

# Collaboration and Distances between German Immunological Institutes<sup>1</sup>

Frank Havemann\*, Michael Heinz\* and Hiltrun Kretschmer\*\*

*\*fh@wissenschaftsforschung.de, michael.heinz@rz.hu-berlin.de*

Institute of Library Science, Humboldt University Berlin, Dorotheenstr. 26, D-10117 Berlin (Germany)

*\*\*kretschmer.h@onlinehome.de,*

COLLNET, Borgsdorfer Str. 5, D-16540 Hohen Neuendorf (Germany)

## Abstract

The hypothesis that in recent years geographical proximity has become less important for research collaboration was tested and confirmed for papers co-authored by researchers at German immunological institutes in 1997 and 2002. Mean and median of distances between German immunological institutes weighted with the number of co-authored papers are greater in 2002 than in 1997. The un-weighted distance distribution of collaboration relations established between these institutes, however, has not changed significantly during the five years. In addition, productivity distributions of institutes have been found to be lognormal.

## Introduction

The global and secular tendency to more and more collaboration in scientific research has been demonstrated in many scientometric and other studies. To establish and maintain collaboration links can be more or less easy, depending on scientific, cultural, political, and geographical barriers which have to be overcome. In recent years some of these barriers have been lowered. The Internet has made telecommunication easy, cheap and fast. We have tested the hypothesis, that in recent years geographical proximity has become less important for establishing research collaboration. For this test we choose a sample of institutions where all other barriers for collaboration mentioned above are nearly absent. We analysed the collaboration between 80 German institutions of immunology, in 1997 and five years later.

## Other Studies

A similar study was made by Sylvan Katz (1994). He observed that in Australia, Canada, and UK the numbers of papers which members of two universities published together decreased exponentially with their distance. Smith & Katz (2000) found, that UK "life sciences showed the largest change in their geographical collaboration pattern, the average distance between collaborating institutions increasing over the time period [1980 - 1994]. The pattern of geographical collaborations in the natural sciences remained quite constant over time." Liang & Zhu (2002) obtained for inter-regional co-operation in China that geographical proximity is an important factor. Nagpaul (2003) studied international research collaboration and concluded that geographical proximity has greater positive impact than thematic proximity and socio-economic proximity.

## Data

The papers taken into account for this study are drawn from the 1997 and 2002 editions of the Science Citation Index (SCI). We analysed the publications of 80 German research institutions which are listed on the website of the German Society for Immunology (Deutsche Gesellschaft für Immunologie, [www.immunologie.de](http://www.immunologie.de)). 52 of the 80 institutes co-authored papers in 1997 or in 2002. Their 1326 mutual distances were determined.

---

<sup>1</sup> We thank Dr. Ute Kretschmer and Theo Kretschmer for preparing the data of immunological institutes.

**Method**

We use numbers of collaboration links and of collaborative papers to characterise the collaboration behaviour. Like Katz (1994) and others we plot numbers of collaborative papers against distances but take also into account how the collaborating institutions are scattered over the country. In addition, Salton's measure  $S_{ij}$  is used as an indicator of collaboration strength. It is defined as the number  $C_{ij}$  of joint publications of a pair of collaborators ( $i$  and  $j$ ) divided by the square root of the product of the number of total publication outputs of the two collaborators,  $T_i$  and  $T_j$ :  $S_{ij} = C_{ij} / (T_i T_j)^{1/2}$ .

**Results**

We have 1311 papers of 57 German immunological institutes in 1997, 48 of them co-authored by researchers at two different institutes. The corresponding figures in 2002 are 1635 papers in total produced by 66 institutes, 87 of the papers written in collaboration. There is only one paper where three of these institutes were involved.

