Available online at www.sciencedirect.com
SCIENCE DIRECT®

J. Eng. Technol. Manage. 22 (2005) 9–30

Journal of ENGINEERING AND TECHNOLOGY MANAGEMENT JET-M
www.elsevier.com/locate/jengtecman

The citation impact of research collaborations: the case of European biotechnology and applied microbiology (1988–2002)

Koen Frenken[a,*], Werner Hölzl[b], Friso de Vor[a]

[a] *Section of Economic Geography, Urban and Regional Research Centre Utrecht (URU), Faculty of Geosciences, Utrecht University, P.O. Box 80115, NL3508TC Utrecht, The Netherlands*
[b] *Department of Economics, Vienna University of Economics and Business Administration, Vienna, Austria*

Available online 26 January 2005

Abstract

Research collaboration is generally motivated by quality enhancement. The networks underlying collaborative knowledge production also serve as vehicles of knowledge diffusion. Both aspects are expected to contribute to the citation impact of publications. We analyse knowledge production in European biotechnology for the period 1988–2002 focusing on the role of research collaboration. Different aspects of research collaboration are taken into account simultaneously to assess their relative importance. We distinguish between the number of contributing authors and addresses as to differentiate between the effect of the collaboration between individuals and between organizations. We further distinguish between different spatial scales of collaboration (national, European, international) and between different institutional types of collaboration (between academia, outside academia, and hybrid). We find evidence that the diffusion of scientific knowledge, as measured by citation rate, is dependent on both intra- and inter-organisational characteristics. An important finding has also been that the further differences in citation impact can be related to the geographical scale of collaboration with the European scale being most successful. Furthermore, country-fixed effects suggest that the European Union, though successful as a geographical scale of collaboration, still harbours many national varieties of knowledge production.

JEL classification: L65; O32; O33

Keywords: Citation impact; Knowledge networks; Diffusion; Institutional collaboration

* Corresponding author.
E-mail address: k.frenken@geog.uu.nl (K. Frenken)
URL: http://econ.geog.uu.nl/frenken/frenken.html.

0923-4748/$ – see front matter
doi:10.1016/j.jengtecman.2004.11.002

1. Introduction

The role of research collaboration in modern scientific knowledge production is complex and multifaceted. Generally speaking, research collaboration enhances the quality of research, which leads papers with more authors to be cited more often (Katz and Martin, 1997). Recent research showed that the networks underlying collaborative knowledge production also serve as vehicles of knowledge diffusion, which would account for some part of their relative higher citation impact (Breschi and Lissoni, 2002; Singh, 2004). Thus, collaboration networks not only contribute to quality of knowledge, but also to its diffusion, which only reifies its importance from a science policy perspective.

Though high-quality empirical data on research collaboration are widely available, the analysis of the impact of research collaboration is by no means straightforward. There are multiple scales at which researchers collaborate, which challenge us to disentangle all relevant aspects in an empirical research design. There are at least three relevant levels of analysis at which collaboration can be studied, and which are, for theoretical reasons, not expected to be independent from one another nor to be fully overlapping either. First, one can distinguish between intra-organisational and inter-organisational collaboration. Second, one can distinguish between intra-regional and inter-regional collaboration (which obviously depends on the definition of regions applied). Third, one can distinguish between intra-disciplinary and inter-disciplinary collaboration (again, depending on the definition of disciplines applied). The three dimensions can be regarded as being orthogonal in that all possible combinations occur in reality.

In this contribution, we deal only with the organizational and geographical dimensions, as we focus on one particular discipline, biotechnology and applied microbiology, which we do not differentiate further into sub-disciplines. Our main interest is to disentangle the geographical and organizational aspects of scientific knowledge production. In particular, we test a number of hypotheses relating the citation impact of scientific articles published between 1988 and 2002 to the size and characteristics of the underlying collaboration network.

2. Research collaboration

Research collaboration has been a proliferating phenomenon in science. The most convenient indicator to describe this process is to count the number of co-authored scientific papers that are published. A recent study found that at the beginning of the twentieth century co-authorships accounted for less than 10 percent of all publications, while at the end of the twentieth century, this percentage had gone up to account for over 50 percent of all publications (Wagner-Doebler, 2001). There is little doubt that this percentage will further increase given that the trend has not levelled off.

