Four-facets study of scholarly communities: artifact, producer, concept, and gatekeeper

Chaoqun Ni
Indiana University Bloomington
1320 East 10th St. Bloomington, IN 47405
chni@indiana.edu

Cassidy R. Sugimoto
Indiana University Bloomington
1320 East 10th St. Bloomington, IN 47405
sugimoto@indiana.edu

ABSTRACT
This short paper compares the clustering of journals through three elements identified by Borgman (1989) as the main variables for bibliometric research: artifacts, producers, and concepts. In addition, this work extends this framework by adding “gatekeepers” as a variable. Fifty-eight journals from Information Science and Library Science (IS/LS) category in Journal Citation Reports were studied and clustered according to these four variables: producer, artifact, concept and gate keeper. The results demonstrate difference in clustering through these variables and the degree of the relationship between the variables. The discussion argues for more holistic measurements of scholarly communication that take multiple variables into account. This work is highly relevant in our assessment-conscious and metrics-driven age.

Keywords
Measurement, clustering, scholarly communication

INTRODUCTION AND REALTED WORK
Bibliometric research is well established as a quantitative method used to study formal channels of scholarly communication. Borgman (1989) first proposed the intersection between bibliometrics and scholarly communication in the form of coordinates, where the axis for bibliometrics method is displayed with three main variables: producers, artifacts, and concepts. According to her definition, the producers of scholarly communities are individual authors or aggregations of authors, artifacts are publication units or aggregations of publications, and concepts expressed by authors’ own words and research terminologies.

The used of separate variables for studying scholarly communication are ubiquitous, from authors (e.g. Liu, Bollen, Nelson, & Van de Sompel, 2005; Zhao & Strotmann, 2008), artifacts (e.g. Jarneving, 2007) to concepts (e.g. Courtial, 1994). One item that has been neglected is the role of the gatekeeper in scholarly communication. Editorial board members serve as gatekeepers a particular discipline, since their main job is to determine which articles are qualified for publication (Budd, 2000). While noted anecdotally, this variable has only recently received quantitative treatment: Ni and Ying (2010) proposed an approach to journal clustering via the interlocking of editorship information, which provides a novel different view or studying scholarly communication.

This paper extends this work by comparing the clustering of journals by all four of these variables: producers, artifacts, concepts, and gate keepers. To this point, most research has focused on a single variable and has created homogeneous networks and analyses. This work seeks to understand the differences inherent in the choice of variables and how that might impact the results of scientometric evaluations. To this end, all the producers, artifacts, concepts and gate keepers are aggregated at the journal level, to compare the same sets of unit of analysis through journal clustering results.

METHODS
Data
There are two sources of data used in this paper: publication data from Web of Science\(^1\) (WoK set) and editorial board member data (editor set) of the 58 journals classified as Information Science and Library Science (IS/LS) journals in the Journal Citation Reports. Only two types of publications, i.e. article and proceeding paper, published within the period of 1955-2010 and indexed by Web of Science were analyzed. Editorial board member information of these 58 journals was either directly collected from the website, or, for some cases when the website was inaccessible, from the hard copy of the journal. All the editorial board members included in this article are those indicated officially as of December, 2009. For more detailed information about the editorial board information, please refer to (Ni & Ding, 2010).

\(^1\) There are 61 journals categorized as Information Science and Library Science journals in 2008 Journal Citation Report, and 3 of them were excluded in this research as they are in languages other than English.
Methods
This paper attempts to characterize the scholarly communities of IS/LS through four different variables: producers, artifacts, concepts and gatekeepers. Operationalizations of these variables were done through venue author coupling (producers), journal co-citation (artifacts), topic modeling (concepts), and interlocking editorship (gatekeepers). These models produce quantitative results of proximities of journals. Statistical techniques commonly used in bibliometrics and scholarly communication studies, hierarchical clustering and multi-dimensional scaling (MDS), were employed to cluster journals, and a comparison using Quadratic Assignment Procedure (QAP) correlation on results obtained from these four different variables was conducted.

RESULT
In Borgman’s 1989 piece, she modeled an intersection of bibliometrics and scholarly communication studies, where three main variables for the research are producer, artifact and concept. In this work, we extend the three-facet model by adding one more variable, gatekeeper, as gatekeepers (in this case editorial board members) are critical for a discipline, in that they filter the knowledge displayed to public. The following section will discuss the results of these individual analyses and compare the results.

