Qualitative and Quantitative Methods in Libraries (QQML) Special Issue
Bibliometrics and Scientometrics:49-60, 2015

Research Collaboration between Stockholm University and other Swedish Academic Institutions: A Bibliometric Study to Support Decisions on Library Collaboration

Per Ahlgren, Johan Hinders, Camilla Lindelöw, Sara Parmhed, and Per Swedberg

University Library, Stockholm University, SE-106 91 Stockholm, Sweden

Abstract: Academic libraries collaborate in several ways. For instance, collaboration can concern standards for indexing and statistics, technical solutions or collection development. A question that a given academic library might ask is with which other academic libraries the library should principally collaborate. In this study, we show how bibliometric methods can be used to generate information that can support decision making with regard to the question at stake. We evaluate the amount of research collaboration between Stockholm University and other Swedish academic institutions across five publishing years, and for the whole considered time period, where research collaboration is operationalized as co-publishing. A dataset of publications obtained from Web of Science, where each publication has at least one Stockholm University address, is used in the study. Co-publishing rates, non-fractionalized and fractionalized, across the publishing years and for the whole for period, for Stockholm University and other Swedish academic institutions, are reported. Further, parts of the outcome of the study are visualized in terms of co-publishing networks.

Keywords: Bibliometrics, Co-publishing, Decision making, Fractionalization, Research collaboration, Stockholm University

1. Introduction

Academic libraries collaborate in several ways. For instance, collaboration can concern standards for indexing and statistics, technical solutions or collection development. A question that a given academic library might ask is with which other academic libraries the library should principally collaborate. In this study, we show how bibliometric methods can be used to generate information that can

Received: 15.6.2014 Accepted: 30.11.2014 ISSN 2241-1925

support decision making with regard to the question at stake. We evaluate the amount of research collaboration between Stockholm University (SU) and other Swedish academic institutions across five publishing years, and for the whole considered time period, where research collaboration is operationalized as co-publishing.

Using co-publishing as a proxy for research collaboration is indeed debated (Katz & Martin, 1997), and could certainly be misleading at the individual level. However, since the aggregation level in this study is institution, the results should be reliable. Clearly, a study as ours gives a limited view on research collaboration. We believe, though, that the results from studies of this kind provide interesting information on the collaboration pattern for a given institution, in our case SU.

The remainder of the paper is organized as follows. Data and methods are described in Section 2, while we report the results in Section 3. In Section 4, the results are discussed, and conclusions are given.

2. Data and methods

In this study, the publication data are from Web of Science (WoS). We searched those citation indices of WoS that we had access to when the search was executed (2012-11-17).The searched indices were SCI-EXPANDED, SSCI, A&HCI, BKCI-S and BKCI-SSH. The following document types were searched: Article, Book, Book Chapter, Proceedings Paper (published in journals) and Review. The used query, which was constructed to retrieve those publications, from the publishing period 2008-2012 and in the indices in question, that contain at least one SU address, retrieved 8,734 publications. With regard to these publications, the following operations were performed:

1. Extraction of organization names (the WoS fields "Address" (C1) and "Reprint author" (RP) were used).
2. Standardization of variant names: different names of a given organization were mapped on the same name. For instance, *Stockholm Univ* and *Univ Stockholm* were both mapped on *Stockholm Univ*.
3. Elimination of duplicate names within publications.
4. Extraction of Swedish organization names.
5. Elimination of publications with only one organization name.
6. Elimination of publications with only one author (remaining, after the preceding point, publications where the author is affiliated to two or more organizations).

After the operations, 3,251 publications remained. Let P be the set of these publications. For a given time period t (t = 2008, …, 2012, 2008-2012), each publication in P such that its publishing year is equal to, or belongs to, t was extracted. Let P_t be the resulting set ($P_{2008\text{-}2012} = P$). The number of publications

in *P* with publishing year equal to 2012 is 648. This number is less than the corresponding numbers for 2010 and 2009. One should bear in mind, though, that the WoS search was executed in November 2012. For each pair (*Stockholm Univ*, *y*), where *y* is a name of a Swedish institution that occurs in the publications in P_t, the *co-publishing frequency* for the pair was calculated the number of publications *p* in P_t such that both *Stockholm Univ* and *y* occur in *p*.