The tendency for knowledge production to become increasingly collaborative can be understood as a consequence of increased division of labour among scientists. Science develops into an ever-increasing number of fields and sub-fields. Consequently, research requires some sort of integration between different knowledge bases, and, more often than not, between the knowledge of different people. In particular, though not exclusively, research collaboration has become the dominant form in problem-oriented and applied sciences, where many heterogeneous pieces of technologies, knowledge and competencies are combined to produce new pieces of technology, knowledge and competencies (Gibbons et al., 1994; Adams et al., 2004).[1] For example, patent analysis showed that the usefulness of R&D alliances between firms tends to increase when firms differ more in their knowledge bases (Rosenkopf and Almeida, 2003). Besides the aspect of division of labour, collaboration also provides economic opportunities to realize economies of scale, for example, with regard to costs of training and research infrastructure (Katz and Martin, 1997).

2.1. Internationalisation and institutional hybridisation

Research collaborations in science are not limited to collaboration between scientists, but also concern collaboration between research institutes. Research networks span across different organizations, and, increasingly, across geographical boundaries among countries (Katz and Martin, 1997) as well as across institutional boundaries among academia, industry and government (Gibbons et al., 1994; Rosenberg, 1990). The blurring of traditional boundaries reflects that knowledge production takes place primarily within international epistemic communities sharing codes of communication and practice, rather than in geographically localised (e.g., nation state) entities or historically institutionalised contexts (academia, industry, and government) per se. The first trend is associated with an increasing internationalisation of scientific research (Wagner, 2002), while the second trend is captured by concepts as innovation systems (Nelson, 1993) or the 'triple helix' (Etzkowitz and Leydesdorff, 2000).

Regarding internationalisation, it has been estimated that the share of international collaborations doubled between 1987 and 1997 to account for 15% of world publications in 1997 (Wagner, 2002). This trend continues, and has been further facilitated recently by the widespread use of the Internet, the emergence of English as standard language in most disciplines, and the rapid fall in the costs of long-distance travel. These processes, however, should not be taken to imply that one should speak of 'the death of distance' in science, and that location would no longer matter in scientific knowledge production. Evidence for the EU shows that the large majority of collaborations still occur within the boundaries of the nation state, and that this bias only slowly decreases (Frenken, 2002). Research at the sub-national level suggests that geographical proximity affects the probability of collaboration. Katz's (1994) study on collaboration in the UK showed that scientific collaboration decreases exponentially with the distance separating partners. A study by Liang and Zhu (2002) on China also found that geographical proximity is one of the important factors determining the pattern of inter-regional collaboration. Danell and Persson (2003) found that the distribution of R&D activities in all sectors is highly concentrated in the three main urban areas of Sweden. Within these three large regions occur the strongest flow of PhDs and the strongest co-authorship links between the sectors. In all, geographical proximity thus provides a good predictor for the frequency of collaboration between research institutes.

[1] In this context, Gibbons et al. (1994) proposed to distinguish the organisation of problem-oriented, applied and interdisciplinary research from traditional disciplinary scientific knowledge production, calling the former "Mode 2" and the latter "Mode 1".

One of the factors explaining the importance of geographically localised networks of knowledge production is said to be the institutional hybridisation of knowledge production. Since collaborations in university–industry–government networks are only partially formalised in contractual arrangements, these are to be supplemented by frequent face-to-face, informal contacts, and the exchange of personnel. The formation and stability of these networks is to an important extent facilitated by geographical proximity between the participants who share the same institutional environment, also termed ‘‘regional innovation systems’’ (Cooke et al., 1997; Fornahl and Brenner, 2003).

