Eight clusters : a dynamic perspective and structural analysis for the evaluation of institutional research performance

Thijs, B.

5 A structural analysis of collaboration between European research institutes

Bart Thijs[1], Wolfgang Glänzel[1,2]

Bart.Thijs@econ.kuleuven.be,

Wolfgang.Glanzel@econ.kuleuven.be

[1]Katholieke Universiteit Leuven, ECOOM,

Dekenstraat 2, B-3000 Leuven, Belgium

[2]glanzw@iif.hu

Hungarian Academy of Sciences, IRPS, Budapest, Hungary

Presented at the 12th International conference on Scientometrics

and Informetrics, ISSI 2009, Rio de Janeiro, Brazil

Submitted to *Research Evaluation*

Collaboration in science has been the subject of many studies. In this study we focus on the influence of the research profile of an institute on its collaborative behaviour and pattern. The classification model developed by the authors provides a helpful tool to identify the specialisms of a research institution and to create groups of likewise institutions that enable to study the relation between science fields and collaborative behaviour. First we could show the effect of research profile on the shares of different types of collaboration. Next, citation indicators are used to investigate the changes on impact and publication strategy over the different types of collaboration. And in the last part we try to find for each group the research profile of the most preferred partners.

5.1 Introduction

An extensive literature on collaborative behaviour of nations, institutes and individual researchers has been build during past decades. Many authors found a positive relation between collaboration and scientific productivity (De Sola Price and Beaver, 1966; Beaver and Rosen, 1978, 1979; Glänzel and De Lange, 2002). These findings were extended by including also citation indicators to establish a relation between collaboration and impact or visibility of scientific output (Schmoch and Schubert, 2008). Other authors investigated collaboration patterns in specific fields or disciplines (Bordons et al. 1996, Gomez, et al., 1999). Most of these papers deal with collaboration at the macro (national) level (eg. Luukkonen et al. 1992, 1993).

With this paper we want to study collaboration behaviour among research institutions and not only look at international but also at domestic extramural collaboration. Especially the influence of the fields in which an institution is active will be the subject of this paper. It is still remains unclear how the research profile of an institution – the specialism of science fields in which an institute is active – influences its collaboration pattern. In order to shed more light on this problem area we proceed from the following model.

A clustering and classification model for institutions was developed earlier by the authors (Thijs & Glänzel, 2008) and provides a unique tool to study the effects of differences in research profiles. In this, a research profile of an institution is defined as a vector in the 'field space', it holds the shares of publications in each of 16 major fields in science. We use the subject category system as developed by Glänzel and Schubert (2003). A cluster analysis using these vectors as input variables divided the institutions into eight different clusters. Each cluster is labelled by the most typical field in which the member institutions are active. These are the clusters 1. Biology (BIO), 2. Agriculture (AGR), 3. a group of institutions with a multidisciplinary focus (MDS), 4. Earth and Space Sciences (GSS), 5. Technical and Natural Sciences (TNS), 6. Chemistry (CHE), 7. General Medicine (GRM) and 8. Specialized Medicine (SPM). The research activities in each cluster are not limited to the field indicated in the label of the cluster and the third cluster is labelled as multidisciplinary as research activities of member institutions are distributed along several fields. This MDS group contains most of the large universities active in almost all fields of science. After the cluster analysis, a discriminant analysis created a classification model that enables us to assign an institution, bases on its own research profile, to one of the eight clusters. This model allows us to classify institutions that were not in the original set used for clustering. In order to distinguish between the results of the cluster analysis and the classification of institutions by the model we will use the term 'group' instead of 'cluster'.

The main goal for the creation of this classification model was to enhance comparability between institutions for evaluative studies and benchmarking exercises (Thijs & Glänzel, 2008, 2009). However, the model proved to be useful for other applications as well. In the present study we use this model to investigate collaboration between institutions.

We formulated seven different questions to get a better understanding of inter-institutional collaboration:

Publication counts:

1. Does the share of collaborative papers of an institution differ amongst groups?

Citation indicators:

2. Do citation indicators on sets of collaborative papers of institutions differ amongst groups?
3. Are citation indicators of sets of co-authored papers of institutions higher than those on papers with no collaboration?
4. Is there a difference between groups in improvement of citation based indicators?

Collaboration patterns:

5. Do pairs of institutes that collaborate have more similar research profiles than those pairs that do not collaborate? This is a plain binary question comparing those that collaborate with those pairs that do not compare.

6. Do institutes that are more similar in research profile have a stronger collaborative relation? This is the reverse version of the previous question checking whether a growing similarity goes with a growing collaboration.
7. Is there a difference in most preferred partners between the groups?

Some of the issues mentioned here are already tackled at the level of collaboration between countries and positive relations were found between collaboration and both publication and citation indicators but this analysis has not yet been done at the meso-level.

5.2 Data sources, processing and methodology

Data were extracted from the yearly updates of the Web of Science database of Thomson Reuters (Philadelphia, PA, USA). Only papers of the document type article, letter, and review indexed in the 2003 to 2005 volumes have been selected. This data has undergone a detailed cleaning and processed to bibliometric indicators.

Subject classification of the publications was based on the field assignment of journals according to sixteen major fields and sixty subfields of science developed in Leuven and Budapest (Glänzel & Schubert, 2003).

