[ξ]

KSZI [ξ] AKTÁK

MTA KIK TTO műhelytanulmányok

2013/2

Bibliographic Coupling and Hierarchical Clustering for the validation and improvement of subject-classification schemes*

Bart Thijs – Lin Zhang – Wolfgang Glänzel

wolfgang.glanzel@econ.kuleuven.be

➲ http://www.mtakszi.hu/kszi_aktak/

Bibliographic Coupling and Hierarchical Clustering for the validation and improvement of subject-classification schemes*

Bart Thijs[1], Lin Zhang[2] and Wolfgang Glänzel[1,3]

[1] *KU Leuven, ECOOM and Dept. MSI, Leuven (Belgium)*

[2] *Department of Management & Economics, North China University of Water Conservancy and Electric Power, Zhengzhou, (China)*

[3] *Library of the Hungarian Academy of Sciences, Dept. Science Policy & Scientometrics, Budapest (Hungary)*

Abstract

An attempt is made to apply bibliographic coupling to journal clustering of the complete Web of Science database. Since the sparseness of the underlying similarity matrix proved inappropriate for this exercise, second-order similarities have been used. Only 0.12% out of 8,282 journals had to be removed from the classification as being singletons. The quality at three hierarchical levels with 6, 14 and 24 clusters substantiated the applicability of this method. Cluster labelling was made on the basis of the about 70 subfields of the Leuven-Budapest subject-classification scheme that also allowed the comparison with the existing two-level journal classification system developed in Leuven. The further comparison with the 22 field classification system of the Essential Science Indicators does, however, reveal larger deviations.

Introduction

The issue of subject classification and the creation of coherent journal sets has been a major topic in our field since the seventies (see e.g., Narin et al., 1972; Narin, 1976). The development of computerised methods and the availability of large datasets have shifted the attention from mapping small or single disciplines to the generation of global science maps (Garfield, 1998). Data available from Thomson Reuters' Journal Citation Reports (JCR) has been used by several authors (Bassecoulard and Zitt, 1999; Leydesdorff, 2004). Unlike in Thomson Reuters' Web of Science (WoS) database, where citations are determined for each paper individually, in the JCR citation data are based on journal information in the papers' reference lists and therefore aggregated to the journal level. However, also WoS data was used at the level of individual publications for the generation of global maps. Jarneving (2005) applied bibliographic coupling to map and to analyse the structure of an annual volume of the Science Citation Index. Janssens et al. (2008; 2009) used a combination of cross-citations and a lexical approach to map journals. Zhang et al. (2010) validated this approach. This paper builds on prior attempts to classify journals relying on computerised techniques. In this study we take a different approach and attempt to build a network among journals based on bibliographic coupling similarities.

The advantage of bibliographic coupling is that there is no delay for the calculation of the link between publications or journals as all data needed are present upon publication or indexing in the database. This also means that link between documents, once established will remain constant over time. Sharing this property with text-based method, new mappings of journals based on bibliographic coupling are able to reflect the current situation as soon as the underlying documents are indexed in the database. However, for this paper and the

* Paper to be presented at the 14th International Conference on Scientometrics and Informetrics (15–19 July, 2013), Vienna (Austria)

development and validation of our methodology we use the 2006-2009 publications set to be able to relate our results to those of previous exercises.
In contrast to the above-mentioned advantages of bibliographic coupling, this method has one drawback which is shared with other citation-based approaches such as co-citation analyses.This disadvantage is a result of the very sparse nature of the link matrix (Janssens, 2007; Janssens et al., 2008). The overwhelming number of document pairs does not share any reference at all and thus a large number of zeros occur in the similarity matrix. This deteriorates the quality of the subsequent clustering and may result in an unrealistic large number of singletons (cf. Jarneving, 2005). As cross-citation data suffers from the same problem, Janssens et al. (2008) introduced a hybrid approach, where they combined citation-based with lexical similarities.
Another solution to overcome the sparseness problem is the use of second order similarities (Janssens, 2007; Ahlgren & Colliander, 2009; Thijs et al., 2013). The objective of the present paper is to demonstrate the applicability of bibliographic coupling as link measure in the mapping of journals as well as to compare the results with those of previous cross-citation and hybrid citation-text based studies.

Data sources

A set of journals was compiled from the Web of Science database (SCI-Expanded, SSCI and AHCI). All journals covered in this database between 2006 and 2009 with at least 100 publications in this period are taken into account. This resulted in a set of 8282 journals. For the calculation of the bibliographic coupling between journals we took the following approach. In total more than 134 million references in 4,753,892 publications could be processed on the basis of uniquely coded reference items. All data was uploaded into an *Oracle* database and regular *SQL* was used to query for joint references between journals. Analyses are run in *Matlab* and visualizations are made with *Gephi* (Bastian et al., 2009).

