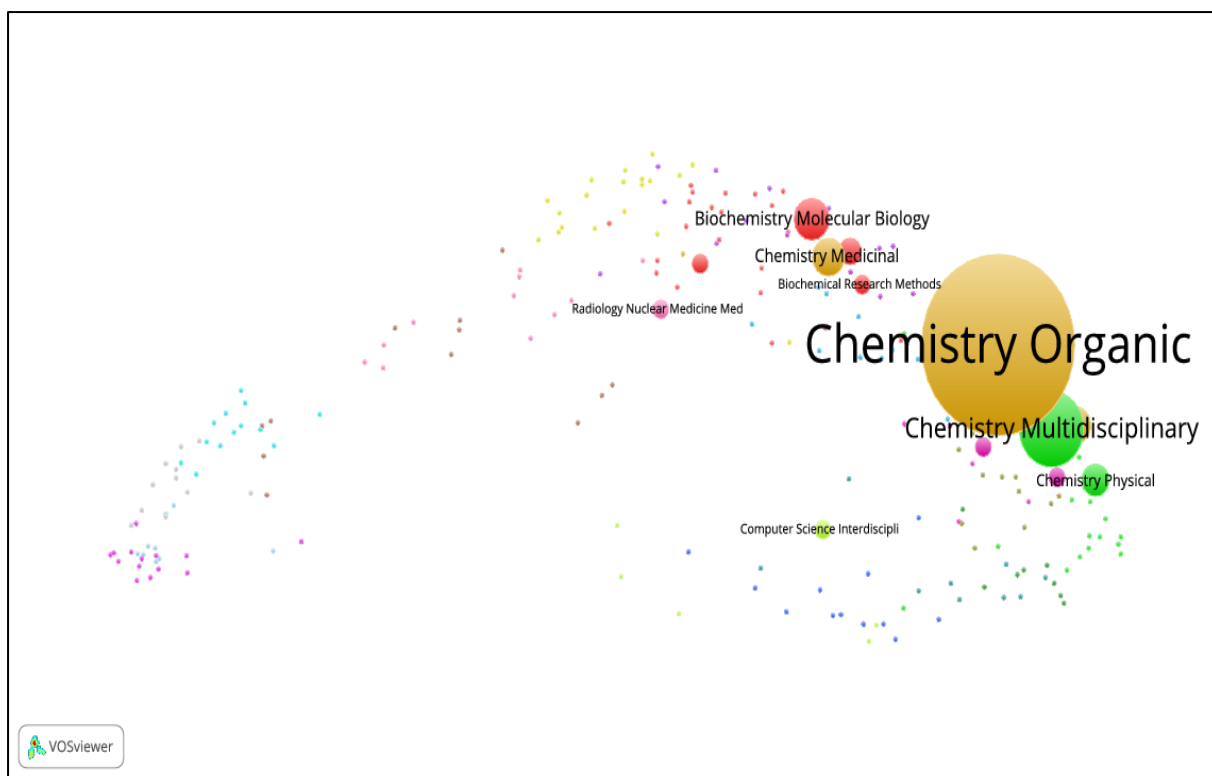


Figure 60. WoS SCs overlay map of PM6's publications

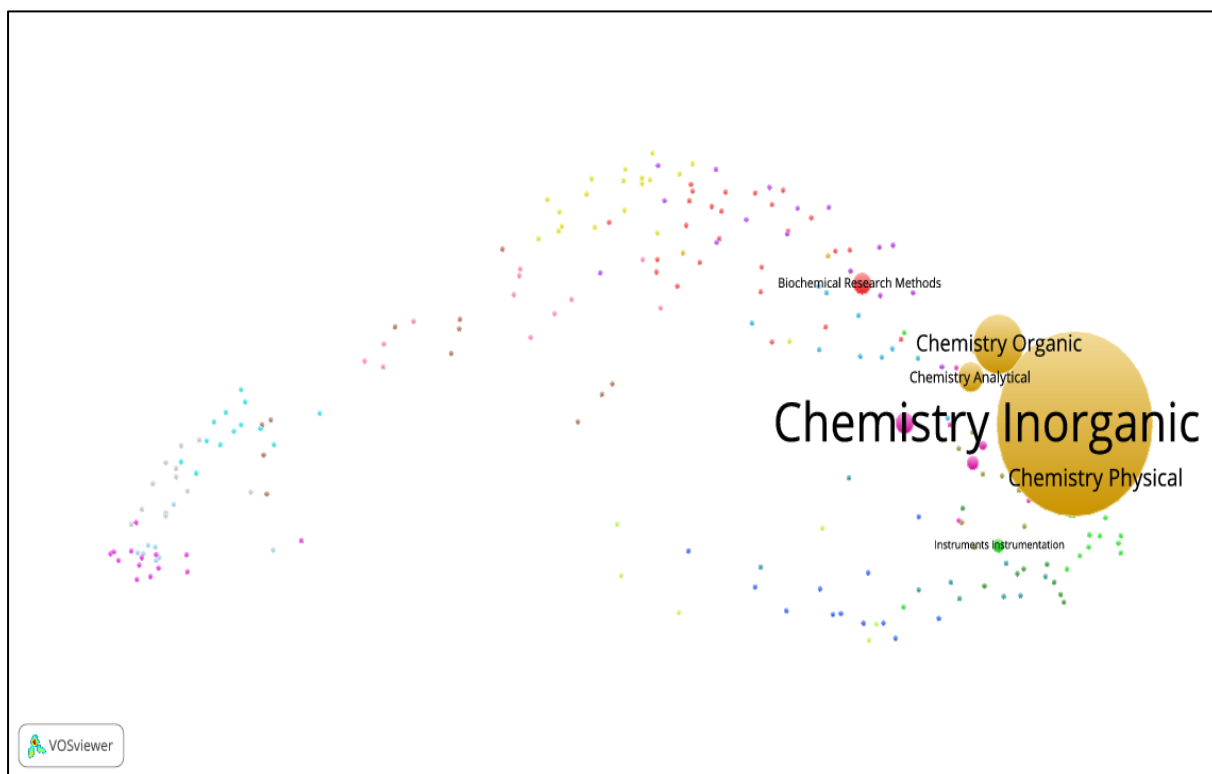


Figure 61. WoS SCs overlay map of PM7's publications

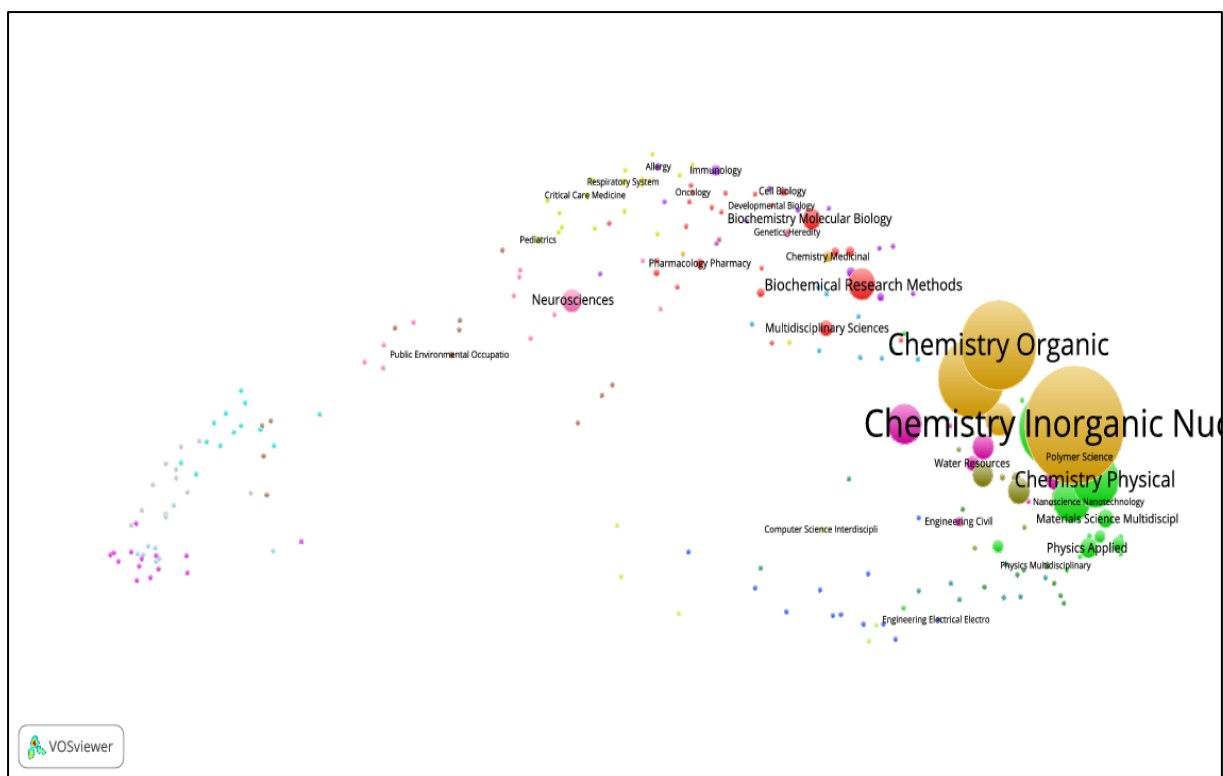


Figure 62. WoS SCs overlay map of panel's publications

Appendix B

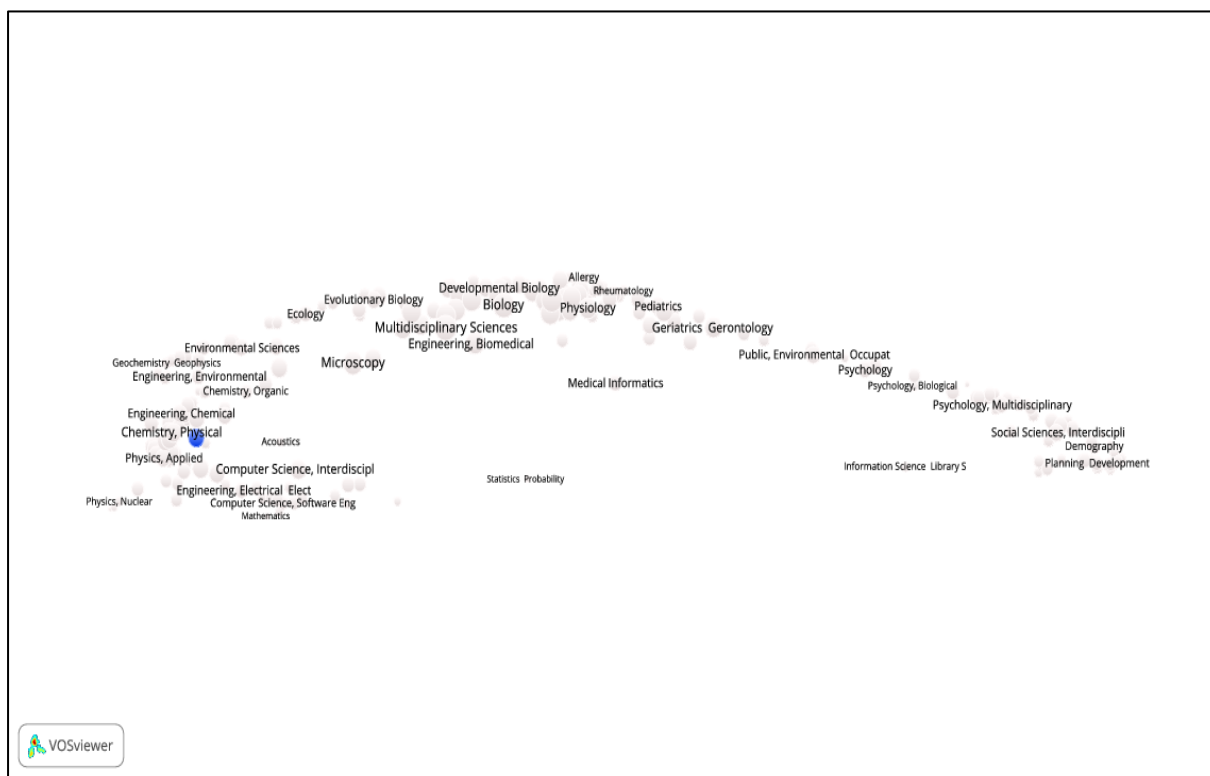


Figure 63. SAPV of the CHEM-A research group's publications in WoS SCs similarity matrix

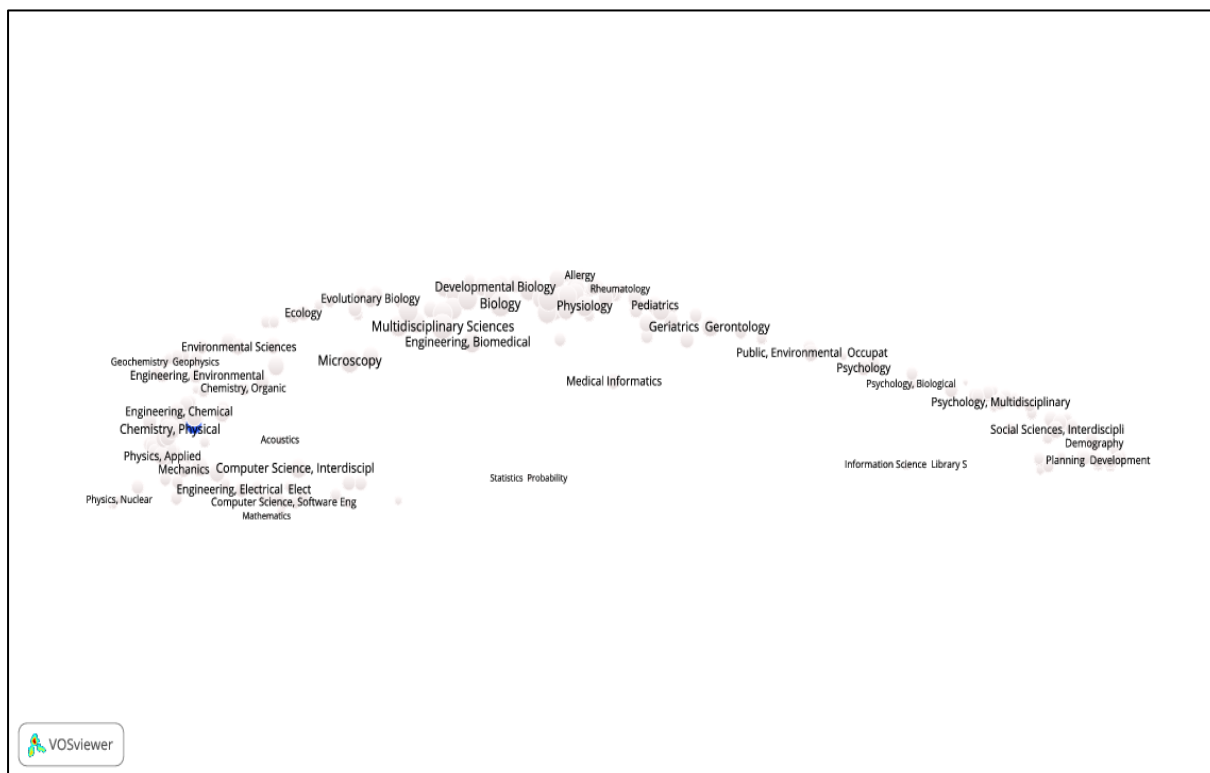
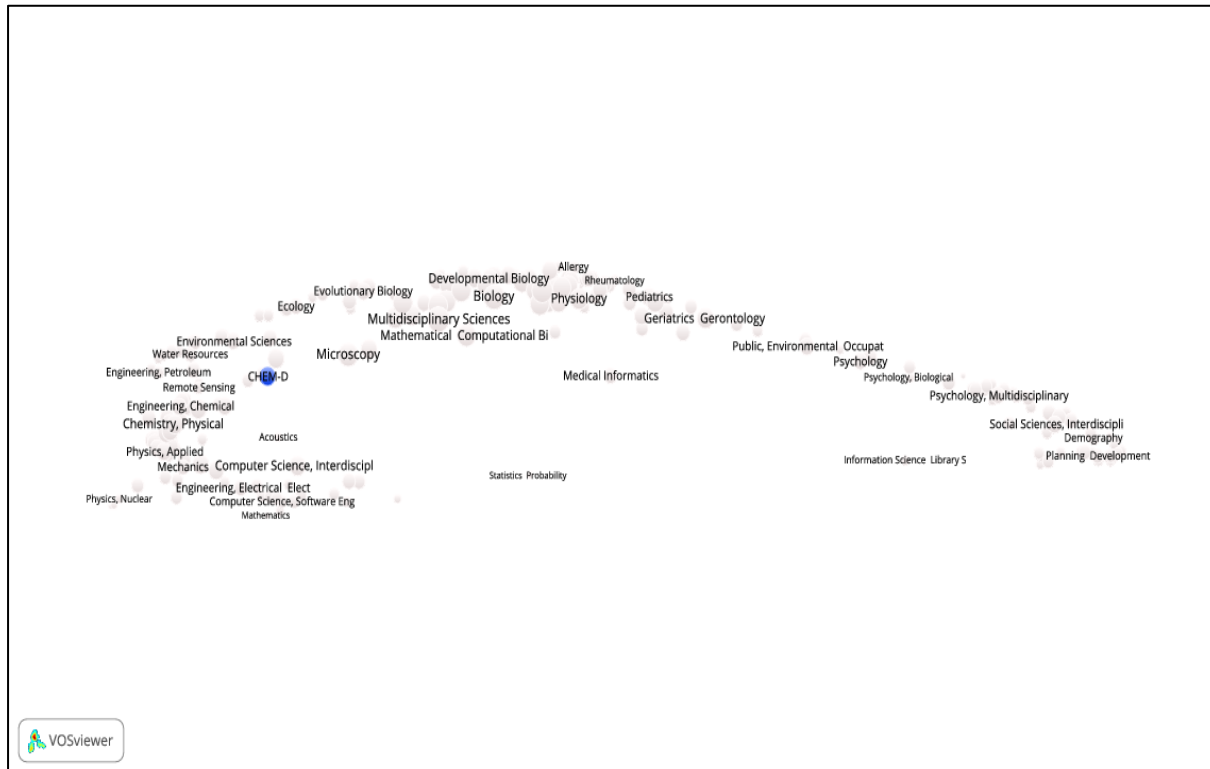
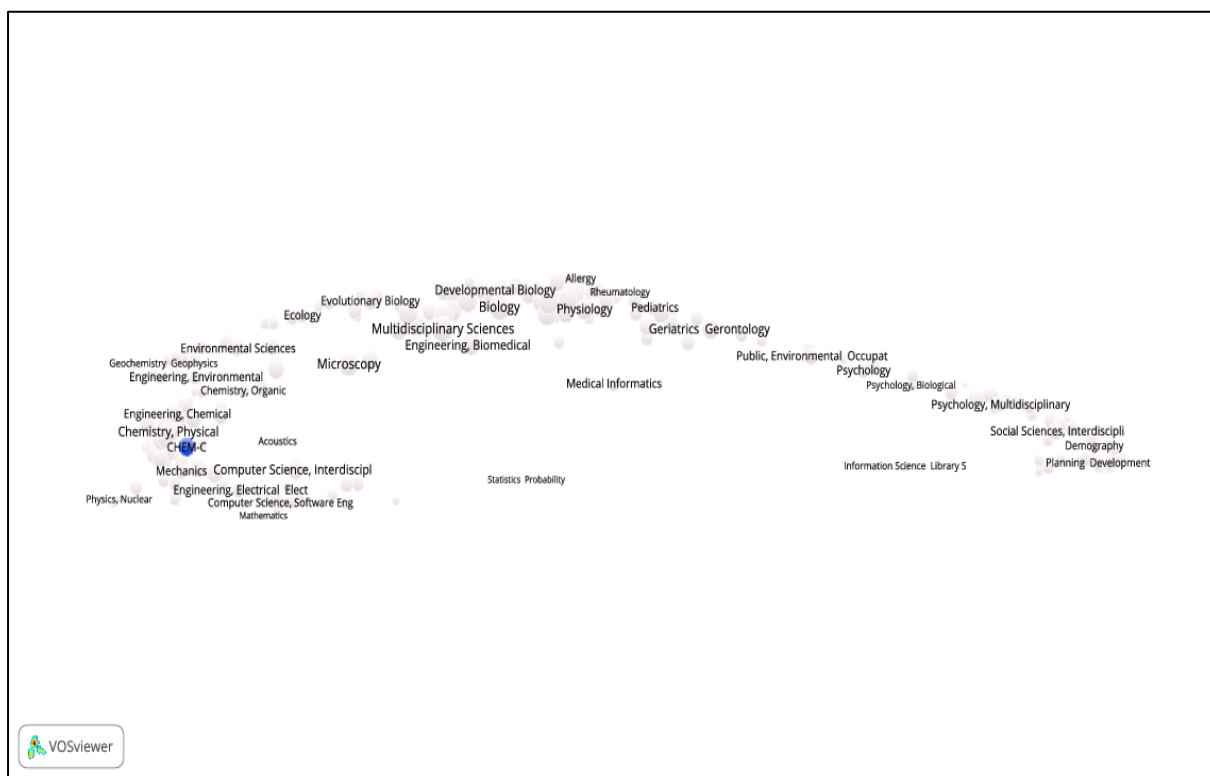