An increase of the co-publishing frequency for a given pair from time period 1 to time period 2 might be explained by a larger number of SU publications for period 2, and not by an increased collaboration propensity. In order to control for the effect of the number of publications, we calculated, for each P_t and for (*Stockholm Univ*, *y*), the *co-publishing rate* for the pair, defined as the co-publishing frequency for (*Stockholm Univ*, *y*) divided by $|P_t|$, i.e., by the number of publications in P_t. (One can observe that the co-publishing rates preserve the ratio between two pairs with respect to the co-publishing frequencies of the pairs.)

Publications in, for instance, particle physics, have on average a large number of involved organizations. Furthermore, such publications are frequent among the 3,251 publications, since SU has several researchers active in this field. This yields that the measured collaboration, in terms of co-publishing, between SU and certain other institutions, like Uppsala University, with several researchers active in particle physics, might be misleadingly high. In the light of this, we carried out a second analysis, where a fractionalization approach was used. By this analysis, a complementary picture of the national collaboration of SU is obtained, where the influence of publications in particle physics, and of other publications with a large number of organizations, is reduced.

Let the *SU fraction* for a publication *p* i $P_t$be $1/\log_2 n$, where *n* is the number of unique organization names in *p* (at least two). For example, the SU fraction for a publication with four organization names is $1/\log_2 4 = 1/2$. We define the *number of fractionalized publications for SU* with respect to P_t, fr_{SU}, as the sum of the SU fractions across all publications in P_t. The *fractionalized co-publishing frequency* for the pair (*Stockholm Univ*, *y*), $fr_{SU,y}$, is defined as the sum of SU fractions for those publications *p* in P_t such that both *Stockholm Univ* and *y* occur in *p*. Regarding a publication where the pair (*Stockholm Univ*, *y*) occurs, the pair is then associated with the SU fraction for the publication. This fraction is less than 1 when the number of organizations is three or more. Thus, in such cases, a co-publishing down-weighting occurs. In order to control for the effect of the number of fractionalized publications (cf. the second paragraph under the numbered list above), we calculated, for each P_t and for (*Stockholm Univ*, *y*), the *fractionalized co-publishing rate* for the pair, defined as $fr_{SU,y}/fr_{SU}$.

By not dividing 1 by *n*, but by the logarithm of *n*, a comparatively less sharp co-publishing down weighting is obtained. Further, the greater the value of the base of the logarithm, the less sharp down-weighting.

For the period 2008-2012, the outcome is given, besides being tabulated, in the form of two weighted networks. We used Pajek for visualization of the networks (de Nooy, Mrvar, & Batagelj, 2011). The Kamada-Kawai algorithm was used for automatic layout generation. A large part of the data processing was done by the aid of Bibexcel, a toolbox for the analysis of bibliographic data (Persson, Danell, & Schneider, 2009).

3. Results

In this section, we put forward the results of the study. The first subsection gives results for the non-fractionalization case, the second results for the fractionalization case.

3.1 Co-publishing without fractionalization

Table 1 reports results for the first analysis, where co-publishing fractionalization was not applied. The table gives, for six time periods, the co-publishing rates (in %) for the 20 Swedish institutions that have the highest co-publishing rates in pair with SU for the period 2008-2012. According to that period, SU has the highest co-publishing rate with Uppsala University, 25.8, followed by Karolinska Institute and KTH Royal Institute of Technology, 23.0 and 19.0 respectively (Table 1, last column).

The ranking of the 20 institutions for year 2012 (the second last column) – the latest of the considered publishing years – is not identical to the corresponding ranking for the period 2008-2012, even if the two rankings show great similarities (rank correlation measured by Kendall's tau = 0.83).

Table 1. Co-publishing rate (in %, rounded to one decimal) for SU in pair with 20 Swedish institutions across six time periods. The institutions are ordered descending after the values for the period 2008-2012. Ranks within parentheses.