In spite of the hybridisation of scientific knowledge production and the increased involvement of industry herein, one does not expect that the organisation of scientific knowledge production will converge to that of industrial knowledge production. Scientific research is qualitatively different from industrial innovation in a number of ways (Dasgupta and David, 1994). First, the tacit component is smaller in scientific knowledge production (although, certainly present), which renders communication and collaboration at a distance much easier compared to innovation processes. Second, the specificity of knowledge (‘appliedness’) is expected to be much smaller in scientific research compared to industrial R&D. Consequently, problem definitions are to a lesser extent determined by the local context (e.g., market demand or government regulation). Rather agendas emerge from a global discourse. Third, the incentive structure in scientific knowledge production is explicitly oriented towards (international) diffusion, while investors in industrial R&D have an incentive to appropriate the results (whatever the mechanism used to achieve this). For these reasons, one should expect scientific knowledge production to be less localised geographically than industrial innovation.[2]

2.2. *Costs and benefits of research collaboration*

Research collaboration comes at a cost and at a risk. In all cases, these are to be assessed against the potential benefits. To assess the benefits of international collaboration, one can distinguish between benefits for individual researchers contributing to research, and benefits for the research team as a whole (Katz and Martin, 1997).

The individual benefits can be researched by looking at an author’s productivity. It is well known since a study by Lotka (1926) that researchers differ considerably in their productivity. The power-law found by Lotka suggests that the number of authors producing n papers is proportional to $1/n^2$. This means that, simply stated, the large majority of authors have a very low productivity, some have a medium productivity, and only few have a high productivity. This result is rather striking if one assumes the input of researchers, in

[2] A number of studies have addressed the characteristics of tacitness, specificity and appropriability of knowledge as variables that explain the degree of geographical localisation of knowledge production (Feldman, 1999). The classic study has been a US patent citation study that found that specificity and appropriability of knowledge as documented in patents contributed significantly to the extent that citation originated from the same region (Jaffe et al., 1993). If one accepts that scientific knowledge production is typically characterised by a low degree of specificity and appropriability, this finding suggests that the degree of localisation of scientific knowledge production is indeed lower compared to industrial R&D. This is also in line with recent evidence showing that patents drawing heavily on published texts, diffuse more quickly and more widely in geographical space (Sorenson and Fleming, in press).

terms of working time, to be rather equal (as people are physically constrained). Part of the explanation for these large differences lies in the productivity benefits of collaboration. The input required to be an author of a paper decreases with the number of co-authors of a paper. As expected, the most productive scientists are also those who collaborate most frequently. In this way, they are able to generate papers with relatively few inputs in terms of time (Katz and Martin, 1997).

More recently, the power-laws found in the production of scientific papers by individual scientists have been understood as reflecting a self-reinforcing success-breeds-success mechanism (Barabasi et al., 2002; Newman, 2004). Successful authors in terms of productivity are more attractive research partners for other authors. Intellectually, the more productive researchers will generally prove to be more rewarding to collaborate with due to experience and access to latest research results. Possibly more important, more productive authors provide access to larger networks of previous collaborators, to resources, and to editorial boards. Furthermore, the reputation of successful authors carries over, to some degree, to collaborators. As a result, already successful researchers attract most collaborators and they can select the best ones among these. As a result, highly productive scientists will tend to work with other highly productive scientists (Katz and Martin, 1997).

The benefits at the level of the research team, i.e., the mutual benefits of collaboration, stem from a variety of resources. As summarised by Katz and Martin (1997) these team benefits range from the cross-fertilisation of ideas that previously were unconnected, better quality control through internal refereeing and a higher rate of acceptance in journals. Researchers contributing to a paper are also important vehicles of knowledge diffusion. Thus, a multi-authored paper may simply be more successful than a single-authored paper, other things being equal, because it diffuses more widely through the personal networks of the contributing authors.[3] All these factors contribute to the fact that the number of citations that a paper receives is positively related to the number of authors. However, since coordination costs increase exponentially with the number of collaborators, one expects decreasing returns to set in once the marginal costs of coordination in the research team become larger than the marginal benefits of a researcher added. This explains why the size of research teams is constrained. One can think of an optimal size of research group, the exact size being different in different disciplines.

Hypothesis 1. A higher number of authors of a paper increases a paper's citation impact.

The costs and benefits of collaboration between scientists holds, *mutates mutandis*, also for inter-organizational collaboration. However, the exact costs and benefits associated with inter-organizational collaboration compared to intra-organizational collaboration are expected to differ. First, the benefits that can be generated from inter-organizational collaboration are expected be higher than of intra-organizational collaboration, as the benefits of the latter type of collaboration is limited by the knowledge base within the

[3] Recent attempts to single out the effect of previous collaboration on citation flows, be it in the context of patents rather than publications, is the study by Breschi and Lissoni (2002) using Italian patent data of the European Patent Office and the study by Singh (2004) using USPTO data.

organization, while the former type of collaboration can be of any kind. Put differently, and in line with Rosenkopf and Almeida (2003), inter-organisational collaboration allows an organization to escape from its narrow knowledge base. However, the costs of inter-organizational collaboration are also expected to be higher than the costs of intra-organizational collaboration as organizational boundaries create the risk of opportunism and misunderstanding as well as higher bureaucratic costs.

