

Development of a journal recommendation tool based upon co-citation analysis of journals cited in Wageningen UR research articles

Marco G.P. van Veller¹ and W. Gerritsma²

¹Wageningen UR Library, Droevendaalsesteeg 2, 6708 PB, Wageningen, the Netherlands

²VU University Library, De Boelelaan 1103, 1081 HV, Amsterdam, the Netherlands

Abstract. Wageningen UR Library has developed a tool based upon co-citation analysis to recommend alternative journals to researchers for a journal they look up in the tool. The journal recommendations can be tuned in such a way to include citation preferences for each of the five science groups that comprise Wageningen UR.

For the development of the tool we have looked at the reference lists of 18,490 research articles published in 2006-2013 by Wageningen UR staff in 2,530 peer reviewed journals covered by Web of Science. The total collected reference lists contains 795,585 references, of which 700,115 have been identified to articles in 10,712 unique journal titles. For the 700,115 references to these articles we have made an inventory of the co-occurrence (co-citation) of journals. The abundance of co-citations for these journals is calculated.

We have included the results of the co-citation analysis of journals in a database that can function as a journal recommendation tool. With the tool, we can retrieve a list with frequencies of co-cited journals based on any journal that has been cited in Wageningen UR research articles. We surmise that frequently co-cited journals are more similar in topic(s) and research field(s) because these journals together have been cited in the same article dealing with a particular topic within a certain research field.

This list of co-cited journals provides suggestions on related journals that share topic(s) and research field(s). The tool presented in this paper is set up in an interactive way since it is based upon articles published by the researchers themselves. In case new articles will refer to other journals (because other topics

Received: 19.4.2015 / Accepted: 1.5.2015

ISSN 2241-1925

© ISAST



and research fields are dealt with in new published articles) this will influence the co-citation analysis and result in recommendation of other journals.

Keywords. Journals, Co-citations, References, Similarity, Research fields

1. Introduction

The digital library of Wageningen UR (University & Research Centre) provides access to online resources such as research reports, MSc or PhD theses, websites, databases (mostly bibliographies), electronic books and online journals for staff, students or other visitors. Of these different resources, especially (scholarly) journals provide an important source of scientific information for users of the digital library. In particular, researchers (e.g. PhD students or Wageningen UR staff) that are interested in scientific outcomes in a certain research field within the realm of Wageningen UR might be interested in a selection of scholarly journals that are of importance for the field they are interested in. Wageningen UR digital library can help these users to make a selection of these journals for further consultation.

A selection of journals important in a specific research field can be made via filtering by subject in the Wageningen UR Library collection of journals (see figure 1). Hereby, via a filtering menu the user has to choose for a particular subject (i.e. a research field) in order to get a list of journals recommended by the library for this field.

The screenshot shows the Wageningen UR Library collection of journals interface. At the top, there is a green header with the Wageningen UR logo and the text "Browse resources & recommendations". Below the header, there is a navigation bar with links for Home, Library, My Library, Journals, and Search result. The main content area is divided into two columns. The left column contains a "Filter by subject" menu with several categories: Basic Sciences, Social Sciences and Humanities, Agrotechnology and Food Sciences, Environmental Sciences and Forestry, Plant and Crop Sciences, and Zoology and Animal Sciences. The right column displays a search interface with a search bar, a search button, and a dropdown menu for "Sort result by:" set to "Importance". Below the search bar, there is a list of search results, including "Proceedings of the National Academy of Sciences of the United States of America [Journal]", "Nature: international journal of science [Journal]", and "Science: a weekly journal devoted to the advancement of science / American Association for the Advancement of Science [Journal]". Each result includes a checkbox, a title, a year range, and a list of subjects. The interface also includes a "Records 1 - 20 / 42551" indicator and various utility icons like alert, export, print, and help.

Figure 1: Filtering by subject in the Wageningen UR Library collection of journals.

The filtering and selection is based upon classification of journals in one or more (out of 1,236) categories. The library staff adds a category (or categories) to each journal when they enter the metadata for this journal in the library system. In order to include the appropriate category for a journal in the library system, the library staff has to examine the aim and scope for each new to include journal. This classification of journals has the consequence that extra time is needed in the workflow for entering new journals in the library system.

For the classification of journals library staff has the option to chose from a fixed list of 1,236 categories. This fixed list of categories is hierarchical (broad subjects with sometimes nested sub-subjects) and based upon the research fields that are of importance to Wageningen UR (i.e. food and food production, living environment and health, lifestyle and livelihood; Group Annual Report Wageningen UR, 2012). Research fields that fall outside the realm of Wageningen UR are not included in the fixed list or only as a broad research field. Sometimes, this results in too broad classifications of journals.

When a researcher wants to select journals for a particular research question from the Wageningen UR Library collection of journals it is important that he or she filters for the subjects that correspond with his or her research question. Sometimes this translation is hard to make because the research question does not fit in the subject categories that are available in Wageningen UR digital library. Also, it can be hard for the researcher to browse to the appropriate subjects in the Wageningen UR digital library because he or she does not know under which more broader subjects the sub-subjects can be found).

For the reasons described above (time needed by library staff for classification of journals, difficulty of classification of journals because of fixed list and difficulty of translation of research questions to subject categories), we developed an alternative supplementary tool for selection and recommendation of journals based upon co-citation analysis of journals included in the reference lists of articles published by Wageningen UR staff and PhD students in 2006-2013. With this tool, starting from a particular journal a researcher gets a presentation of the most similar journals in the Wageningen UR Library collection of journals. Hereby, the more a journal is co-cited with the journal the researcher started with, the more similar both journals are. This paper describes the methodology and development of this tool.

2. Materials and methods

The analysis of similarity between journals is based upon information from the references in research articles that are published by Wageningen UR staff and PhD students. For the analysis of co-cited journals in this article we used the references from research articles published in 2006-2013. Every year t the analysis of co-cited journals is extended with the analysis of reference lists of articles published in year $t-1$. For articles published in the years $t-2$ and $t-3$ it is checked if they were included in the analysis of co-cited journals in year $t-1$. If not, the references from these articles are also included in the analysis in year t . The sources and methodology for finding similar journals are schematically presented in figure 2.