Citations were calculated for a three year citation window, the year of indexing in the database and the two subsequent year. This means for publications of 2003, citations from 2003, 2004 and 2005, for the 2005 publications the window stretches from 2005 to 2007.

Label	Description	Share
BIO	Biology	6.5%
AGR	Agriculture	3.8%
MDS	Multidisciplinary	32.5%
GSS	Geo & Space Sciences	3.6%
TNS	Technical & Natural Sciences	13.2%
CHE	Chemistry	3.3%
GRM	General & Research Medecine	9.5%
SPM	Specialised Medicine	27.7%

Table 34. Eight clusters of institutes with label, description and share

Publications were assigned to countries and institutions according to the address in the by-line of the paper using a 3-step assignment procedure. For this study we selected fifteen European countries (EU15 without Greece but including Switzerland). We have excluded those institutes with less than 20 publications in the publication window 2003-2005. We do not make a distinction between academic or non-academic organizations, or between private or public institutions.

The remaining 1736 institutes were assigned to one out of eight different groups based on their research profile using the classification model developed by the authors (Thijs & Glänzel, 2008). Table 34 lists the groups, their label and description together with their share in the total set of institutes.

Each group can be characterized by an average research profile of the groups member institutions. Table 35 presents these average profile by group. As journals can be assigned to more than one field the sums of shares can be more than 100%.

Field	BIO	AGR	MDS	GSS	TNS	CHE	GRM	SPM
Agriculture (A)	**20.0%**	**66.7%**	6.0%	6.2%	4.9%	4.6%	0.2%	0.6%
Arts & Humanities (U)	0.0%	0.0%	0.3%	0.0%	0.2%	0.1%	0.1%	0.1%
Biology (Z)	**57.2%**	**25.1%**	9.6%	1.8%	2.3%	1.6%	4.1%	5.2%
Biomedical Research (R)	7.2%	5.2%	11.2%	0.1%	1.5%	1.2%	9.3%	9.8%
Biosciences (B)	14.6%	5.2%	**15.5%**	0.5%	2.6%	2.1%	7.0%	5.9%
Chemistry (C)	5.5%	14.3%	**16.8%**	2.9%	**27.8%**	**86.6%**	0.6%	1.1%
Engineering (E)	1.4%	5.1%	9.0%	4.0%	**35.0%**	7.3%	0.6%	0.7%
General & Internal medicine (I)	9.0%	1.2%	12.7%	0.2%	0.8%	0.1%	**64.8%**	**31.2%**
Geo- & Space science (G)	7.1%	5.5%	4.5%	**87.4%**	7.7%	1.3%	0.0%	0.1%
Mathematics (H)	0.8%	0.7%	5.7%	0.5%	6.9%	0.7%	0.1%	0.2%
Neuroscience & Behavior (N)	1.8%	0.5%	9.1%	0.0%	1.1%	0.1%	2.6%	4.8%
Non-internal medicine specialties (M)	7.7%	4.0%	**16.8%**	0.3%	2.5%	0.9%	**33.5%**	**61.9%**
Physics (P)	1.4%	3.1%	10.5%	4.7%	**37.0%**	10.8%	0.4%	0.6%
Social Sciences I (S)	0.6%	0.3%	1.7%	0.1%	0.4%	0.0%	0.7%	1.5%
Social Sciences II (O)	0.3%	0.5%	1.3%	0.0%	0.3%	0.2%	0.0%	0.1%
Multidisciplinary (X)	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Table 35. Average research profile per group as expressed by the average share of each field

5.2.1 *Collaborative behaviour*

First we have to define different types of collaboration:

- *International*: At least two different countries appear in the by-line of the paper. When looking at the collaborative behaviour this means that these papers can also be assigned to at least one other country next to the country of the institute,
- *Within Europe*: there appears at least one other country from the selected European countries in the by-line of the paper,
- *Outside Europe*: at least one country is not a selected European country,
- *Extramural domestic*: at least two institutes from the same country appear on the by-line,

- *No extramural*: there is only one country mentioned in addresses on the paper and the paper could be assigned to only one institute. This type might, however, include so-called intramural collaboration between two or more research teams from the same institute.

The first four types of collaboration do not exclude each other, 'International' collaboration does imply at least 'Within Europe' or 'Outside Europe'. 'Extramural domestic' collaboration does not imply nor exclude any of the three other types of collaboration .'No extra-mural collaboration' does exclude all four other types.

5.3 Results

5.3.1 Publication counts

To answer the first question the share of papers in each type of collaboration is calculated for each of the institutions. Then the average share is calculated within each of the eight different groups. Table 36 shows these group averages for each type of collaboration. Thus, institutions in the fourth group, GSS, have, on average, 67.2% of all their papers with international collaborators, 47.1% of their papers with at least one European collaborator and 16.2% with no extramural partners.

Statistical tests are needed to give a clear answer on the first question. We want to know whether these observed differences in the average shares are indeed statistically significant. Non-parametric testing is appropriate here as the sample sizes of the eight different groups are not alike (see table 34) and the distribution of these shares is not normal in all eight groups (Conover, 1999 or Sheskin, 2004). We have applied a Kruskal-Wallis test. The results are shown in table 37. Within each type of collaboration, there is a significant difference in the average shares between the eight groups.

