Journal cross-citation matrices reconsidered. Tracing the role of individual journals in the communication network

Zhang Lin[1,2] Wolfgang Glänzel[1,3]

14 June 2008

Abstract

This study aims at uncovering structural patterns of information flow among scientific journals on basis of bibliometric means and statistical methods of data analysis. Beyond measuring the individual journals position within the communication network, we shed light on their cognitive background as well. Language barrier and lacking internationality proved one of the main hindrances for integration into the communication network, and on the other hand, it was the document type that hinders journals from establishing self-links. Against our expectations, we have found a clear divergence between strongly interlinked and high-entropy journals.

1 Introduction

Yet in the era of e-communication scientific journals remain the most important channels of documented scholarly communication in the sciences. The information flow among actors at different levels of aggregation can successfully be measured and characterised by scientometric transaction matrices, particularly, by cross-citation matrices. Although this idea of using transaction matrices to analyse the structure of the underlying network is not new (e.g. Schubert and Glänzel 2006), so far little effort has been made to analyse the role of individual journals by means of bibliometric measures derived from the properties of a given transaction matrix. The present study aims at uncovering structural patterns of information flow among scientific journals on basis of bibliometric means and statistical methods of data analysis. Furthermore the attempt is made, beyond measuring the individual journals position within the communication network, also to shed light on their cognitive background as well. In particular, we have addressed the following questions. Which are the journals with extraordinarily high or low shares of self-citations (or self-references) and what are their characteristics? Which are the journals inclined to "in-links" much more than "out-links", and which are the opposite cases? Are there some strong asymmetric links between two journals, namely, one journal relied on another journal strongly in the aspect of citations or references? Which journals have many strong links with other journals? Do these form real nodes in the network and can they be distinguished from other journals with high citation entropy? Finally, what can be said about subject-related issues, document types, multi-disciplinarity, internationality and language barriers in the context with the above questions?

1 K.U. Leuven, Steunpunt O&O Indicatoren and Dept. MSI, Leuven (Belgium)
2 WISE Lab, Dalian University of Technology, Dalian (China)
3 Hungarian Academy of Sciences, Institute for Research Policy Studies, Budapest (Hungary)

H. Kretschmer & F. Havemann (Eds.): Proceedings of WIS 2008, Berlin
Fourth International Conference on Webometrics, Informetrics and Scientometrics & Ninth COLLNET Meeting
Humboldt-Universität zu Berlin, Institute for Library and Information Science (IBI)

2 Method

In contrast to other approaches (e.g. Leydesdorff 2004a, 2004b, 2006) our study is not based on the Journal Citation Reports (JCR). Although the JCR immediately provides transaction matrices, the use of the WoS allows the determination of citation on a paper-by-paper basis and the exclusions of not relevant documents (see Data section). Moreover our approach allows extending the analysis to the Arts & Humanities Citation Index (AHCI) since the JCRs are issued only for the sciences and social sciences.

Besides the original cross-citation matrix $\mathbf{A} = \{a_{ij}\}$ containing the (directed) citation links a_{ij}, i.e., the number of citations the journal i receives from journal j, we have also used the symmetrised matrix $\mathbf{A' = (A + A^T)}/2$. While the original (asymmetric) matrix is used for the analysis of deviations of citation patterns from reference patterns of individual journals, the symmetric matrix serves for the identification of general nodes in the network.

In particular, the following three groups of measures will be applied for each individual journal.

I. Measures of journal self-link:

1) Journal self-citation ratio (SCR)[4]

$$SCR_i = \frac{SC_i}{TC_i}$$

2) Journal self-reference ratio (SRR)

$$SRR_i = \frac{SC_i}{TR_i}$$

These two measures are used to study extent of a journal's relative isolation as well other document-specific characteristics in the communication network of scientific journals. According to Schubert and Braun (1993), extremely low share of self-references are typical of review journals. On the other hand, a high share of self-references does not necessarily imply a high share of self-citations as well. Among others, specialisation and cross-disciplinarity can affect the symmetry of self-references and –citations.

3) Journal self-link Salton (SLS)

$$SLS_i = \frac{SC_i}{\sqrt{TC_i * TR_i}}$$

This indicator corrects for the possible asymmetry of the previous two measures.