Figure 64. SAPV of the CHEM-B research group's publications in WoS SCs similarity matrix



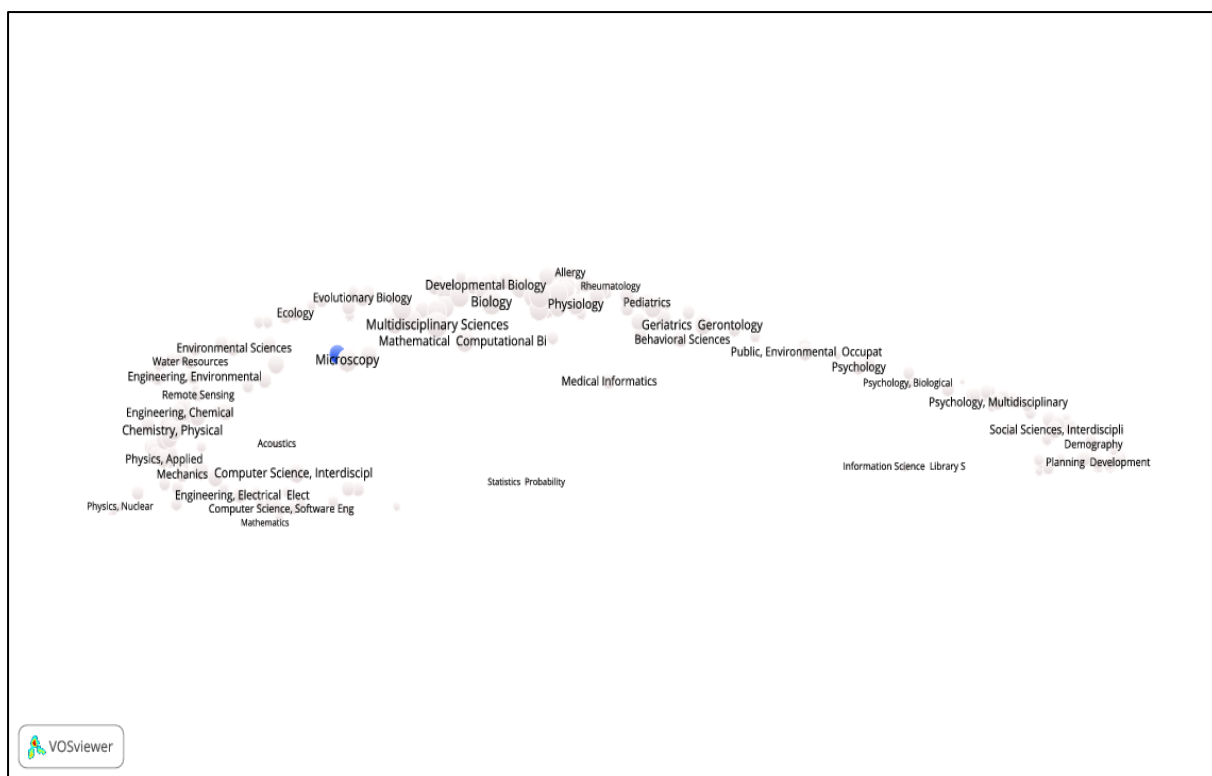


Figure 67. SAPV of the CHEM-E research group's publications in WoS SCs similarity matrix

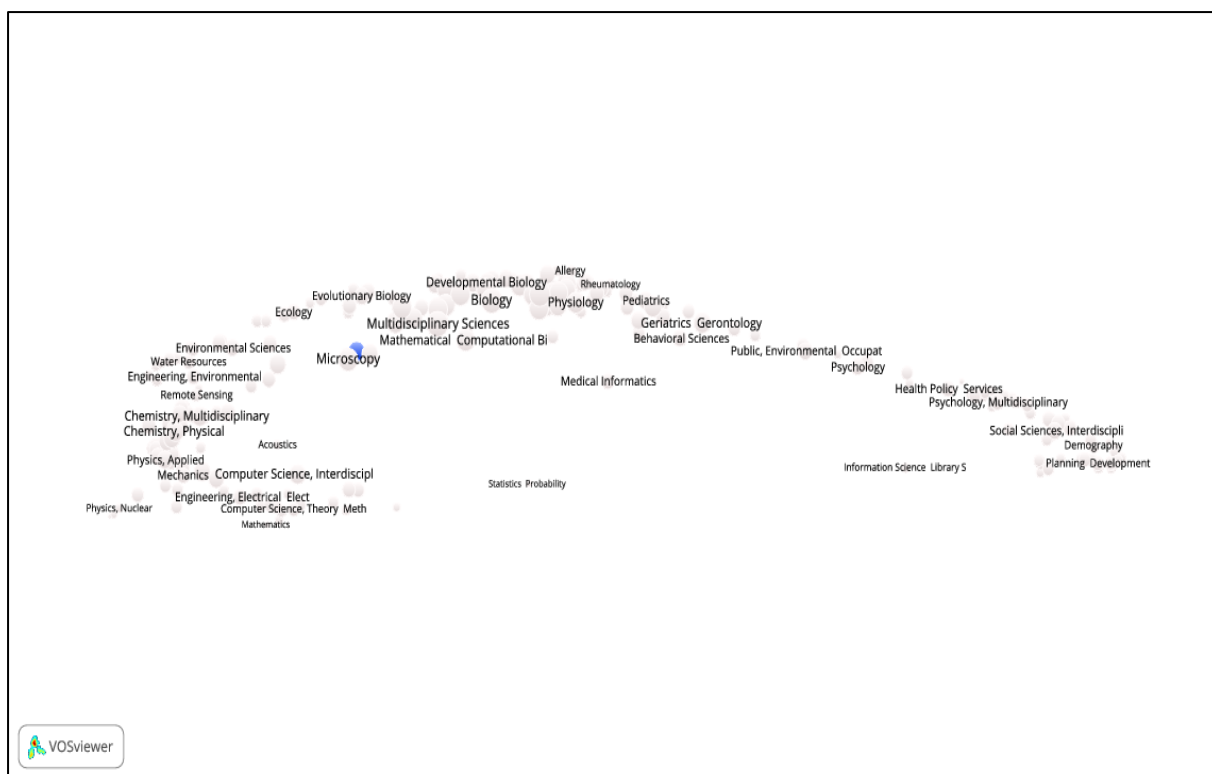


Figure 68. SAPV of the CHEM-F research group's publications in WoS SCs similarity matrix

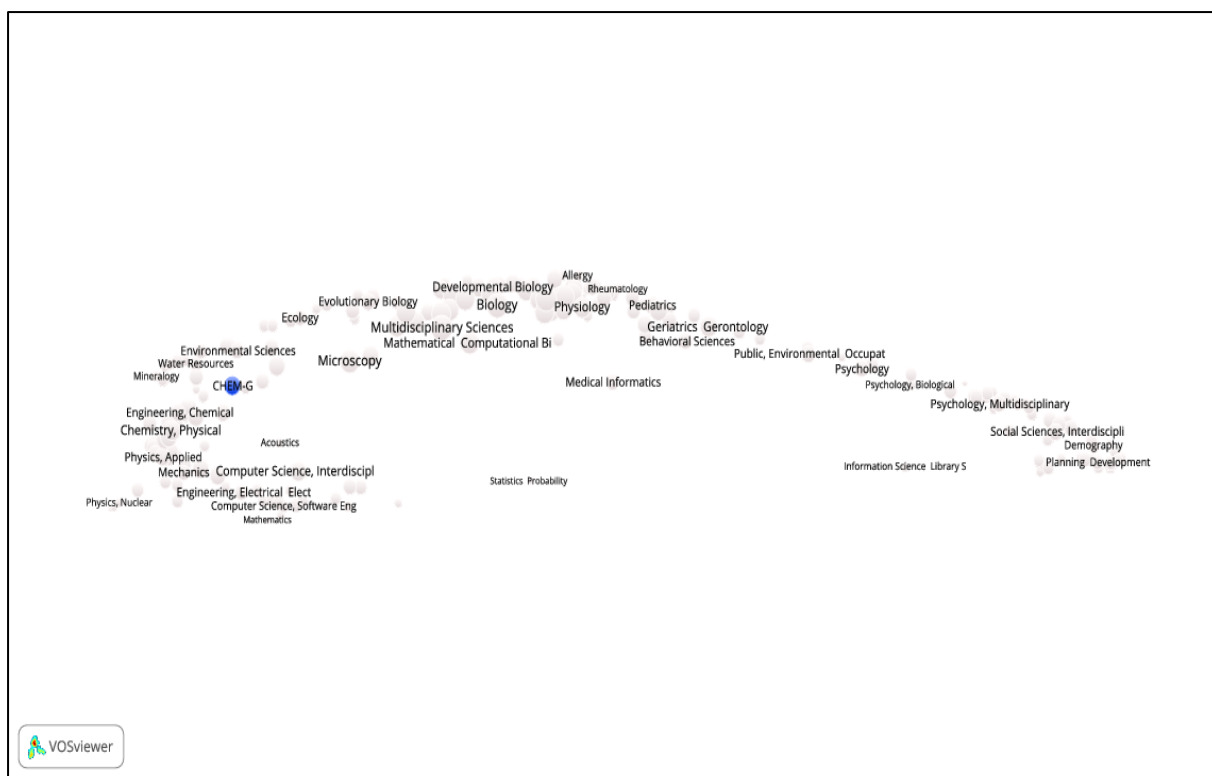


Figure 69. SAPV of the CHEM-G research group's publications in WoS SCs similarity matrix