Methods

This section describes the choices that have been made for our journal mapping. In order to enhance comparability with the earlier studies (Janssens et al., 2009; Zhang et al., 2010) we adopted the same clustering technique, namely Ward’s hierarchical clustering. A short description of this method will follow later. The goal of this paper is, however, to make it possible to create a mapping based on bibliographic coupling and covering all selected journals.
Analogously to document mapping based on bibliographic coupling, all items that appeared in the reference lists of papers published in the journal are taken into account., As references appear only once in the reference list of a paper, a binary approach was chosen assigning the values 0 or 1 according as the reference was shared or not by the two papers. We followed the same principle for journals since weighting according to multiple occurrences of shared references at the journal level resulted in just marginal deviations from the binary approach.
Figure 1 presents an example of reference links between two journals. Journal *A* has published 3 articles with six references in total but two papers refer to the same article (*R*4). Journal *A* has thus 5 distinct references. Journal *B* has 4 papers with six references in total, each pointing at a distinct publication. Journal *A* and *B* share 3 distinct references.

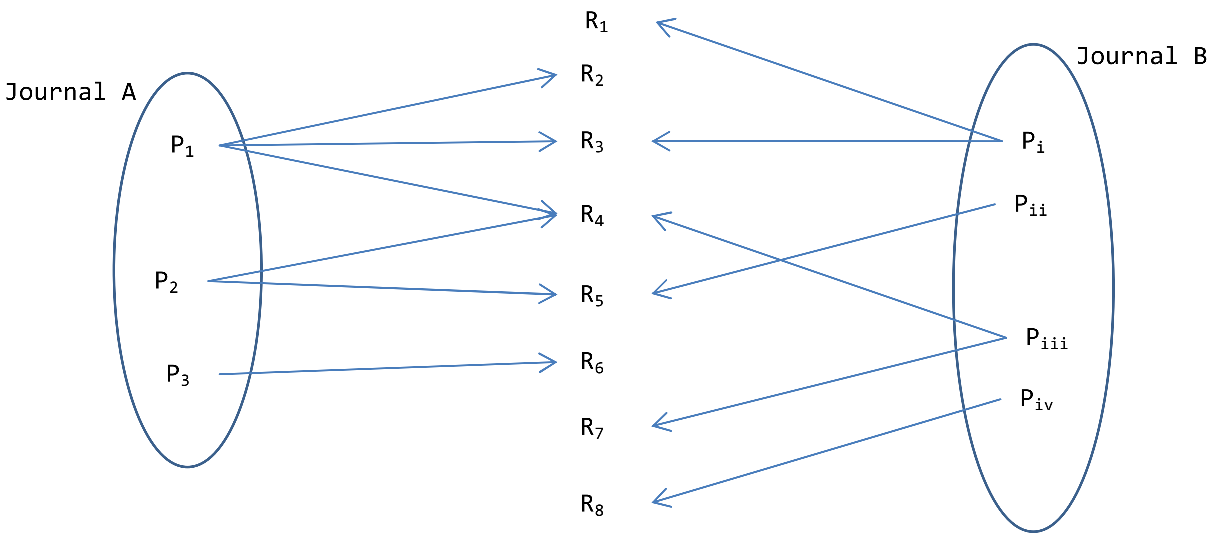

Figure 1. Graphic representation of bibliographic coupling between journals

To express the strength of a link between two journals we calculated a first order similarity based on Salton's cosine measure. The mathematical derivation and interpretation of this similarity measure in the framework of a Boolean vector space model can be found in (Sen & Gan, 1983; Glänzel & Czerwon, 1996). As bibliographic coupling tends to produce very sparse similarity matrices we applied a second order similarity to reduce this effect. While the first-order similarity is based on the angle between two reference vectors, the second-order similarity is calculated as the cosine of the angle of two vectors holding the first order similarity between two journals. After the calculation of the second-order similarities, ten journals were removed from the set as they appeared to be singletons without any link to the other journals in the set. The network thus included 8272 journals in total.

Hierarchical clustering with Ward's agglomeration method was used to create a hard clustering of all the journals. Given the rather limited set of entities to be clustered, Ward's method already proved its validity in many studies. This method does not provide any automated optimum number of clusters so that the decision was made on the basis of the dendrogram and the silhouette statistics (Rousseeuw, 1987). As Ward assumes distance measures instead of similarities we converted the similarities to distances before clustering.