II. Measures of journal links strength:

4) Citation links strength between journal i and j (CL_{ij})

$$CL_{ij} = \frac{a_{ij}}{\sqrt{TC_i * TR_j}}$$

These indicators measure the strength of the citation links between two individual journals in the asymmetric matrix, which are directional as a cite from journal i to journal j differs from a cite from j to i. Through these indicators, we can compare the emission (the referencing) and the reception (the citedness) side of individual journals, and can also detect some extremely asymmetric links between two journals.

5) Symmetrised link strength between journal i and j (SL_{ij})

$$SL_{ij} = \frac{a_{ij} + a_{ji}}{\sqrt{[TC_i + TR_i] * [TC_j + TR_j]}}$$

The indicators are the citation links between two individual journals in symmetric matrix. If a journal has many strong links with other journals, one can assume that this journal plays an important part as node in the journal-citation network.

6) Journal citation entropy according to

[4] Here i and j denote journals, SC_i the number of self citations of journal i, TC_i the total number of citations of journal i, TR_i the total number of references of journal i and a_{ij} the number of citations of journal i receives from journal j.

symmetrised links (EL)

$$EL_i = - \sum_{j=1}^{n} \frac{a_{ij} + a_{ji}}{TC_i + TR_i} * \log_2 \frac{a_{ij} + a_{ji}}{TC_i + TR_i}$$

This indicator measures in how far references/citations spread among other journals but unlike the previous indicators, it does not measure centrality. Therefore, indicators 4 and 5 complement each other.

3 Data

The data have been collected from the *Web of Science* (WoS) of Thomson Scientific (Philadelphia, PA, USA) for the period 2002-2006. We have selected all journals covered in this period. Altogether 9527 journals were found having publications in this period. Only four document types, namely *articles, notes, letters* and *reviews* were taken into account. Citations have been summed up from the publication year till 2006.

4 Results

In the cross-citation matrix covering 9527 periodicals and serials, there are more than 2,000,000 unique citation relations among different journals out of the total 90,000,000 possible citation relations (2.41%). The cross-citation matrix is thus extremely scarce; more than 97% of the cells in the matrix are empty. We have found an average of 8.8 citations per journal-journal relation. The most citations are received from *Journal of Biological Chemistry* (398013 citations). 80% of the total citations are received from the 14% most cited journals. Though this huge journal cross-citation matrix can be analyzed from different views and through various ways, in this study, we have our main interests in analyzing journal self-citations, journals with relatively more strong links and those having high citation entropies .

4.1 Analysis of journal self-citations

Self-citations made up a considerable share of the journal cross-citations under study. The total number of self-citations was up to almost 2,500,000, which constituted approximately 13% of all citations. 6441 journals, that is, 67.6% of all journals received the most citations from themselves.

Before finding the extremely high or low self-cited journals, we had set the following thresholds:

During the period of 2002-2006:

- The total number of publications of the journal (TP) was at least 50, namely, TP $\geqq$ 50;
- The journal's total number of citations was at least 30, namely, TC $\geqq$ 30.

Eventually, we got 7835 journals, i.e., 82.2% of all journals, which met the above thresholds.

Out of these journals, there were 136 journals having self-cited ratio over 60%. There are various factors for journals having high self-cited rate. For some journals which are leaders in their fields, the factor may be the consistently high quality of the papers they publish. We have found different causes for this phenomenon. In particular, some journals are highly specialised and there are also journals having high share of self-citations resulting from their editorial practice.

In terms of the 136 highly self-cited journals, all of the top 10 journals were found relatively national journals, namely, for one particular journal, more than half of its authors came from a particular country or region, or over 50% of its languages were native languages (not English). Roughly three quarters of the 136 journals could be considered as relatively national. The result indicates that lacking internationality and language barrier proved one of the main hindrances for integration into the communication network. The share of science journals, social science journals and humanity journals were respectively 60%，33% and 15%. The sum of these shares was over 1 as some journals were assigned to more than one category. In terms of the specific categories, there was a weak correlation between self-cited rate and specific categories. On one hand, the population of highly cited journals was spread throughout different categories; on the other hand, there was no indication that journals in large or broadly defined categories had lower overall self-cited rate compared to journals in

smaller categories. So the lack of large number of closely related literatures for journals in narrowly defined categories was not a significant reason for high self-cited rate. Finally low visibility and low impact can be considered a third factor affecting foreign citations.