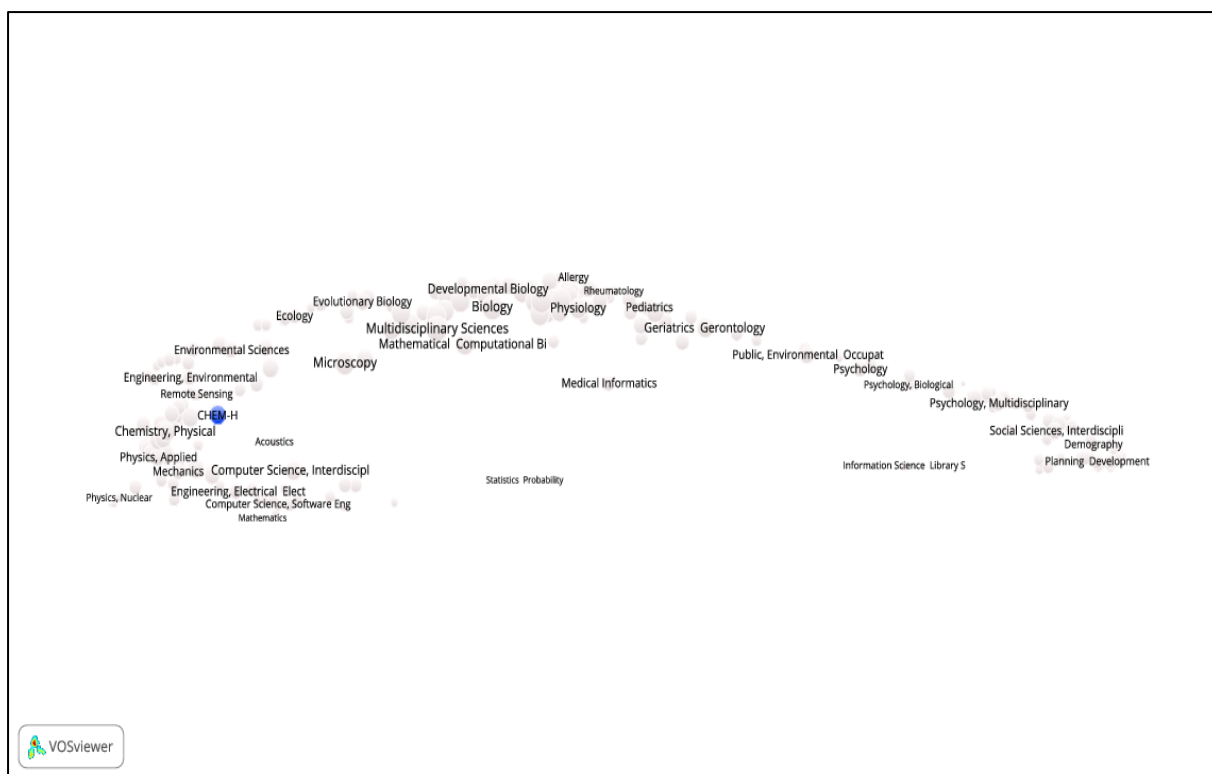
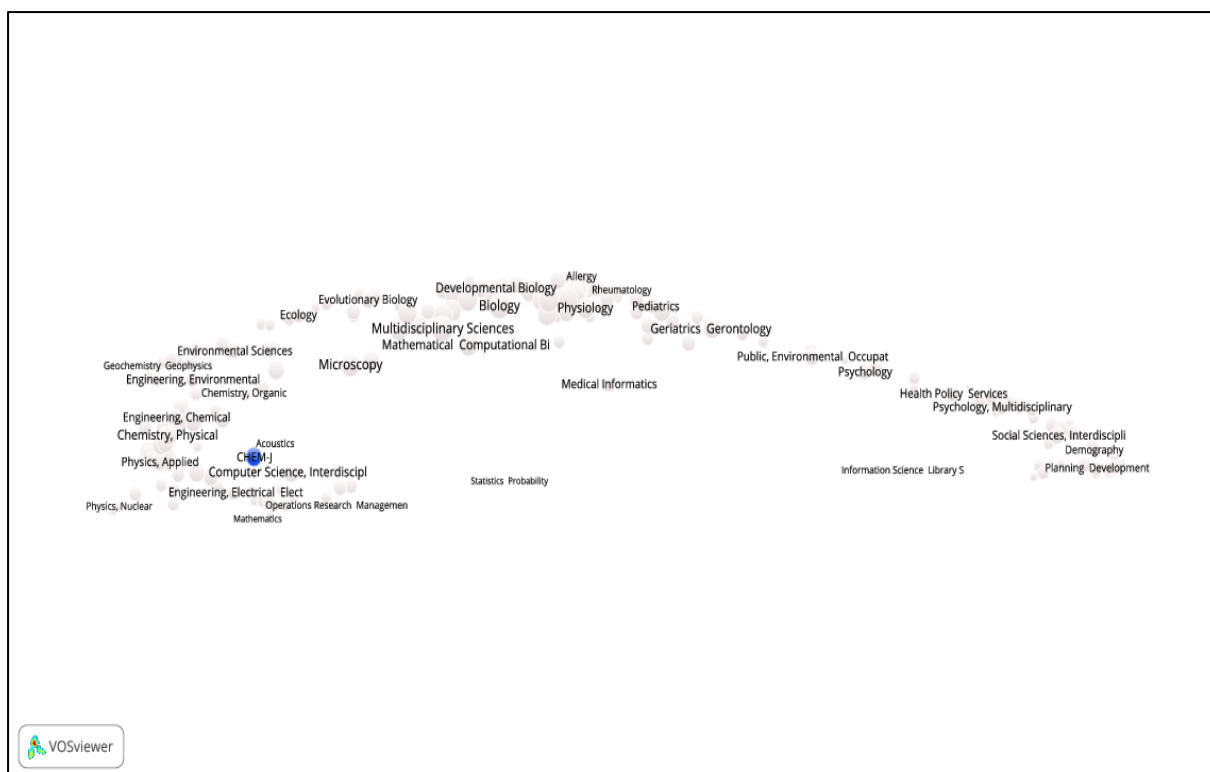
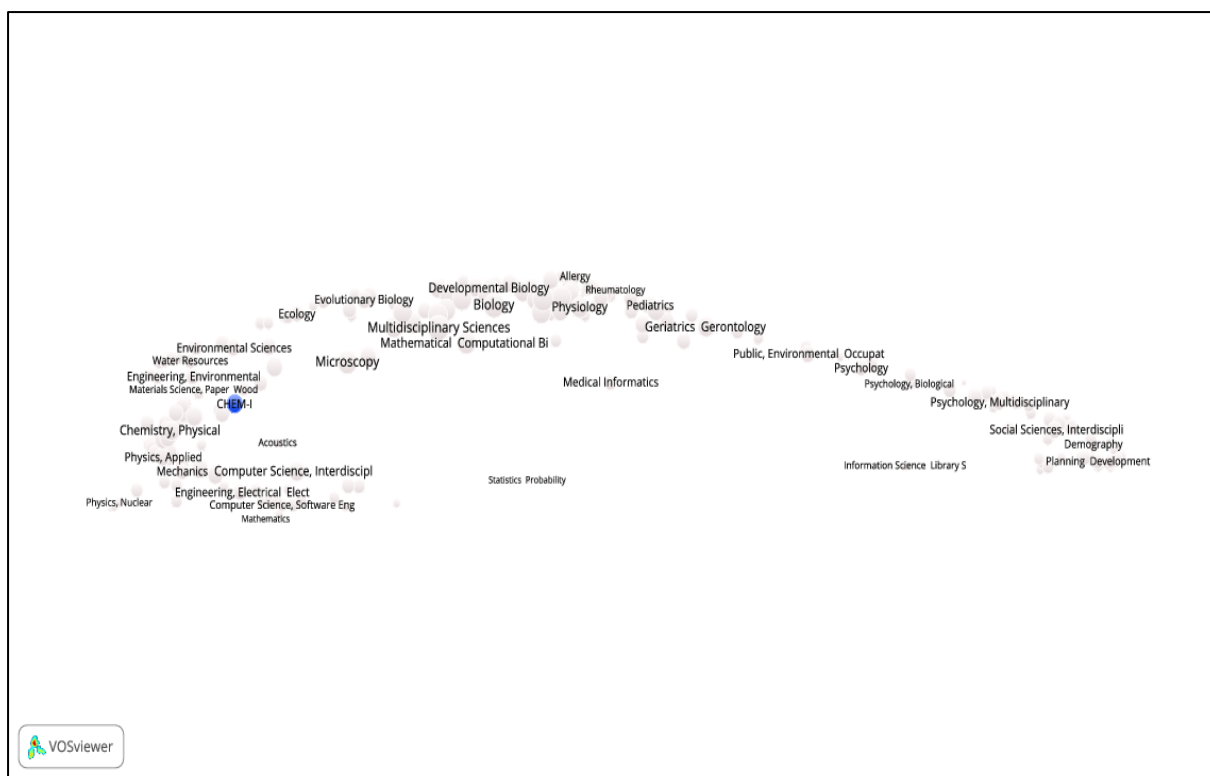


Figure 70. SAPV of the CHEM-H research group's publications in WoS SCs similarity matrix



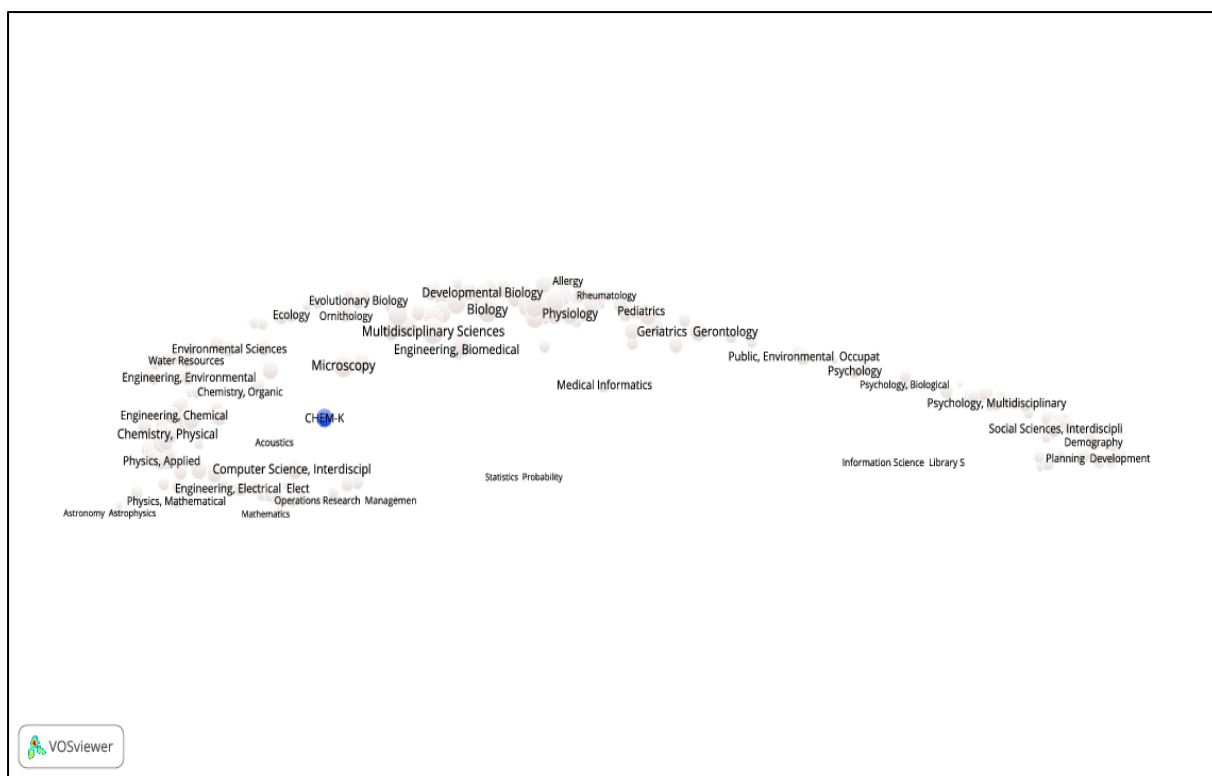


Figure 73. SAPV of the CHEM-K research group's publications in WoS SCs similarity matrix

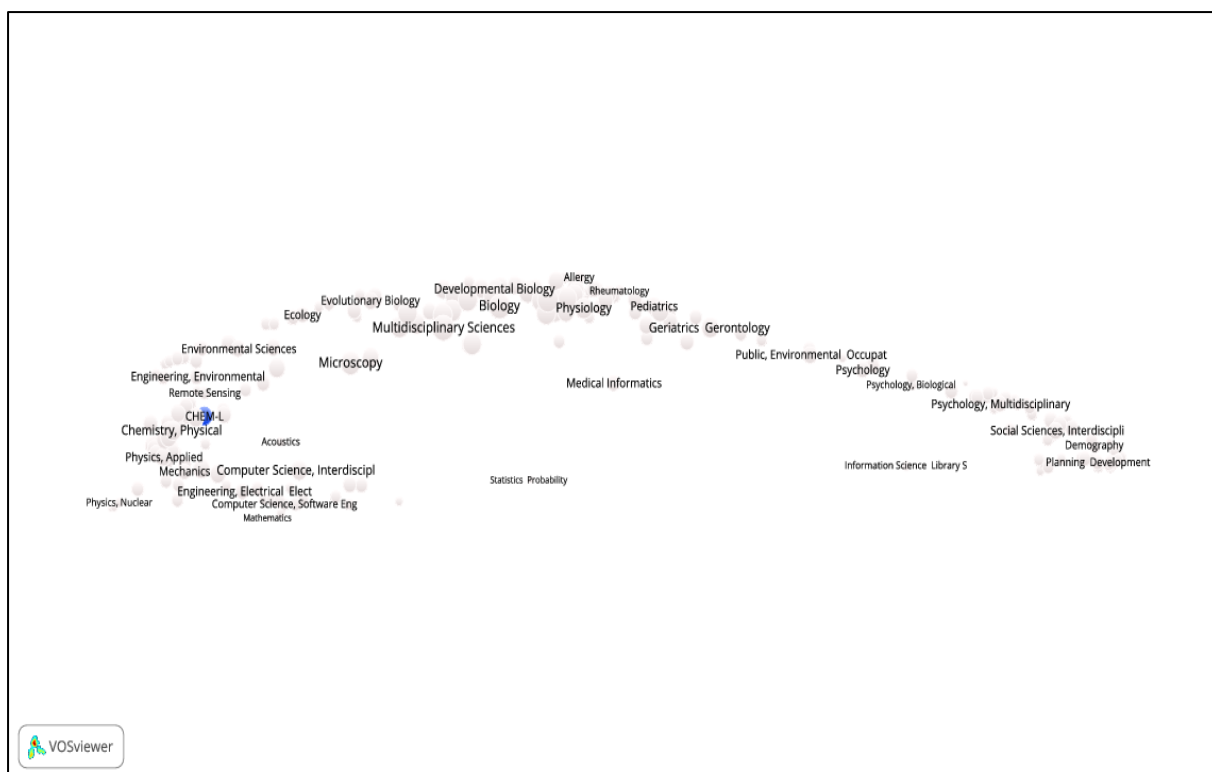


Figure 74. SAPV of the CHEM-L research group's publications in WoS SCs similarity matrix

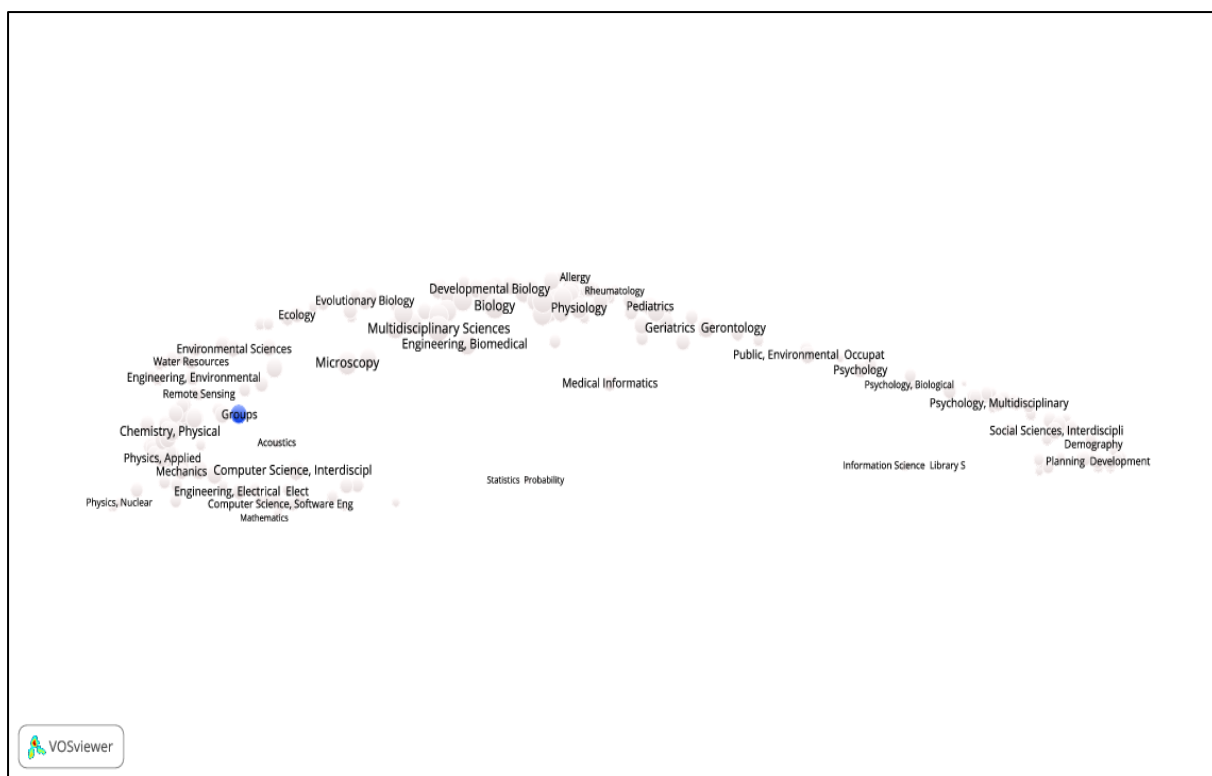


Figure 75. SAPV of the Chemistry research group's publications in WoS SCs similarity matrix