Results

In this section we present the results of the clustering and discuss the validity of the partitioning of journal set. As pointed out in the previous section, a dendrogram and a silhouette-value plot were used to select an appropriate number of clusters. The two diagrams are presented in Figure 2 and Figure 3. Three different levels were chosen. The dendrogram holds strong arguments for a six cluster partitioning while the silhouette plot shows a first peak at 7 clusters. For the highest hierarchical level in the following analysis we use the six cluster solution. At a lower level, the silhouette plot suggests the solutions with 14 and 24 clusters, respectively. Both will be described in subsequent subsections.

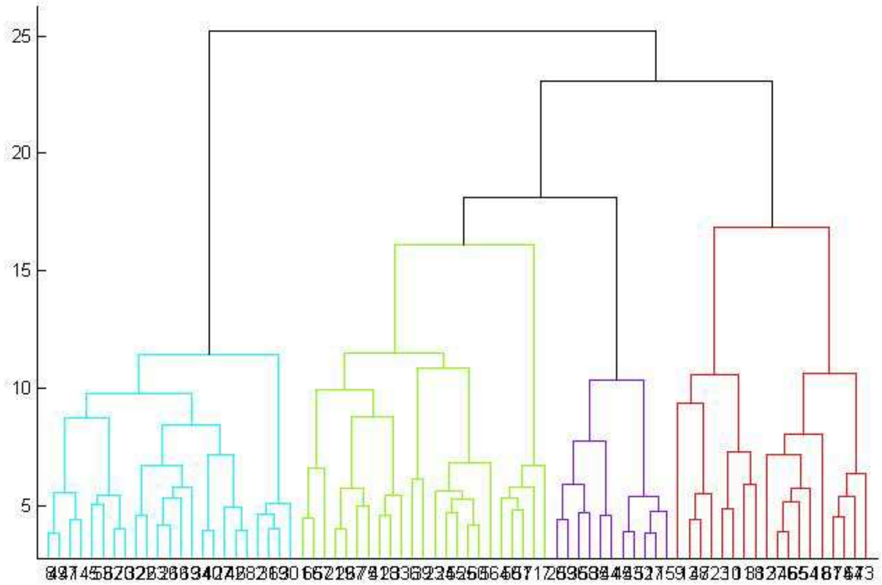

Figure 2. Dendrogram for hierarchical clustering of the 8272 journals based on Ward's method [Data sourced from Thomson Reuters Web of Knowledge]

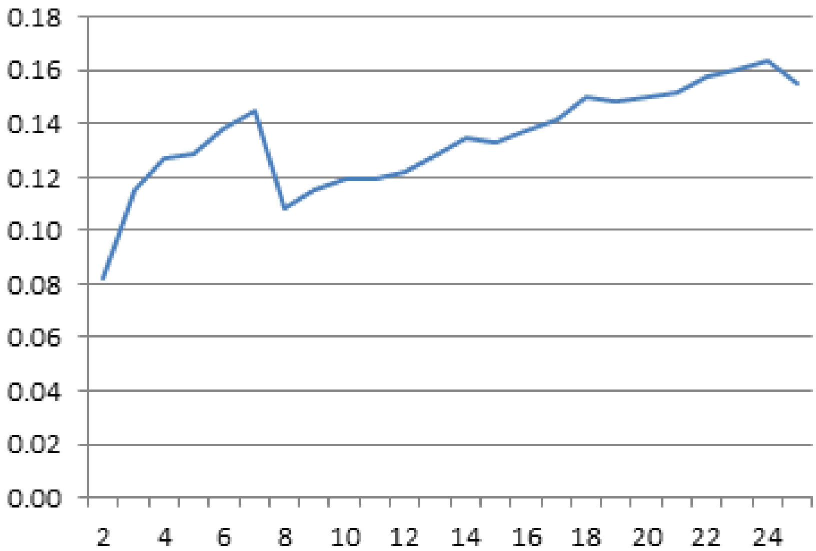

Figure 3. Mean Silhouette values for solutions of 2 up to 25 clusters, with local maxima at 7, 14 and 24 clusters [Data sourced from Thomson Reuters Web of Knowledge]

For the evaluation of the specific cluster solution we can rely on the silhouette graphs presented in Figure 4. Each graph presents the silhouette values of the journals in the respective cluster. For each journal a silhouette value is calculated. These values range between 1 and -1 where positive values indicate an appropriate clustering of the journals. Journals are grouped by cluster and ordered from highest silhouette value to lowest. As a consequence the graph gives a good profile of the quality of each cluster. A larger area at the positive side of the vertical axis thus represents a better partitioning. The most favourable situation is found in the six-cluster solution. Here most journals are assigned to the appropriate cluster and only the second cluster has a larger share of negative values (cf. left-most diagram in Figure 4).