	Collaboration				
	International	Within Europe	Outside Europe	Extram. Dom.	No Extram.
BIO	46.5%	28.3%	28.1%	46.8%	25.3%
AGR	37.5%	23.2%	21.4%	46.3%	30.6%
MDS	39.5%	22.7%	24.3%	41.3%	34.4%
GSS	67.2%	47.1%	46.7%	46.4%	16.2%
TNS	43.3%	26.2%	25.9%	41.9%	30.7%
CHE	33.7%	21.3%	16.0%	42.8%	33.7%
GRM	31.8%	23.1%	18.0%	61.3%	26.2%
SPM	24.0%	15.5%	14.4%	56.6%	33.0%

Table 36. Average share of each type of collaboration over 8 classes

Collaboration	Kruskal test df=7
International	501.13***
Within Europe	282.90***
Outside Europe	491.08***
Extramural Domestic	303.77***
No Extramural	140.50***
*** p-value less than 0.0001	

Table 37. Kruskal-Wallis test for average of share of collaboration

The highest share of international collaboration is in the Geo & Space Sciences group. This group also has subsequently the lowest share of papers with no extramural collaboration. These results are easily understood when looking at the peculiarities and international character of Space Sciences (earlier described by Latour, 1987; or Beaver and Rosen, 1978). Figure 10 shows the clear difference in distribution of shares of international collaboration for each group.

The lowest share of international papers is found in the Specialized Medicine (SPM) group with only 24% of its total of papers. This group has, however, a quite large percentage of 'Extramural domestic' collaboration. Some typical institutes in this group are 'Salzburger Landeskliniken' in Austria, the Danish 'Bilharziasis Laboratory' or the 'National Institute of Occupational Health' or the 'Our Lady's Hospital for Sick Children' in Ireland. It seems that for these institutions, which are more directed towards the local community, the international collaboration is less important than in other groups. It confirms also the conclusion of Frame and Carpenter (1979) that there is a link between the position of a field on the basic – applied dimension and their rank in terms of degree of international collaboration. The other medical group 'GRM' has the largest percentage of the 'Extramural Domestic' type of collaboration.

In five of the eight groups about 33% of papers has no extramural collaboration. In the GRM and BIO-group the share is about 25% and as mentioned above the GSS has one out of six papers with no extramural collaboration. Thus we can conclude with a positive answer on our first question: Indeed, there is a significant difference in the share of co-authored papers among our eight groups for each of the types of collaboration.

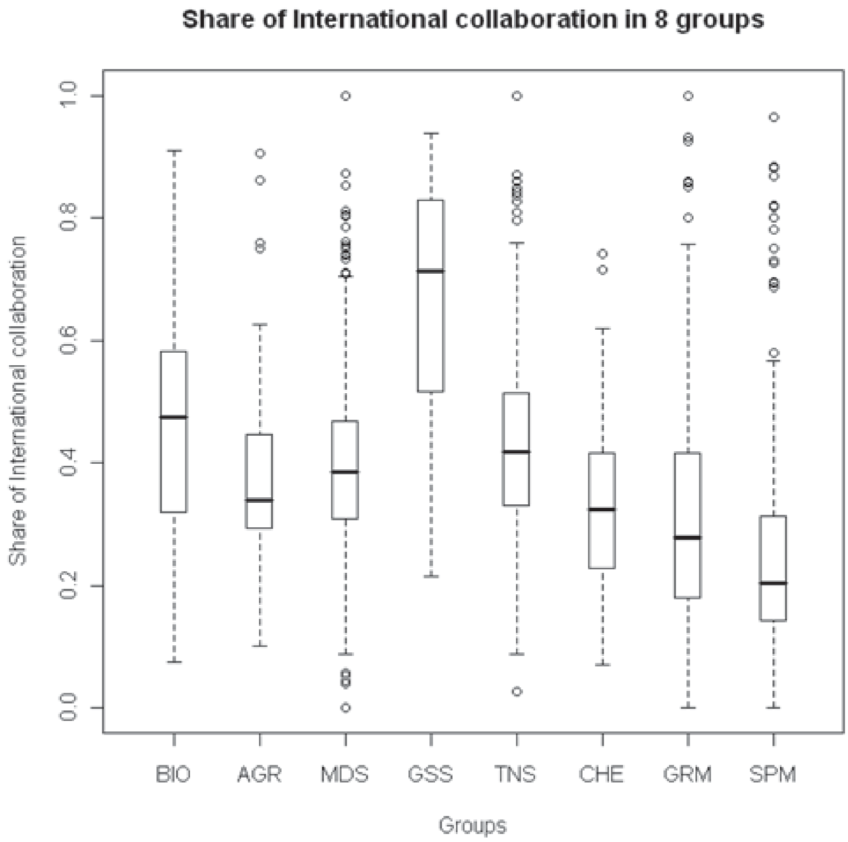

Fig. 10. Boxplot with distribution of share of international collaboration

5.4 Citation indicators

The next three questions deal with the impact of collaboration on citation indicators. In this paper three different citation indicators are used, RCR, NMCR and NMCR/RCR (introduced by Braun, Glänzel and Schubert, e.g., Schubert and Braun, 1986 and Braun and Glänzel, 1990).