Institution	2008	2009	2010
Uppsala Univ	26.8 (2)	23.0 (2)	20.8 (1)
Karolinska Inst	27.9 (1)	24.7 (1)	18.6 (2)
KTH Royal Inst Technol	16.2 (4)	17.1 (3)	17.6 (3)
Lund Univ	19.1 (3)	8.8 (4)	11.5 (4)
Umea Univ	5.7 (6)	8.0 (5)	5.6 (8)
Univ Gothenburg	6.3 (5)	4.8 (8)	6.6 (7)

Swedish Univ Agr Sci	4.5 (7)	5.5 (7)	3.9 (9)
Linnaeus Univ	1.4 (11.5)	7.2 (6)	8.7 (5)
Chalmers Univ Technol	1.4 (11.5)	2.3 (9.5)	7.1 (6)
Linkoping Univ	2.0 (10)	2.3 (9.5)	1.7 (11)
Sodertorn Univ	3.9 (8)	1.6 (12)	1.7 (11)
Univ Orebro	2.5 (9)	2.1 (11)	1.7 (11)
Mid Sweden Univ	0.4 (15)	0.2 (18)	1.0 (14)
Malardalen Univ	0.8 (13)	0.7 (13)	1.4 (13)
Stockholm Sch Econ	0.0 (19)	0.2 (18)	0.7 (15)
Lulea Univ Technol	0.2 (17)	0.2 (18)	0.4 (16)
Jonkoping Univ	0.4 (15)	0.2 (18)	0.1 (17.5)
Karlstad Univ	0.4 (15)	0.2 (18)	0.1 (17.5)
Univ Gavle	0.0 (19)	0.5 (14.5)	0.0 (19.5)
Swedish Natl Def Coll	0.0 (19)	0.5 (14.5)	0.0 (19.5)

Table 1 (continued).

Institution	2011	2012	2008-2012
Uppsala Univ	26.9 (1)	31.5 (1)	25.8 (1)
Karolinska Inst	22.1 (2)	23.5 (2)	23.0 (2)
KTH Royal Inst Technol	20.0 (3)	23.0 (3)	19.0 (3)
Lund Univ	13.3 (4)	18.2 (4)	14.0 (4)
Umea Univ	5.4 (5)	7.7 (5)	6.4 (5)
Univ Gothenburg	4.3 (7)	5.9 (6)	5.5 (6)
Swedish Univ Agr Sci	4.7 (6)	4.3 (7)	4.6 (7)
Linnaeus Univ	2.6 (9)	0.8 (14)	4.2 (8)
Chalmers Univ Technol	2.2 (10)	2.5 (9)	3.2 (9)
Linkoping Univ	2.7 (8)	3.9 (8)	2.5 (10)
Sodertorn Univ	1.4 (12.5)	0.8 (14)	1.8 (11)
Univ Orebro	1.4 (12.5)	0.9 (11.5)	1.7 (12)
Mid Sweden Univ	1.7 (11)	2.0 (10)	1.1 (13)
Malardalen Univ	0.9 (16)	0.9 (11.5)	1.0 (14)
Stockholm Sch Econ	1.0 (14.5)	0.3 (16.5)	0.5 (15)
Lulea Univ Technol	1.0 (14.5)	0.2 (18.5)	0.4 (16)
Jonkoping Univ	0.0 (20)	0.8 (14)	0.3 (17.5)
Karlstad Univ	0.5 (17)	0.2 (18.5)	0.3 (17.5)

Univ Gavle	0.2 (19)	0.3 (16.5)	0.2 (19)
Swedish Natl Def Coll	0.4 (18)	0.0 (20)	0.2 (20)

Moving averages (2 years) over co-publishing rates, for the 12 Swedish institutions that have the highest co-publishing rates in pair with SU for the period 2008-2012 are displayed in Figure 1. For instance, the value for SU and another institution for year 2009 is equal to the average over the co-publishing rates for the pair across the years 2008-2009, while the value for 2010 is equal to the average over the rates across the years 2009-2010. SU and KTH Royal Institute of Technology show an increasing trend. This is also the case for SU and Lund University, from year 2010, while the opposite is true for SU and Linnaeus University, as well as for SU and Karolinska Institute regarding the years 2009-2011.