Four-facet overview of IS/LS
Table 1 provides the detailed information of IS/LS from four variables of characterizing scholarly communities discussed above. From the perspective of producers, it was found that there are 50,673 authors who have ever published in these 58 journals within the time period examined. Interestingly, about 32,299 (63.74%) of them only published once in IS/LS journals, and about 47,477 (93.69%) of them published no more than 5 times in IS/LS journals. The number of publications, i.e. artifacts in IS/LS communities, is 73,629. The number of single-authored publications is about twice as many as multi-authored ones: 49,632 single-authored (67.4%) and 23,997 multi-authored (32.6%). The concepts of IS/LS were examined through the number of unique words in both title and abstract after deleting some stop words. As for the gate keepers, there are in total 1,785 available seats occupied by 1,561 editorial board members.

Table 1. Four-facet overview of IS/LS

<table>
<thead>
<tr>
<th>Variables</th>
<th>Sub-items</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Producer</td>
<td>#unique-authors</td>
<td>50,673</td>
</tr>
<tr>
<td></td>
<td>#unique author per journal</td>
<td>1,170</td>
</tr>
<tr>
<td></td>
<td>#authors per paper</td>
<td>1.8</td>
</tr>
<tr>
<td>Artifact</td>
<td>#single-authored papers</td>
<td>49,632</td>
</tr>
<tr>
<td></td>
<td>#multi-authored papers</td>
<td>23,997</td>
</tr>
<tr>
<td></td>
<td>#papers per journal</td>
<td>1,269</td>
</tr>
<tr>
<td></td>
<td>#total citation in 2009</td>
<td>42,315</td>
</tr>
</tbody>
</table>

Producer (submission)
According to Borgman (1989), the producers of scholarly communities can be individual authors or aggregations of individual authors at different levels. In this project, we regard the journal as having an author profile, determined by the aggregation of authors who publish within it. Venue- Author-Coupling (VAC) method, which measures the relationship between two journals via the authors they share, was employed to identify the scholarly communities of IS/LS. VAC method was derived from document bibliographic coupling, and studies journals based on the assumption that if an author published on two journals, then these two journals would share some similarity. The number of shared authors for two journals indicates their similarity. Here journals were clustered through the number of authors they share. Figure 1 displays the cosine similarity among journals in a network view. Fifty-eight journals are clustered into four clusters. Journals in Management Information System (MIS) cluster are colored yellow, research-oriented IS/LS journals blue, practice-oriented IS/LS journals red, and specialized IS/LS journals green. It can be found that MIS cluster is separated from other three clusters, and journals within the specialized IS/LS cluster are not well connected within their cluster.

Figure 1. Network view of journal clusters by producers (cosine>=0.2)

2 For detailed information about VAC, please refer to (Ni, Sugimoto & Jiang, 2011).
3 In all of the following network views, cosine similarity was used as a proximity measure between journals. Journals are colored according to clustering result by hierarchical clustering technique, which we do not have enough space to include. Some network views only display lines with value larger than 0.2 to make them more readable.
Artifact (linking)
Artifacts of scholarly communication are the products of scholarly communications (Borgman, 1989). In our case, articles are aggregated at the journal level. Artifacts are the embodiment of ideas, and co-citation relationship can indicate some relationship between artifacts. Therefore, journal co-citation analysis was conducted to reveal the relationship between journals in IS/LS. This clustering technique also yielded four groups. Nine journals which are in MIS cluster by producers are also in one group by co-citation here, also named MIS cluster. There is also a research-oriented IS/LS journal cluster by co-citation, consisting of 14 journals.

Figure 2. Network view of journal cluster by artifact

Concept (writing)
The concept of scholarly communication can be expressed either by authors' own words or by researchers' terms (Borgman, 1989). In this paper, we used a topic modeling technique based on Latent Dirichlet Allocation (LDA) model, which identifies the probability that each journal belongs to a certain topic based on article titles. The cosine similarity was obtained based on the probability distribution of journal in a certain topic. Figure 3 displays the clustering result by coloring nodes in a network. It can be found that MIS is still a separate cluster, though in this result it is more related to specialized and computer-related IS/LS journals (green). Blue colored journals are those research-oriented IS/LS journals, though some of them are different compared with the results by producer and artifact.