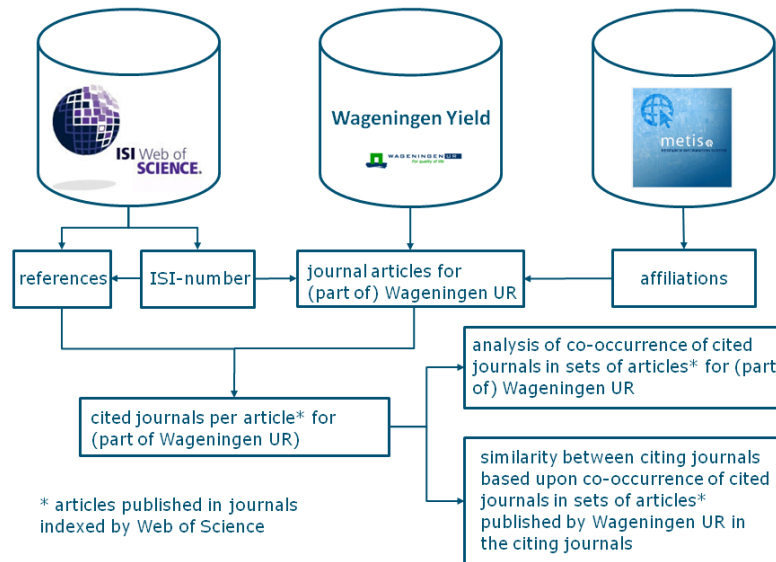


Figure 2: Sources and methodology for the analysis of similarity between journals based upon co-citations.

The sources and methodology for the analysis of similarity between journals partly correspond with the steps that Wageningen UR Library makes to measure journal usage based upon the abundance of journals in reference lists of articles published by Wageningen UR staff and PhD students (see Veller, 2013).

As input for the analysis, we collected all research articles published by researchers of Wageningen UR from our institutional repository (Wageningen Yield: WaY). Via a connection with our Current Research Information System (Metis), the repository contains updated information

on work relations (affiliation data) for the authors (i.e. Wageningen UR staff and PhD students) of the scientific output. For articles published in ISI- journals (i.e. journals covered by Web of Science) the unique identifier to the corresponding record with bibliographic information in Web of Science is also stored in the repository.

From the research articles we collected, we made a further selection comprising articles published in ISI-journals. For each of these articles, via the unique identifier to the corresponding bibliographic record in Web of Science, we collected all references. The collected references are subject to further analysis. Since we are interested in co-cited journals we only selected references to journal articles for further analysis. Criteria for the identification of these references are journal titles, indications of volume and issue data or listings of DOIs in the reference data obtained from Web of Science. For the selected references that represent journal articles we made an inventory of the journals in which they were published.

All information for the references is included in a database (see Figure 3) whereby each record represents a reference collected from the research. Besides information on the referred (or cited) journal, each record contains the unique identifier (ISI_NR; see Figure 3) to the bibliographic record, in Web of Science, of the research article from which we collected the reference. This unique identifier enables us to make an analysis of cited journals per research article. Also, via the unique identifier it is possible to get information from the institutional repository (Wageningen Yield) on the journal where the research article was published in. Via a connection between the institutional repository and the Current Research Information System (Metis) it is possible to get information on the research groups that published the research article from which we collected the reference. This information enables us to make a co-citation analysis of journals for only certain research groups (or science groups; explained below) of Wageningen UR.

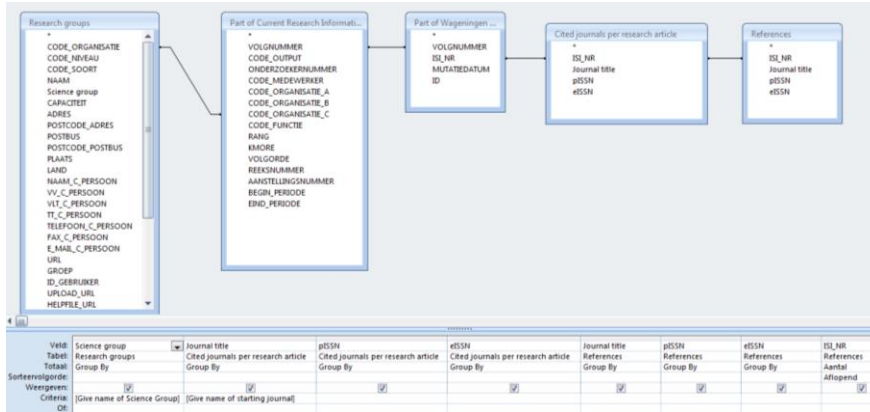


Figure 3: Schematic representation of relations between tables in a database with references that is made to make an analysis of co-cited journals.

With the database described above we can make an analysis of co-occurrence of cited journal (i.e. co-cited journals) in sets of articles for (a part of) Wageningen UR. In figure 4, via an example for the journal Agricultural Systems, it is illustrated how this analysis is done.

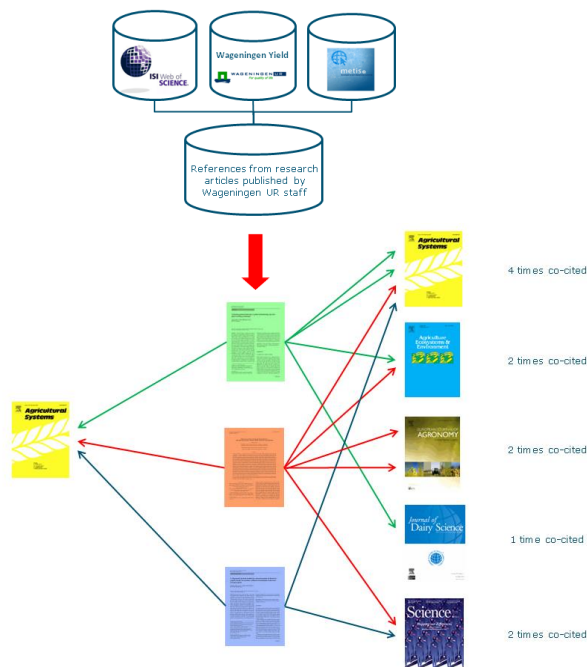


Figure 4: Example of the analysis of journals co-cited with the journal Agricultural Systems in three articles collected from Wageningen Yield.