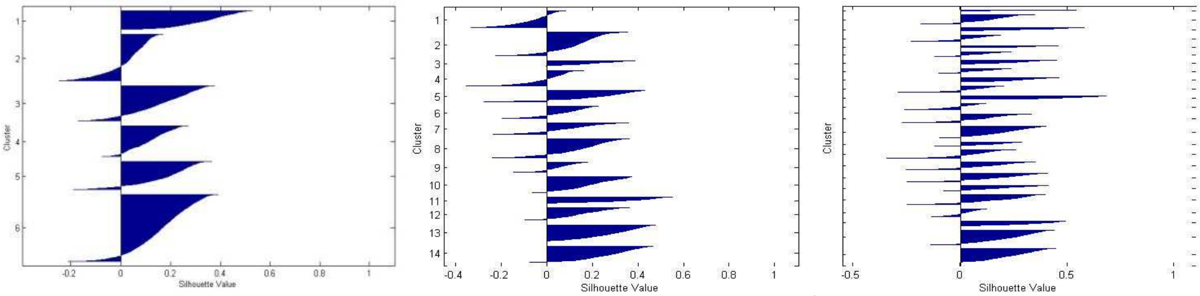

Figure 4 Silhouette values of three distinct clustering solutions with 6, 14 and 24 groups (from left to right) [Data sourced from Thomson Reuters Web of Knowledge]

Cluster Description

Unlike in lexical or hybrid citation-textual methods, where clusters can be labelled and described using the textual component, e.g., the best terms or keywords, pure citation-based approaches are put at a severe disadvantage if the content of the clusters have to be described. In order to find an acceptable solution, we decided to use the journal-based subject-classification scheme developed in Leuven (Glänzel & Schubert, 2003). This solution proved most advantageous since both clustering and classification scheme are based on journal assignment. Table 1 presents the hierarchical structure of the three level partitioning. For each cluster the number of journals is mentioned. The labels for the higher levels can be deduced from the lowest level. These labels are taken from the Leuven classification system . The label from the most prominent subject category has been assigned to the corresponding cluster.
Another way to describe the cluster is by using core journals. This notion can be analogously defined as core documents introduced by Glänzel & Czerwon (1996) and extended by Glänzel & Thijs (2011). In this particular application, a core journal can be identified as journal with at least *n* links with other journals of at least a given strength *r* on the second order similarity measure. For the identification of core journals in each cluster we set the number of strong links to at least half the set of journals in the cluster. As we are using second order similarities this choice is not unreasonable. The value of the strength is chosen such that 12 journals within each cluster comply with both criteria. This means that for more dense clusters the choice of appropriate *r*-value is higher than in clusters where the journals are not as strongly linked. Cluster 21 labelled as ‘Arts & Humanities’ is such a cluster where a lower value of *r* was required to retain twelve journals. This is a result of the specific citation behaviour in the humanities, where citations play a somewhat different role than in the sciences (cf. Glänzel & Thijs, 2011). A list of selected core journals for each cluster is given in Table 2.
Concerning the results, two striking observations could be made. Above all, chemistry is at each level a separate cluster. One might expect that at the highest level, chemistry is merged with Physics but we found different patterns. The second noteworthy observation concerns cluster 17 (Public Health & Nursing). This is a cluster within the ‘Psychology – Neuroscience’ cluster at the highest, six-cluster level. In other partitions or subject classification systems this is attributed to Non-Internal Medicine.

Table 1. Hierarchical structure of the three level partitioning with labels *l(i)* and number of journals *n(i)* according to the level with 6, 14 and 24 clusters [Data sourced from Thomson Reuters Web of Knowledge]

l(6)	*n(6)*	*l(14)*	*n(14)*	*l(24)*	*n(24)*	*Leuven subfield*
I	*n=691*	**n**	*n=691*	**24**	*n=691*	Chemistry; Material Science
II	*n=1704*	**c**	*n=268*	**19**	*n=268*	Geosciences; Geography
		d	*n=632*	**15**	*n=226*	Physics; Astronomy & Astrophysics;
				16	*n=406*	Engineering; Classical Physics
		k	*n=272*	**22**	*n=272*	Pure Mathematics
		l	*n=532*	**1**	*n=80*	Statistics & Probability
				2	*n=452*	Computer Science; Applied Mathematics
III	*n=1285*	**g**	*n=487*	**7**	*n=207*	Neuroscience; Neurology
				8	*n=280*	Psychology; Psychiatry
		h	*n=798*	**17**	*n=381*	Public Health; Nursing
				18	*n=417*	Social Psychology; Therapy; Counseling
IV	*n=1128*	**i**	*n=428*	**21**	*n=428*	Arts & Humanities
		j	*n=700*	**3**	*n=170*	Management; Marketing; Innovation
				4	*n=337*	Sociology; Social & Political Sciences; Law
				11	*n=193*	Economics; Accounting;
V	*n=1032*	**e**	*n=492*	**20**	*n=492*	Biology
		f	*n=540*	**9**	*n=225*	Agriculture; Plant Science
				10	*n=315*	Microbiology; Biotechnology; Food Science
VI	*n=2432*	**a**	*n=712*	**5**	*n=137*	Veterinary Sciences; Animal Sciences
				6	*n=251*	Immunology; Respiratory Medicine
				12	*n=324*	Non-Internal Medicine;
		b	*n=1007*	**13**	*n=432*	Haematology; Oncology; Surgery; Radiology
				14	*n=575*	Internal Medicine; Cardiovascular Medicine
		m	*n=713*	**23**	*n=713*	Biosciences; Biomedical Research