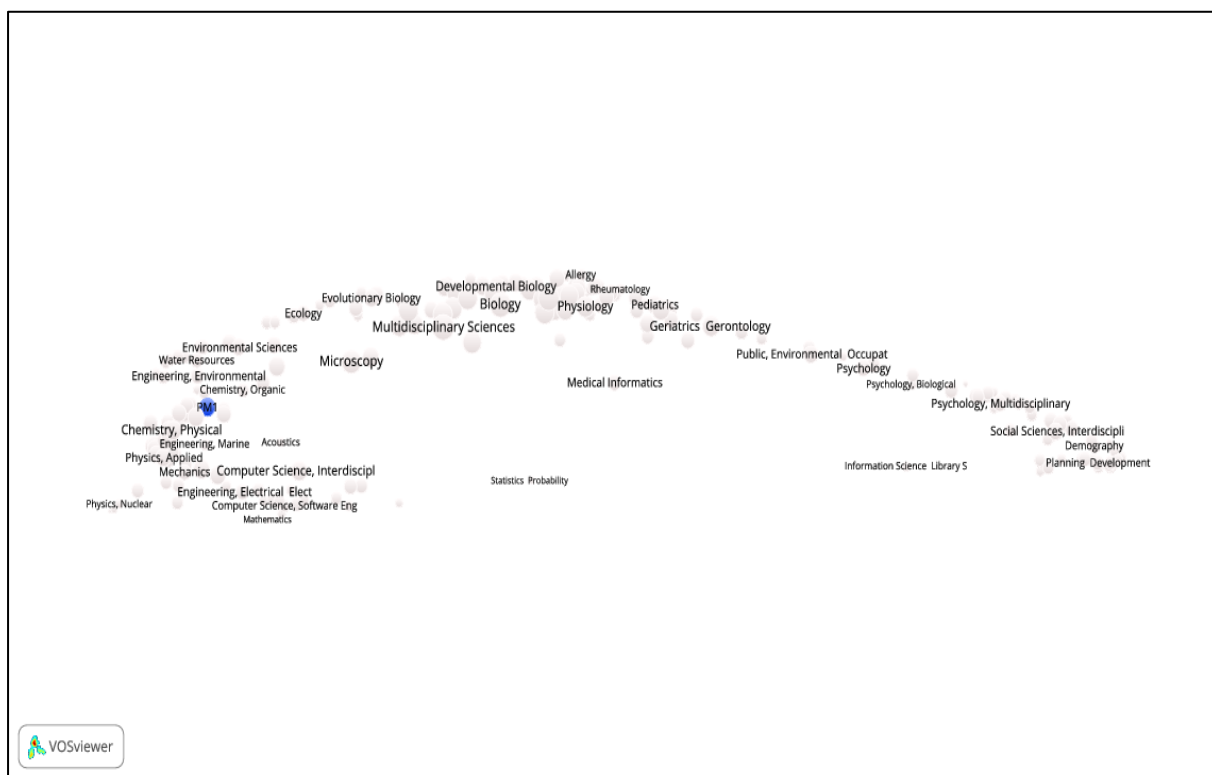


Figure 76. SAPV of the PM1's publications in WoS SCs similarity matrix

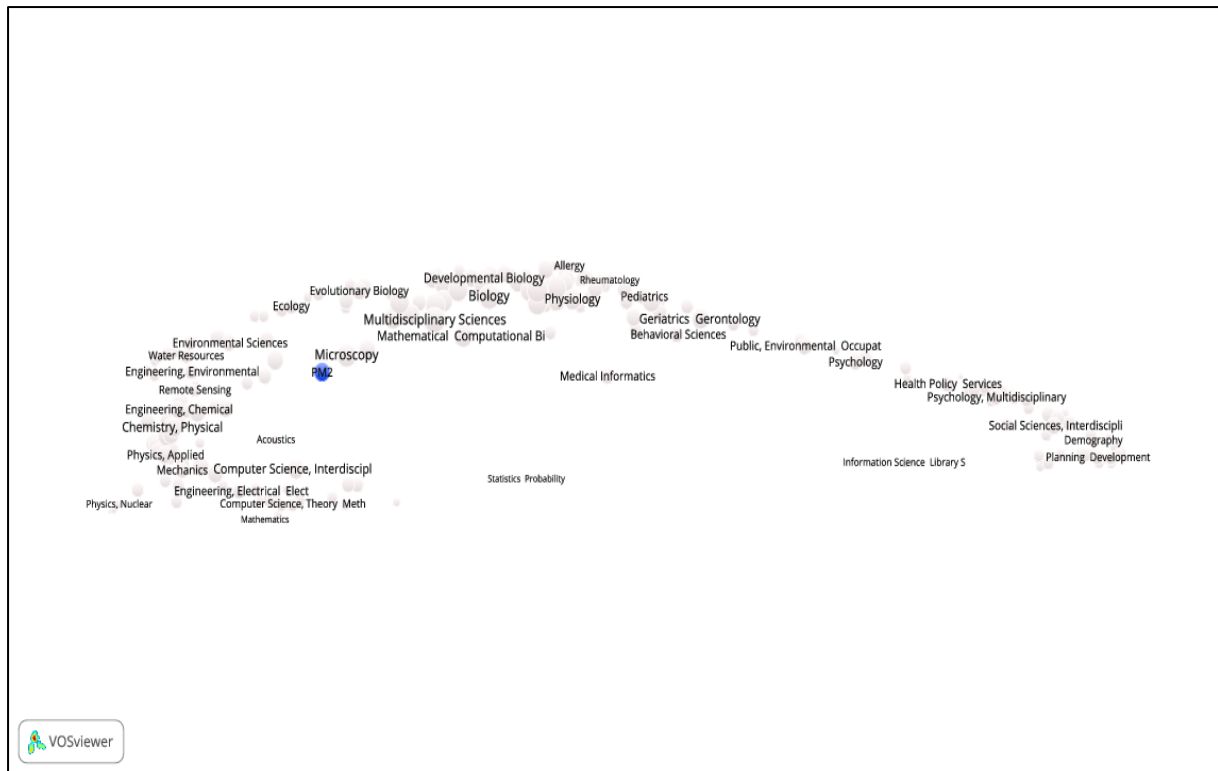


Figure 77. SAPV of the PM2's publications in WoS SCs similarity matrix

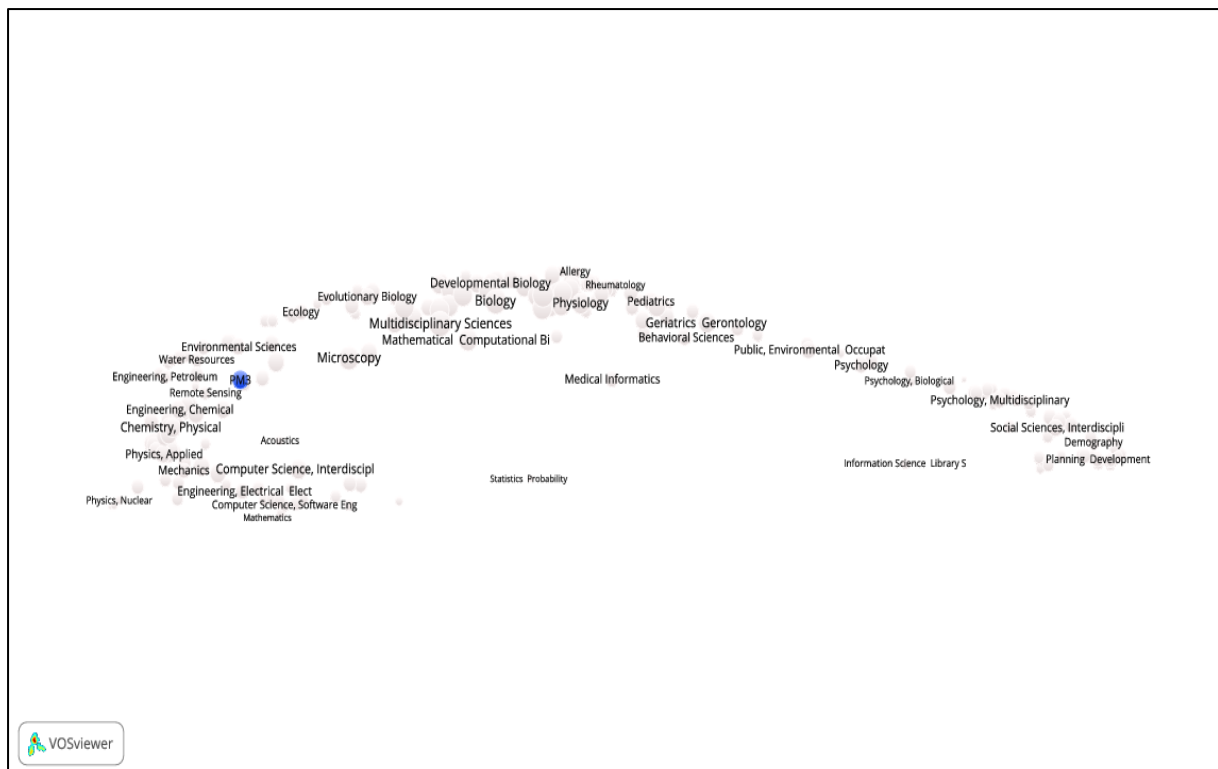
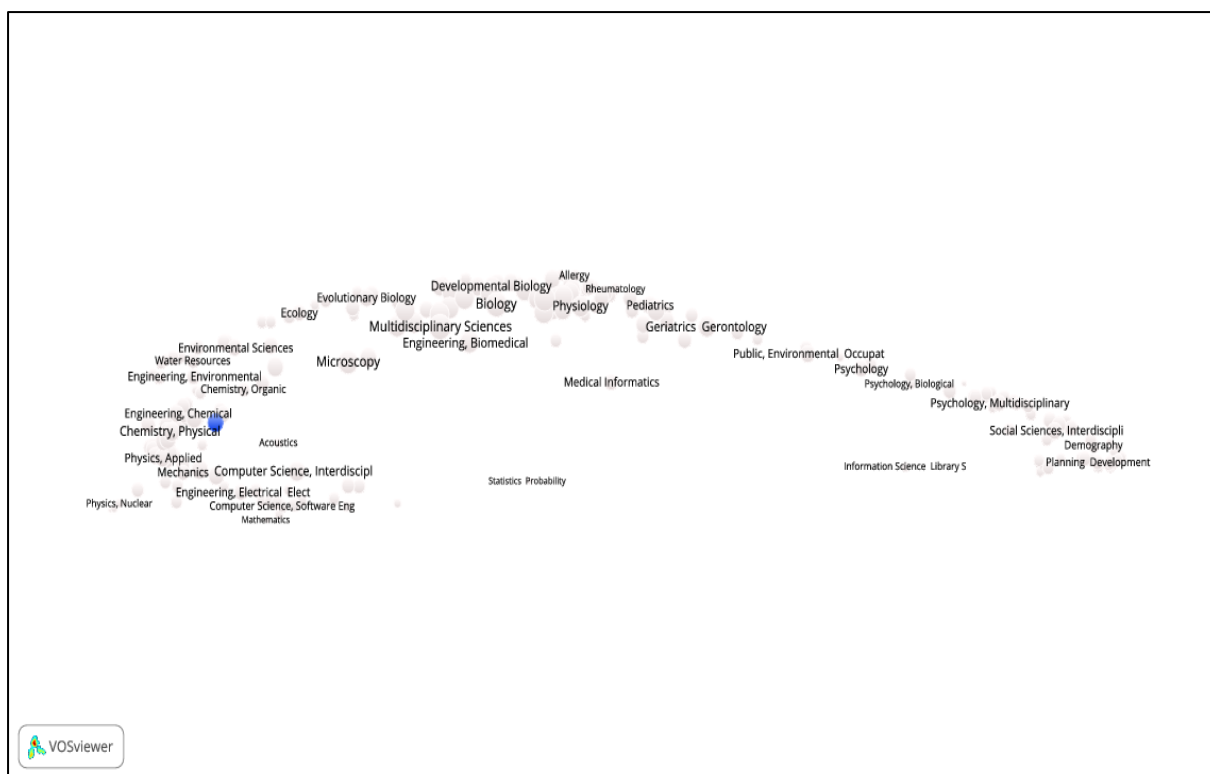
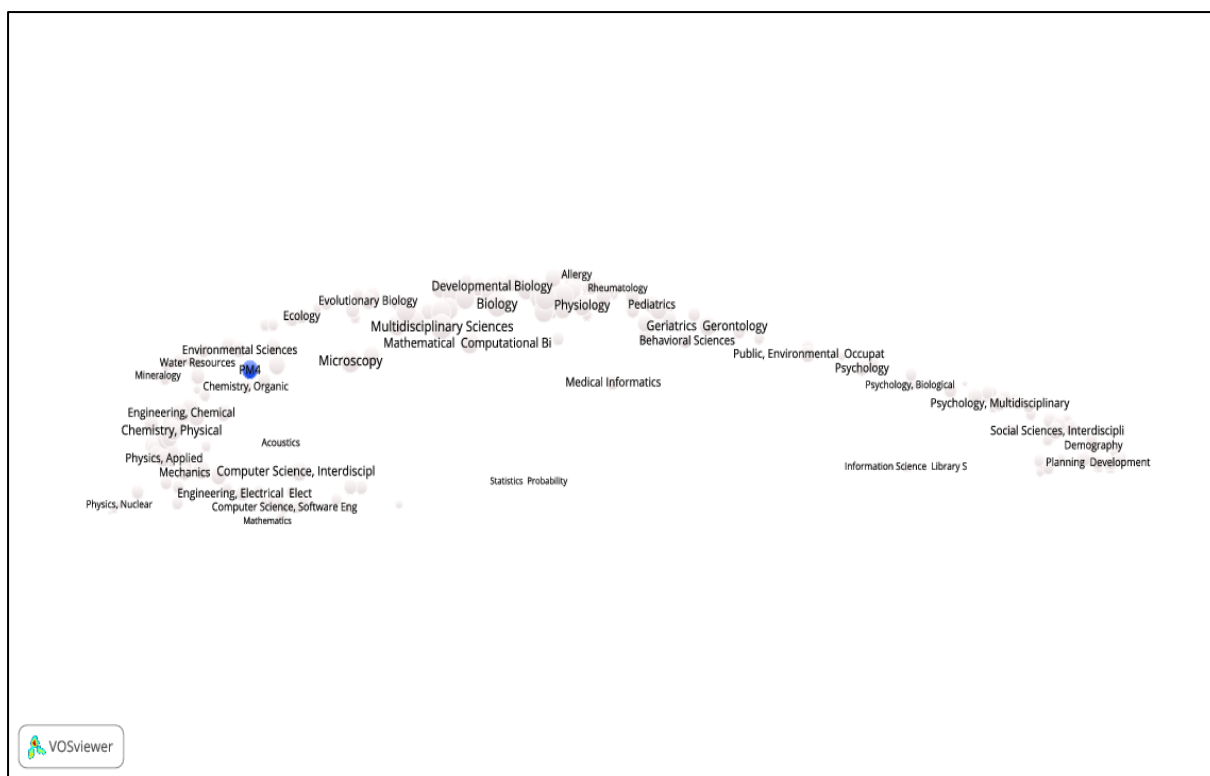