First the Relative Citation Rate (RCR) is discussed. The RCR-value compares the observed citation rate of the institutes in a 3-year citation window with the expected citation rate. This expected citation rate is calculated on the basis of the journals in which papers are published. Values above 1 indicate that the observed citation rate of the institute lies above the expected rate.

Table 38 presents the average RCR values of different sets of papers calculated within each group. Within each group, averages are calculated over all member institutions. The highest values can be seen in the Geo & Space Sciences (1.19) and the medical classes GRM (1.25) and SPM (1.14). The last row gives the result of a Kruskal-Wallis test to check whether the values are different over the groups within each type of collaboration. We can savely conclude that the RCR value within each set of collaborative papers differs among groups.

	Collaboration					
	All Publications	International	Within Europe	Outside Europe	Extram. Dom.	No Extram.
BIO	1.09	1.19	1.28	1.28	1.11	0.93
AGR	1.11	1.24	1.32	1.38	1.14	0.99
MDS	1.05	1.18	1.26	1.22	1.13	0.90
GSS	1.19	1.32	1.38	1.42	1.33	0.78
TNS	1.03	1.10	1.16	1.15	1.08	0.89
CHE	1.00	0.99	1.02	1.11	0.93	0.97
GRM	1.25	1.67	1.87	1.87	1.37	0.82
SPM	1.14	1.51	1.66	1.59	1.27	0.75
Kruskal Test	65.91***	183.11***	166.33***	130.91***	120.35***	128.69***

Table 38. RCR in different types of collaboration

The next question formulated in the Introduction concerns the effect of collaboration on the RCR. The positive relation between scientific cooperation and citation impact has a large literature, but was mainly studied at the macro level and comparing ‘international collaboration’ with ‘no international collaboration’. The comparison of RCR-values of the different sets of papers by each individual institutions enters the meso-level. This approach adds the extramural domestic collaboration to the analysis.

Table 39 gives for the percentage of institutions for which the RCR-value of the set of collaborated papers is higher than the papers without collaboration. In the last column the RCR-values of the international papers are compared with the papers with domestic collaboration. In the first row, the share within the total set of 1736 institutes is given. A t-test is used to test whether this share is higher than 50%. About 80% of all institutes have a higher RCR value in both the set with international collaboration and the set with extramural domestic collaboration than the set with no extramural collaboration. A t-statistic with value 35.82 and 1735 degrees of freedom indicates that the share of institutions for which the RCR of the internationally co-authored papers is higher than the papers without extramural collaboration is statistically far above 50%. This means that we can give a positive answer on the third question on the basis of a relative citation indicator. For each combination of types of collaboration, the test indicates significant differences.

Table 39 gives these shares also for each separate group (e.g. 85.5% of all institutions in the GSS group have a RCR value for their international papers that is higher than their papers without collaboration). We see in the first column that the shares of institutes with a higher RCR-value on international collaboration is well above 50% for seven groups. Only within the Chemistry group the share is not significantly different from 50%. The same pattern can be found in the second column. It is clear that collaboration is associated with a higher RCR value.

	International vs. No Extram	Extram. Dom. vs. No Extram.	International vs. Extram. Dom.
Total	82.60%	79.55%	65.67%
t-test	35.83	30.52	13.74
BIO	78.76%	73.45%	61.85%
AGR	74.24%	68.18%	63.64%
MDS	83.56%	79.68%	63.07%
GSS	85.48%	85.48%	46.77%*
TNS	76.96%	70.43%	57.82%
CHE	52.63%*	43.86%*	57.89%
GRM	85.37%	84.76%	78.66%
SPM	88.49%	88.49%	72.19%
Kruskal	58.21	91.74	42.54
* not significant different from 50%			

Table 39. Share of institutes with a higher RCR-value in the first vs second type of collaboration

In the third column of table 39 we see that the share of institutes with an RCR value 'International' higher than the RCR value of 'Extramural Domestic' is still more that 50% but not as high as the two other columns. For the GSS-group, the observed value is even slightly below 50%, but not significantly. In general, collaboration itself - no matter if it is domestic or abroad- seems enough to increase this citation indicator but international collaboration gives an extra boost. The institutes from the medical clusters form an exception, international collaboration really changes the performance.

Furthermore, Kruskal-Wallis tests are used to test for significance of differences in shares among groups. The results are reported in the last row of table 6. In the three presented comparisons (cf. last row of Table 39) we found significant deviations among groups, indeed.

Comparing collaboration 'Outside' vs. 'Within Europe' (i.e., without and within selected European countries, respectively) did not reveal much difference. About 50% of the institutes have a RCR value of the Outside Europe collaboration set that is higher than the Within Europe value.

This means that for the RCR-value we can give three positive responses on the above mentioned questions about citation indicators:

- The RCR value does differ among groups on each sets of collaborative papers.
- The RCR value is for each type of collaboration higher than on papers with no collaboration. This is not true in the CHE-group.
- There is a difference between groups in improvement of the RCR value.