We start with a journal for which we want to determine the journals that it has been co-cited with most. In this example we use the multidisciplinary journal *Agricultural Systems*. For this journal we collect the research articles that have made citations to articles published in it. The research articles are selected from the database described above which contains information on cited journals per research article and which has been derived from establishing connections between our institutional repository (Wageningen Yield), Current Research Information System (Metis) and Web of Science. In this example, we collect three research articles (green, red and blue).

For the three research articles we count in the references per journal the number of times it co-occurs with the journal *Agricultural Systems*. In case the same journal is present in the references more than once, every occurrence is counted. For the green research article we count four references to three different journals, for the red research article we count five references to four different journals and for the blue research article we count two references to two different journals. Every co-occurrence of a journal with *Agricultural Systems* is a co-citation between this journal and *Agricultural Systems*.

Adding up the number of times each journal is present (in co-occurrence with *Agricultural Systems*) in the references of the three research articles shows the following:

- *Agricultural Systems* is co-cited most with itself; four times
- *Agriculture, ecosystems & environment, European journal of Agronomy and Science* each are co-cited twice with *Agricultural Systems*
- *Journal of Dairy Science* is co-cited with *Agricultural Systems* only once.

With this analysis it is possible to make an ordering in a set of journals based upon the number of times they are co-cited. Journals that most have been co-cited with a particular starting journal (*Agricultural Systems* in the example above) can be selected and recommended to a user that is interested in the starting journal.

From the analysis of the occurrence of cited journals in sets of research articles for (part of) Wageningen UR it is possible also to determine a measure of similarity between the journals in which these research articles have been published (see figure 2). We have derived a similarity measure based upon Jaccard similarity coefficient (Jaccard, 1901). In appendix 1 with this paper we describe the calculation of a similarity measure between journals in which Wageningen UR staff and PhD

students published research articles. The similarity is based upon relative abundances of cited journals in the references of the research articles published in the (citing) journals for which the similarity is calculated. The more the relative abundances of co-cited journals in the references of articles published in two citing journals are, the higher the similarity between these citing journals is.

$$S_{gh} = 1 - \frac{\sum_{j=1}^{n_j} \left| \frac{r_{jg}}{\sum_{j=1}^{n_j} r_{jg}} - \frac{r_{jh}}{\sum_{j=1}^{n_j} r_{jh}} \right|}{2}$$

S_{gh} =Similarity between journal g and journal h.

r_{jg} = abundance of a cited journal j in the references of articles published in a particular citing journal g.

n_j =number of citing journals

r_{jh} = abundance of a cited journal j in the references of articles published in a particular citing journal h

The similarities between the journals can be visualized with VOSviewer (Van Eck & Waltman, 2015). For the cluster analysis and mapping we used the following parameter settings:

- Mapping attraction: 2
- Mapping repulsion: 1
- Clustering resolution: 1.00
- Minimum cluster size: 1
- Normalization method 1

3. Results

For the co-citation analysis we collected 18,490 research articles published by Wageningen UR staff and PhD students in 2,530 ISI journal covered by Web of Science. The research articles were published in the period 2006-2013 and represented 57% of the total scientific output by Wageningen UR in the same period.

From the research articles we collected 795,585 references of which 88% was identified as references to journal articles. The 700,115 references to journal articles represented references to articles published in ISI as well as articles published in non-ISI journals. In total 10,712 journal are collected from the references. For these journals we made an inventory on their co-occurrence in the references of the research articles.

Figure 5 shows the number of citations for each of the 10,712 journals that we collected from the reference lists of the Wageningen UR published ISI-journal articles in 2006-2013. The journals on the horizontal axis are sorted in descending order by the number citations to journal articles.

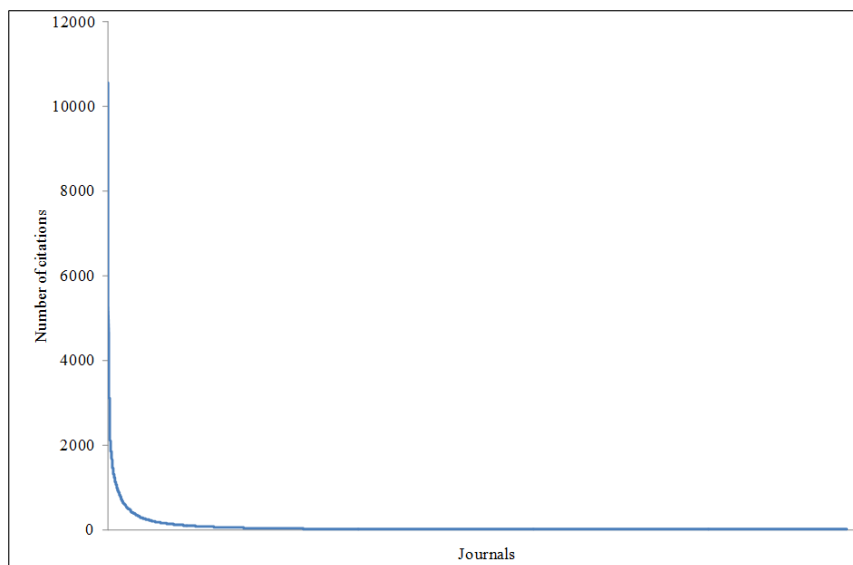


Figure 5: Number of citations per journal for the 10,712 journals that are inventoried from the references of the Wageningen UR research articles published between 2006 and 2013.