Figure 78. SAPV of the PM3's publications in WoS SCs similarity matrix



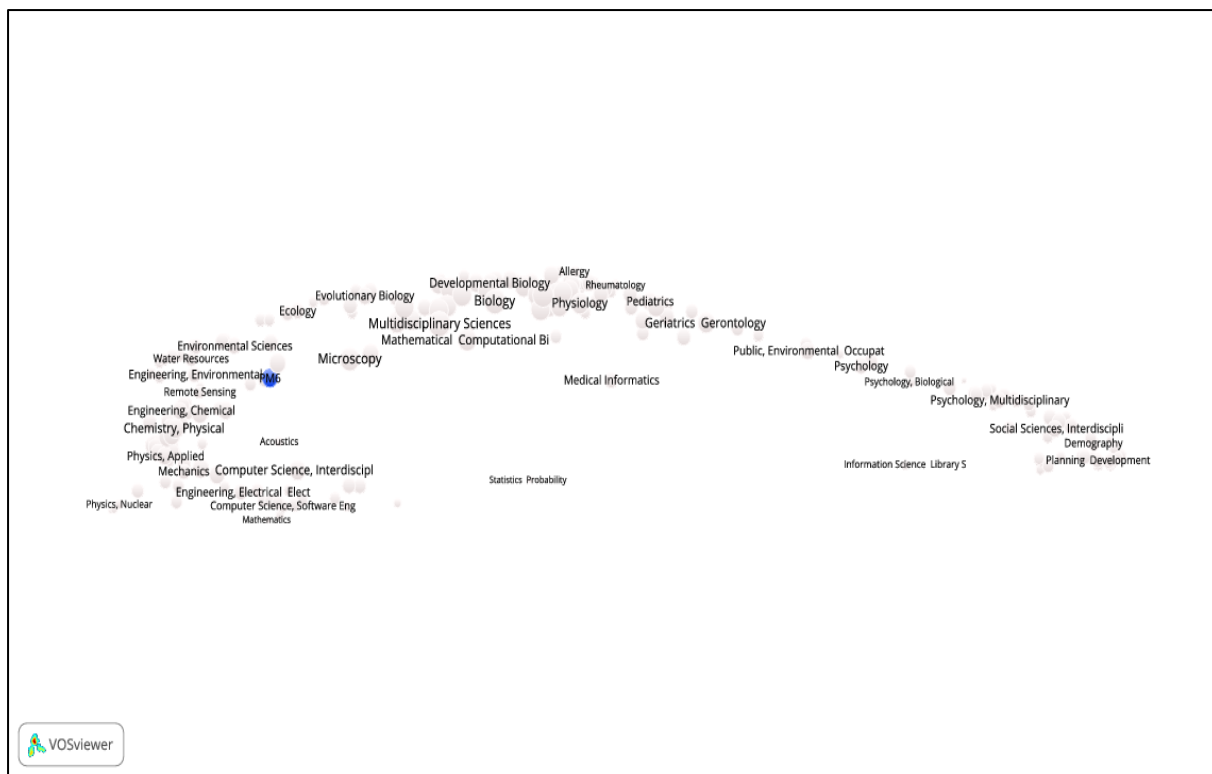


Figure 81. SAPV of the PM6's publications in WoS SCs similarity matrix

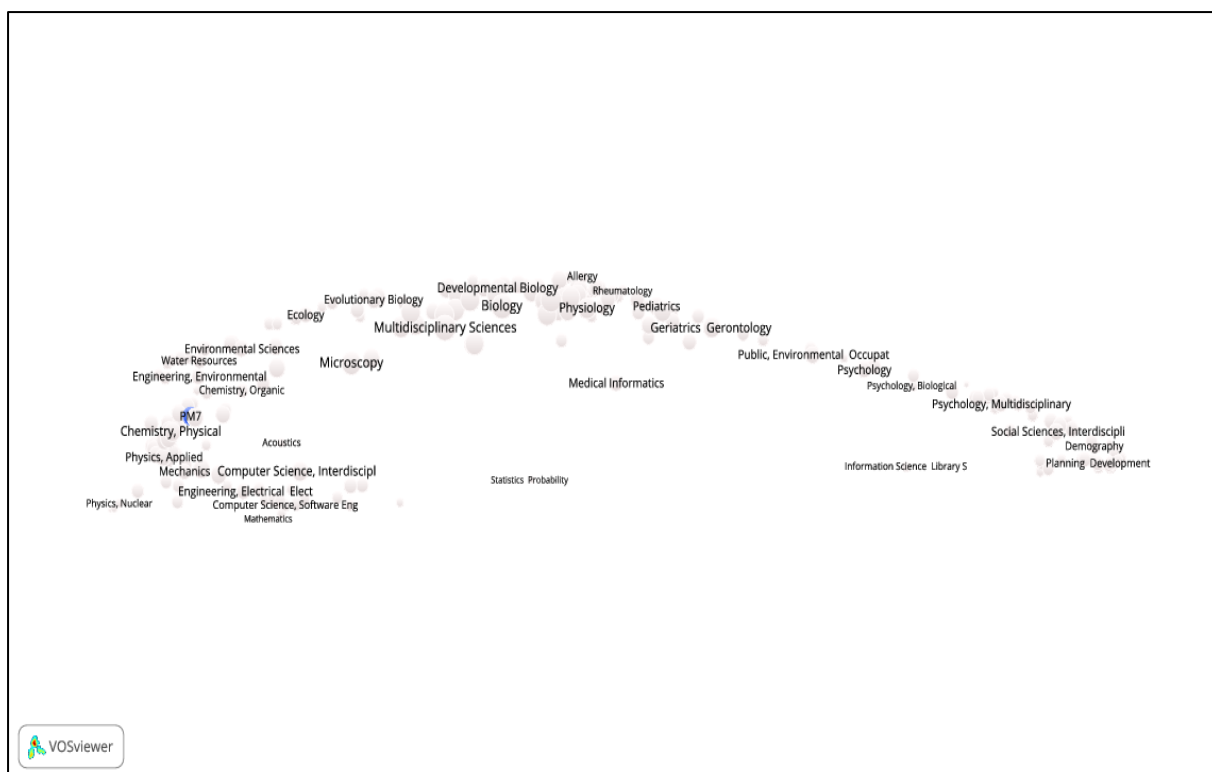


Figure 82. SAPV of the PM7's publications in WoS SCs similarity matrix

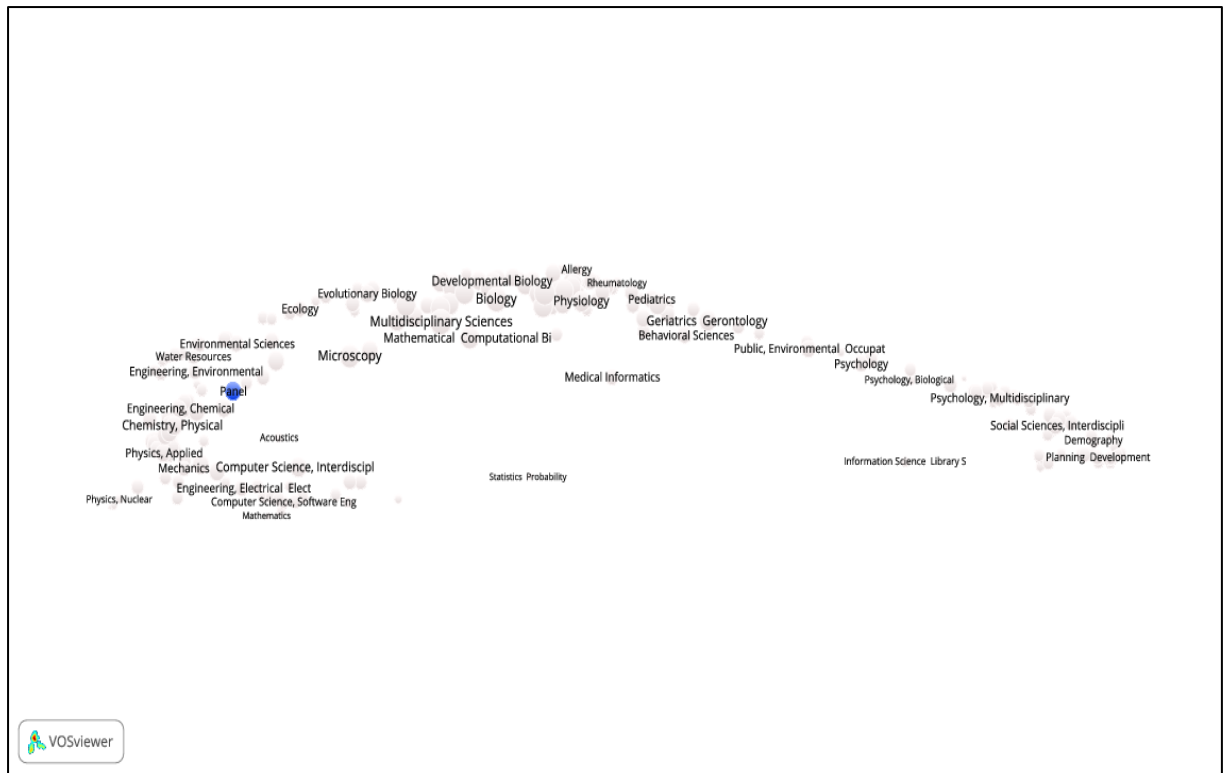


Figure 83. SAPV of the panel publications in WoS SCs similarity matrix

Appendix C

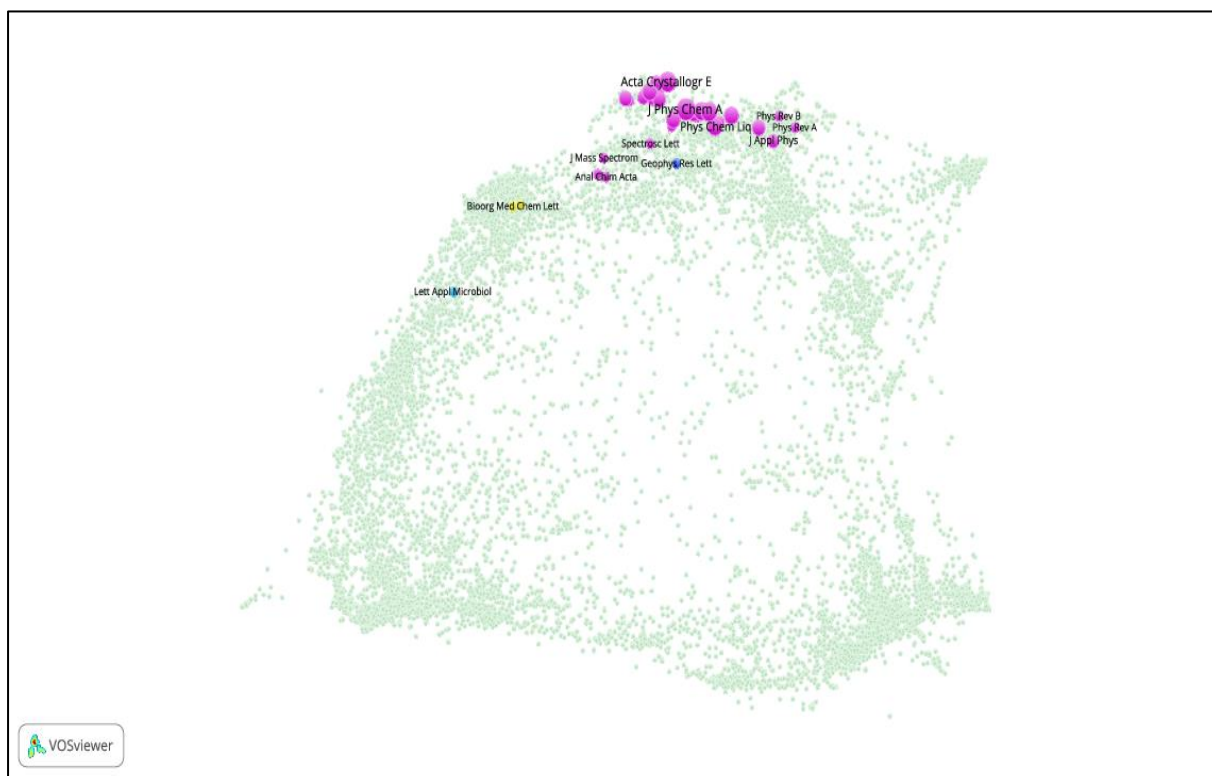


Figure 84. Journal overlay map of CHEM-A research group's publications

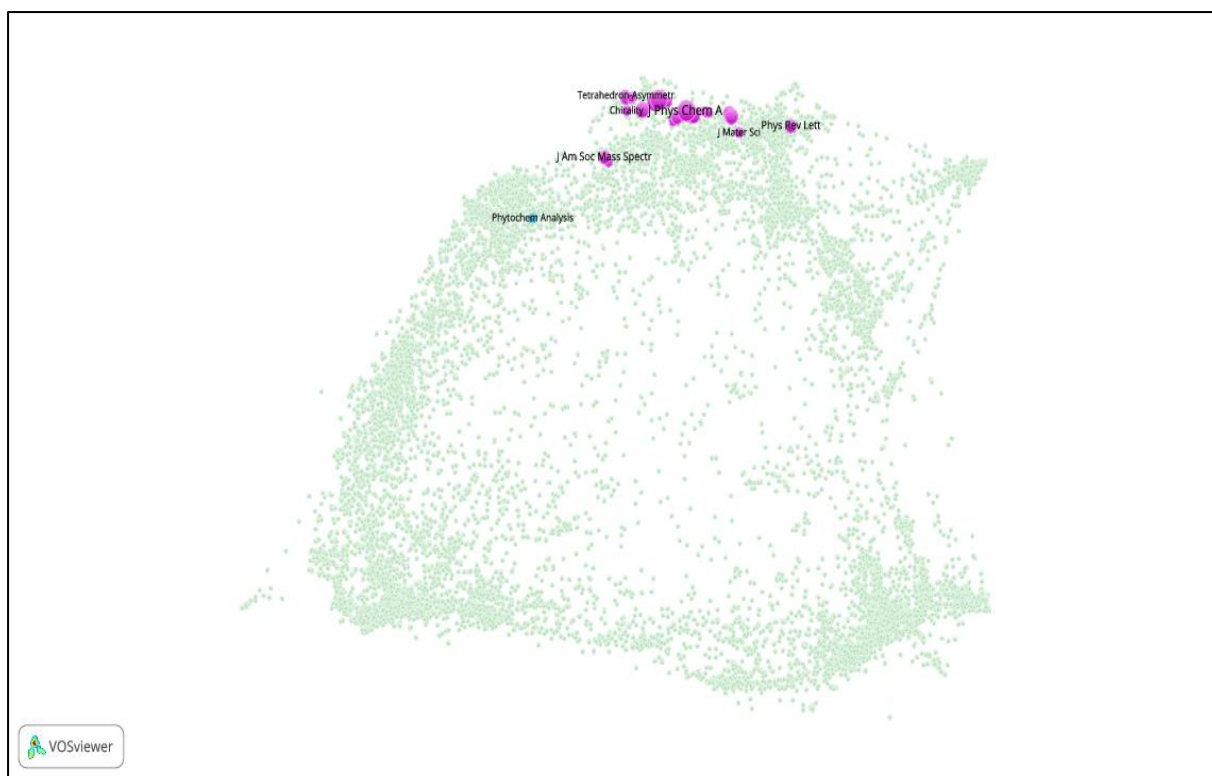


Figure 85. Journal overlay map of CHEM-B research group's publications