Cluster Structure

To visualise relations between the 24 clusters we created an additional map. Figure 5 shows these relations. The link between the clusters is based on bibliographic coupling. Also for this map we used a binary approach just as we did for the journals. The map was drawn in *Gephi* using the 'Force Atlas 2' layout method. The thickness of the link represents the similarity. The colours represent the six cluster solution. Here we see the central position of the chemistry cluster between physics, biology and life sciences (especially biosciences and biomedical research). Given the strong links with the three groups the separation of chemistry from physics seems justified.

Cluster 17 (Public Health – Nursing) is linked to several (psychology – neuroscience clusters) medical clusters. This position of the topic is interesting and deserves more attention.

Table 2. Three core journals per cluster (selection does not imply any ranking) [Data sourced from Thomson Reuters Web of Knowledge]

#	*Journal title*	#	*Journal title*
1	biometrika canadian journal of statistics-revue canadienne de statistique computational statistics	13	annals of surgical oncology diseases of the esophagus world journal of gastroenterology
2	elektronika ir elektrotechnika ieee transactions on industrial informatics ieee transactions on systems man and cybernetics part a-systems and humans	14	american journal of the medical sciences annals of medicine clinical and investigative medicine
3	california management review ieee transactions on engineering management journal of business research	15	canadian journal of physics central european journal of physics chinese physics letters
4	china quarterly environment and planning c-government and policy environmental politics	16	acta mechanica sinica advances in engineering software comptes rendus mecanique
5	archivos de medicina veterinaria arquivo brasileiro de medicina veterinaria e zootecnia polish journal of veterinary sciences	17	applied nursing research bmc health services research contemporary clinical trials
6	clinical and vaccine immunology fems immunology and medical microbiology international journal of immunopathology and pharmacology	18	american psychologist canadian journal of behavioural science-revue canadienne des sciences du comportement canadian psychology-psychologie canadienne
7	annals of neurology brain research brain research bulletin	19	canadian journal of earth sciences comptes rendus geoscience earth-science reviews
8	biological psychology developmental neuropsychology international journal of psychophysiology	20	african zoology biological invasions israel journal of zoology
9	annals of applied biology botanical studies journal of horticultural science & biotechnology	21	american historical review new literary history critical inquiry
10	applied biochemistry and biotechnology biotechnology and bioprocess engineering engineering in life sciences	22	archiv der mathematik bulletin des sciences mathematiques chinese annals of mathematics series b
11	canadian journal of economics-revue canadienne d economique economic inquiry australian economic review	23	acta biochimica et biophysica sinica advances in experimental medicine and biology biochemical and biophysical research communications
12	journal of burn care journal of dental research physikalische medizin rehabilitationsmedizin kurortmedizin	24	acta chimica sinica acta physico-chimica sinica chemical journal of chinese universities-chinese