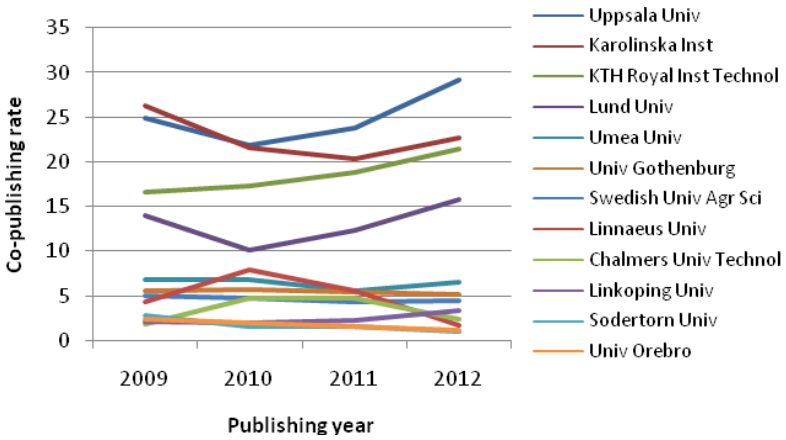

Figure 1. Moving averages (2 years), over co-publishing rates (in %), for SU in pair with 12 Swedish institutions.

The network in Figure 2 (a star network) visualizes the outcome for the period 2008-2012 and the 20 institutions represented in Table 1. The nodes represent institutions, and the width of the links corresponds to co-publishing rates. From the figure it is clear that SU has the strongest relations, in terms of co-publishing and for the whole studied period, to, in turn, Uppsala University, Karolinska Institute, KTH Royal Institute of Technology and Lund University.

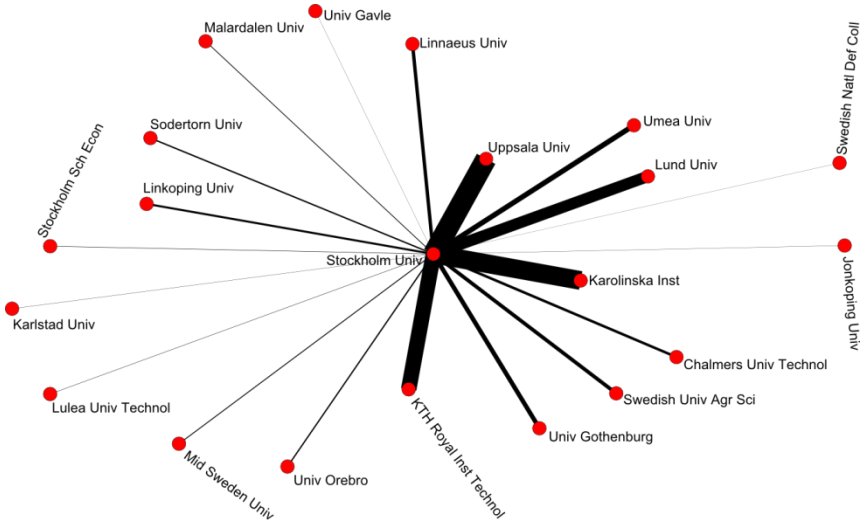

Figure 2. Network with SU and those 20 institutions that have the highest co-publishing rates in pair with SU for the period 2008-2012.

3.2 Fractionalized co-publishing

In Table 2, results for the second analysis, where co-publishing fractionalization was applied, are reported. The table gives, for six time periods, the fractionalized co-publishing rates (in %) for the 20 Swedish institutions that have the highest fractionalized co-publishing rates in pair with SU for the period 2008-2012. The institutions represented in Table 2 are the same that are represented in Table 1. However, there are differences compared to when fractionalization is not applied. For the period 2008-2012, Karolinska Institute, and not Uppsala University, has the strongest relation to SU (25.2 and 17.2, respectively). When fractionalization is used, the rank for Chalmers University increases from 9 to 11; which in turn means that this institution has a weaker relation to SU - relative to the other 19 institutions – compared to when fractionalization is not applied.

However, the rank correlation between co-publishing rate regarding non-fractionalization on one hand, and fractionalization on the other, is strongly positive (Kendall's tau = 0.87; period 2008-2012; 20 institutions). The scatterplot of Figure 3 visualizes for 20 institutions, and for the period 2008-2012, the non-fractionalized and the fractionalized co-publishing rates. It is evident from the plot, for example, that Uppsala University has the strongest relation to SU, followed by Karolinska Institute, when fractionalization is not used, whereas the opposite is the case when fractionalization is used.