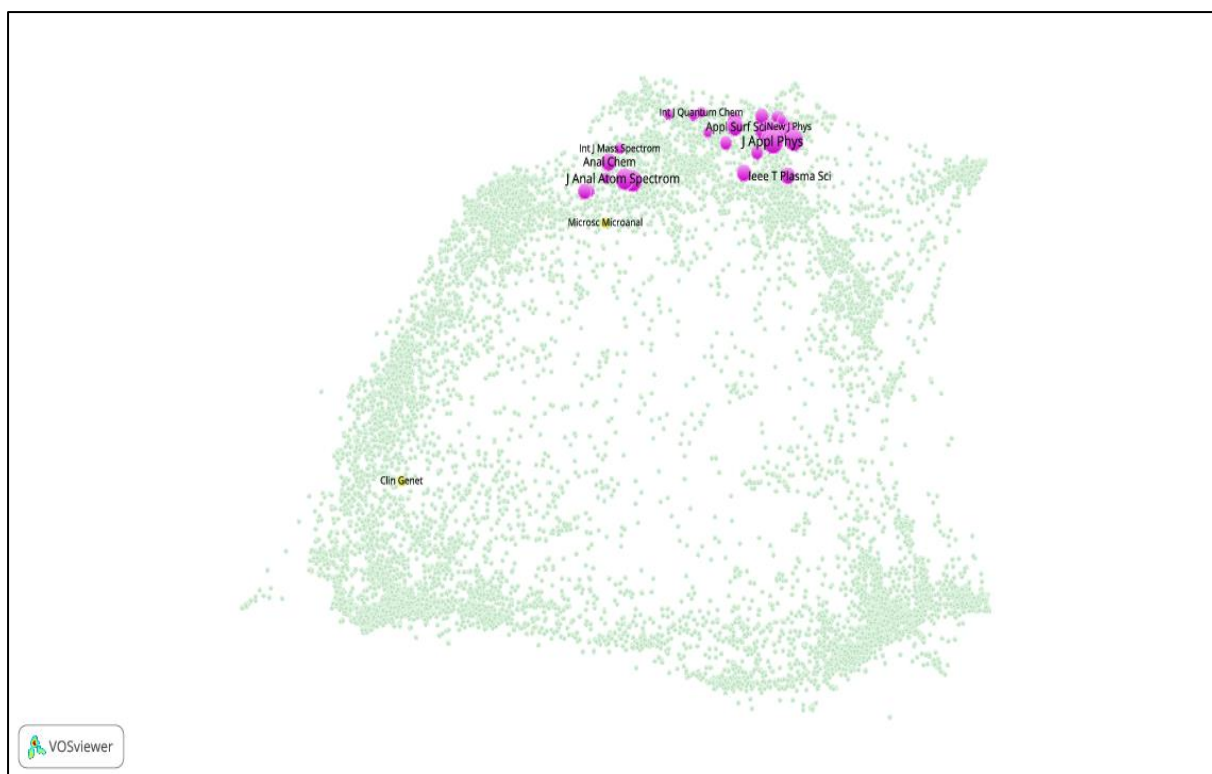


Figure 86. Journal overlay map of CHEM-C research group's publications

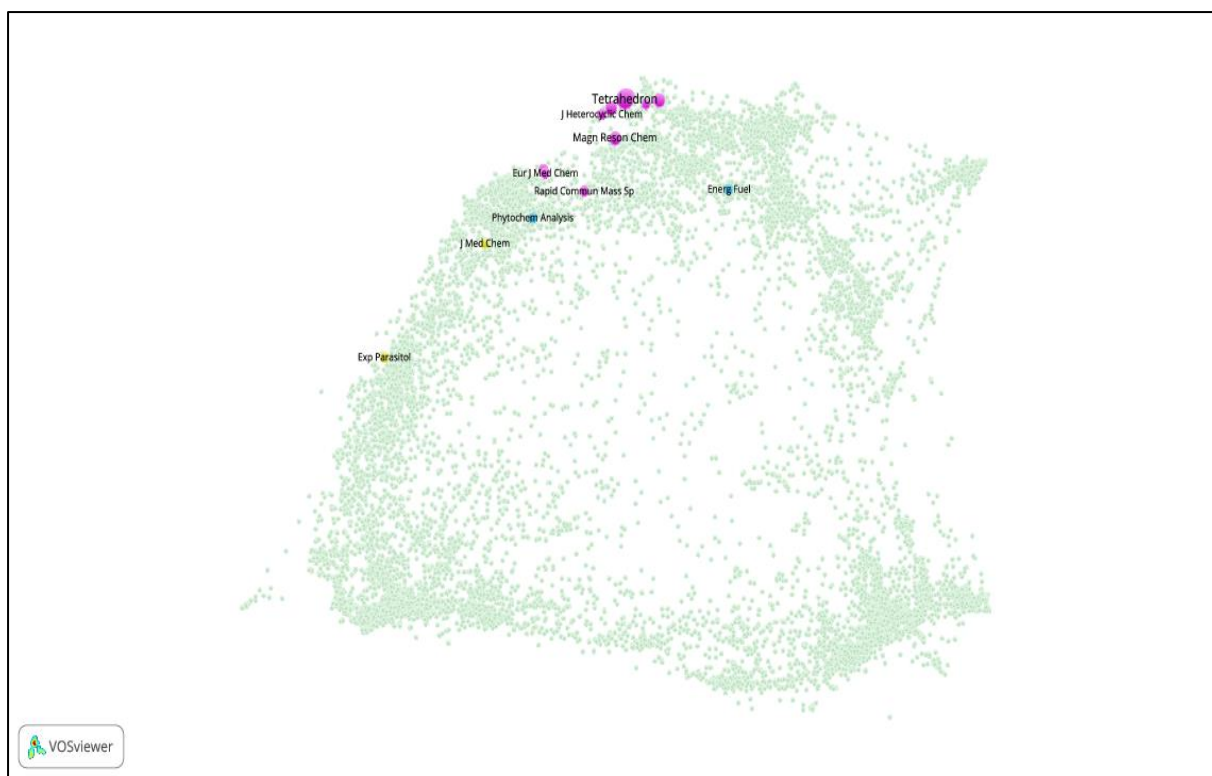


Figure 87. Journal overlay map of CHEM-D research group's publications

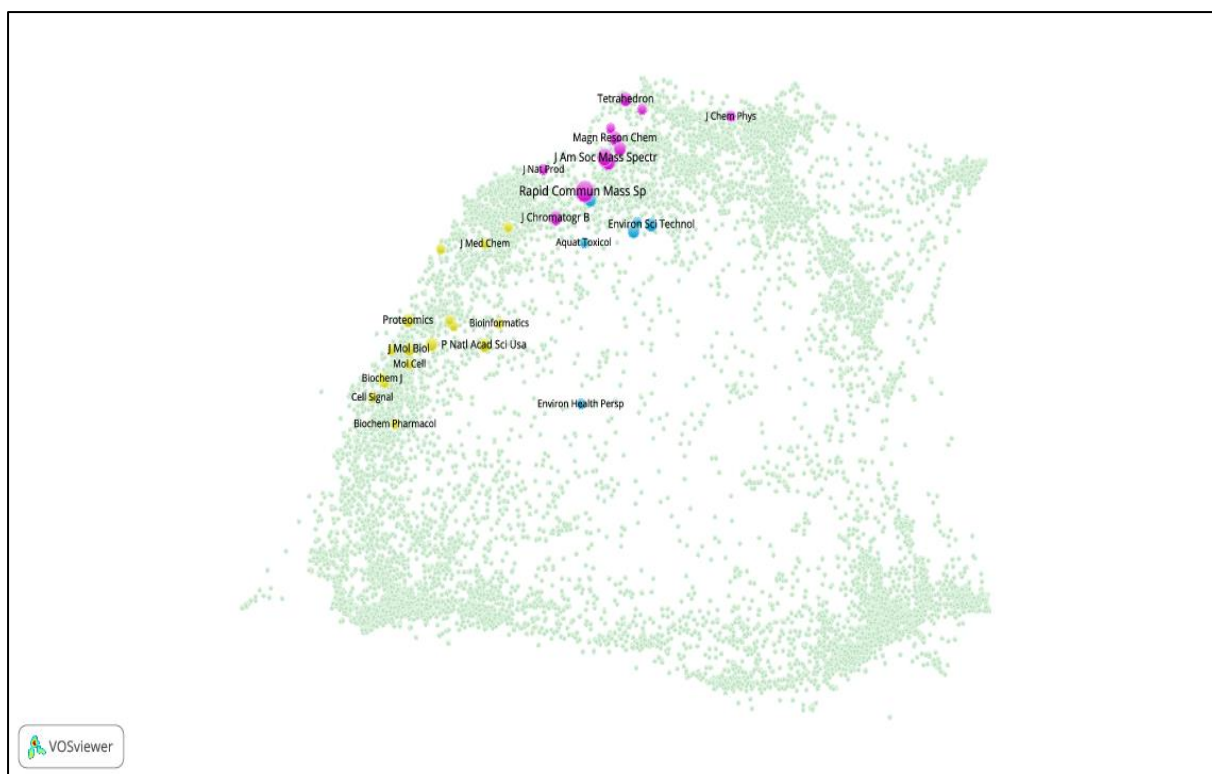


Figure 88. Journal overlay map of CHEM-E research group's publications

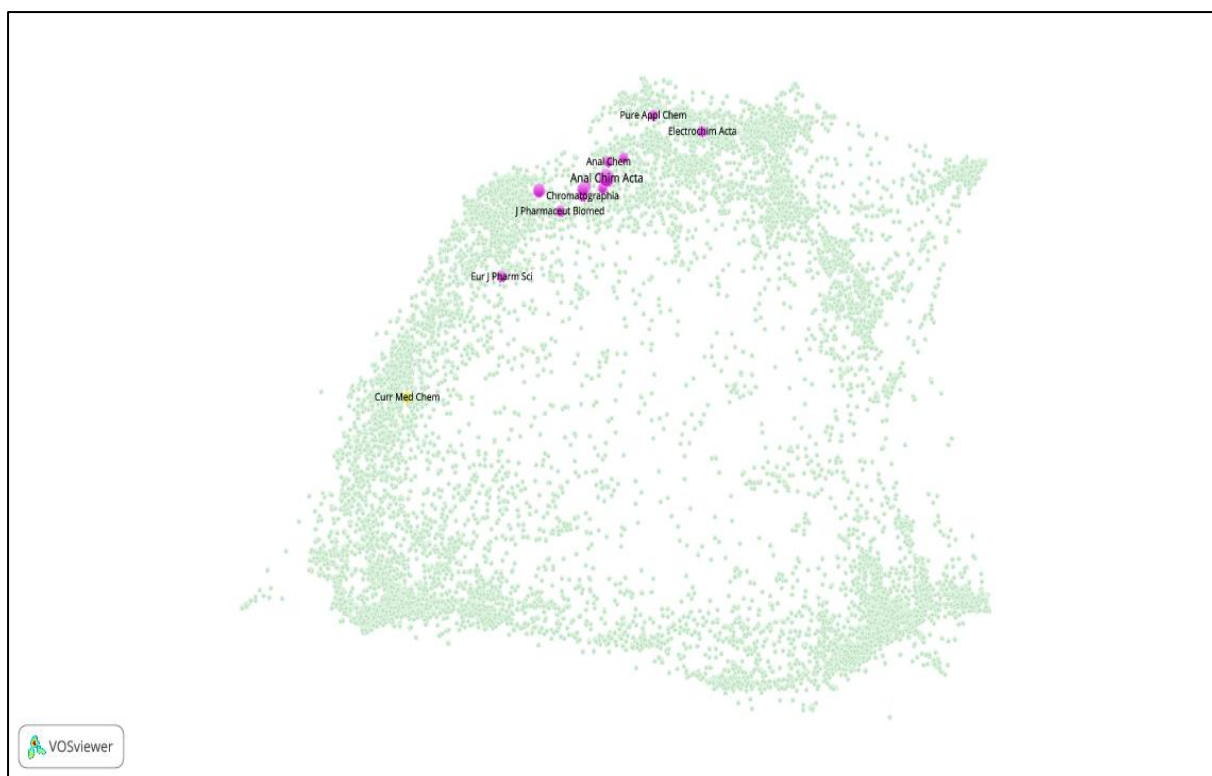


Figure 89. Journal overlay map of CHEM-F research group's publications

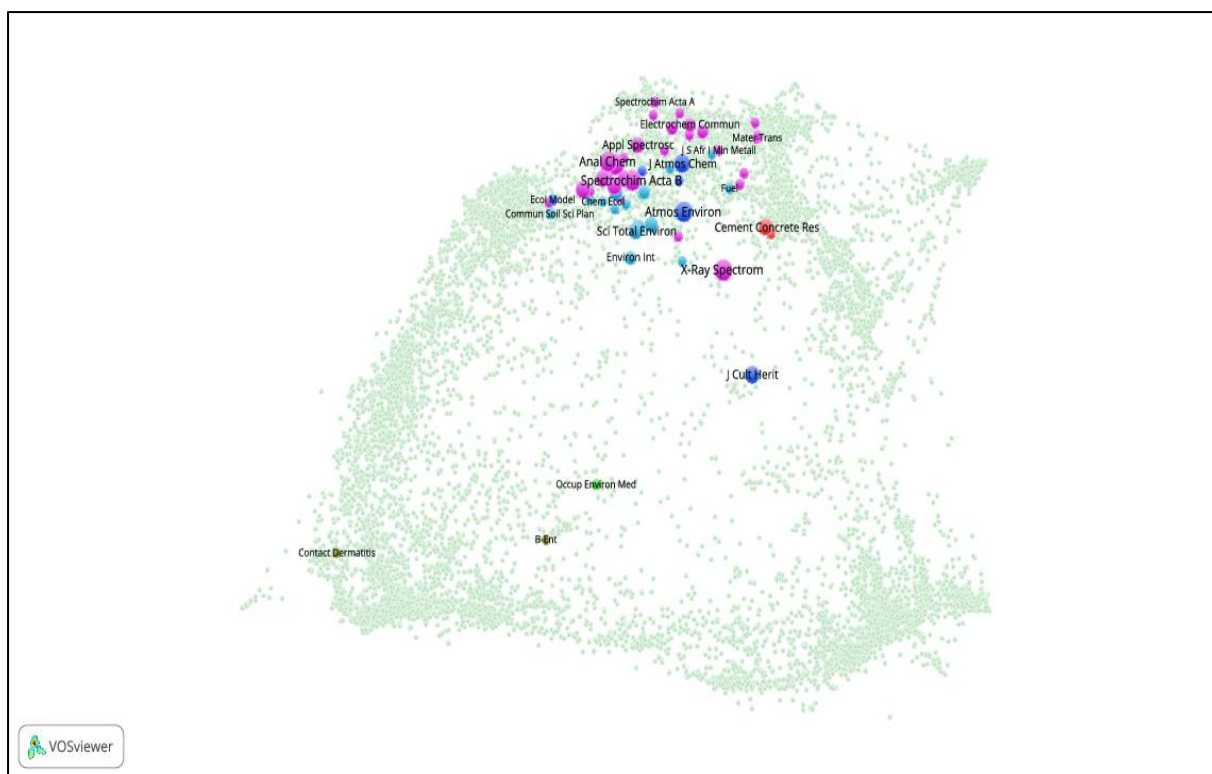


Figure 90. Journal overlay map of CHEM-G research group's publications

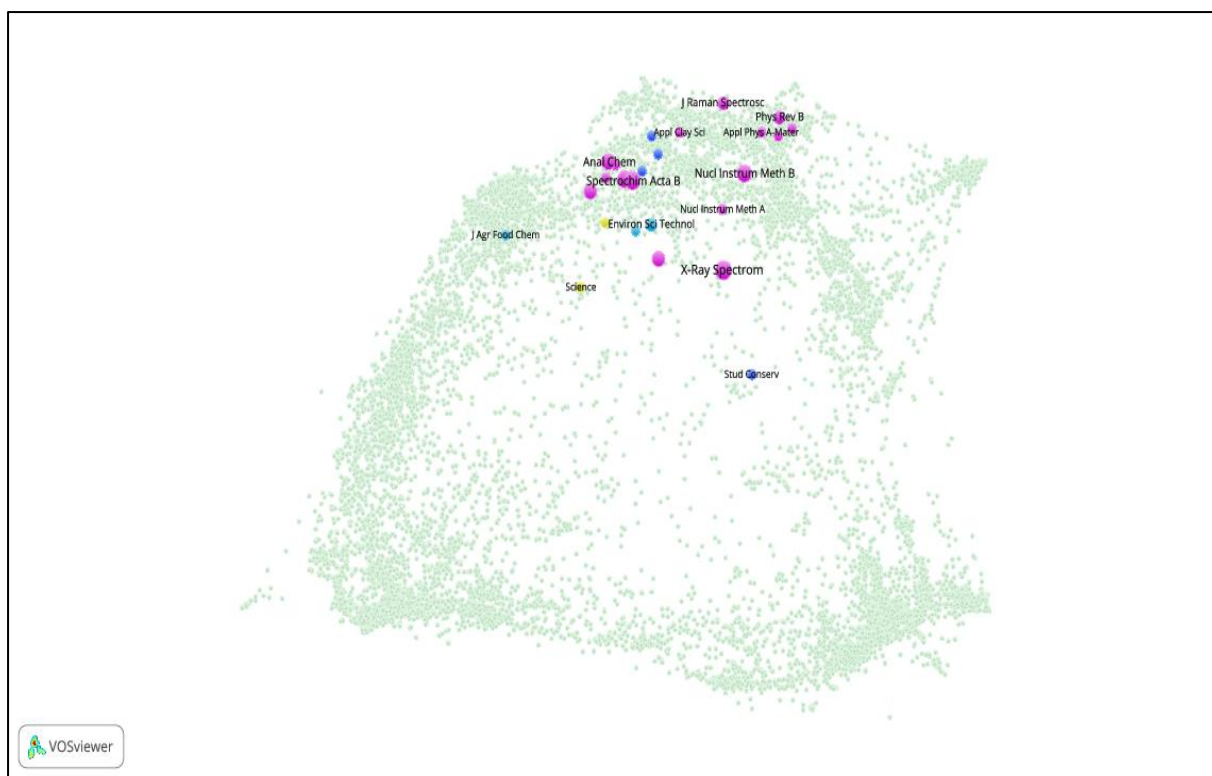


Figure 91. Journal overlay map of CHEM-H research group's publications