*Productivity of Institutes and the Trend towards More Collaboration*

The productivity of institutes is lognormally distributed—if institutes without papers in SCI are excluded. In both years the log data pass the Kolmogorov-Smirnov (KS) test for the Gaussian distribution (Table 1, Fig. 1). For the collaborative productivity the KS test-values  $D\sqrt{n}$  are above the threshold .89. These distributions are even skewer then the lognormal one. The trend towards more mutual collaboration appears in higher growth rates of data in the sample of collaborative papers (Table 2). This trend can also be read off from Table 3 where all collaborative shares show an upward trend.

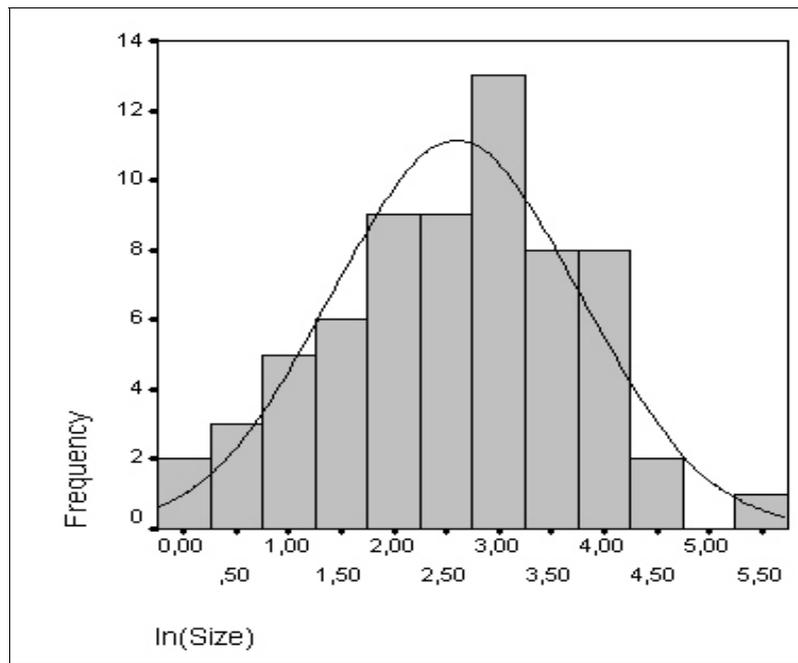


Figure 1: Distribution of 66 German immunological institutes with regard to the logarithms of their numbers of papers (“Size”) in journals filed in SCI 2002. The curve displays the best fitted lognormal distribution.

*Geographical Proximity*

How important is geographical proximity for German immunological institutes to establish collaboration links? To answer this question we first determined the distribution of all  $1326 = 52 \cdot 51 / 2$  distances between those 52 institutes which co-authored papers in 1997 or in 2002. Then we compared this distance distribution with those of links actually achieved in 1997 and 2002. All three distribution functions are displayed in Fig. 2A. For both years a clear bias toward smaller distances can be realised. The distribution functions of actually achieved links have smaller medians and means than the function of possible links (Table 4).

Table 1. Distributions of  $n$  German immunological institutes with regard to their numbers of papers in journals filed in SCI 1997 and 2002. The samples of collaborative papers include all those with authors from at least two of the institutes. KS test-value  $D\sqrt{n}$  was calculated with normal distribution of log data using mean and standard deviation of the observed log data.

| Sample of papers | SCI  | $n$ | Geometric mean | KS test value $D\sqrt{n}$ |
|------------------|------|-----|----------------|---------------------------|
| all              | 1997 | 57  | 11.60          | 0.850                     |
|                  | 2002 | 66  | 13.35          | 0.477                     |
| collaborative    | 1997 | 31  | 2.08           | 1.172                     |
|                  | 2002 | 50  | 2.63           | 1.210                     |

Table 2. Comparison of growth rates (percentages) 1997–2002 of data from the sample of all papers and of collaborative papers, respectively

| Sample of papers | Growth rates of          |                  |
|------------------|--------------------------|------------------|
|                  | number $n$ of institutes | number of papers |
| all              | 16%                      | 25%              |
| collaborative    | 61%                      | 81%              |