Hypothesis 2. A higher number of organizations contributing to a paper increases a paper's citation impact.

Following our previous discussion on institutional hybridisation of knowledge production, one can expect that collaborations between academia are more successful in terms of citation impact that collaborations involving firms or government institutes. University scientists are more often concentrating on basic knowledge, which is applicable in many contexts. Researchers working for firms and governments, by contrast, are expected to produce more specific and more applied knowledge useful to their organization. This does not mean that scientists working outside academia do not have incentives to publish (Rosenberg, 1990), but rather that the knowledge produced is more applied compared to scientists working in academia.

Hypothesis 3. Collaboration within academia increases a paper's citation impact compared to collaboration involving non-academic organizations.

The success of collaborative research not only depends on the number of authors, but also on the spatial scale of collaboration. Narin et al. (1991) were the first to find that internationally co-authored papers are cited significantly more than nationally co-authored papers (about twice as much). This can be understood, again, as balancing costs and benefits. Choosing research partners internationally greatly enlarges the probability of a complementary knowledge bases as well as the geographical scale of diffusion. However, costs and risks of international collaboration are much higher compared to national collaboration given differences in national institutions and, possibly, language and culture as well.

Hypothesis 4. International collaboration increases a paper's citation impact compared to national collaboration.

3. Research design

To understand the impact of research collaboration on knowledge diffusion empirically, the difficulty is to take into account all facets of research collaboration. In previous research, most contributions concentrated on selected aspects only, for example, the collaborative nature of knowledge production, the geography of knowledge production, or the institutional hybridisation of knowledge production. Relatively few studies have attempted to disentangle the organisational, spatial, and institutional aspects of modern scientific knowledge production. In the following, we attempt to take different aspects of research collaboration into account simultaneously to assess their relative importance in a multiple regression analysis. In order to do so, we take the paper as the unit of analysis and characterise the paper by its year of production, the number of authors, the number of addresses, the geographical scale of collaboration (if any), the institutional scale of collaboration (if any), and, the countries of origin of participating organizations. We also construct a variable to indicate whether a top institute participated to control for reputation and scale effects, where a top institute is defined as an organization producing more than ten articles during the period 1988–2002 in the field of biotechnology and applied microbiology. These variables are expected to explain jointly the citation impact of a paper.

3.1. Data

The main data source in scientometrics in general (and used in our study) is the Web of Science, a product offered by the Institute of Scientific Information (ISI, http://www.isinet.com/). Web of Science contains information on publications in all major journals in the world. It covers three databases: the Science Citation Index (SCI) including natural science journals, the Social Science Citation Index (SSCI) including social science journals, and the Arts and Humanities Citation Index (A&HCI) including journals belonging to the arts and humanities. Using Web of Science, one can construct data on a specific discipline in a relatively straightforward way. Once a list of journals is obtained that is representative for the scientific discipline in question, publications belonging to a discipline can be simply retrieved by using the set of journals as a query. We analysed publications for the field of biotechnology and applied microbiology using a list available from the Institute for Scientific Information at the above mentioned website (see the Appendix A for the list of journals).

Focusing on the European science system, we selected all papers that contained at least one address for a country in the European Union for the period 1988–2002 (data before 1988 were not available to us). In total 19,769 papers were published in the field of biotechnology and applied microbiology in this period containing at least one EU address. The data were retrieved on the 24th of May 2003 from the Web of Science. Definitions of variables are given in Appendix B. The variables ‘univ’, ‘nuniv’ and ‘hybrid’ deserve further explanation. The classification of collaboration as being between academic institutes (univ), between non-academic institutes (nuniv) or hybrid depends crucially on the definition of an academic institution. Whether or not an organization counts as an academic institute is made dependent on whether the primary goal of the organization is the advancement of science. All universities, colleges, schools, and non-profit organisations with the advancement of science as its primary goal have been categorised as academic. Non-academic organisations include firms, hospitals, and government research agencies that have primary goals other than advancing science, though very often pursue the advancement of science as a secondary goal.[4]

[4] All entries have been manually checked using the information on the websites of the organization.