For the following analysis the Normalized Mean Citation Rate (NMCR) is used. This indicator compares the observed citation rate of the institutes in a 3-year citation window with the subfield based expected citation rate. This expected citation rate is calculated on the basis of the subfields in which papers are published. Besides the sixteen major fields, journals (and thus papers) are also assigned to sixty subfields on the basis of the subject classification scheme developed by Glänzel and Schubert. According to a recent study by Glänzel et al. (2009) this field-depth is sufficient for building subject-normalized indicators. For each subfield, an expected citation rate is calculated for the same 3-year citation window as for the observed citation rate for the papers. The ratio between the observed rate of the papers and the expected rate of the subfields indicates whether or not a set of papers attract more citations than one might expect based on the subfield. A value greater than 1.0 indicates more observed citations than expected. Results are presented in table 40 and table 41. These are in line with the results of the RCR-value. The Kruskal-Wallis tests show that the NMCR differs – for each type of collaboration- among the eight groups.

	Collaboration					
	All Publications	International	Within Europe	Outside Europe	Extram. Dom.	No Extram.
BIO	1.14	1.30	1.41	1.44	1.23	0.93
AGR	1.12	1.31	1.49	1.42	1.25	0.99
MDS	1.24	1.53	1.61	1.62	1.35	0.91
GSS	1.34	1.58	1.72	1.75	1.58	0.78
TNS	1.11	1.26	1.35	1.31	1.26	0.89
CHE	0.99	1.06	1.13	1.12	1.01	0.97
GRM	1.25	1.99	2.22	2.37	1.41	0.83
SPM	1.05	1.75	1.97	1.92	1.23	0.75
Kruskal Test	66.11***	126.18***	113.52***	117.15***	50.99***	128.68***

Table 40. NMCR in different types of collaboration

In each group, except CHE, the observed average value for collaboration is higher than the NMCR value for the 'no extramural collaboration' set. In four groups (BIO, AGR, GSS and CHE) we see that for about 50% of the institutes the value of NMCR is higher in international collaboration than in 'extramural domestic'. This means that for these groups, the effect of international collaboration does not essentially differ from the effect of domestic collaboration. The citation rates, standardized by field, are equal in both international collaboration and in extramural domestic collaboration. In the other four groups (MDS, TNS, GRM and SPM) there are more institutes with a higher NMCR value in the international collaboration set than in 'extramural domestic'. In the last row the results of the Kruskal-Wallis test are given. In the three presented comparisons we found significant differences in these shares among the eight groups.

We can conclude, that for this indicator, too, we have three positive answers on our questions:

- The average NMCR value is different between groups on each sets of collaborative papers.
- The average NMCR value is for each type of collaboration higher than on papers with no collaboration.
- There is a difference between groups in improvement of the NMCR value.

However, the difference between international and extramural domestic collaboration is not so pronounced as it was in the case of the RCR indicator.

	International vs. No Extram	Extram. Dom. vs. No Extram.	International vs. Extram. Dom.
Total	85.65%	81.51%	70.85%
t-test	42.37	33.81	19.11
BIO	79.64%	74.34%	53.10%*
AGR	78.78%	74.24%	60.60%*
MDS	88.16%	84.45%	70.60%
GSS	91.94%	93.55%	50.00%*
TNS	83.04%	80.00%	63.04%
CHE	59.65%*	52.63%*	52.63%*
GRM	86.58%	80.49%	84.76%
SPM	88.28%	83.68%	80.54%
Kruskal	46.17	48.85	86.66
* not significant different from 50%			

Table 41. Share of institutes with a higher NMCR-value in the first vs second type of collaboration

The last indicator (NMCR/RCR) is defined as the ratio between the journal-based and the subfield-based expected citation rates. A value above 1 means that the journal-based expectation is greater than the subfield-based citation rate since the latter one appears in the denominator (cf. Glänzel et al., 2009). As journals belong, at least partially, to the given subfield, a NMCR/RCR value greater than

1 corresponds to the situation where the set of papers in question is published in more visible journals, that attract more citations than the subfield on average. A value lower than 1 thus means that the papers are on the average published in journals rather belonging to the lower segment of the subfield. This value gives some indication of publication strategy of an institution or group.

	Collaboration					
	All Publications	International	Within Europe	Outside Europe	Extram. Dom.	No Extram.
BIO	1.04	1.09	1.10	1.12	1.11	0.94
AGR	1.00	1.07	1.13	1.03	1.09	0.89
MDS	1.17	1.29	1.28	1.32	1.19	1.13
GSS	1.13	1.20	1.23	1.23	1.19	0.90
TNS	1.06	1.14	1.17	1.13	1.12	0.92
CHE	0.98	1.05	1.08	0.98	1.04	0.88
GRM	0.99	1.18	1.17	1.25	1.02	0.82
SPM	0.91	1.13	1.16	1.15	0.96	0.79
Kruskal Test	233.14***	50.48***	166.33***	66.17***	210.5***	200.41***

Table 42. NMCR/RCR in different types of collaboration

Table 42 gives the average values for each group with the different types of collaboration. The Multidisciplinary group has the best ratio for all papers (1.17). The values for non-domestic collaboration (International, Inside or Outside Europe) are even higher. This means that these institutions publish in the journals with the highest visibility in the fields. The ratio for the papers with extramural domestic or no extramural collaboration is still quite above 1 (1.19 and 1.13). The MDS-group is the only group with a value above 1 for the no extramural collaboration. For five groups (BIO, AGR, GSS, TNS, CHE) the values in the four types of collaboration are about the same. This means that we could suspect that the type of collaboration has no essential impact on the publication strategy. Only papers with no extramural collaboration are published in journals with the lowest visibility..