Figure 3. Network view of journal cluster by concept (cosine>=0.2)

Gate Keeper (examining)
In the previous research of scholarly communication, the importance of gatekeepers of a discipline, in our case editorial board members of journals, has been mentioned, but few studies have provided evidence of their real roles in exploring scholarly communities. According to Ni and Ding (2010), the phenomenon that an editor sits on more than one journals' editorial boards is called interlocking editorship. The number of editors that two journals share can be viewed as an indicator of similarity between the two journals. Of all the journals studied, 10 journals (pink in figure 4) do not share editors with any other journals, i.e. they are isolates in terms of interlocking editorship. The rest of the 48 journals are clustered into four clusters, where those nine journals in yellow are still in a single cluster, MIS. Different from all the other results obtained above, the result by editors had a small new cluster, consisting of The Information Society, Journal of Health Communication, Journal of Computer-Mediated Communication, Social Science Computer Review and Telecommunication Policy, which are all communication-related research journals.

Figure 4. Network view of journal cluster by gate keeper

Comparison
Similarities and differences are evident in the clustering results presented about. First, the same nine journals are clustered as a single cluster, MIS, by four different variables. Conceptually (by concepts), the MIS cluster

4 Detailed analysis for the reason why these 10 journals do not share editors is not included due to space. Please refer to (Ni & Ding, 2010).
seems to be closely related to specialized and computer-related research journals. Second, some journals, e.g. IPM, JDOC, JASIST, Scientometrics, Journal of Informetrics, Journal of Information Science, ARIST, and CJLS, seem to be in the same cluster in the results by four different variables. This may indicate their stable proximity in all four aspects. A similar pattern can be seen with library practice journals which consistently appear together: Library Quarterly, LCATS, College Research Library, Journal of Academic Libraries, and RUSQ. Third, concepts and editorship information provide more clusters than citations and authors. In the case of editorship, a communication cluster is apparent.

Finally, Quadratic Assignment Procedure (QAP) correlation was used to calculate the correlation between cosine similarity matrices obtained from four variables. QAP has been commonly used by social network analysis to test the correlations between two or more networks with the same set of nodes, though originally proposed and employed in mathematics. Table 2 displays the QAP result. It shows that only the similarities obtained from interlocking editorship and journal co-citation are strongly correlated. The VAC similarity also significantly correlated with co-citation and topics, but weakly.

<table>
<thead>
<tr>
<th></th>
<th>Cocite</th>
<th>VAC</th>
<th>Editor</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cocite</td>
<td>1</td>
<td>0.346</td>
<td>0.986</td>
<td>0.188</td>
</tr>
<tr>
<td>VAC</td>
<td>0.346</td>
<td>1</td>
<td>0.23</td>
<td>0.388</td>
</tr>
<tr>
<td>Editor</td>
<td>0.986</td>
<td>0.23</td>
<td>1</td>
<td>0.177</td>
</tr>
<tr>
<td>Topic</td>
<td>0.188</td>
<td>0.388</td>
<td>0.177</td>
<td>1</td>
</tr>
</tbody>
</table>

*All the correlations are significant at 0.05 level, but p-values are not provided due to space.

**DISCUSSION AND CONCLUSION**

This paper extend the current three-variable model for scholarly communication research by Borgman by adding one more variable, the gate keeper of a scholarly discipline. The result demonstrates the clustering of journals using all four variables and identify the similarities and differences in clustering techniques. All clustering techniques are significantly correlated, although editorship and co-citation appears to be the strongest relationship among the four.

Moreover, the QAP results indicate that the editorship and co-citation data create nearly identical clustering of journals. Yet, editorship and topics and co-citation and topics are the weakest relationships. This brings interesting implications to the scholarly communication process: why might citations follow editorship relationship more than topical ones? Is there a causal relationship to be explored (i.e., do editors influence the citation patterns; do citation patterns influence the appointment of editors)? Furthermore, topics seem to be more correlated with authorship—a finding that demonstrates VAC can be useful in identifying topic areas of journals.

However, the conceptual implications of this research are that any single-variable clustering of the scholarly landscape will be inherently skewed. Although correlated, each variable produces a slightly different view on the relationship between units. In a metrics-heavy, indicator-drive environment, those conducting research assessments would do well to note these biases and potential for error. Future research should focus on ways in which we can more appropriate combine variables for holistic measurements of scholarly communication. In addition, as the web changes the ways in which we communicate, we should evaluate if there are new variables to be explored and how these relate to our current approaches.

**REFERENCES**