3.2. Descriptive statistics, trends and country profiles

Table 1 provides the descriptive statistics of the database. The variable to be explained is the total number of citations received at the time of the database construction (24 May 2003). From the data, Figs. 1–3 are constructed to provide some descriptive insight in the

Table 1
Descriptive statistics

	Descriptive statistics				
	N	Minimum	Maximum	Mean	S.D.
Citation	19769	0	675	8.56	15.417
Author	19769	1	8	3.38	1.517
Address	19769	1	29	1.52	.858
Top	19769	0	1	0.344	
Univ	19769	0	1	0.260	
Nuniv	19769	0	1	0.036	
Hybrid	19769	0	1	0.145	
Nation	19769	0	1	0.224	
EU	19769	0	1	0.068	
Inter	19769	0	1	0.137	
AUS	19769	0	1	0.028	
BEL	19769	0	1	0.034	
DEN	19769	0	1	0.028	
FIN	19769	0	1	0.025	
FRA	19769	0	1	0.170	
GER	19769	0	1	0.180	
GRE	19769	0	1	0.015	
IRE	19769	0	1	0.014	
ITA	19769	0	1	0.070	
LUX	19769	0	1	0.000	
NET	19769	0	1	0.079	
POR	19769	0	1	0.028	
SPA	19769	0	1	0.105	
SWE	19769	0	1	0.054	
UK	19769	0	1	0.243	
y1988	19769	0	1	0.038	
y1989	19769	0	1	0.041	
y1990	19769	0	1	0.045	
y1991	19769	0	1	0.061	
y1992	19769	0	1	0.061	
y1993	19769	0	1	0.062	
y1994	19769	0	1	0.070	
y1995	19769	0	1	0.070	
y1996	19769	0	1	0.077	
y1997	19769	0	1	0.060	
y1998	19769	0	1	0.066	
y1999	19769	0	1	0.082	
y2000	19769	0	1	0.082	
y2001	19769	0	1	0.099	
y2002	19769	0	1	0.085	
Valid *N*	19769				

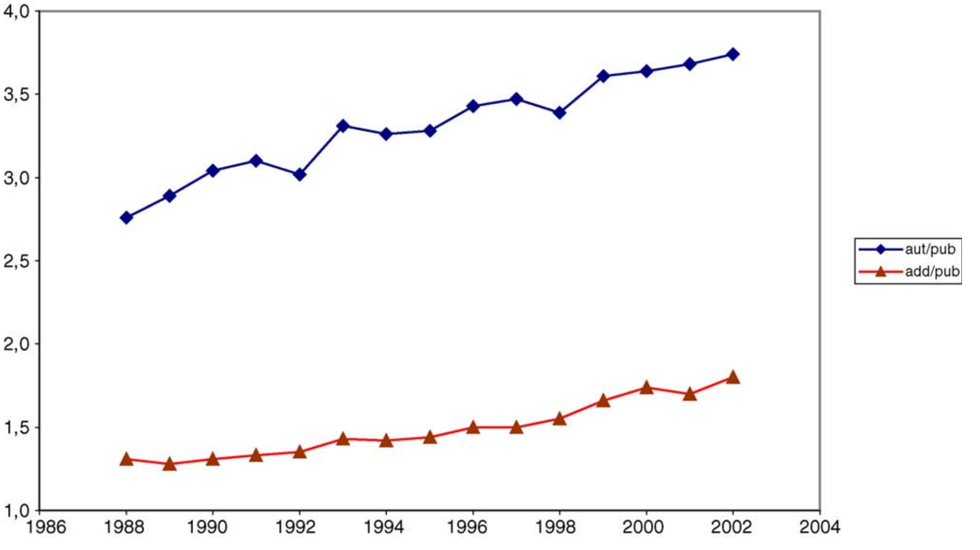

Fig. 1. Average number of authors per publications and average number of addresses per publications, per year.

rising importance of collaborations and the patterns in collaboration per country in the European Union. Fig. 1 shows the average number of addresses and authors per publication for the EU. It is clear that the number of addresses per publication has steadily risen in the past fifteen years approaching 1.8 addresses per publication in 2002 in the EU. This means that inter-organisational networks are proliferating rapidly in the field of biotechnology and applied microbiology.