In table 43, the NMCR/RCR-value is compared between different sets of publications. The pattern in this table is different from the two previous tables. The low value for Kruskal-Wallis test reported in the last row of the second column indicates that there is no significant difference between the groups. About 80% of all institutes, in each group has a value in the 'Extramural Domestic' set which is higher than the value in the 'No Extramural Collaboration'. More striking are the results in the last column. Only in the 'Multidisciplinary' and the Medical groups most of the institutions have a higher value in the 'International' set than in the 'Extramural Domestic' set. This is in line with the findings of Bordons et al. (1996) who investigated collaboration in biomedical sciences. The shares of about 50% for other five groups confirm what was seen in Table 42, as soon as these institutes collaborate with extramural partners the publication strategy tends towards journals with a higher impact.

	International vs. No Extram	Extram. Dom. vs. No Extram.	International vs. Extram. Dom.
Total	84.39%	81.85%	67.40%
t-test	39.47	34.43	15.46
BIO	76.99%	72.57%	44.25%*
AGR	80.30%	81.82%	50.00%*
MDS	86.04%	83.75%	69.61%
GSS	85.48%	82.26%	41.94%*
TNS	86.09%	86.09%	56.09%*
CHE	66.67%	75.43%	49.12%*
GRM	87.80%	81.71%	78.66%
SPM	84.73%	80.54%	79.71%
Kruskal	22.34	12.83**	120.61
* not significant different from 50%			
** p-value more than 0.05			

Table 43. Share of institutes with a higher NMCR/RCR-value in the first vs second type of collaboration

For this last indicator the answer is three times positive on the questions about citation indicators but it seems that the differences between the groups are once more smaller than in the previous cases.

- The average NMCR/RCR value is different between groups on each sets of collaborative papers.
- The NMCR/RCR value is for each type of collaboration higher than on papers with no collaboration.
- There is a difference between groups in improvement of the NMCR/RCR value. But this difference is not observed with all types of collaboration.

5.5 Collaboration patterns

After the sections on publication and citation indicators let us shift the attention towards the collaboration patterns between institutions. In this section we want to investigate the role of the research profile of an institution on the choice of collaborators.

For the fifth question – whether collaborating institutions have a more similar profiles – we have to define a measure for likeliness between institutions. For each of the 1736 institutions we have calculated the research profile, a vector containing 16 shares on the different fields. For each pair of institutes it is possible to calculate a cosine based on the Cauchy-Schwarz Inequality (Steele, 2004) between those vectors. It is defined as

$$\cos(\theta) \equiv \frac{\langle \underline{x}, \underline{y} \rangle}{|\underline{x}| \cdot |\underline{y}|}$$

or, in our case, with vectors containing 16 fields in Sciences, Social Sciences and Art & Humanties the similarity between two institutions is defined as

$$\cos(\theta) \equiv \frac{(\sum_{i=1}^{16} x_i y_i)^2}{(\sum_{i=1}^{16} x_i^2)(\sum_{i=1}^{16} y_i^2)}$$

The value of this cosine varies from 0 for complete dissimilarity to 1. Using this formula we calculated the similarity between all pairs of institutes. For each of them we also determined whether they collaborated in the period 2003-2005. This allows us to calculated the average similarity over those institutes that collaborated and those that do not.

Figure 11 shows the distribution of the shares in a boxplot. It is clear that there is a difference between collaborating and non-collaborating pairs. The average similarity for collaborating pairs is 0.68 and for the non collaborating the average similarity measure is 0.38

A Kruskal-Wallis test (value is 80475.1, df=1, p<2.2e-16) was applied which indicated that the difference is significant. This allows us to conclude that the research profile of collaborating institutes is more alike than those pairs of institutes that have no collaborative relation. At this point it is difficult to draw conclusions on the causality of the relation.

5.5.1 Measuring collaboration preferences

However, we want to dig deeper into the relation between similarity and collaboration so that we try to answer a subsequent question (6th): ‘Do institutes that have more similar research profiles also collaborate more?’.

In order to be able to answer this question only the pairs that collaborate are selected and the Jaccard-index is used to indicate the strength of the collaborative relation.

The Jaccard index, also known as the Jaccard Similarity Index, is defined as

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

with A and B denote the set of papers assigned to two institutes. For this indicator the value 1 also indicates complete similarity, hence all papers would be joint papers of both institutions.As both the similarity of research profiles and the Jaccard-index of collaboration are continuous variables a Pearson correlation is suitable as test of significance of the relation.

R=0.17 (t=56.02, df=98350, p-value<2.2e-16)

The Pearson correlation of 0.17 together with the p-value suggests to reject any assumption of independence but the correlation is indeed very low. So, the answer to question six is positive, however, this low value is in fact not yet satisfactory in answering the question concerning the most preferred collaboration partners.