Table 2. Fractionalized co-publishing rate (in %, rounded to one decimal) for SU in pair with 20 Swedish institutions across six time periods. The institutions ordered descending after the values for the period 2008-2012. Ranks within parentheses.

Institution	2008	2009	2010
Karolinska Inst	30.4 (1)	25.4 (1)	18.9 (1)
Uppsala Univ	17.1 (2)	17.8 (2)	15.7 (2)
KTH Royal Inst Technol	10.5 (4)	12.1 (3)	12.7 (3)
Lund Univ	10.8 (3)	5.3 (7)	9.4 (4)
Umea Univ	6.1 (5)	8.8 (4)	6.3 (6)
Univ Gothenburg	6.1 (6)	5.3 (6)	7.0 (5)
Swedish Univ Agr Sci	4.4 (8)	5.6 (5)	4.2 (8)
Linkoping Univ	2.0 (10)	2.3 (10)	1.7 (12)
Linnaeus Univ	1.1 (12)	2.6 (9)	3.6 (9)
Sodertorn Univ	4.7 (7)	2.0 (11)	2.4 (10)
Chalmers Univ Technol	1.4 (11)	1.6 (12)	4.6 (7)
Univ Orebro	2.6 (9)	2.9 (8)	2.0 (11)
Mid Sweden Univ	0.3 (15)	0.1 (19.5)	1.4 (14)
Malardalen Univ	0.9 (13)	1.1 (13)	1.7 (13)
Stockholm Sch Econ	0.0 (19)	0.2 (17)	1.0 (15)
Lulea Univ Technol	0.1 (17)	0.1 (19.5)	0.4 (16)
Swedish Natl Def Coll	0.0 (19)	0.9 (14)	0.0 (19.5)
Karlstad Univ	0.4 (14)	0.2 (18)	0.2 (18)
Jonkoping Univ	0.3 (16)	0.3 (16)	0.2 (17)
Univ Gavle	0.0 (19)	0.5 (15)	0.0 (19.5)

Table 2 (continued).

Institution	2011	2012	2008-2012
Karolinska Inst	24.5 (1)	28.5 (1)	25.2 (1)
Uppsala Univ	16.8 (2)	18.9 (2)	17.2 (2)
KTH Royal Inst Technol	13.9 (3)	13.4 (3)	12.7 (3)
Lund Univ	7.2 (4)	8.2 (4)	8.1 (4)
Umea Univ	5.5 (5)	8.1 (5)	6.9 (5)
Univ Gothenburg	4.0 (7)	6.6 (6)	5.7 (6)
Swedish Univ Agr Sci	4.6 (6)	4.5 (7)	4.7 (7)
Linkoping Univ	3.0 (9)	4.4 (8)	2.7 (8)
Linnaeus Univ	3.0 (8)	0.8 (14)	2.3 (9)
Sodertorn Univ	1.8 (11)	1.0 (13)	2.3 (10)

Chalmers Univ Technol	1.7 (12)	1.8 (10)	2.3 (11)
Univ Orebro	1.7 (13)	1.1 (12)	2.0 (12)
Mid Sweden Univ	2.3 (10)	2.0 (9)	1.3 (13)
Malardalen Univ	1.0 (14)	1.1 (11)	1.2 (14)
Stockholm Sch Econ	1.0 (15)	0.3 (17)	0.5 (15)
Lulea Univ Technol	1.0 (16)	0.2 (18.5)	0.4 (16)
Swedish Natl Def Coll	0.5 (17)	0.0 (20)	0.3 (17)
Karlstad Univ	0.4 (18)	0.2 (18.5)	0.3 (18)
Jonkoping Univ	0.0 (20)	0.4 (15)	0.2 (19)
Univ Gavle	0.2 (19)	0.3 (16)	0.2 (20)

