Topological analysis of interdisciplinary scientific journals: Which journals will be the next Nature or Science?

Yongjun Zhu, Erjia Yan, and Il-Yeol Song
College of Computing and Informatics
Drexel University, 3141 Chestnut Street, Philadelphia, PA 19104, USA.
{zhu;ey86;song}@drexel.edu

Abstracts

Identifying prestigious interdisciplinary journals is very significant for researchers. By publishing research works in prestigious journals, researchers can better propagate their works and get spotlights. Even though the quality of a papers is not represented by the journal that published the paper, it is a general concern of researchers that how to identify a set of good journals to submit their papers. Nature and Science are the two journals that have been considering as the two top interdisciplinary journals worldwide. In this paper, we propose a method for identifying journals that have the potential to become the next Nature and Science through topological analysis of interdisciplinary scientific journals.

1. Introduction

Without doubt, the world's most prestigious interdisciplinary journals are Nature and Science. There are many reasons that make these journals prestigious. To name a few, impact factor, history, etc. are some of the reasons that explain the prestige of Nature and Science. While these two journals are prestigious, people, especially scholars could be interested in other journals that resemble these journals or have highest possibilities to become journals that are as prestigious as these two journals. There are many reasons why discovering such journals are meaningful and important. For example, researchers can selectively choose these journals to submit papers. This could be a strategic way because these papers have high potential to become the next Nature and Science and this could be helpful in many cases such as tenure evaluation.

Discovering such journals is not easy because there are many factors that contribute to the prestige of Nature and Science, and we are not able to capture all. In addition, in terms of impact factor and history, there are other journals that surpass Nature and Science, but not as prestigious as Nature and Science. This means we cannot focus on readily available factors to do the analysis. A good alternative is using citation data. By using citation data, we can discover the topology of scientific journals in terms of citation patterns. Through these patterns, we can identify journals that resemble Nature and Science.

2. Data and Method

Granted citation dataset from Elsevier is used. This dataset contains journal-level citation instances among all indexed fields during the past 15 years (1997-2011) with a two-year citation window. The dataset contains 4,287,565 citation records, 128625 journal descriptions, and 324 subject area descriptions.

Journal-to-journal citation data is used to monitor knowledge flow. Citation from journal A to journal B means the knowledge flow from journal B to journal A because journal A imports knowledge from journal B by citing journal B (Figure 1). Thus, by monitoring citations, we can detect patterns of knowledge flow among journals. In turn, we can identify journals that have similar patterns with Nature or Science.

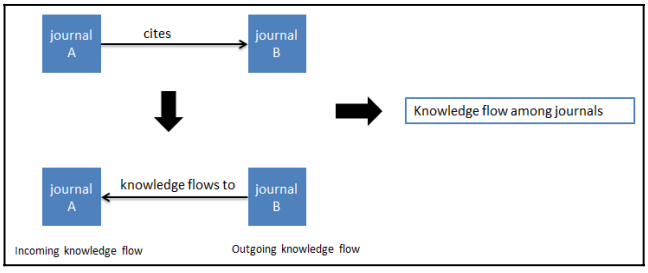

Figure 1: Citation and knowledge flow

As shown in Figure 1, two kinds of citation exist: incoming and outgoing citation. Thus, we can construct two kinds of citation matrices. Incoming citation is used to capture the status of cited journals, i.e., how cited journals are viewed by citing journals. Outgoing citations is used to capture knowledge composition of citing journals (Figure 2). Thus, we can capture two different patterns from these matrices.

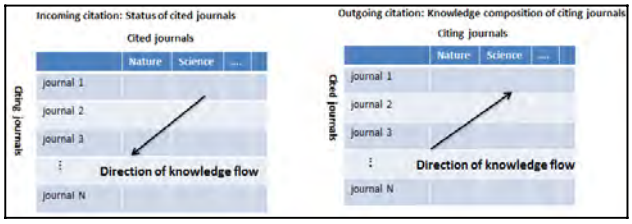

Figure 2: Incoming and outgoing citation matrices

The number of columns in two matrices is the same as the number of interdisciplinary journals. In the dataset used in this study, all journals have ASJS code, which is used for classifying journals based on their topics. In our study, this code should be "1000", because Nature and Science are interdisciplinary journals and "1000" represent interdisciplinary journals. Total 76 interdisciplinary journals indexed by Elsevier were used. The number of rows of these matrices is the same as the number of journals indexed by Elsevier, i.e., 128,625 journals were used. The study also focuses on temporal evolution of the topology of scientific journals by dividing the dataset into five groups (i.e., 1997/1999, 2000/2002, 2003/2005, 2006/2008, and 2009/2011) and using ten matrices (i.e. 5 incoming citation matrices and 5 outgoing citation matrices).

In the area of bibliometrics, citation data have been used to identify author clusters from author co-citation networks [4], identify journal clusters [5], etc. To the best of our knowledge, there was no study which tried to identify journals that resemble Nature or Science using citation data.