Identifying most preferred partners

In this section we return to the eight groups of institutions in order to find for each group of institutions the most preferred partners. In table 44 the share of combinations of two groups are given. These shares are calculated within one group, e.g. 86.4% of all collaborative papers from an institutions in the 'Biology' group are in collaboration with an institution from the 'Multidisciplinary' group.

These shares have to be compared with the expected distribution. This expected distribution would be the distribution if each institution would choose it partners independently from their research profile and thus from their group membership. This expected distribution has to be same for all the eight groups. It is calculated by taking all papers with at least two institutes and counting the number of papers in each group. See table 45 for this distribution.

	BIO	AGR	MDS	GSS	TNS	CHE	GRM	SPM
BIO	14.2%	28.2%	6.4%	3.8%	2.4%	2.1%	1.6%	2.9%
AGR	8.0%	6.8%	1.6%	1.0%	0.9%	0.8%	0.2%	0.3%
MDS	86.4%	77.8%	67.3%	88.3%	83.4%	90.2%	93.7%	92.1%
GSS	1.5%	1.3%	2.5%	17.4%	2.2%	0.4%	0.0%	0.1%
TNS	9.3%	12.8%	24.3%	22.1%	23.8%	45.0%	2.7%	4.1%
CHE	0.4%	0.5%	1.1%	0.2%	1.9%	3.2%	0.1%	0.1%
GRM	1.4%	0.5%	6.3%	0.1%	0.6%	0.4%	16.8%	12.1%
SPM	9.5%	3.6%	22.4%	0.9%	3.5%	2.1%	43.8%	31.3%
Chi-Square df=7	8626.4	5665.4	5207.6	7433.2	12579.6	1248.1	9796.3	187993.0

Table 44. Preference of collaborators.

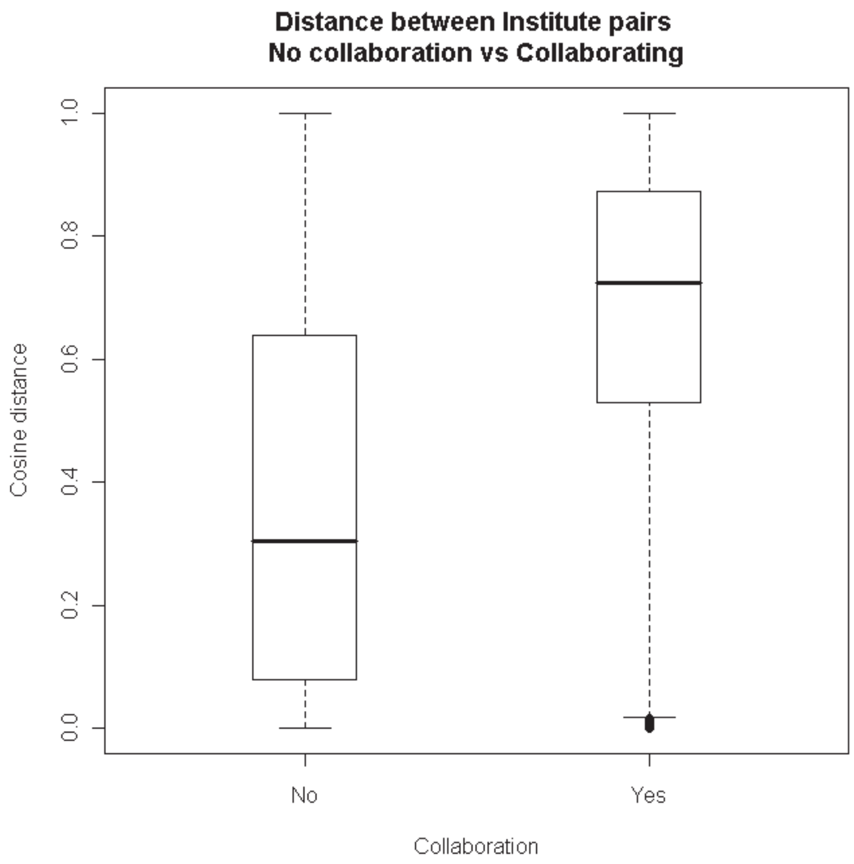

Fig. 11. Boxplot comparing distance between institutes

Expected Distribution	
BIO	5.6%
AGR	1.6%
MDS	76.2%
GSS	2.2%
TNS	22.2%
CHE	0.9%
GRM	5.1%
SPM	18.5%

Table 45. Expected distribution of co-authored papers.

Chi-Square tests (see table 44) confirm that for each group the observed distribution significantly deviates from the expected distribution (df=7; p<0.0001). Comparing the observed shares with the expected ones leads to these observations::

- For each group, institutions from the multidisciplinary group are the most likely partners,
- All shares of collaboration within groups are higher than expected,
- This observation does not hold for the Multidisciplinary group, here the within group collaboration is lower,
- For the seven specialized groups, the share of collaboration with the MDS group is the highest, it is also higher than expected. This allows is to conclude that members of the multidisciplinary group are the most preferred partners.
- Also closely related groups are more collaborative than other groups. These closely related groups are those groups that in the hierarchical clustering of the classification model are grouped together at a higher level of aggregation. These are the related groups AGR-BIO, TNS-CHE, GRM-SPM.
- For the multidisciplinary group the observed share of collaboration with the 'Specialized Medicine' cluster is 3.9% larger than the expected one. This is the highest increase for the Multidisciplinary group.