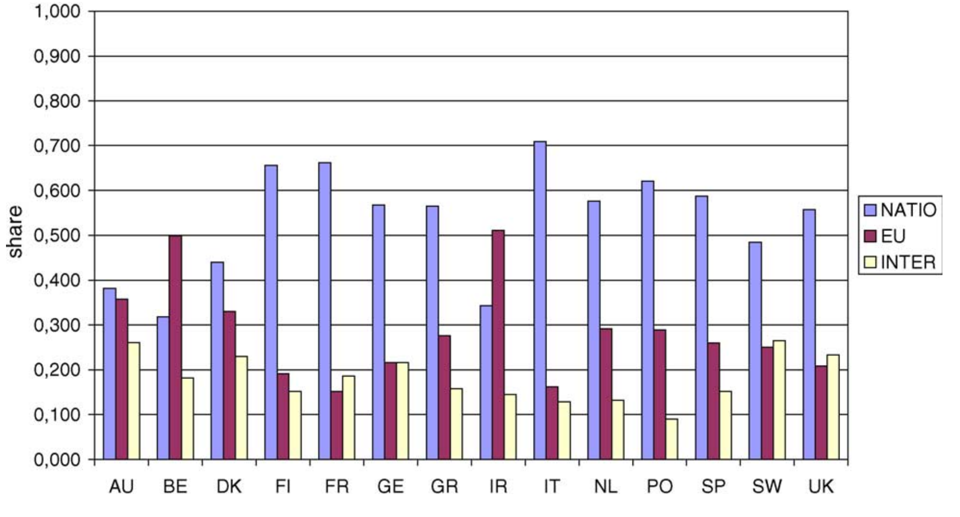

Fig. 2. Geography of collaborations for all members of the European Union (all years).

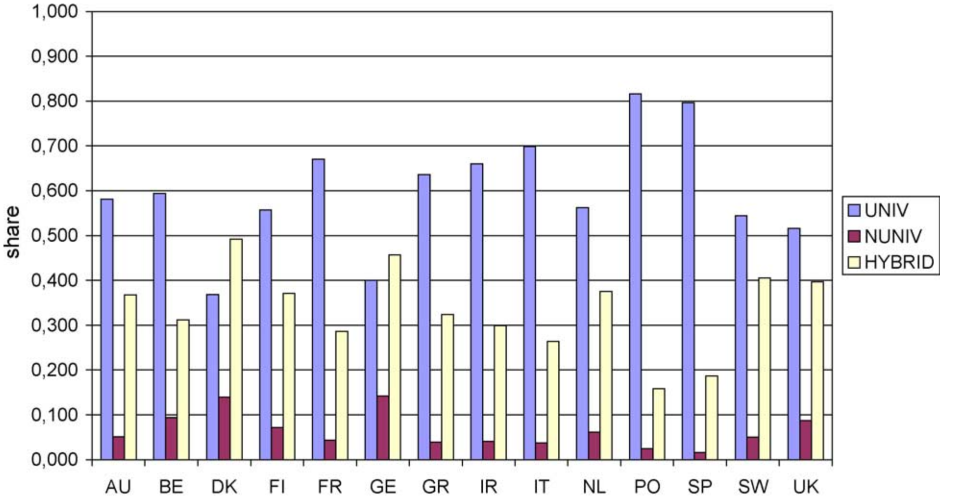

Fig. 3. Collaboration types for all members of the European Union (all years).