In contrast to the highly self-cited journals, relatively low self-citation share is normally considered as one of the characteristics of high-quality journals, especially for the journals having a large number of citations. A relatively low self-citation rate can also be due to different factors. Among the 7835 journals, there were 279 journals of which self-cited rate not exceeding 1%. For these extreme cases, most of them were international journals and about one third had published a large share of reviews and/or letters (more than 40% among their documents). It was the document type that hinders journals from establishing self-links. This result demonstrates that self-citation is more common among original articles than among review papers or letters. Though the latter documents are normally written by experts with considerable previous research work in the reviewed field, they apparently avoid extensive self-citation. This result is in line with earlier observations by Schubert and Braun (1993).

Some similar analyses have been applied in the self-reference ratio. We found 51 journals with self-reference rate at least 60% and 696 journals with the rate over or equal to 30%. There are also several reasons for a journal having high self-reference share. For instance, a paper relating to a particular topic has more chances to be submitted to a journal which has already published many such papers in that particular field and consequently a big share of references in the submission will come from that journal. Among the top 100 journals with high self-reference ratio, more than 40% were international journals. The correlation between nationality and high self-reference rate was not as strong as it with high self-cited rate. More than half of the journals were dedicated to social science or humanities (social science 33% and humanities 28%), and only 47% to science. Compared to the analysis of highly self-cited journals, humanity journals have much more inclination for having large share of self-references.

In terms of the journals with low self-reference rate, most of them were international journals. Different from lowly self-cited journals, there was not so strong correlation between the document types and low self-reference shares. Taken the 114 journals of which self-reference rate were 0 into account, only 15 journals had published a large share of reviews and/or letters (more than 40% among their documents).

4.2 Analysis of the direction of journal citations

The duality of referencing and citation is a central feature of citation processes. It is of great interests to compare and analyze the emission (the referencing) and the reception (the citedness) side of journals. In our study, the network of journals citation links was first represented by directed graphs based on the unsymmetrical journal citation matrix. The analysis of comparing two different directions according to out-links (references) and in-links (citations) will reveal interesting cognitive aspects of information transferring among scientific journals.

About one third of the journals had more citations than references. This means that roughly two thirds of journals were prone to “emission” than “receiving”.

For the analysis of extreme cases, we respectively selected top 100 journals having big value of (TC-TR)/(TC+TR) and (TR-TC)/(TC+TR), under the threshold of TC+TR ≥ 500. The first group of journals were called “receivers” and the latter ones were “emitters”. In the first group, categories relatively spread around. Journals assigned to “Computer Science” were somewhat predominant (13%), four times than its share among all the journals (3%). While for the latter kind of extreme cases, the categories were much more concentrated. Journals of “Clinical Medicine” and “Biology & Biochemistry” occupied 44% and 10%, but “Computer Science” disappeared. We can conclude that Computer Science journals act as “receivers”, while “Clinical Medicine” journals are rather emitters. Among the extreme “receivers”, we found several high impact journals of multidisciplinary: *Nature, Science, Proceedings of the National Academy of Sciences of the United States of America*. For such high-quality and high-impact journals, it’s quite understandable that they receive much

more citations than the references they emit.

As a result of the directional analysis, we found 70 journal pairs with strongly asymmetric links, namely, the journal pairs met the following thresholds: 1) $TP_i \geqslant 50$; $TP_j \geqslant 50$; 2) $TC_i \geqslant 30$; $TC_j \geqslant 30$; 3) $a_{ij} \geqslant 20$; 4) $CL_{ij} - CL_{ji} \geqslant 0.1$ (See formula 4). In such cases, journal *i* heavily relies on journal *j* in terms of citations or, journal *j* strongly relies on journal *i* with respect to references.

Taken the 70 asymmetric pairs of journals into consideration, we found that over half of journals were assigned to social science or humanities. The subject categories were relatively concentrated. There were 14 pairs in "Social Science, General" and 13 pairs in "Psychiatry & Psychology". Only few journals were related to the large fields such as biology or medical science. Of course, in large fields like the life sciences the great number of journals numerically decreases the chance of having strong links with particular journals.