As figure 5 shows, most journals received only one or a few citations. In order to include 90% of all citations we have to select the journals that received 45 or more citations. This selection contains 1885 cited journals. For these most cited journals, per journal the number of co-cited journals (with 45 or more citations) varies from maximum 1861 (for the journal *Science*) to minimum 14 (for the *American potato journal*). The distribution of the number of co-cited journals for the most cited journals is presented in figure 6.

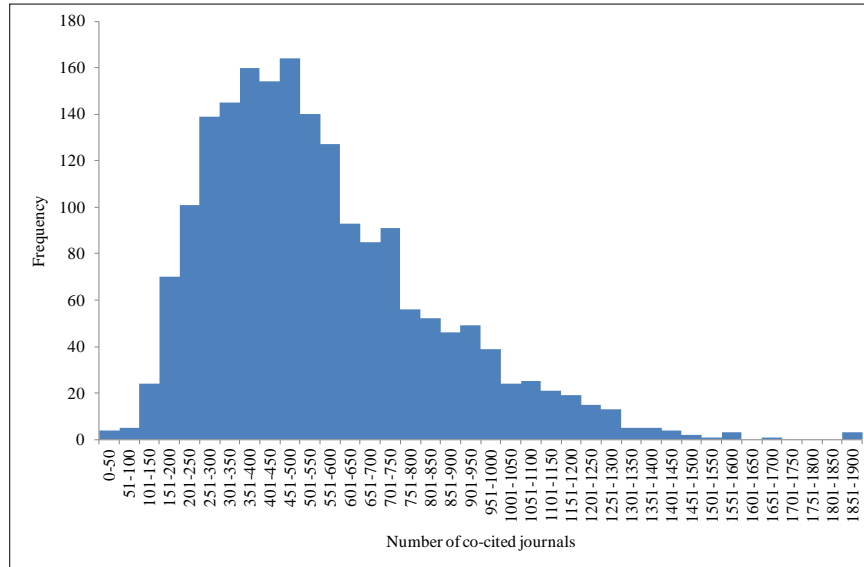


Figure 6: Distribution of the number of co-cited journals for journals that were inventoried from references lists of Wageningen UR research articles and that were cited 45 or more times between 2006 and 2013.

The histogram above shows that most journals were co-cited with 351-500 journals. About half of the most cited journals were co-cited with 15-500 journals. However, the number of co-citations per combination of two journals is not taken into account in these calculations.

For all references selected from the Wageningen research articles published in 2006-2013, 2,874,865 pairwise combinations of co-cited journals are inventoried. The maximum number of co-citations between two journals is found for the British Medical Journal, which is co-cited with itself 21,095 times. Half of the pairwise combinations of journals are co-cited only once in the references.

When we consider only journals that have received at least 45 citations between 2006 and 2013, 1,035,081 pairwise combinations of co-cited journals are found. Still in this selection, the largest amount (31%) of pairwise combinations of journals were co-cited only once and 77% of all pairwise combinations of journals were co-cited less than 10 times. When we select only journals that were co-cited 10 or more times, we find 241,622 pairwise combinations. The distribution of the number of co-citations for these combinations is presented in figure 7.

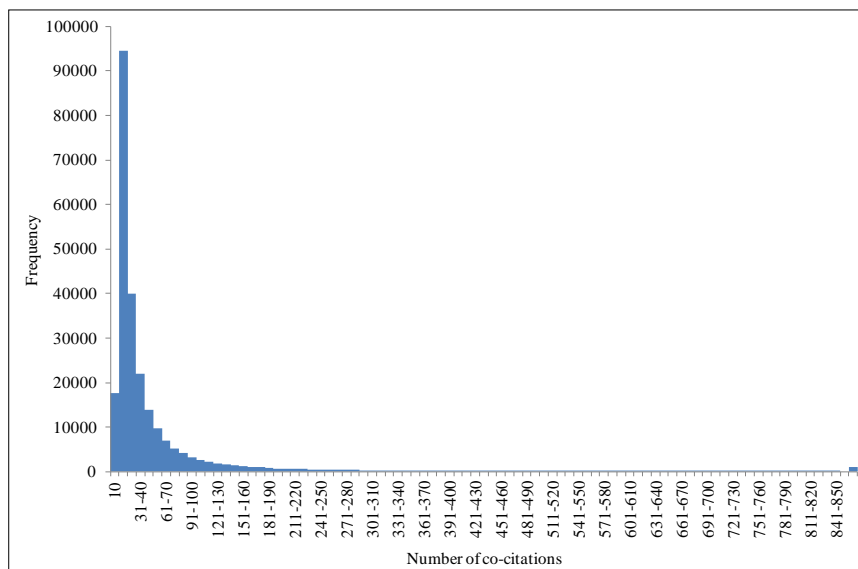


Figure 7: Distribution of the number of co-citations per pairwise combination of journals that were inventoried from references lists of Wageningen UR research articles and that were cited 45 or more times between 2006 and 2013. Only combinations of journals that are co-cited at least 10 times are presented.

Figure 7 shows that by far most combinations of journals were co-cited only 11-20 times; their share is 39% of all combinations. Together, the pairwise combinations of journals that are co-cited 10-100 times comprise 90% of all co-citations. From figure 6 it follows that most journals have been co-cited in Wageningen UR research articles with high numbers of other journals; 90% of all inventoried journals have been co-cited with 125 or more journals. These findings suggest that, starting from a particular journal, high numbers of co-cited journals may be found in combination with low numbers of co-citations for most journals co-cited with the journal we started with. Sorting of the journals co-cited with the starting journal based upon the number of co-citations enables us to make a selection of relatively few journals that are highly co-cited with the journal we started with. We will illustrate this with the journal *Agricultural Systems*, also used in the example illustrated in figure 4.

For the journal *Agricultural Systems*, 334 co-cited journals are inventoried from the references in the Wageningen UR research articles published in 2006-2013. The number of co-citations for the 10 most co-cited journals with *Agricultural Systems* are shown in table 1.

Table 1: Numbers of co-citations for top 10 of most co-cited journals with Agricultural Systems. Inventory based upon references collected from Wageningen UR research articles published in 2006-2013. Only journals cited more than 44 times are included.