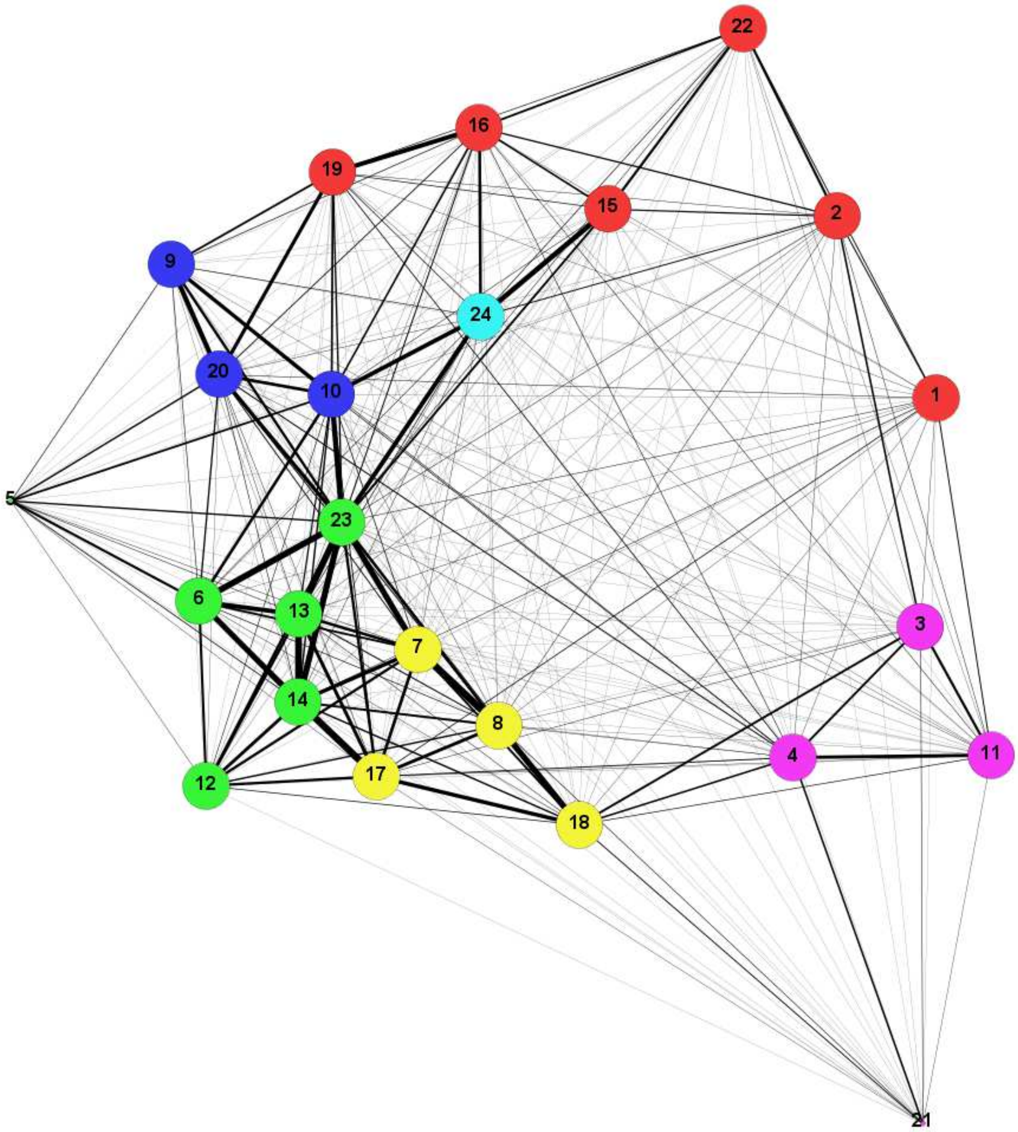

Figure 5. Map with 24 clusters based on bibliographic coupling
[Data sourced from Thomson Reuters Web of Knowledge]

Comparison with the Leuven classification system

The partitioning in 14 clusters is suitable for comparison with the 15 main fields in the Leuven classification system. In this latter system a sixteenth field exists, namely the multidisciplinary sciences but this has been omitted from this analysis for obvious reasons. An important difference between the two systems is that the Leuven classification allows multiple assignments of journals to fields. With the applied Ward methodology this is not possible for the clustering developed in this paper. Despite these multiple assignments we used the Jaccard Index to measure the concordance between the two journal classifications. The results are presented in Table 3. For most fields a good mapping with one of the fourteen clusters can be found. Fields 'Biosciences' and 'Biomedical Research' are jointly mapped on cluster 'm' which explains the reduction by one field. But journals assigned to the field 'Non Internal Medicine Specialties' are spread across four clusters ('a', 'b', 'g', 'h') according to the 14-cluster solution (see column *l*(14) in Table 1). 'Neurosciences & Behaviour' is split into two clusters ('g' and 'h'), both these have also a link to 'Non internal medicine'. Cluster ‘h’ also has a link to social sciences. In this last cluster we see the common focus in medicine, psychology and social and community issues. Most of the journals assigned to the field ‘General, Regional & Community Issues’, that have no relevance to medicine or psychology, are assigned to cluster ‘j’.