In terms of the internationality, among the top 10 pairs, 90% were national journals and in the top 70 pairs, over 60% journals were found rather national. Considering the extreme cases that we have found, we can make a conclusion that the national and social science or humanity journals indeed occupied a high share among the journal pairs with strong asymmetric links.

4.3 Analysis of journals' citation entropies and strong links

Diametrically opposing the cases of isolation in section 4.1 (high share of self-citations or self-references), in this part we mainly compared another two important functions journals may hold within the communication network. Here we used the symmetrised matrix $\mathbf{A'} = (\mathbf{A} + \mathbf{A}^T)/2$ for mapping cognitive links independently of the direction of transaction, where $\mathbf{A}^T$ denotes the transposed matrix of **A**. For all the journals under study, we had set the thresholds of $TP \geq 50$ and $TC \geq 30$ (during 2002-2006).

4.3.1 Analysis of journal strong links

When links between journal *i* and journal *j* meet the thresholds of $a_{ij} + a_{ji} \geq 20$; $SL_{ij} \geq 0.025$ (see formula (5)), they were considered as "strong links", where self-links had been excluded.

According to the above definition, we counted the number of strong links for each journal. The biggest number of strong links was found from *Lecture Notes in Computer Science* (LNCS), in which there were 43 strong links. In figure 1, we can observe the descending exponential trend of the number of journals according to the increasing number of strong links. However, *Lecture Notes in Computer Science* is actually not a journal; it is a serial, i.e., a book series in which a large amount of paper appear annually. Therefore the extreme case of LNCS had been excluded.

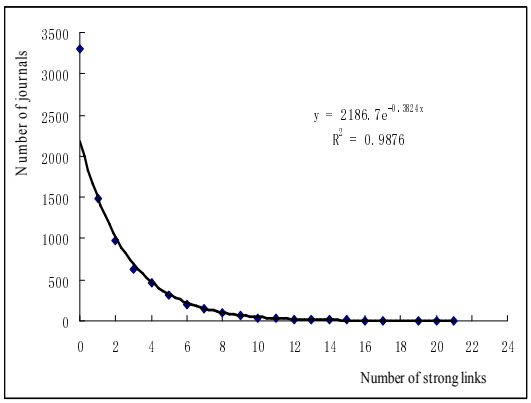

Figure 1: Graph of number of journals according to different number of strong links

There were 173 journals having at least 10 strong links and 990 journals having no less than 5 strong links. In table 1, we listed the top 30 journals with at least 15 strong links, where SL means number of strong links.

Out of the 173 journals with at least 10 strong links, 101 journals (58.4%) were international while the other 72 journals (41.6%) were considered to be American, as more than half of their authors were from USA. There was no one national journal apart from American, and among all the 173 journals, authors from USA took the most shares of authors in 94% journals, which indicated the absolutely leading place of the American authors among the top journals with

more strong links. Over 80% of the top 173 journals were science journals. The only one humanity journal was *Journal of Archaeological Science*, which had 11 strong links. Among these journals we had mainly found traditional core journals in the fields they represent. Most fields of the sciences, applied sciences but also fields in the social sciences were represented.

Table 1: Top 30 journals with more SL

Journal	*SL*	*Journal*	*SL*
LECT NOTES COMPUT SC	43	ASTROPHY S J	16
OPHTHALMO LOGY	21	CIRCULATI ON	15
AM J OPHTHALMO L	20	AM ECON REV	15
PHYS REV D	19	J IMMUNOL	15
IEEE J SEL AREA COMM	18	EUR J OPER RES	15
ENVIRON PLANN A	17	J PERS SOC PSYCHOL	15
J BONE JOINT SURG AM	17	INT J SOLIDS STRUCT	15
COLUMBIA LAW REV	17	J AM CHEM SOC	15
J AM ACAD DERMATOL	16	MAR ECOL-PROG SER	15
J UROLOGY	16	J APPL PSYCHOL	15
LECT NOTES ARTIF INT	16	PHYS REV LETT	15
INT J HEAT MASS TRAN	16	INORG CHEM	15
PROG HUM GEOG	16	TETRAHED RON LETT	15
J FINANC	16	INT J PROD RES	15
PHYS LETT B	16	PHYS REV B	15

4.3.2 Analysis of journal entropies

The entropy of each journal measures the degree of its isolation, or the distribution of its links to other journals. According to formula (6) in the methods part, we calculated the EL for each of the 7835 journals under study.