Table 3. Development of the collaboration network of 80 German immunological institutes: Numbers of collaborative papers, institutes and links and their shares in all papers (1311 and 1635), in all institutes (80) and in all possible links ( $3160 = 80 \cdot 79/2$ )

| Items                    | 1997 |     | 2002 |     |
|--------------------------|------|-----|------|-----|
|                          | Nr.  | %   | Nr.  | %   |
| collaborative papers     | 48   | 3.7 | 87   | 5.3 |
| collaborating institutes | 31   | 39  | 50   | 63  |
| Collaboration links      | 32   | 1.0 | 70   | 2.2 |

Table 4. Distance distribution of all possible collaboration links between 52 collaborating German immunological institutes, and of actual links in 1997 and in 2002

| Sample       | Median   | Mean     | Stand. Dev. | Skewness |
|--------------|----------|----------|-------------|----------|
| all possible | 388.0 km | 408.1 km | 214.4 km    | 0.20     |
| 1997         | 296.5 km | 318.8 km | 254.6 km    | 0.81     |
| 2002         | 295.5 km | 321.1 km | 245.2 km    | 0.28     |

We made KS tests whether the actual links behave as random samples from the distribution of all possible links in the sample of 52 collaborating institutes. The null hypothesis can be rejected with less than one percent failure probability in 2002 and with less than five percent in 1997. Then we asked whether the collaboration strength of already collaborating institutes depends on geographical distance. We used Salton's measure as an indicator of collaboration strength between a pair of collaborators. We tested whether the distance of collaborating institutions lowers the collaboration strength between them. On average, the stronger collaborations can be found with a somewhat higher probability for not so distant partners (Table 5). The null hypothesis of independence can be rejected with 4.4% failure probability. Corrected Cramer's  $C_{\text{corr}} = 0.23$  ( $C_{\text{corr}} = 0$  for independent and  $C_{\text{corr}} = 1$  for totally correlated variables). So, there is some weak dependence of collaboration strength on geographical distance.

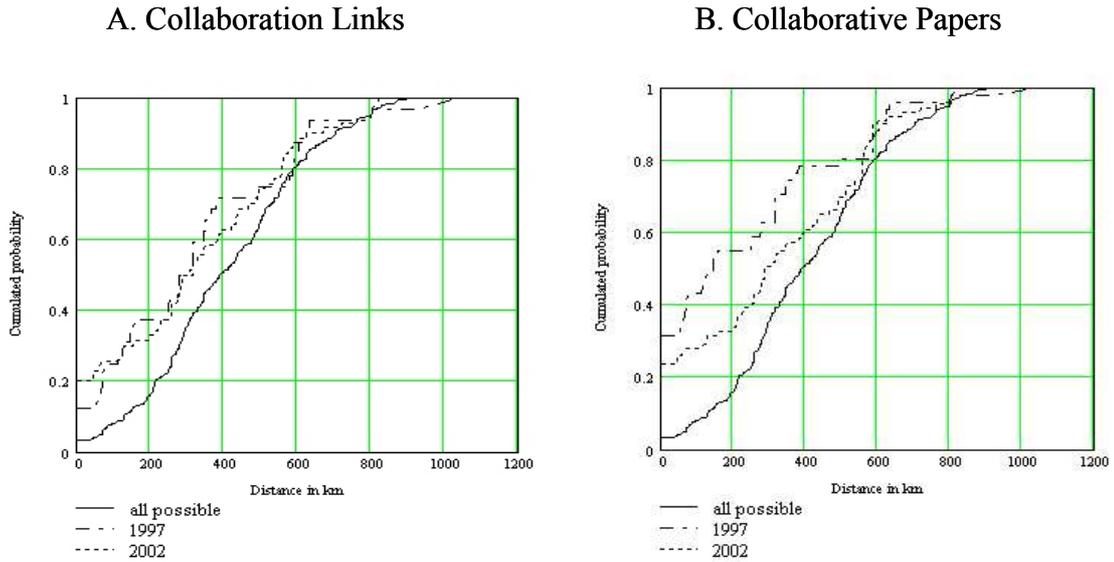


Figure2: A. Distance distribution functions of possible collaborative links between 52 German immunological institutes, and of actual links in 1997 and in 2002 (left graph).  
 B. Distance distribution functions of collaborative papers of German immunological institutes in 1997 and in 2002 (right graph; the distance distribution functions of all possible collaborative links between 52 German immunological institutes is shown for comparison).