Table 3. Concordance measured with Jaccard Index between 14 clusters and the Leuven subject classification system in 15 disciplines [Data sourced from Thomson Reuters Web of Knowledge]

	a	b	c	d	e	f	g	h	i	j	k	l	m	n
Agriculture & Environment	0.03		0.05	0.02	0.08	0.30				0.05			0.01	
Biosciences	0.02	0.02			0.07	0.04	0.02						0.32	0.01
Chemistry				0.08		0.06							0.04	0.48
Engineering			0.02	0.15		0.01		0.01		0.03	0.02	0.35	0.01	0.05
Geosciences & Space Sciences			0.37	0.12	0.02					0.03				0.01
Mathematics				0.02						0.03	0.44	0.18		
General & Internal Medicine	0.08	0.36					0.01	0.03					0.04	
Non-Internal Medicine Specialties	0.19	0.19				0.01	0.13	0.16					0.02	
Neurosciences & Behaviour					0.01		0.28	0.21						
Economical & Political Issues			0.02					0.03	0.08	0.43				
Physics				0.26							0.04	0.03		0.12
Biomedical Research	0.04	0.09				0.02	0.04	0.02					0.20	
General, Regional & Community Issues	0.04		0.01				0.02	0.15	0.08	0.16				
Arts & Humanities							0.03	0.02	0.55	0.02				
Biology	0.13				0.34	0.14							0.04	

Comparison with ESI

A 24 cluster solution can be compared with the 22 categories from the classification of Thomson Reuters’ Essential Science Indicators (ESI). Unlike most classification schemes, this classification system provides just like our cluster solutions a structure, where each journal is assigned to only one single category. This means that we can calculate the concordance between the two classification systems. The appendix presents the distribution of journals across both systems. Janssens et al. (2009) showed very low mean silhouette values for the ESI category system in a space with respectively textual distances, cosine similarities of cross-citation vectors and combined distances. As can be seen from the table in the Appendix the same situation occurs here as well. Also in the present study, not all clusters have a unique counterpart in the ESI classification system and vice versa (cf. Janssens et al., 2009). Notably, the ESI fields *clinical medicin*e and *engineering*, *mathematics* and *social sciences, general* are almost uniformly spread over numerous clusters.

Conclusions

The application of the second-order similarities proved to be surprisingly stable, and resulted in high-quality cluster solutions. Notably the six-cluster solution provided the best result. The number of singletons, that had to be removed, was marginal: Only ten journals representing 0.12% out of the 8282 journals had to be removed from the classification. The main advantage of this method is that clustering can be made as soon as a new database volume is available. The only issue is the lacking cluster labelling that cannot directly be obtained from the method. As a substitute, intellectual classification schemes can be used as reference system. Cluster labelling was made on the basis of the Leuven-Budapest subject-classification scheme that also allowed the comparison with the existing two-level journal classification system developed in Leuven. In all, the results have been found to provide a well-balanced hierarchical system of 6–14–24 clusters.

The further comparison with the 22 field classification system of the Essential Science Indicators does, however, revealed some striking deviations. These concerned, above all, the fields of clinical medicine, engineering, mathematics and the social sciences. New

developments in computer science, neuroscience and psychology as well as in public health (cf. Glänzel & Thijs, 2011) do certainly contribute to such growing deviation.
The main objective of this study was to analyse whether the proposed methodology is appropriate for multi-level journal clustering and to what extent the solutions fit in the framework of traditional subject classification. Further comparison with other solutions such as cross-citation and hybrid methods will be part of future research.

Acknowledgement

Lin Zhang acknowledges the support from the National Natural Science Foundation of China under Grant 71103064.

References

Ahlgren, P., & Colliander, C. (2009). Document–document similarity approaches and science mapping: experimental comparison of five approaches. *Journal of Informetrics*, 3(1), 49–63.

Bassecoulard, E. & Zitt, M., (1999). Indicators in a research institute: A multi-level classification of scientific journals. *Scientometrics*, 44(3), 323–345.

Bastian M., Heymann S. & Jacomy M. (2009). Gephi: an open source software for exploring and manipulating networks. International AAAI Conference on Weblogs and Social Media.

Garfield, E. (1998). *Mapping the world of science*. The 150 Anniversary Meeting of the AAAS, Philadelphia, PA, http://www.garfield.library.upenn.edu/papers/mapsciworld.html.

Glänzel, W. & Czerwon, H.J. (1996), A new methodological approach to bibliographic coupling and its application to the national, regional and institutional level. *Scientometrics*, 37(2), 195–221.

Glänzel, W., Schubert, A. (2003), A new classification scheme of science fields and subfields designed for scientometric evaluation purposes. *Scientometrics*, 56 (3), 357–367.

Glänzel, W. & Thijs, B. (2011), Using 'core documents' for the representation of clusters and topics. *Scientometrics*, 88 (1), 297–309.

Janssens, F. (2007). *Clustering of scientific fields by integrating text mining and bibliometrics*. Ph.D. Thesis, Faculty of Engineering, KU Leuven, Belgium. http://www.hdl.handle.net/1979/847

Janssens, F., Glänzel, W. & de Moor, B. (2008), A hybrid mapping of information science. *Scientometrics*, 75(3), 607–631.