The traditional core journal in medical science *Lancet* had been found having the highest EL (9.780) amongst all the 7835 journals. The following five journals were *Annals of the New York Academy of Sciences* (9.777), *New England Journal of Medicine* (9.746), *Science* (9.642), *Brazilian Journal of Medical and Biological Research* (9.564) and *Nature* (9.511).

In figure 2 we can see that most journals had EL in the range of [5, 8]. The average EL of the 7835 journals was 6.256.

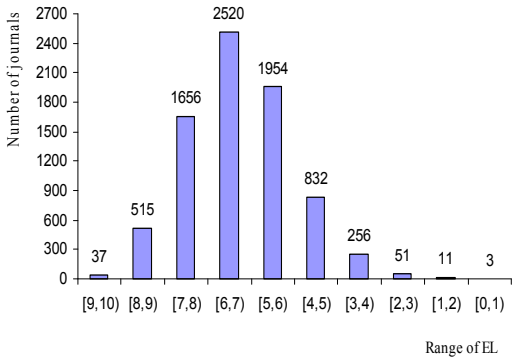

Figure 2: Number of journals according to different EL

Among the top 100 journals having high EL, 59 journals were international and 18 journals were considered to be relatively national apart from USA (See table 2). Normally national journals were supposed to be relatively more isolated, namely, high share of self-links and low degree of links diffusivity. But these 18 journals in table 2 had exactly opposing characteristics. Their high EL indicated that they had rather broad communication network.

In terms of categories, 98 journals of the top 100 journals belonged to sciences, only *TLS-THE Times Literary Supplement* to Humanity, and *European Journal of Epidemiology* to Social Science. Among the science subjects, 93% were related to biology or medical sciences, while almost all the rest belonged to "multidisciplinary sciences", including the top journals *Science*, *Nature* and *Proceedings of the National Academy of Sciences of the United States of America*. This result was again in line with the broad, cross and penetrative attributes of biology and medical sciences, as well as "multidisciplinary sciences". Taken into account the large number or scope of journals in these subjects, these journals had much more advantages for extending their communication network.

Table 2: National journals with high EL

Journal name	*EL*	*Nationality*
BRAZ J MED BIOL RES	9.56	Brazil
J KOREAN MED SCI	9.40	South Korea
YONSEI MED J	9.27	South Korea
BRIT MED BULL	9.21	England
ISRAEL MED ASSOC J	9.10	Israel
BRIT MED J	9.08	England
POSTGRAD MED J	9.07	England
CR BIOL	9.06	France
CHINESE MED J-PEKING	9.04	China
SWISS MED WKLY	8.96	Switzerland
PROG NAT SCI	8.87	China
J FORMOS MED ASSOC	8.87	Taiwan
B EXP BIOL MED+	8.86	Russia
INTERNAL MED	8.85	Japan
EUR J MED RES	8.84	Germany
ANN ACAD MED SINGAP	8.84	Singapore
J BIOSCIENCES	8.80	India
ACTA BIOCHIM POL	8.77	Poland

4.3.3 Comparison and extending analysis of journals with more strong links and those with high entropy

Against our expectations, we had not found a large overlap between journals with more strong links and those with high EL. Taken the 990 journals with at least 5 strong links into account, most journals had quite common EL. Only a few journals had relatively high EL, which were shown in table 3, where SL means number of strong links.

Table 3: Journals with more SL and high EL

Journal names	*SL*	*EL*
CIRCULATION	15	8.16
NEUROLOGY	10	8.28
GASTROENTEROLOGY	10	8.24
AM J HUM GENET	10	8.21
ZOOTAXA	6	8.58
CHEST	6	8.23
GUT	6	8.20
CANCER RES	6	8.08
AM J PSYCHIAT	6	8.00
PEDIATRICS	5	8.43
COMP BIOCHEM PHYS A	5	8.10
ENDOCRINOLOGY	5	8.02
ARCH GEN PSYCHIAT	5	8.02
AM J CLIN NUTR	5	8.01

The journals in Table 3 were quite outstanding in both of the spreading of their links and the number of the strong links. Almost all the journals were the core journals in fields related to medical science or biology. As known to all, biology and medical sciences were the most comprehensive and far-ranging fields, which include a great number of journals in Web of Science. Obviously, journals in these fields have much more choices and chances to communicate with other journals. Therefore, the super excellence of the 17 journals behaving in the communication networks was due to their own quite high qualities on one hand, while on the other hand, also had correlations with the comprehensive characteristics of the fields they belonged to.