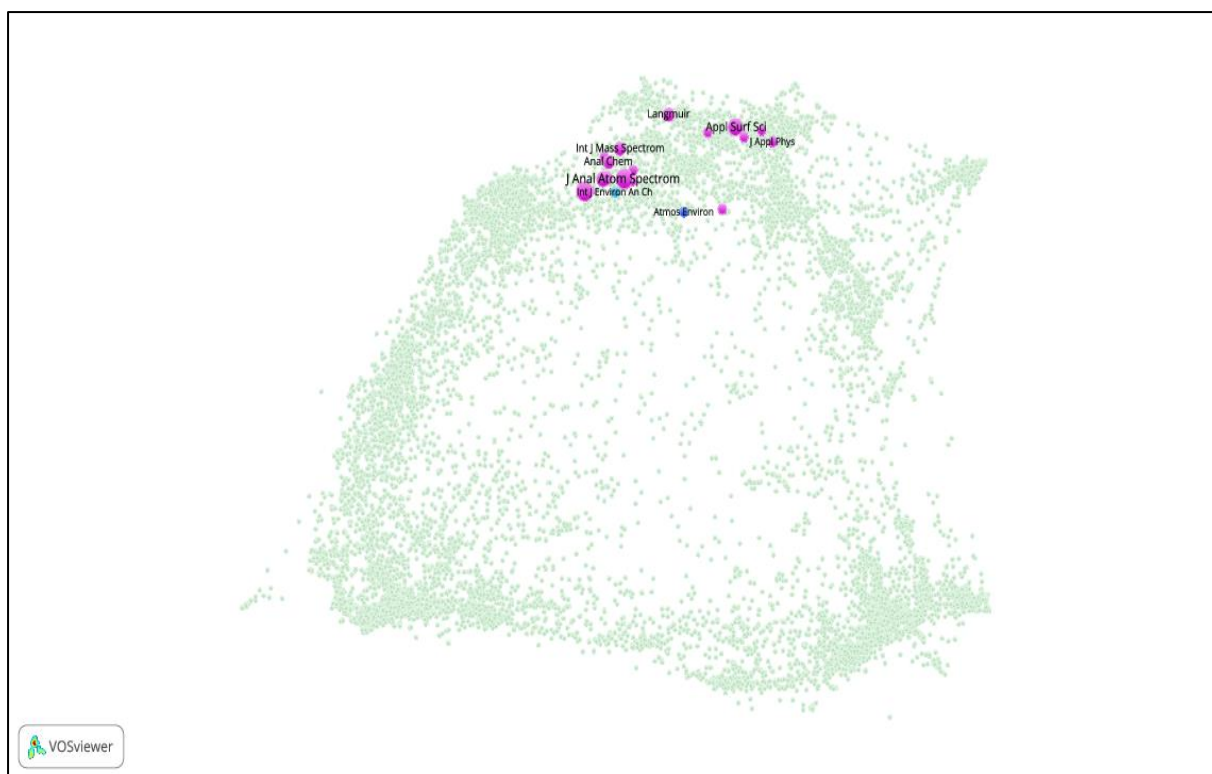


Figure 92. Journal overlay map of CHEM-I research group's publications

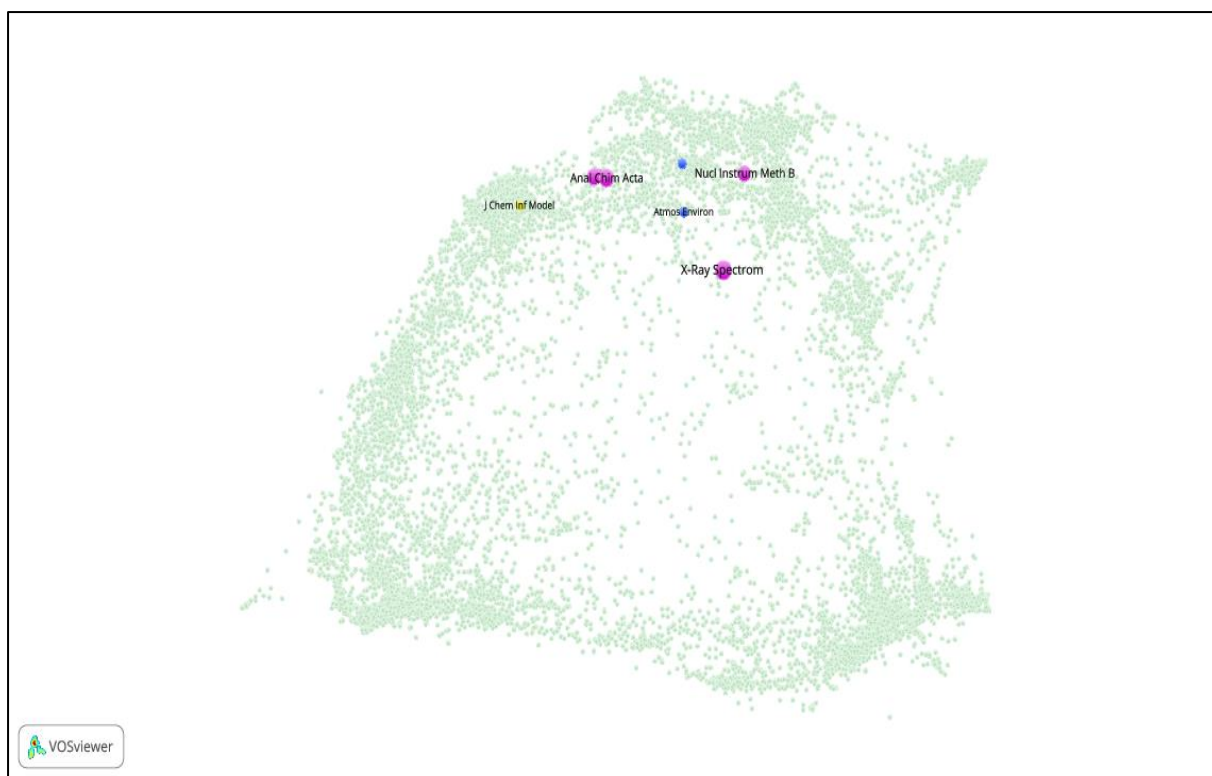


Figure 93. Journal overlay map of CHEM-J research group's publications

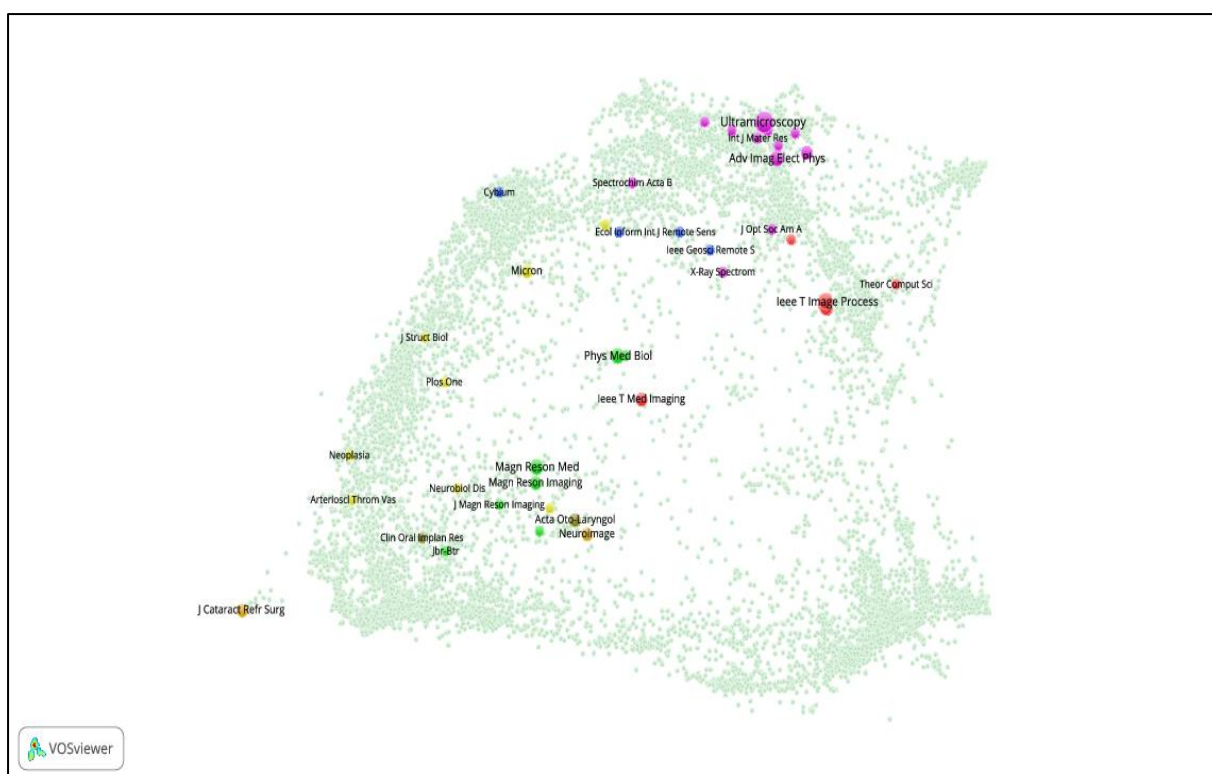


Figure 94. Journal overlay map of CHEM-K research group's publications

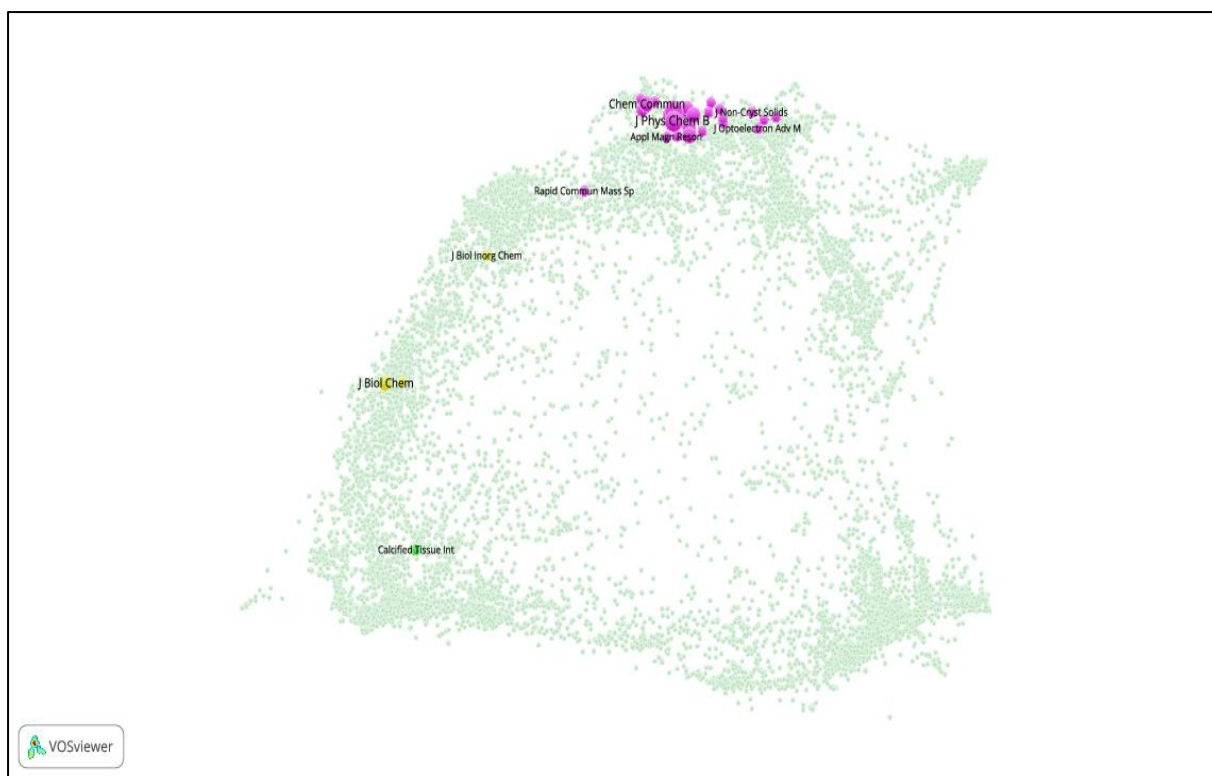


Figure 95. Journal overlay map of CHEM-L research group's publications

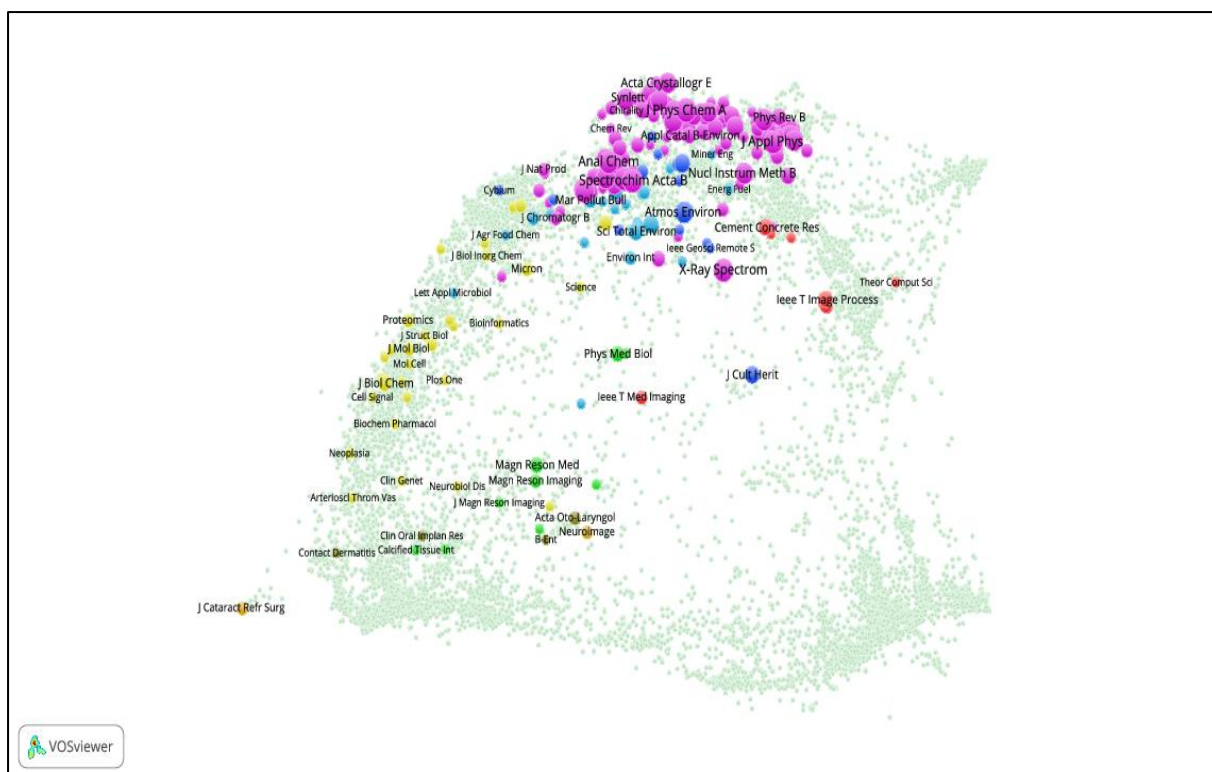


Figure 96. Journal overlay map of Chemistry research groups' publications

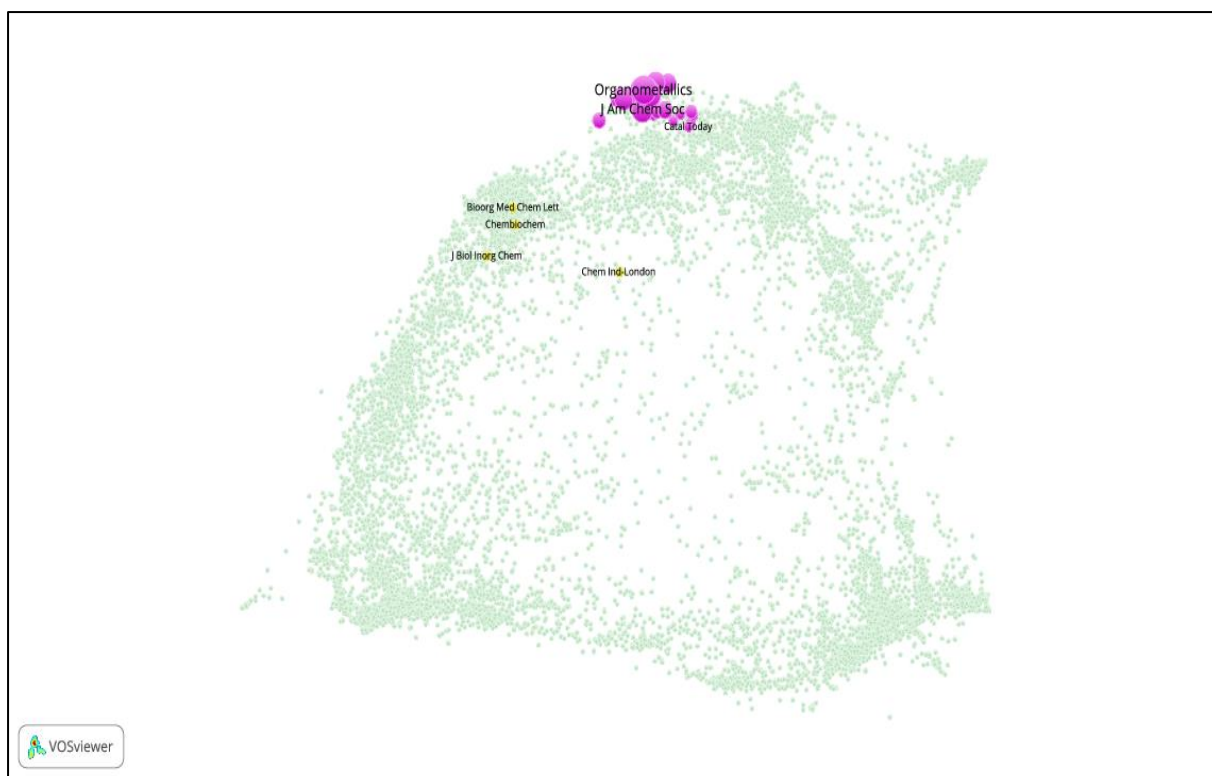


Figure 97. Journal overlay map of PM1's publications