Table 5. Collaboration strength  $S$  and distances between collaborating institutions. The expectation values of the  $\chi^2$  test are in brackets.

| Distance | $S > .05$ | $S < .05$  | Sum |
|----------|-----------|------------|-----|
| < 100 km | 13 (8.78) | 15 (19.22) | 28  |
| > 100 km | 19 (23.2) | 55 (50.78) | 74  |
| Sum      | 32        | 70         | 102 |

*Has Geographical Proximity Become Less Important?*

Has geographical proximity become less important to establish collaboration links within the five years from 1997 to 2002? To answer this question we made a KS test, whether the data of the two years behave as random samples drawn from one unknown theoretical distribution. Here the null hypothesis cannot be rejected (test value  $D [n_1 n_2 / (n_1 + n_2)]^{1/2} = 0.56$ ). Thus, the five year period under consideration brought no statistically significant change. A real change can be observed, however, if the amount of work done in collaboration is considered. For measuring this amount each link between two institutes was weighted by the number of papers co-authored by researchers of the two institutes. In this case, mean and median of distances are significantly greater in 2002 than in 1997 (Figure 2B and Table 6). The corresponding KS test value is 1.27. It far exceeds the one of the unweighted calculation. Rejecting the null hypothesis has less than 10% failure probability.

Table 6. Distance distribution functions of collaborative papers of immunological institutes.

| Sample | Median | Mean     | Stand. Dev. | Skewness |
|--------|--------|----------|-------------|----------|
| 1997   | 145 km | 237.8 km | 255.1 km    | 1.04     |
| 2002   | 285 km | 317.0 km | 247.8 km    | 0.18     |

## Discussion

The hypothesis that in recent years geographical proximity has become less important for *establishing* research collaboration within a country cannot be confirmed for the case of German immunological institutes in 1997 and 2002. For both years the partner institutes are not chosen randomly from all collaborating German immunological institutes but with a statistically significant bias toward short distances. This bias is not smaller in 2002 than in 1997. In contrast, the *numbers of collaborative papers* became less dependent on geographical proximity of co-authors. Mean and median of distances between German immunological institutes weighted with the number of co-authored papers are significantly greater in 2002 than in 1997.

## Further Studies

Further questions could be asked regarding the collaboration and distances between our sample institutes. For example, one could compare the distance distributions of collaboration links with the corresponding distribution of all 80 institutes (for this aim further distances had to be determined). As a next step we will take also earlier and later years into consideration. We expect, that with these data the hypothesis can even more clearly confirmed because the Internet was used by researchers since the early nineties. Our assumption that the sample of German immunological institutes is homogenous with regard to their research subjects can be checked by comparing the spectra of journals in which they publish their results. This we must leave to a further study, too.

## References

- Katz, J.S. (1994). Geographical proximity and scientific collaboration. *Scientometrics* 31, 31–43.
- Liang, L. M. & Zhu, L. (2002). Major factors affecting China's inter-regional research collaboration: Regional scientific productivity and geographical proximity. *Scientometrics* 55, 287–316.
- Nagpaul, P. S. (2003). Exploring a pseudo-regression model of transnational cooperation in science. *Scientometrics* 56, 403–416.
- Smith, D. & Katz, J. S. (2000). *Collaborative Approaches to Research*. HEFCE Fundamental Review of Research Policy and Funding, Final Report. Received from the website <http://www.sussex.ac.uk/Users/sylvank/pubs>, file: collc.pdf, date: 2005-1-3.