In terms of journals with high entropies, the case was even more unexpected. Out of all the top 100 journals with high entropies, there was none journal having more strong links. The most number of journals' strong links was only 2, from *Toxicology Letters*. 90 journals had no strong links at all. The divergence between strongly interlinked and high-entropy journals was somewhat surprising, which revealed that it may be difficult for journals which spread links in a large scope to have more strong links to some particular journals.

In line with our expectation, all of the top 100 journals with high entropies had a low value of self-Salton (see formula (3)), as journals which had a very broad communication network were not supposed to contribute much to the links to themselves. Nevertheless, the analysis of the correlations between number of strong links and self-Salton was somewhat surprising. Among the 990 journals with no less than 5 strong links, 99 (10%) journals had the value of self-Salton over 0.3. The share of journals with relatively high self-Salton (over 0.3) was even somewhat higher

than that among all the 7835 journals under study. So we can make a conclusion that high self-link share was the obstacle for journals having high EL, but was not the hindrance for journals having more strong links.

In order to extend the journal entropy analysis, we applied another way for measuring the entropies of each journal, namely, we measured the distribution of journal links to disciplines, instead of to journals.

Several subject classification schemes are used in bibliometrics, for instances, the ISI Subject Categories, the 22 fields of the Essential Science Indicators (ESI), and the subject classification scheme of the Steunpunt O&O Indicatoren (see Glänzel and Schubert 2003). The first system has many multiple assignments of journals, which might strongly distort the entropy. On the other hand, the ESI 22 fields are too large for measuring entropies, as information flows takes mainly place within the large fields. Therefore the 68 sub-fields of the SOOI classification seem to be a good choice. Nevertheless we should mention that SOOI scheme is not unique either. The problem arises that journals with multiple assignment self-citations automatically produce links among different disciplines. Taken into account the big share of journals self-citations, this will have a notable impact on the result of entropy analysis. To avoid the problem above, we excluded all the journal self-citations when calculating the entropy.

For the journal cross-citation matrix, we kept the journals in the columns and replaced the journals appearing in the rows by the disciplines they are assigned to. Finally we got a "journal-discipline" cross-citation matrix. Here, for simplicity, we call this entropy on basis of "journal-discipline" links as "Dis_EL". We applied formula (6) by changing "journal 'j'" to "discipline 'j'". As expected, journals' Dis_EL were much lower than EL in journal cross-citation matrix since dispersion of information over fields is assumed to be lower than that over journals. The average Dis_EL among the 7835 journals was only 3.19, almost half of the latter. Most journals (6115 journals) have Dis_EL in the range of [2, 4]; only 20 journals' Dis_EL were over 5. The top 3 journals were found all multidisciplinary journals, respectively *Scientific American* (5.49), *American Scientist* (5.35) and *New Scientist* (5.35). Half of the top 10 journals were assigned to multidiscipline, 4 journals to humanities and only one journal (*TCE*) to science. This result is much different from the analysis of journal high EL in the previous part, where science journals were extremely dominant.

Among the top 100 journals with high Dis_EL, there are a quarter multidisciplinary journals, one fifth to biosciences, 18% to biomedical research and 16% to humanities. The big share of multidisciplinary journals and humanity journals does not tell against our hypothesis. Spreading links among different categories is a trait of multi-disciplinarity. In the case of the humanity journals, topics tend to be relatively diversified and are spread over a broader academic area, thus their Dis_EL is supposed to be relatively high.

