Scopus and Web-of-Science 2012 compared in terms of aggregated journal-journal citation relations: Global maps and interactive overlays
Leydesdorff, L.A.; de Moya-Anegón, F.; de Nooy, W.

Published in:
"Context Counts: Pathways to Master Big and Little Data": proceedings of the science and technology indicators conference 2014 Leiden: 3-5 September 2014 in Leiden, the Netherlands

Citation for published version (APA):
Proceedings of the science and technology indicators conference 2014 Leiden

“Context Counts: Pathways to Master Big and Little Data”

3 - 5 September 2014 in Leiden, the Netherlands

Electronic version - Edited by Ed Noyons
Published by Universiteit Leiden - CWTS 2014
Scopus and Web-of-Science 2012 compared in terms of aggregated journal-journal citation relations: global maps and interactive overlays

Loet Leydesdorff*, Félix de Moya-Anegón** and Wouter de Nooy*

*loet@leydesdorff.net; W.deNooy@uva.nl
Amsterdam School of Communication Research (ASCoR), University of Amsterdam, Kloveniersburgwal 48, 1012 CX Amsterdam (The Netherlands)

**felix.moya@scimago.es
CSIC, SCImago Research Group, Calle Albasanz 26, Madrid 28037 (Spain)

Introduction
We compare the networks of aggregated journal-journal citation relations as provided by the Journal Citation Reports (JCR) 2012 of the Science and Social Science Citation Indexes (SCI and SSCI) with similar data for 2012 based on Scopus. First, we develop base maps and overlays for the two sets separately. Second, we match journal names across databases to assess the overlap.

Data
The data for Scopus 2012 was extracted from the Scopus database (1996-2012) in October 2013 (Leydesdorff et al., in press). Since single citations are aggregated in the JCR under “All others,” we discarded these values and pursued the analysis with the 2,688,731 remaining links which contain 36,748,156 citation relations.

Table 1: Descriptive statistics of the data.

<table>
<thead>
<tr>
<th></th>
<th>Scopus 2012</th>
<th>JCR (SCI + SSCI) 2012</th>
<th>JCR SCI</th>
<th>JCR SSCI</th>
</tr>
</thead>
<tbody>
<tr>
<td>N of journals</td>
<td>20,172 *</td>
<td>10,936</td>
<td>8,471</td>
<td>3,047</td>
</tr>
<tr>
<td></td>
<td>(6,672,033)</td>
<td>2,350,491</td>
<td>2,122,083</td>
<td>253,320</td>
</tr>
<tr>
<td>Citation links</td>
<td>2,688,731 **</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sum of citations</td>
<td>(40,731,458)</td>
<td>37,759,948</td>
<td>35,721,660</td>
<td>2,454,015</td>
</tr>
<tr>
<td>Self-citations</td>
<td>36,748,156 **</td>
<td>3,248,968</td>
<td>3,049,332</td>
<td>298,637</td>
</tr>
</tbody>
</table>

* The N of journals is 20,554 for the period 1996-2012
** corrected for single citation links.

2 We are grateful to Lykle Voort of the Amsterdam computer center SARA for his support. Some of this work was carried out on the Dutch national e-infrastructure with the support of the SURF Foundation.
JCR data were harvested from two JCR files for the SCI 2012 and SSCI 2012, respectively. The two files were first merged. The category “All others” is denoted as missing values.

**Methods**
The mapping method is analogous to the one applied previously to the aggregated set of Scopus 1996-2012 data published by Leydesdorff et al. (in press), and to the map based on JCR 2011 used by Leydesdorff et al. (2013). However, the two maps for 2012 (and the underlying matrices) can also be compared to each other.

Table 2: Statistics used for the visualization in VOSviewer

<table>
<thead>
<tr>
<th></th>
<th>JCR-WoS 2012</th>
<th>Scopus 2012</th>
</tr>
</thead>
<tbody>
<tr>
<td>Giant component</td>
<td>10,549</td>
<td>18,160</td>
</tr>
<tr>
<td>After correction for visual outliers</td>
<td>10,546</td>
<td>18,154</td>
</tr>
<tr>
<td>N of clusters (Blondel et al., 2008)</td>
<td>12</td>
<td>65</td>
</tr>
<tr>
<td>N of clusters (VOSviewer)</td>
<td>11</td>
<td>47</td>
</tr>
<tr>
<td>Modularity Q</td>
<td>0.557</td>
<td>0.694</td>
</tr>
</tbody>
</table>

**Global maps**
Figure 1 shows the base map for the 10,546 journals (96.4%) included in the largest component of JCR 2012. The shape and coverage is very similar to the map for 2011 (Leydesdorff et al., 2013, Fig. 1 at p. 2575). This reproduction of a base map in two different years—using the same methods—provides confidence in the validity of the technique and the reliability of the data.
Figure 1: Citing patterns of 10,546 journals in JCR 2012 visualized as a base map; cosine > .2; colors correspond to 11 communities distinguished by VOSviewer; available for webstart at http://www.vosviewer.com/vosviewer.php?map=http://www.leydesdorff.net/journals12/jcr12.txt

The map based on Scopus data 2012 (Figure 2) is also not so different from the previously published map based on aggregated Scopus data 1996-2012 (Leydesdorff et al., in press: Figure 3). The tail of the humanities journals at the bottom right is lacking from the JCR-based maps, while the A&HCI is not included in JCR.
**Interactive overlay maps**

The base maps can be used to position sets of documents (e.g., portfolios) in terms of the disciplinary composition. The routines provide Rao-Stirling diversity values for the sets under study relative to the respective maps.

In previous studies, we used datasets generated by Rafols *et al.* (2012) in which the Science and Technology Policy Research Unit (SPRU) at the University of Sussex was compared with the London Business School (LBS). These same sets of documents are used as the example in this study (e.g., Figure 3).
Figure 3: Scopus-based overlay map 2012 of journal publication portfolios from 2006 to 2010 of the Science and Technology Policy Research Unit SPRU at the University of Sussex (N = 268).

Table 3: Rao-Stirling diversity for SPRU and LBS documents (2006-2010) in both the 2011 and 2012 maps based on annual JCR data, and the two Scopus maps.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SPRU</td>
<td>0.2170</td>
<td>0.2175</td>
<td>155</td>
<td>0.1219</td>
<td>0.1489</td>
<td>268</td>
</tr>
<tr>
<td>LBS</td>
<td>0.0918</td>
<td>0.0922</td>
<td>348</td>
<td>0.0863</td>
<td>0.0917</td>
<td>715</td>
</tr>
</tbody>
</table>
Overlap between databases
Using fuzzy-string matching and ISSN numbers, we were able to match 10,524 journal names between the two sets. An Excel file with lists of matched and unique journals in Scopus and WoS, is available online at http://www.leydesdorff.net/journals12/all_journals.xlsx.

Figure 4: Citation relations among shared and unique journals in JCR (left side) and Scopus (right side).

The 10,524 journals matched between JCR and Scopus comprise 96.3% of all JCR journals and 51.2% of all journals in Scopus. Citation flows point from journals that are unique to Scopus to journals shared by both databases (Figure 4), suggesting that the shared journals are the more important ones. Citation flows are more balanced between shared and unique journals in JCR.

Conclusion
The basemaps are available for interactive usage at http://www.leydesdorff.net/journals12 (WoS) and http://www.leydesdorff.net/scopus12 (Scopus). The user can overlay downloads from either Scopus or WoS, and generate maps in VOSviewer. In the full paper, we add a network analysis of the two citation matrices; we also compare journal ranks in these two environments.

References