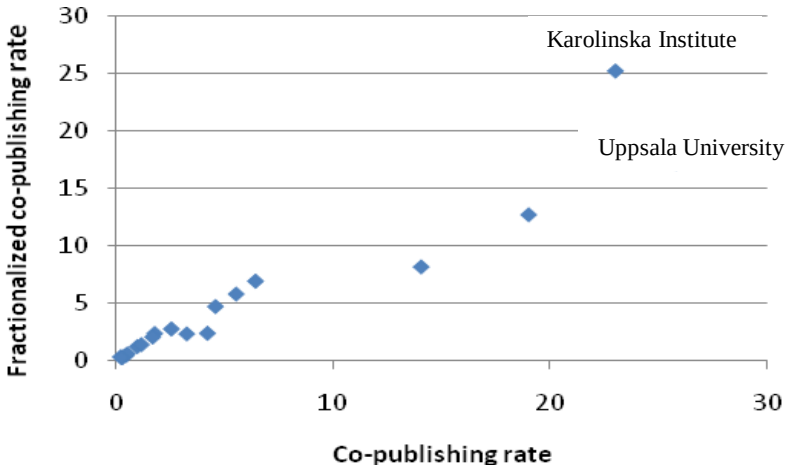

Figure 3. Scatterplot for co-publishing rate and fractionalized co-publishing rate for the period2008-2012 (20 institutions).

In Figure 4, moving averages (2 years), over fractionalized co-publishing rates, are given for the 12 institutions that have the highest fractionalized co-publishing rates in pair with SU for the period 2008-2012. The institutions that are represented in Figure 4 are the same that are represented in Figure 1. SU and KTH Royal Institute of Technology show an increasing trend, as in the non-fractionalization case. SU and Uppsala University have an almost constant average across the four years, in sharp contrast to the outcome for the non-fractionalization case (Figure 1). The variation across the four years is small also for SU and Lund University, again in sharp contrast to the outcome for the non-fractionalization case.

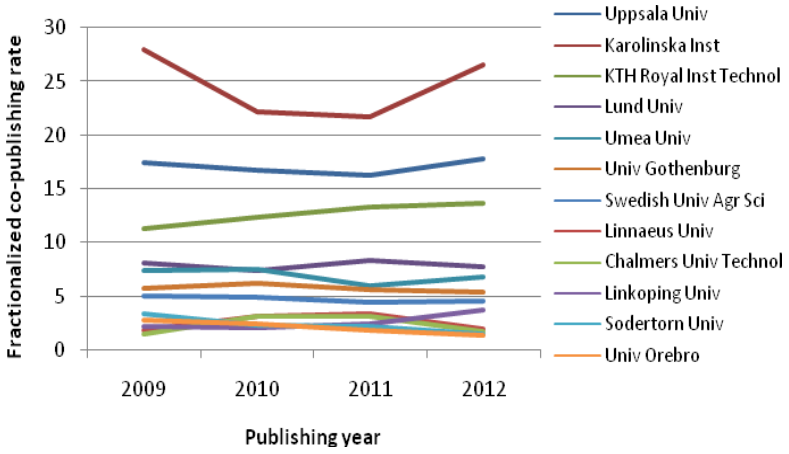

Figure 4. Moving averages (2 years), over fractionalized co-publishing rates (in %), for SU in pair with 12 Swedish institutions.

The network in Figure 5 visualizes the outcome for the period 2008-2012 for the 20 institutions represented in Table 2.

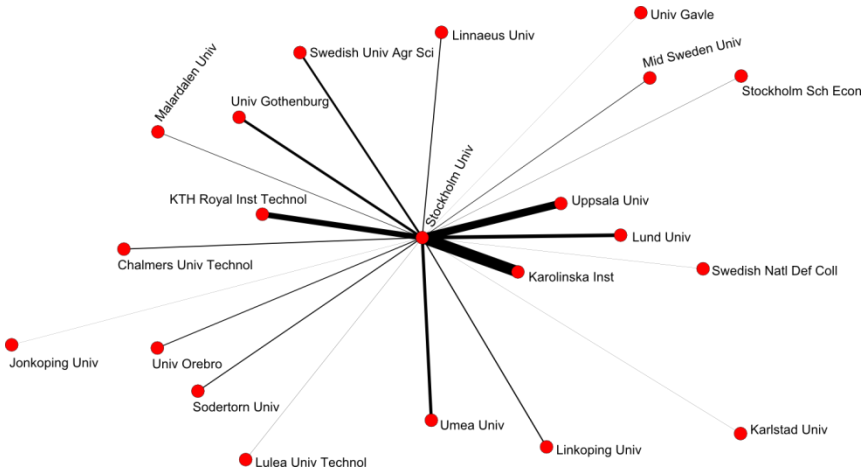

Figure 5. Network with SU and those 20 institutions that have the highest fractionalized co-publishing rates in pair with SU for the period 2008-2012.