Three kinds of statistical methods were used: Principal component analysis [1], Multidimensional scaling [2], and Cluster analysis (Hierarchical clustering) [3]. Principal component analysis was used to discover journals that share the same component with Nature and Science. Multidimensional scaling is used to discover journals that are located closest to Nature and Science. Cluster analysis was used to discover journals that share the same cluster with Nature and Science. By synthesizing the results of these three analyses, we can get more meaningful results.

3. Results

As incoming and outgoing citations signify different characteristics, we report the results of each analysis separately. Incoming citation analysis is used to identify journals that resemble Nature or Science at the perspective of citing journals whereas outgoing citation analysis is for identifying journals that resemble Nature or Science in terms of knowledge composition because they tend to cite journals that also cited by Nature or Science.

3.1 Outgoing Citation Analysis

Figure 3 shows the result of PCA across five time periods.

1997/1999		2000/2002		2003/2005		2006/2008		2009/2011	
Journal	Factor loading	Journal	Factor loading	Journal	Factor loading	Journal	Factor loading	Journal	Factor loading
Science	0.959	Science	0.943	Science	0.953	Science	0.917	Science	0.895
Nature	0.924	Nature	0.917	Nature	0.936	Nature	0.943	Nature	0.939
PNAS	0.862	Scientist	0.882	Scientist	0.902	Scientist	0.873	PNAS	0.83
Chinese Science Bulletin	0.736	PNAS	0.848	PNAS	0.822	PNAS	0.832	Scientist	0.797
American Scientist	0.72	Chinese Science Bulletin	0.72	Complexity	0.543	Natures Sciences Societes	0.621	Scientific American	0.792
Current Science	0.717	Current Science	0.639	Chinese Science Bulletin	0.499	Interdisciplinary Science Reviews	0.557	Complexity	0.619
Scientist	0.616	American Scientist	0.532	Current Science	0.442	Complexity	0.482	American Scientist	0.58
New Scientist	0.414			Ohio Journal of Sciences	0.438	Scientific American	0.448	Science Technology and Society	0.509
Natures Sciences Societes	0.421			American Scientist	0.426	Chinese Science Bulletin	0.44	Natures Sciences Societes	0.507
Interciencia	0.409					Journal of the Indian Institute of Science	0.402	Current Science	0.494
								Science Progress	0.472

Figure 3. PCA results of outgoing citation analysis

In each of five time periods, the component in which Nature and Science have the highest factor loading was selected and other journals with factor loadings greater than 0.4 in the same component were selected. Results showed that *PNAS* and *Scientist* are the two journals that appeared 5 times, and journals such as *American Scientist*, *Current Science*, and *Chinese Science Bulletin* appeared 4 times.

Multidimensional scaling analysis was performed for each time period, and journals that close to Nature and Science in terms of distance were selected. Figure 4 shows selected journals for the time period of 2009/2011. They are also connected with Nature or Science with red lines to show approximate distance.

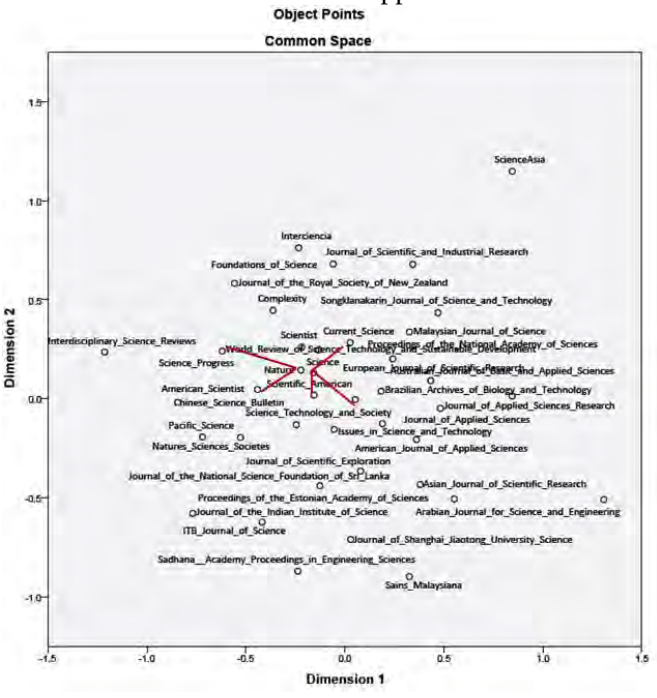

Figure 4: MDS result for outgoing citations for 2009/2011

Based on five MDS maps, we can easily identify

journals that are closely located with Nature or Science. We can divide them into two groups based on the results. Group A consists of two journals: *PNAS* and *Scientist* that are highly close to Nature or Science with the smallest distance. Group B consists of four journals: *American Scientist*, *Current Science*, *Scientific American*, and *Chinese Science Bulletin* that are close to Nature or Science.

Finally, hierarchical clustering was performed for each time period. Those journals that have the same cluster with Nature and Science were identified. For hierarchical clustering, we need to manually cut the dendrogram. Thus, we selected journals that join the cluster of Nature and Science at the early stage by using some heuristics. Figure 5 shows the results of hierarchical clustering for the last time period-2009/2011. As we can see from Figure 5, journals such as *PNAS*, *Scientist*, and *Scientific American* were clustered in the same group with Nature and Science.