Co-cited journal	Number of co-citations with Agricultural Systems in research articles published by:					
	Wageningen UR	ASG	AFSG	ESG	PSG	SSG
Agricultural Systems	1950	227	39	544	1100	745
Agriculture, ecosystems and environment	1310	147	23	524	724	246
Field crops research	609	17	13	106	508	78
European journal of agronomy	468	32	17	136	368	92
Journal of dairy science	431	246	5	9	37	214
Nutrient cycling in agroecosystems	424	61	4	114	297	42
Science	359	33	11	155	170	68
Plant and soil	322	16	7	104	218	17
Agronomy journal	305	1	4	81	227	22
Ecological economics	296	52	4	124	85	134

ASG=Animal Sciences group, AFSG=Agrotechnology and Food Sciences Group, ESG=Environmental Sciences Group, PSG=Plant Sciences Group, SSG=Social Sciences Group

The column “Wageningen UR” in table 1 shows the number of co-citations with Agricultural Systems for journals cited in research articles published by Wageningen UR as a whole. The other five columns (“ASG”, “AFSG”, “ESG”, “PSG” and “SSG”) show the number of co-citations with Agricultural Systems for journals cited in research articles published by sub-divisions (research groups) of Wageningen UR. From table 1 it follows that Agricultural Systems have been co-cited most with

itself, followed by the journal *Agriculture, ecosystems and environment*. The other co-cited journals have lower numbers of co-citations. Figure 8 shows for the 334 journals co-cited with *Agricultural Systems* how many times each journal was co-cited.

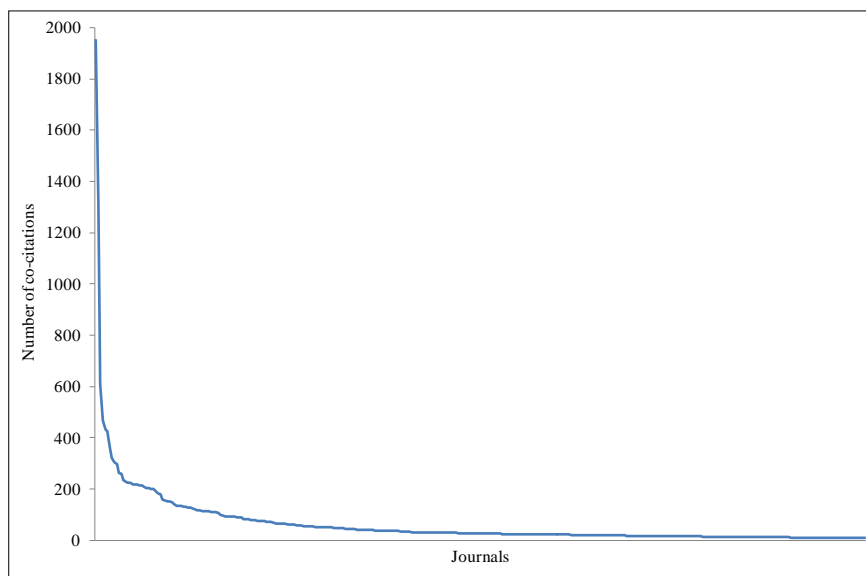


Figure 8: Number of co-citations per journal for the 334 journals co-cited with *Agricultural Systems* in the references of the Wageningen UR research articles published between 2006 and 2013.

Figure 8 shows that by far most of the 234 journals have been co-cited with *Agricultural Systems* only a few times; by selecting only 33 most co-cited journals, we collect 50% of all co-citations. After we order 234 journals by the number of co-citations, we are able to select the journals that most have been co-cited with *Agricultural Systems* (see table 1). When we consider only the 10 most co-cited journals, we find that they represent different research fields (e.g. Field crops research – plant sciences, *Journal of dairy science* – animal sciences or *Ecological economics* – social sciences), underlining the multidisciplinary character of the journal *Agricultural Systems*. The five columns “ASG”, “AFSG”, “ESG”, “PSG” and “SSG” in table 1 list the number of co-citations of journals with *Agricultural Systems* inventoried from research articles from each of these five science groups. Each science group represents a main research theme of Wageningen UR and contains research groups and chair groups of Wageningen UR that perform research in part of these research themes. When we sort the journals in table 1 according to the number of co-citations inventoried from research articles published

by each of these science groups we get different listings of top 10 most co-cited journals (see table 2).

Table 2: Top 10 of most co-cited journals with Agricultural Systems for the five science groups (ASG, AFSG, ESG, PSG and SSG) of Wageningen UR. Inventory based upon references collected from research articles published by each of the science groups in 2006-2013. Only journals cited more than 44 times are included.

ASG	AFSG	ESG	PSG	SSG
Journal of dairy science	Agricultural systems	Agricultural systems	Agricultural systems	Agricultural systems
Agricultural systems	Biomass and bioenergy	Agriculture, ecosystems and environment	Agriculture, ecosystems and environment	Agriculture, ecosystems and environment
Agriculture, ecosystems and environment	Weed research	Remote sensing of environment	Field crops research	American journal of agricultural economics
Livestock production science	Acta horticultrae	Science	European journal of agronomy	Journal of dairy science
Journal of animal science	Agriculture, ecosystems and environment	Ecological modelling	Nutrient cycling in agroecosystems	Agricultural economics
Animal feed science and technology	Postharvest biology and technology	European journal of agronomy	Agronomy journal	Ecological economics
Aquaculture	American journal of agricultural economics	Geoderma	Plant and soil	World development

Livestock science	Management science	Ecological economics	Science	Environmental modelling & software
Nutrient cycling in agro-ecosystems	Agroforestry systems	Agroforestry systems	Agricultural and forest meteorology	Livestock production science
Ecological economics	European journal of agronomy	Nutrient cycling in agro-ecosystems	Soil and tillage research	European journal of agronomy

ASG=Animal Sciences group, AFSG=Agrotechnology and Food Sciences Group, ESG=Environmental Sciences Group, PSG=Plant Sciences Group, SSG=Social Sciences Group

Table 2 shows differences in the ordering of most co-cited journals. Journals that are important (e.g. Journal of dairy science) in a research theme (e.g. animal sciences) that is studied by a particular science group (e.g. ASG=Animal Sciences group) are co-cited most with Agricultural Systems. As a consequence, by the inventory of only references for research articles studied by part of Wageningen UR, we get most co-cited journals for that part of Wageningen UR. Starting from a certain science group of Wageningen UR, the most co-cited journals with Agricultural Systems (in the references of research articles published by this science group) can be recommended as journals that are most similar to Agricultural Systems in the realm of the research theme studies by this science group.