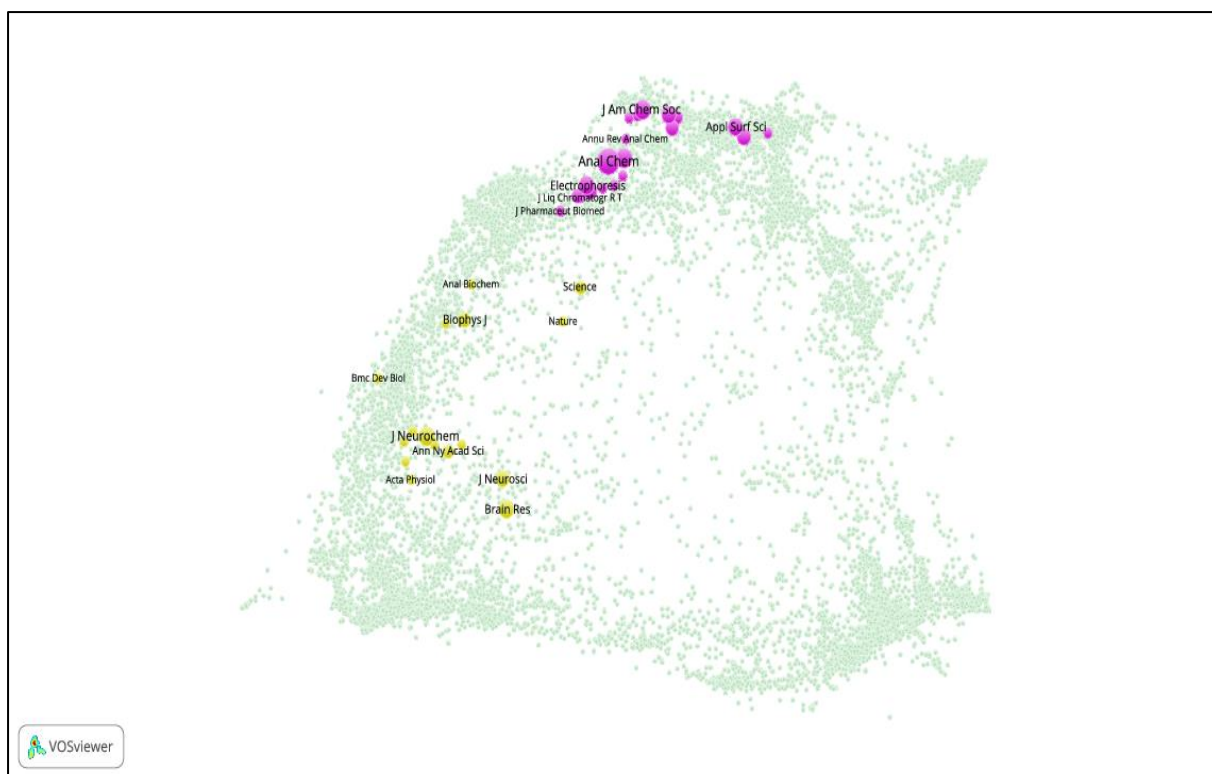


Figure 98. Journal overlay map of PM2's publications

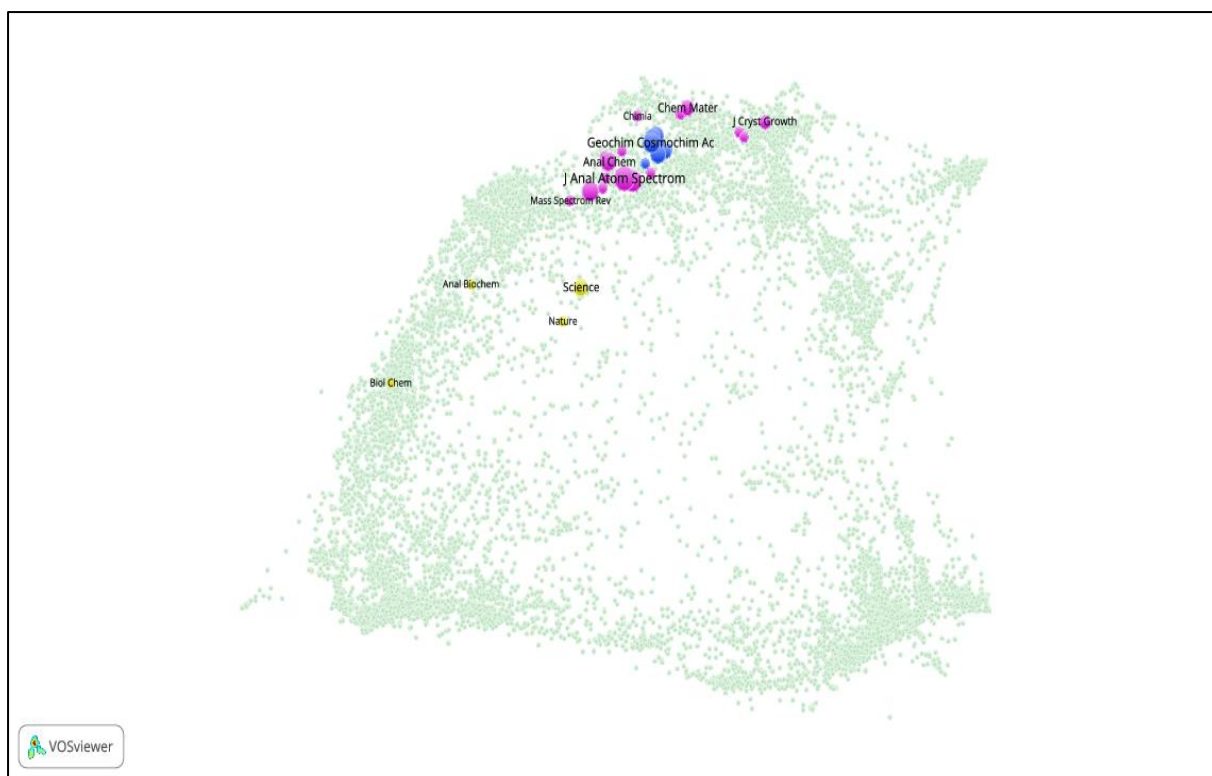


Figure 99. Journal overlay map of PM3's publications

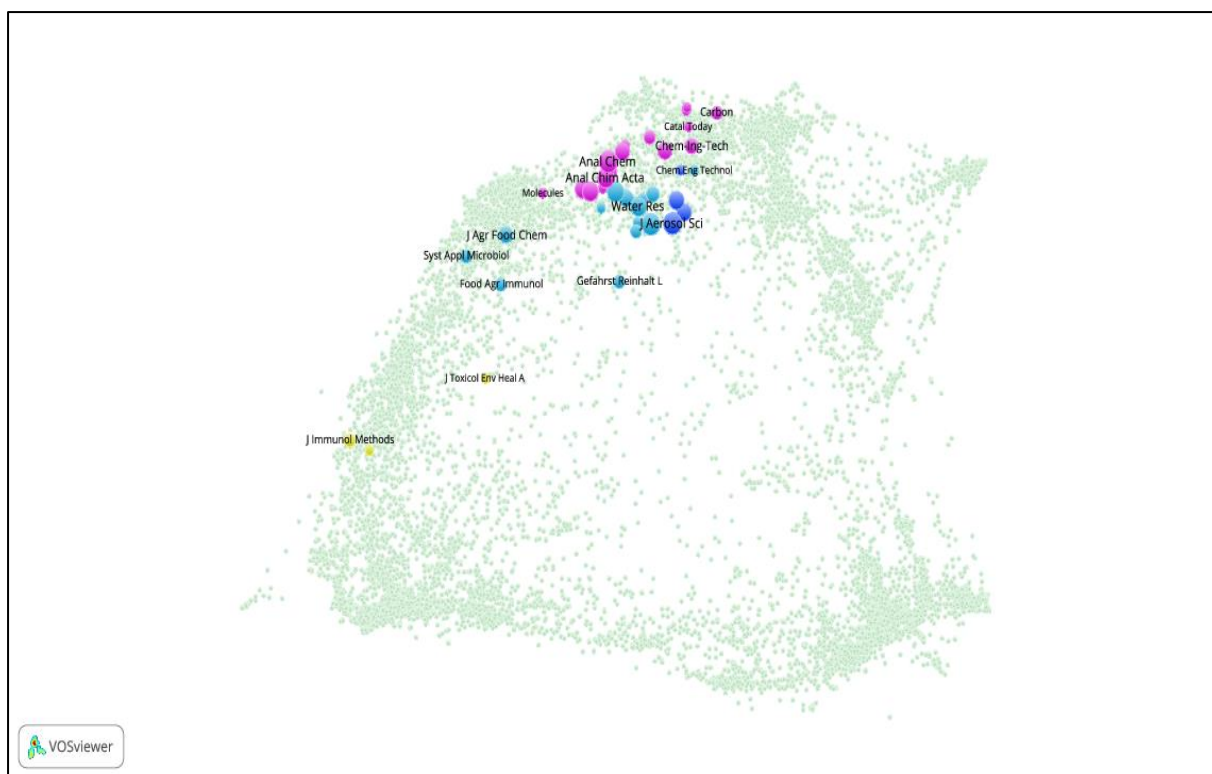


Figure 100. Journal overlay map of PM4's publications

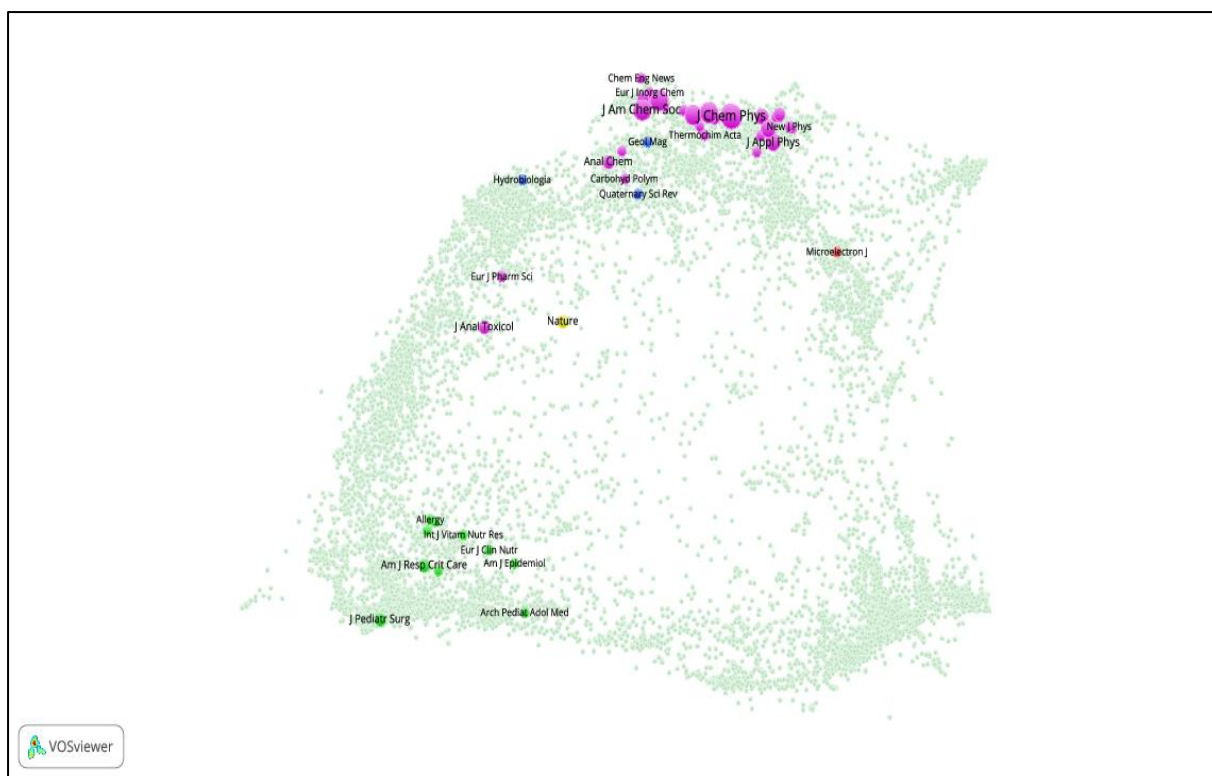


Figure 101. Journal overlay map of PM5's publications

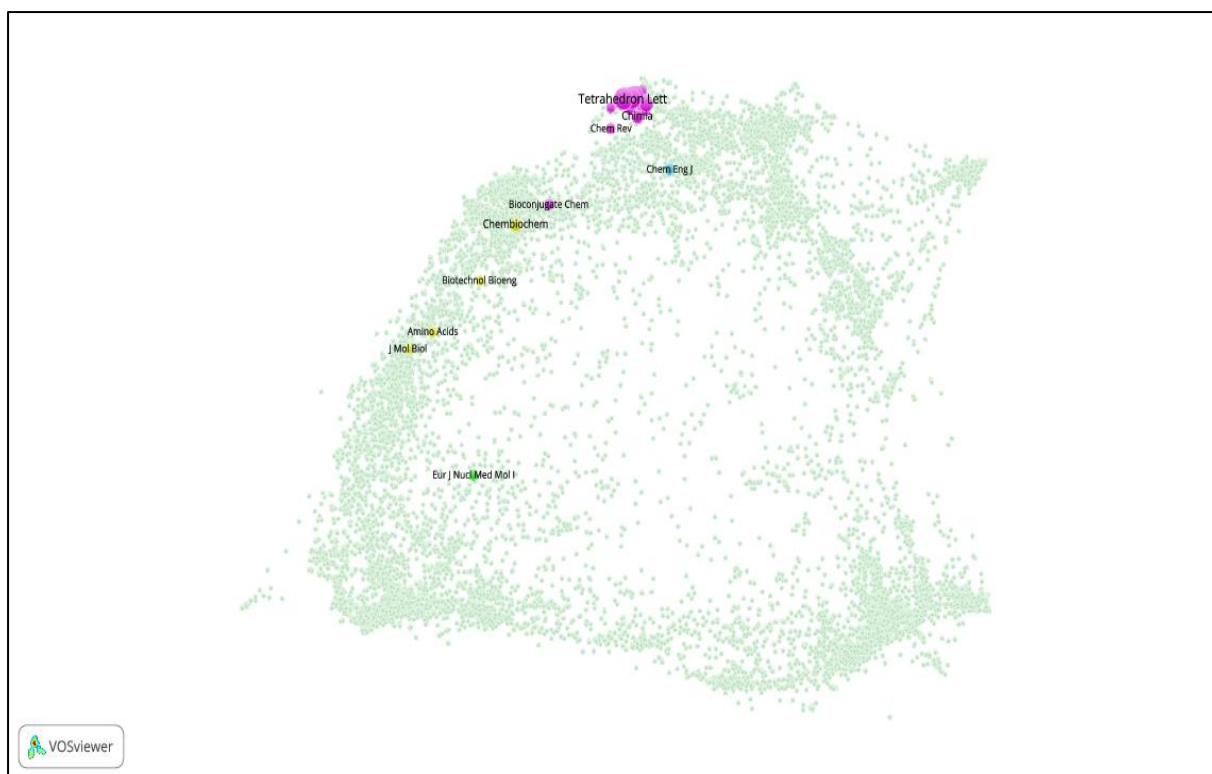


Figure 102. Journal overlay map of PM6's publications

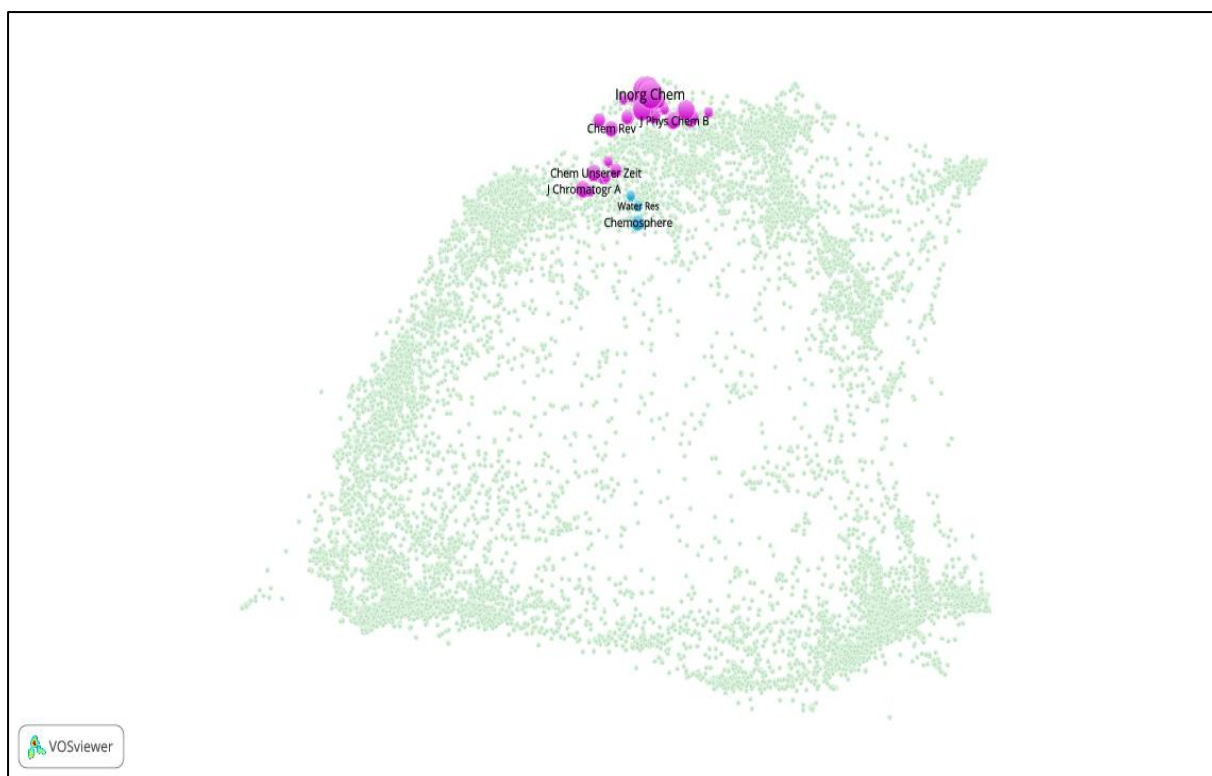


Figure 103. Journal overlay map of PM7's publications

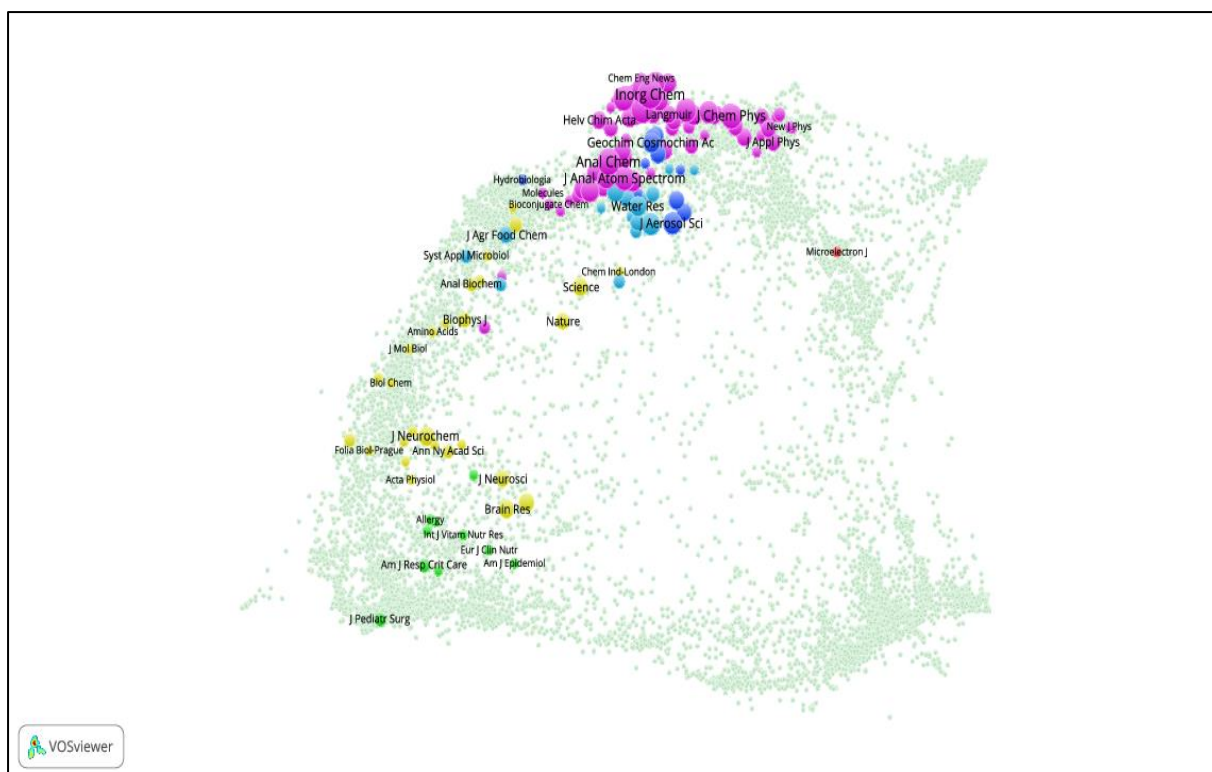


Figure 104. Journal overlay map of the panel's publications