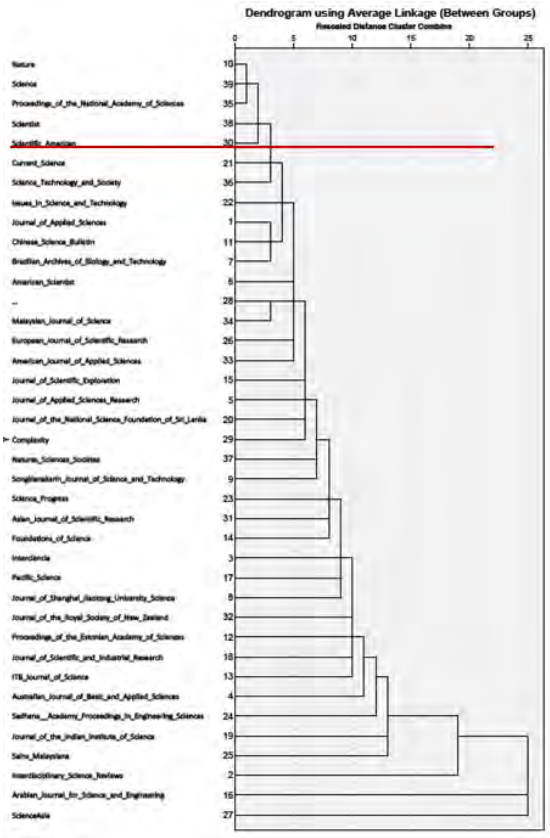

Figure 5: The result of hierarchical clustering for outgoing citation for 2009/2011

Based on results of hierarchical clustering, we can also divide them into two groups. Group A consists of two journals: *PNAS*, *Scientist* that are clustered with Nature and Science at the very early stage, and Group B consists of four journals: *American Scientist*, *Current Science*, *Chinese Science Bulletin*, and *Natures Sciences Sociétés* that are clustered with Nature and Science at early stage.

In summary, based on the results of three different analyses, five journals that are closest to Nature and Science were detected, and they were divided into two groups based on similarity (i.e., group A has greater similarity). *PNAS* and *Scientist* are in group A whereas *American Scientist*, *Current Science*, and *Chinese Science Bulletin* are in groups B.

3.2 Incoming Citation Analysis

Similar analyses were done by focusing on incoming citation. We also used three statistical methods to identify journals that resemble Nature and Science. Due to the page limit, we show the overall results that obtained by combing the results of three analyses. Based on the results of three different analyses, three journals that are closest to Nature and Science were detected, and they were also divided into two groups based on similarity. Group A includes *PNAS* whereas Group B includes *The Scientist* and *Scientific American*.

PNAS[1] and *The Scientist* are the two journals that are closest to Nature and Science. They appeared in results of both incoming and outgoing citation analyses. Because *PNAS* was included in group A in both analyses, it is the single closest journal to Nature and Science followed by *The Scientist*. PNAS is a journal published by National Academy of Science of the United States, and it is a very prestigious journal. This means the proposed methods are effective and self-explanatory.

4. Conclusions

In this study, we aim to identify journals that resemble Nature and Science under the research question of "Which journals are the next Nature or Science" by exploring journal-to-journal citation data from 1997 to 2011. Three different analyses (i.e., Principal component analysis, Multidimensional scaling, and Cluster analysis) were used to answer the research question. Incoming citation and outgoing citation data were analyzed separately in order to draw a more meaningful conclusion.

The results showed that *Proceedings of the National Academy of Sciences (PNAS)* is the closest journal to Nature and Science, and we can conclude that *PNAS* will be the next Nature and Science. Other journals such as *Scientist*, *American Scientist*, *Current Science*,

[1] http://www.pnas.org/

Chinese Science Bulletin, and *Scientific American* are similar to Nature and Science in some degree, but still need more concrete evidence. The results also showed that topological analysis of journals is a reasonable and applicable way to capture similarity of journals. Especially, one can focus on temporal evolution to find out the changing landscape of topology.

While the methods and results of this study is good enough to make a solid conclusion, there are also some limitations. The study focused on 76 interdisciplinary scientific journals that are indexed by Elsevier, and omitted journals that are not indexed by Elsevier, which makes the results less comprehensive. As a feature work, we plan to enlarge our dataset and construct a unified metric that can rank the possibility to become the next Nature and Science rather than depending on human judgement.

5. References

[1] Jolliffe, I. (2002). Principal component analysis. John Wiley & Sons, Ltd.

[2] Joseph B. Kruskal, & Myron Wish. (1978). Multidimensional scaling (Vol. 11). Sage.

[3] Johnson, S. C. (1967). Hierarchical clustering schemes. Psychometrika, 32(3), 241-254.

[4] White, H. D., & McCain, K. W. (1998). Visualizing a discipline: An author co-citation analysis of information science, 1972-1995. Journal of the American society for information science, 49(4), 327-355.

[5] Carpenter, M. P., & Narin, F. (1973). Clustering of scientific journals. Journal of the American Society for Information Science, 24(6), 425-436.