5.6 Conclusion

In this paper we tried to give an answer to several questions on the relation between collaboration and performance or productivity of European research institutions within the perspective of different research profiles. In earlier studies (Thijs & Glänzel, 2008, 2009) we could show that among our eight groups of institutes there are differences in publication and citation indicators. The first section of this paper makes it clear that differences can be found between these groups in choice of location of their collaborators. Institutes in the Geo and Space Sciences have 67% of their papers with an international partner while medical institutes have a share that is less than the half of this. However, for these medical institutes, extramural domestic collaboration is much more important. These two groups are the only ones where the share of this type of collaboration is more than 50%.

In a second section we looked at three different citation indicators to study the effect of the different types of collaboration on the impact of research and the publication strategies. We found collaboration itself is positively related on the visibility of research. Only in the CHE group this relation was absent. Extramural collaboration was enough to raise the citation scores but international collaboration added to this effect. There was not much difference in citation scores whether the collaboration was within Europe or with institutes outside Europe. Striking was the effect in the medical groups where a large difference in citation indicators between international and extramural domestic collaboration was observed. Combining this with the observation of large share of domestic collaboration for medical groups leads to the assumption of a deviating publication behavior for these two medical groups (GRM & SPM) as opposed to the six other groups in these types of collaboration.

In the last section we tried to identify the research profile of the most preferred partners for each group. First it was shown that collaborating institutions have profiles that are more alike than those that do not collaborate. However, a rather weak correlation was found between collaborating intensity and profile similarity among collaborating institutes. Looking at combinations of groups, it became clear that for each of the specialized groups the multidisciplinary institutes are the most preferred partners, even more than expected. Also the institutes from the same group and the related group (AGR-BIO, TNS-CHE, GRM-SPM) are likely partners. The strong link between two medical groups and the multidisciplinary one was striking. Future research with a detailed and more in depth study can be devoted to the relation between these groups.

References

Beaver, D., Rosen, R. (1978), Studies in scientific collaboration: Part I-Professional origin of scientific co-authorship, *Scientometrics*, 1 (1), 65-84.

Beaver, D., Rosen, R. (1979), Studies in scientific collaboration: Part III-Professionalization and the natural history of modern scientific co-authorship, *Scientometrics*, 1 (3), 231–367.

Bordons, M., Gomez, I., Fernandez, M.T., Angeles Zulueta, M., Mendez, A. (1996), Scientific Collaboration in Biomedical Research, *Scientometrics*, 37 (2), 279-295.

Braun, T., Glänzel, W. (1990), United Germany: The new scientific superpower? *Scientometrics*, 19 (5-6), 513–521.

Conover, WJ. (1999), *Practical nonparametric statistics*, 3rd edition. New York: John Wiley & Sons.

De Sola Price, D., Beaver, D. (1966), Collaboration in an invisible college, *American Psychologist*, 21, 1011-1018.

Frame, J.D., Carpenter, M.P. (1979), International Research Collaboration, *Social Studies of Science*, 9, 481-497.

Glänzel, W., De Lange, C. (2002), A distributional approach to multinationality measures of international scientific collaboration. *Scientometrics*, 54, 75–89.

Glânzel, W., Schubert, A. (2003), A new classification scheme of science fields and subfields designed for scientometric evaluation purposes, *Scientometrics*, 56 (3), 357–367.

Glänzel, W., Schubert, A., Thijs, B., Debackere, K. (2009), Subfield-specific normalized relative indicators and a new generation of relational charts: Methodological foundations illustrated on the assessment of institutional research performance. *Scientometrics*, 78 (1), 165188.

Gomez, I., Fernandez, MT., Sebastian, J. (1999), Analysis of the structure of international scientific cooperation networks through bibliometric indicators, *Scientometrics*, 44 (3), 441–457.

Latour, B. (1987), *Science in action*,Milton Keynes: Open University Press.

Luukkonen, T., Persson, O., Sivertsen, G. (1992), Understanding Patterns of International Collaboration, *Science, Technology, & Human Values*, 17 (1), 101-126.

Luukkonen, T., Tijssen, R.J.W., Persson, O. and Sivertsen, G. (1993), The Measurement of International Scientific Collaboration, *Scientometrics*, 28 (1), 15-36.

Schmoch, U., Schubert, T. (2008), Are international co-publications an indicator for quality of scientific research? *Scientometrics*, 74 (3), 361-377.

Schubert, T., Braun, T. (1986), Relative Indicators and relational charts for comparative-assessment of publication output and citation impact. *Scientometrics*, 9 (5-6), 281-291.

Sheskin, DJ (2004), *Handbook of parametric and nonparametric statistical procedures. 3rd edition*. Boca Raton: Chapman & Hall.

Thijs, B., Glänzel, W., (2008), A structural analysis of publication profiles for the classification of european research institutes, *Scientometrics*, 74 (2), 223-236.

Thijs, B., Glänzel, W., (2009), A structural analysis of benchmarks on different bibliometrical indicators for European research institutes based on their research profile. *Scientometrics*,.79 (2), 377–388.