Table 4: Journals with high EL and Dis_EL

Journals	*Dis_EL*	*EL*
ANN NY ACAD SCI	4.71	9.78
SCIENCE	5.01	9.64
BRAZ J MED BIOL RES	4.80	9.56
NATURE	4.93	9.51
MICROSC RES TECHNIQ	4.97	9.40
J KOREAN MED SCI	4.69	9.40
BIOMED PHARMACOTHER	4.78	9.30
ADV EXP MED BIOL	4.69	9.27
MED SCI MONITOR	4.68	9.21
BRIT MED BULL	4.70	9.21
TLS-THE TIMES LITERARY SUPPLEMENT	4.90	9.10
CR BIOL	4.78	9.06
THESCIENTIFIC-WORLDJOURNAL	4.91	8.93
PROG NAT SCI	5.21	8.87
SCI AM	5.49	8.87

The two different groups of high-entropy journals were almost disjoint; only 15 journals overlapped (see table 4). The top journals *Science, Nature* and other multidisciplinary journals, as well as core journals in other fields were found. Besides maintaining numerous links to other

journals, these journals also extend links to far-ranging disciplines.

Analogously to the previous analysis, there were no overlaps between "high Dis_EL" group and group of strongly interlinked journals. We have actually not found any journal among the top 100 journals with high Dis_EL having numerous strong links.

5 Discussions

Our study shows that high share of journal self-references, high discipline based entropy and relatively low impact and centrality makes it difficult to build reliable citation indicators for the arts and humanities. This has again confirmed that further studies should be restricted to the sciences, applied sciences and parts of the social sciences (Glänzel and Schoepflin 1999; Leydesdorff 2004b). Language barrier and lacking internationality (as reflected by their publications and not by the constitution of their editorial boards) proved one of the main hindrances for integration into the communication network. On the other hand, it was the document type that hinders journals from establishing self-links. These findings did, however, not strike us unexpectedly. The divergence between strongly interlinked and high-entropy journals was, on the other hand, somewhat unexpected. Most journals with more strong links had quite normal EL and all the journals with high EL had few strong links. Several high-impact multidisciplinary journals were citing and cited by numerous other journals but the strength of the links with those journals were rather moderate.

This study has shown that a well-defined set of structure measures can be used to analyse the 'communication characteristics' of individual journals within the cross-citation network. Centrality, isolation and entropy are influenced by many factors such as the subject area, document types, number of publications, specialisation, inter-disciplinarity, language barrier, national or international orientation, visibility and 'quality' and other related issues. In future studies we intend to deepen the network analysis by applying different clustering techniques based on a hybrid citation-textual approach in order to improve journal subject classification.

Acknowledgement

Zhang Lin appreciates Prof. Liang Liming from Henan normal university in China for her original idea for writing this paper, and also would like to thank Prof. Ronald Rousseau for his kind suggestions, as well as the supporting from the National Natural Science Foundation of China under Grant 70773015, 70431001, 70620140115, National Social Sciences Foundation under Grant 07CTQ008, Project of DUT under Grant DUTHS1002. We also wish to thank Bart Thijs from Steunpunt O&O Indicatoren for his kind help in collecting data.

References

Glänzel, W and Schoepflin, U (1999). A bibliometric study of reference literature in the sciences and social sciences. *Information Processing and Management* 35 (1), 31-44.

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

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

Leydesdorff, L (2004b). Top-down decomposition of the Journal Citation Report of the Social Science Citation Index: graph- and factor-analytical approaches. *Scientometrics* 60 (2), 159-180.

Leydesdorff, L (2006). Can scientific journals be classified in terms of aggregated journal-journal citation relations using the Journal Citation Reports? *Journal of the American Society for Information Science and Technology* 57 (5), 601-613.

Narin, F., Carpenter M. and Berlt, N.C (1972). Inter-relationships of scientific journals. *Journal of the American Society for Information Science* 23 (5), 323-331.

Price, D.D (1981). The analysis of square ma trices of scientometric transactions. *Scientometrics* 3 (1), 55-63.

Pinski, G., Narin, F (1976). Citation influence for journal aggregates of scientific publications:

theory, with application to the literature of physics, *Information Processing and Management* 12 (5), 297-312.

Schubert, A. and Glänzel, W (2006). Cross-national preference in co-authorship, references and citations. *Scientometrics* 69 (2), 409-428.

Schubert, A. and Braun, T (1993). Reference-standards for citation based assessments. *Scientometrics* 26 (1), 21-35.