From the analysis of occurrence of cited journals in the references of Wageningen UR research articles published in 2006-2013 we also have calculated similarities between 2,522 journals in which these research articles have been published. The calculation of the similarities between the 2,522 citing journals based upon 10,940 cited journals is described in appendix 1. The similarities have been calculated with a script (provided in appendix 2) in the language R for statistical computing and graphics (R Core Team, 2015) and listed in a 2,522 by 2,522 similarity matrix. From the matrix with VOSviewer (Van Eck & Waltman, 2015), a data visualization has been created with the parameter settings listed in materials and methods. The network for the 2,522 citing journals and mapping of the number of research articles published by Wageningen UR in 2006-2013 in these journals is shown in figure 9.

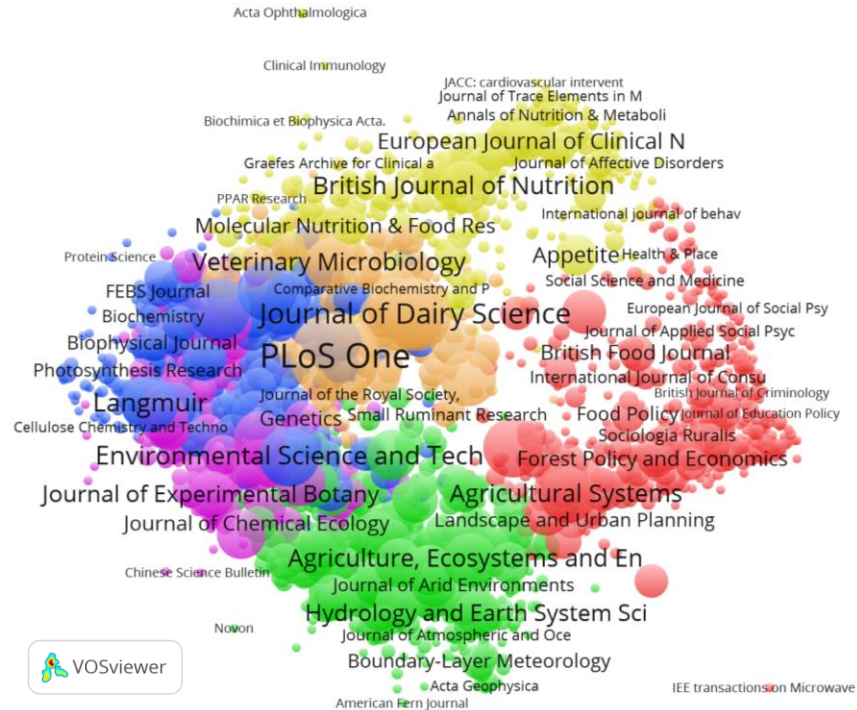


Figure 9: Network of 2,522 journals based upon a similarity analysis on citations of journals in research articles published by Wageningen UR in 2006-2013.

The network in figure 9 visualizes the relationships between the (citing) journals based upon similarity in citations to the same journals from research articles published by Wageningen UR staff in these (citing) journals. The more two journals are placed together in the network, the more similar they are. Besides similarity in citing to the same journals also similarity in the relative abundance of citations to the same journals is taken into account. The size of the bubbles in figure 9 corresponds with the size of the article output in each of the journals by Wageningen UR staff and PhD students in 2006-2013. The six clusters that are shown in figure 9 broadly correspond with six research themes (all in the context of agricultural and life sciences):

- red cluster: social sciences
- green cluster: environmental sciences
- purple cluster: plant sciences
- blue cluster: biotechnology and chemical sciences
- orange cluster: animal sciences
- yellow cluster: nutrition and food sciences

By changing the cluster resolution in VOSviewer it is possible to get more clusters of journals. By changing the cluster resolution setting to 100.0, setting the minimal cluster size to five and calculation of the clusters and mapping we obtained the same network as show in figure 9 but then with 234 clusters varying in size from five to 31 journals.

When we consider the journal Agricultural Systems, in this new clustering it will be placed with nine other journals in one cluster (see figure 10). The journals that are clustered with Agricultural Systems are listed in table 3.

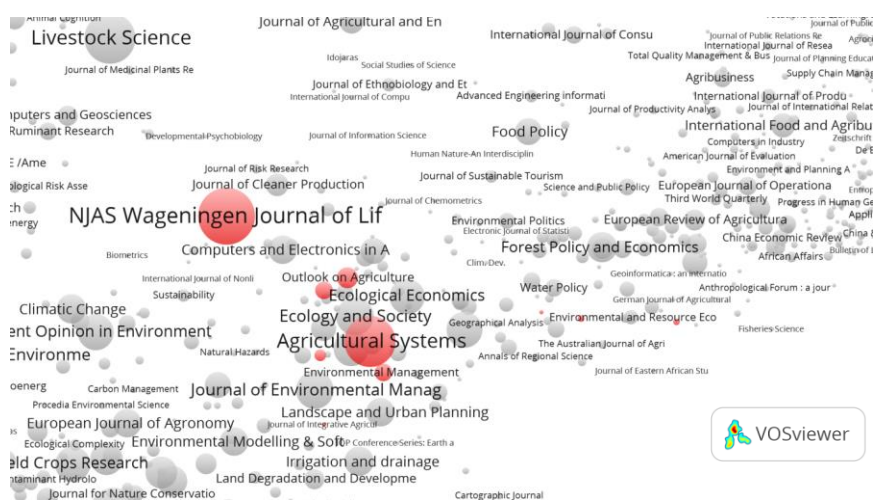


Figure 10: Detail of the network in figure 9 with the cluster where Agricultural Systems is placed in highlighted in red.