4. Discussion and conclusions

We have studied collaboration, in terms of co-publishing, between SU and other Swedish academic institutions. Non-fractionalized and fractionalized co-publishing rates for SU in combination with other institutions are reported and parts of the outcome of the study are visualized in terms of co-publishing networks.

With a non-fractionalization approach, a paper with two organizations is treated, with respect to the weight of the collaboration pairs, in the same way as a paper with, say, 50 organizations, although the strength of the collaborations probably is less in the latter case. The average number of organizations per publication varies between fields, with particle physics as an extreme case, where publications with a large number of organizations (and of authors) are not uncommon. Instead of leaving the extreme cases out, we used a logarithmic approach to down weight cases where the number of organizations is greater than 2. An advantage of a logarithmic approach to fractionalization is that it is possible to adjust the sharpness of the down-weighting by varying the base of the logarithm (higher values of the base give less sharp down weighting).

When fractionalization is used, the co-publishing rate for SU, in pair with Uppsala University and KTH Royal Institute of Technology, drops considerably, while the rates for SU in pair with Karolinska Institute and smaller universities rise. Since KTH Royal Institute of Technology is specialized in technology, and Uppsala University has several researchers active in particle physics, it is reasonable to assume that SU and KTH Royal Institute of Technology/Uppsala University have collaborations including many organizations.

Questions about choosing collaboration partners based on proximity, rather than (or together with) fitness for purpose, have been raised in the literature (see e.g. Cronin, 2008). In our data, geographical proximity plays a role: among the seven institutions with the highest co-publishing rates with SU (for the period 2008-2012), four are located in the Stockholm/Uppsala area, irrespective of if fractionalization is used or not.

Bibliometric methods have been proposed for library collection development for a considerable amount of time (for recent examples, see Ashrafi et al., 2012; Zainal & Zainab, 2011). We have not, however, found much evidence for the use of bibliometric methods in relation to library collaboration, although some efforts exist (Perrault, 2004). The present study focuses on finding possible library collaborators for SU Library. The library collaboration would be related to research collaboration, and could include Swedish institutional archives, as well as the building of the national Swedish research database, SwePub. We point out that bibliometric results are not the only information that can be used to support decisions on library collaboration. The results of bibliometric studies should be used in conjunction with other relevant information. Finally, future

studies could treat co-publishing patterns within certain fields (subjects), with regard to a given university and other institutions. In a study along these lines, the WoS subject categories might be utilized.

References

Ashrafi, F., Mohammadhassanzadeh, H., Shokraneh, F., Valinejadi, A., Johari, K., Saemi, N., . . . Ashayeri, H. (2012). Iranians' contribution to world literature on neuroscience. *Health Information & Libraries Journal, 29*(4), 323-332.

Cronin, B. (2008). On the epistemic significance of place. *Journal of the American Society for Information Science & Technology, 59*(6), 1002-1006.

de Nooy, W., Mrvar, A., & Batagelj, V. (2011). *Exploratory social network analysis with Pajek* (Rev. and expanded 2nd ed.). Cambridge: Cambridge University Press.

Katz, J. S., & Martin, B. R. (1997). What is research collaboration? *Research Policy, 26*(1), 1-18.

Perrault, A. H. (2004). The role of WorldCat in resources sharing. *Collection management, 28*(1-2), 63-75.

Persson, O., Danell, R., & Schneider, J. (2009). How to use Bibexcel for various types of bibliometric analysis. In F. Åström, R. Danell, B. Larsen & J. Schneider (Eds.), *Celebrating Scholarly Communication Studies: A Festschrift for Olle Persson at his 60th Birthday* (pp. 9-24). Leuven, Belgium: International Society for Scientometrics and Informetrics.

Zainal, H., & Zainab, A. N. (2011). Biomedical and health sciences publication productivity from Malaysia. *Health Information & Libraries Journal, 28*(3), 216-225.