Janssens, F., Zhang, L., De Moor, B. & Glänzel, W., (2009) Hybrid clustering for validation and improvement of subject-classification schemes. *Information Processing and Management*, 45(6), 683-702.

Jarneving, B. (2005). *The combined application of bibliographic coupling and the complete link cluster method in bibliometric science mapping*. Ph.D. Thesis, University College of Borås/Göteborg University, Sweden.

Leydesdorff, L. (2004), Clusters and maps of science journals based on bi-connected graphs in Journal Citation Reports. *Journal of Documentation*, 60(4), 371–427.

Narin, F., Carpenter, M., & Nancy, C. (1972). Interrelationships of scientific journals. *Journal of the American Society for Information Science*, 23(5), 323–331.

Narin, F. (1976), *Evaluative Bibliometrics: The Use of Publication and Citation Analysis in the Evaluation of Scientific Activity*. Computer Horizons, Inc., Washington, D.C.

Rousseeuw, P.J. (1987). "Silhouettes: a Graphical Aid to the Interpretation and Validation of Cluster Analysis". Computational and Applied Mathematics 20: 53–65

Sen, S.K. & Gan S.K. (1983), A mathematical extension of the idea of bibliographic coupling and its applications. *Annals of Library Science and Documentation*, 30, 78–82.

Thijs, B., Schiebel, E., & Glänzel, W. (2013), Do second-order similarities provide added-value in a hybrid approach? *Scientometrics*, DOI 10.1007/s11192-012-0896-1, in press.

Zhang, L., Janssens, F., Liang L.M. & Glänzel, W. (2010). Journal cross-citation analysis for validation and improvement of journal-based subject classification in bibliometric research. *Scientometrics*, 82(3), 687–706.

Appendix

Distribution of journals across 24 clusters and 22 ESI fields [Data sourced from Thomson Reuters Web of Knowledge]

ESI field	*24 cluster solution*																							
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*	*21*	*22*	*23*	*24*
agricultural sciences	2	7	3	4	2	5	1	4	26	33	4	3	4	11	2	2	9	4	4	6	1	2	8	7
biology & biochemistry	4	5	6	8	3	3	6	4	7	23	2	3	10	22	7	8	9	6	4	23	11	6	100	21
chemistry	1	11	8	8	8	4	6	6	7	22	5	8	10	8	10	15	4	16	9	13	23	8	33	148
clinical medicine	8	36	22	39	8	68	50	26	18	17	15	119	178	214	26	37	87	42	23	30	50	14	99	74
computer science	1	73	7	4	2	5		4	2	2		1	8	6	5	19	5	6	2	7	4	11	7	9
economics & business	1	7	22	10		5	2	5	2	9	53	11	6	8	4	5	9	13	8	12	13	7	9	19
engineering	11	127	12	19	3	11	2	15	6	28	9	10	22	29	17	84	14	21	16	26	14	27	28	45
environment/ecology	3	4	2	8	3	3	1		16	29	2	2	7	4	1	9	4	8	14	44	8	4	8	7
geosciences	2	3	1	7	2	3	2	5	2	8	2	7	2	11	7	20	13	4	66	12	8	10	9	18
immunology	1	3		2	1	25		1	1	2	1		3	7		2	4	1	1	1	2	3	5	
materials science	1	9	4	5	1		4	1	5	8	4	5	2	6	5	17	3	8	4	9	12	5	10	71
mathematics	24	27	2	8	1	4	3	4	2	2	4	4	7	5	5	16	11	10	6	8	16	74	12	20
microbiology		1	1		1	12	2	3	1	19		2	6	2	1	4	2	1	2		4	3	10	1
molecular biology & genetics	1		2	4	3	7	3	5	3	3	1	3	8	1	8	5	2	2	3	16			88	4
multidisciplinary				1	1	1		1			1		1		1	3	1			3	3		4	3
neuroscience & behavior		5	1	2		1	70	32	2	3	2	7	7	5		2	4	3	1	8	3	1	7	6
pharmacology & toxicology	2	1	1	1	1	2	1	2	1	3		2	5	8	2	3	6	4	2	4	2	4	55	10
physics		15	2	7	2	3	2	4	9	6	1	14	4	2	49	16	2	8	7	10	14	13	7	29
plant & animal science	1	9	8	19	49	9	6	6	46	13	6	9	8	13	5	20	13	19	23	152	17	9	42	21
psychiatry/psychology	4	9	8	11	2	4	9	76	2	8	9	13	6	14	5	10	17	104	5	9	15	5	13	12
social sciences, general		26	21	119	11	18	2	28	7	16	28	21	19	17	8	28	70	89	18	26	48	17	20	37
space science			1		1		2		1		1			3	20	4	1		2	1	1	1	4	1