Table 3: Journals that are placed with the journal Agricultural Systems in one cluster by VOSviewer with cluster resolution 100.0 and minimal cluster size set to 5.

Journal
Cahiers d'Etudes et de Recherches Francophones. Agricultures
Environmental Management
Expert Systems with Applications
International Journal of Agricultural Sustainability
Journal of Integrative Agriculture
Journal of Sustainable Agriculture
NJAS Wageningen Journal of Life Sciences
Outlook on Agriculture
Soft Computing

This alternative clustering may function as a data-driven alternative classification of the journals (see introduction).

4. Discussion and conclusions

Starting from a particular journal, the analysis described in this paper enables us to obtain an ordered list of journals that co-occur with it in the references of Wageningen UR research articles published in 2006-2013. The journals from this list that most have been co-cited with the journal we started with can be selected as journals that most have been consulted in combination with the journal we started with by Wageningen UR staff and PhD students when writing their research articles.

Co-citation of journals in the references of a research article implies that articles from these cited journals have been important for the author of the citing research article. Since each article deals with a particular topic within a research field we assume that the references of this citing article also (partly) deal with the same topic within this research field. Consequently, we surmise that frequently co-cited journals are more similar in topic(s) and research field(s). We think the higher the frequency of co-citations for journals is, the stronger also the similarity in topic(s) and research field(s) for these journals is. For the analysis discussed in this paper we only used research articles published by Wageningen UR staff and PhD students in 2006-2013. Therefore, the similarities between journals we derived in this paper should be considered in the context of the research topics dealt with by Wageningen UR staff and PhD students in these research articles.

In this paper we have derived a network for (citing) journals in which Wageningen UR staff published in 2006-2013 based upon similarities in both (cited) journals and relative abundance of citations to articles in these journals. In the network (shown in figure 9) similar journals are positioned close to each other. When screening the network often journals in the same research field (derived from their titles) are placed close to each other as similar journals (see for an example figure 11). Also, when listing the journals that frequently have been co-cited with a journal in a particular research field, lists of journals in the same research field are obtained. These findings suggest that indeed frequently co-cited journals are often similar in the research field(s) they deal with.



Figure 11: Detail of network in figure 9 with journals in the research field of nutrition.

We have included the results of the co-citation analysis of journals, we discussed in this paper, in a database. This database can function as a journal recommendation tool. Via a set of queries on the tables in the database we obtain for a particular starting journal the most (co-cited) similar journals. These similar journals can be recommended in addition to the starting journal for various applications:

- Alternative journal for publishing of rejected papers. A researcher may select with the journal recommendation tool for an alternative journal that is similar to the journal where he or she submitted a manuscript to which subsequently was rejected.
- Additional journals for literature search on a topic in a particular research field. Starting from a particular journal a researcher may be interested in very similar journals that can be selected with this journal recommendation tool. The most similar journals may be further consulted by the researcher for interesting articles the topic in a particular research field where he or she is interested in.
- Alternative journals for journals the library provides no access to. In case a researcher wants to consult a journal the library provides no access to, the tool provides suggestion to alternative journals the library does provide access to.
- Additional journals the library can subscribe to. If the tool selects unsubscribed journals with high numbers of citations,

these journals may be considered by the library for subscription. The high numbers of citations indicate that there is some need for these journals by Wageningen UR researchers. Also when using the tool one can be directed to information to recommend a purchase of a particular journal selected with the tool but to which the library does not subscribe.

With the journal recommendation tool it is possible to base a recommendation of journals only on research articles published by a part of Wageningen UR (i.e. a science group). Via this selection of research articles from which we obtain information of cited journals we are able to obtain journal recommendations based upon co-citations only for the output of a certain science group that is more homologous in the research fields that are dealt with. Especially for journals that have a multidisciplinary character choosing for the output of a certain science group may gain in specificity with journals that are recommended with respect to the research field dealt with by the science group.

The journal recommendation tool we presented in this paper is set up in an interactive way since it is based upon articles published by the Wageningen UR researchers. In case the same Wageningen UR researchers would use this tool they get a recommendation for a particular journal based upon the citation behaviour of Wageningen UR researchers themselves. In case Wageningen UR staff and PhD students publish new research articles these may refer to other articles in other journals because other topics and research fields are dealt with in the new published articles. The information on journal citations for these new research articles will be added to the tables in the database that is used to recommend journals based upon co-citations. This addition can result in the recommendation of other journals based upon the co-citations. This makes this tool flexible and sensitive to new topics and research fields that may become of more importance for Wageningen UR researchers.

The journal recommendation tool described in this paper will be implemented in the Wageningen UR Library collection of journals. After implementation, the user (e.g. a researcher) of the Wageningen UR digital library will find information on similar journals (based upon co-citation analysis) in the metadata of a journal that has been cited in Wageningen UR research articles.

Acknowledgments

The authors thank Ellen Fest, Marc Loman and Peter van Boheemen for helpful discussions and Peter van der Togt for help with establishing database connections. Also, MvV wishes to thank Opeyemi Emmanuel for inspiration on the writing of this paper.

References

- Eck, N.J. van & Waltman, L. (2015). VOSviewer version 1.6.1. Center for Science and Technology Studies. Leiden University. Available at: <http://www.vosviewer.com/>.
- Jaccard, P. (1901). Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bulletin de la Société Vaudoise des Sciences Naturelles*, 37: 547–579.
- R Core Team (2015). R: A language and environment for statistical computing. R Foundation for statistical Computing, Vienna, Austria. Available at: <http://www.R-project.org/>.
- Veller, M.G.P. van (2013). Analysis of journal usage by Wageningen UR staff members via article references. *Qualitative and quantitative methods in libraries*, 2: 231-244.
- Working Group Annual Report Wageningen UR (2012). Annual report Wageningen UR 2011. Wageningen UR. Wageningen. 168 pp.

Appendix 1: Calculation of the similarity between journals based upon the abundance of cited journals in the references of articles published by Wageningen UR

For the calculation of the similarity a distinction is made between journals that contain Wageningen UR articles (i.e. citing journals) and journals that received citations from these articles (i.e. cited journals). The similarity is calculated for the citing journals.

The calculation of the similarity starts with a table that lists the abundance of cited journals in the references of articles published by Wageningen UR staff in the citing journals. The abundance is measured by counting how many times each cited journals is mentioned in the reference lists of all Wageningen UR articles published in the citing journal.

The total number of citing journals is n_g and the total number of cited journals is n_j . Each cell in the table below contains the abundance r of a cited journal j in the references of articles published in a particular citing journal g .

	Citing journal 1	Citing journal 2	...	Citing journal n_g
Cited journal 1	r_{11}	r_{12}	...	r_{1n_g}
Cited journal 2	r_{21}	r_{22}	...	r_{2n_g}
...	
Cited journal n_j	r_{n_j1}	r_{n_j2}	...	$r_{n_jn_g}$

In the next step the abundance of journal citations is fractionalized by dividing it by the total number of article references that are found in all published articles for each citing journal. The results of this fractionation is shown in the next table.

	Citing journal 1	Citing journal 2	...	Citing journal n _g
Cited journal 1	$\frac{r_{11}}{\sum_{j=1}^{n_j} r_{j1}}$	$\frac{r_{12}}{\sum_{j=1}^{n_j} r_{j2}}$...	$\frac{r_{1n_g}}{\sum_{j=1}^{n_j} r_{jn_g}}$
Cited journal 2	$\frac{r_{21}}{\sum_{j=1}^{n_j} r_{j1}}$	$\frac{r_{22}}{\sum_{j=1}^{n_j} r_{j2}}$...	$\frac{r_{2n_g}}{\sum_{j=1}^{n_j} r_{jn_g}}$
...
Cited journal n _j	$\frac{r_{n_j1}}{\sum_{j=1}^{n_j} r_{j1}}$	$\frac{r_{n_j2}}{\sum_{j=1}^{n_j} r_{j2}}$...	$\frac{r_{n_jn_g}}{\sum_{j=1}^{n_j} r_{jn_g}}$

In general, the calculation of the fractionized abundance of cited journal j for citing journal g can be represented as follows:

$$\frac{r_{jg}}{\sum_{j=1}^{n_j} r_{jg}}$$

The similarity between journals (with Wageningen UR articles from which the citations has been made) is calculated via the Jaccard similarity coefficient (Jaccard, 1901). In general, the similarity between journals g and h can be represented as follows:

$$S_{gh} = 1 - \frac{\sum_{j=1}^{n_j} \left| \frac{r_{jg}}{\sum_{j=1}^{n_j} r_{jg}} - \frac{r_{jh}}{\sum_{j=1}^{n_j} r_{jh}} \right|}{2}$$

The pairwise similarity values are listed in a n_g x n_g similarity matrix.

	Journal 1	Journal 2	...	Journal n _g
Journal 1	1	S ₁₂	...	S _{1n_g}
Journal 2	S ₂₁	1	...	S _{2n_g}
...	1	...
Journal n _g	S _{n_g1}	S _{n_g2}	...	1

Appendix 2: R-script for calculation of the similarity matrix

Below we give the R-script we have used for calculating the similarity matrix for 2,522 citing journals in which Wageningen UR published their research articles based upon the abundance of 10,940 cited journals in the references of the research articles. The similarity matrix is used as input to VOSviewer to create a network for the 2,522 citing journals. Commentary lines in this script start with “#”.

```
#Import datafile with three rows. The first
#column consists of journals (identified by 2522
#unique numbers) in which Wageningen UR (WUR)
#has published. The second consists of journals
#(via 10940 unique numbers) to which citations
#have been made in the WUR publications
#(organizational units in columns and journals
#in rows. The data represent the number of
#references per journal. The third column
#consists of relative shares that have been
#brought to each cited journal per journal in
#which WUR pubs have been found. Similarity is
#calculated for the journals with WUR
#publications. Similarity is based upon a
#modification on the Jaccard index. Only select
#the data and copy them to the clipboard. The
#clipboard contents is copied to the dataframe
#journals. You will have to type in the
#statement below in order to keep the data in
#the clipboard for import.
```

```
journals<-read.table("clipboard")
JacSim<-matrix(,2522,2522)
m<-matrix(0,10940,2)
```

```
#Calculate similarities and place them in the
#matrix JacSim. Calculate the similarities
#between journals in which WUR has published by
#alphabetically ordering these journals and
#replacing their names by a number. This script
#has been developed for 2522 journals in which
#WUR pubs have been found. The journals are
#compared for co-citations in cited journals.
#The list of cited journals in this script is
```



```
#number 1 to 10909 after having ordered them
#alphabetically.
for (k in 1:2522) {
  print (k)
  m[, ]<-0
  a<-
  data.frame(journals[2][journals[1]==k], journals
s[3][journals[1]==k])
  a<-aggregate(a[2], by=a[1], sum)
  m[a[,1],1]<-a[,2]
  for (l in 1:k) {
    print (l)
    if (l==k) {
      JacSim[k,l]<-1
    }
    else
    {
      b<-
      data.frame(journals[2][journals[1]==l], jo
urnals[3][journals[1]==l])
      b<-aggregate(b[2], by=b[1], sum)
      m[b[,1],2]<-b[,2]
      JacSim          [k,l]<-1-(sum(abs(m[,1]-
m[,2])))/(sum(m[,1],m[,2]))
      JacSim          [l,k]<-1-(sum(abs(m[,1]-
m[,2])))/(sum(m[,1],m[,2]))
      m[,2]<-0
    }
  }
}

#Make sure that matrix JacSim only contains
#positive values.
JacSim<-abs(JacSim)

#Write matrix JacSim as .csv file and store it
#in the directory DATA on the D-drive.
write.csv(JacSim,
"D:\\\\DATA\\\\Jacsimilariteiten.csv")
```