The inclusion of country fixed effects using country dummies is motivated by the idea that, despite internationalisation, scientific systems are still primarily nationally organised (Banchoff, 2002). This can be understood from the reluctance of national governments and national science councils to transfer political power and funds to supranational levels. Indeed, the large majority of funds spent on scientific and industrial research are still allocated nationally or via multi-lateral agreements (e.g., CERN, ESA). The European Commission still plays only a modest role (Banchoff, 2002). Given the national specificity of science policy, one can reasonably hypothesise national systems in Europe differ in terms of citation impact. Partly, this can be related to the difference in organisational structures and the patterns of collaboration, for which we control in the regression analysis below.[5]

To characterise the different countries in the European Union, we look at the geographical scale of countries' collaborations (Fig. 2) and the institutional types of countries' collaborations (Fig. 3). In both cases, we see a North–South divide with Latin countries being differently organised than Nordic countries. Latin countries are mainly nationally oriented in their collaboration (Portugal, Spain, France and Italy), while Nordic countries are more internationally oriented. Finland is an outlier as a Nordic country with a strong national orientation. Concerning the institutional nature of collaboration, we observe, again, Portugal, Spain, France and Italy with a similar pattern. In these countries, the share of collaborations within academia is highest and little interaction exists with institutes outside academia. Germany, Denmark, UK, Sweden and The Netherlands have

[5] In this respect, it is worth mentioning that Europe differs to an important extent from the United States and Japan. In a recent study, Verbeek et al. (2003) analysed citation patterns in patents to science for biotechnology. Their results suggest that Europe is (still) lagging behind the United States in translating science and technology into market applications and subsequent revenue generation.

the largest involvement of non-academic institutes, which network relatively often with one another and with academia.

Summarising, collaboration in European biotechnology is an increasingly important trend in biotechnology. EU countries differ in the distribution of collaboration in geographical scales and institutional scales. Two country clubs are visible with research in Latin countries being primarily organised at the national levels and within academia, and the Nordic countries being more internationalised and having a greater involvement of research institutes outside academia (see also Frenken and Van Oort, 2004).

4. Analysis

4.1. The model

The dependent variable is the number of citations of a paper, which can take on only nonnegative integer values. Therefore, we model the citations as events in a count process. The most common method to specify a count process is the Poisson model (see Winkelmann, 2000, for a survey of count data models). The basic form of the Poisson process assumes that the arrival rate of the events is a time independent constant. A citation process is said to follow the Poisson distribution, if (i) the probability of a citation in a short time interval is proportional to the length of the interval, (ii) the probability of more than one occurrence of the citation in a short interval is negligible, (iii) that the citations are independent of one another and (iv) that a certain mean number of citations is for the paper and for the other papers which share the same property. In the case of citations, these assumptions are likely to be met.

The Poisson regression model, however, implies a strong restriction. In this model, it is assumed that the conditional mean and variance of the event given the covariates are equal. Many count data sets do not satisfy this assumption. In the case of overdispersion, when the variance exceeds the mean, this can cause unduly small standard errors of the estimated parameters as the model underestimates the amount of dispersion in the outcome. The negative binomial regression model addresses this failure of the Poisson regression model by adding a quadratic term to the variance representing overdispersion. The basic idea is that a random effect represents unobserved heterogeneity and that if the random effect would be observable the data would be Poisson. We use the NEGBIN II model (Cameron and Trivedi, 1986), which has the variance-mean ratio linear in the mean. In this specification is assumed that the random effect has a gamma distribution with parameters $a = b = 1/\alpha$ where α is the variance of the unobservable effect. α is called overdispersion parameter. The larger α, the greater the overdispersion. For $\alpha = 0$, the negative binomial model converges into the Poisson model. The estimates of the parameters are derived by maximizing the respective log-likelihood functions.

A number of tests were performed for specification purposes. First, we examined whether the data would be subject to multicollinearity. We used the method of highest condition index following Belsley et al. (1980). The condition indices are reported for each specification in Table 2 in the row headed condition index. The highest condition index is always below the critical value of 30 (Belsley et al., 1980). Therefore, we can exclude that

Table 2
Regression results

Specification	1	1′	2	3	4	5
Author	0.0759** [0.0068]	0.0907** [0.0268]	0.0765** [0.0069]	0.0768** [0.0069]	0.0767** [0.0069]	0.0937** [0.0069]
Author^2		−0.0020 [0.0036]				
Address	0.0400** [0.0120]	0.0344+ [0.0186]	0.0782** [0.0223]	0.0483* [0.0205]	0.0645** [0.0230]	0.0857** [0.0218]
Address^2		0.0009 [0.0022]				
Top	0.1620** [0.0193]	0.1615** [0.0194]	0.1674** [0.0194]	0.1645** [0.0195]	0.1691** [0.0195]	0.1111** [0.0206]
Univ			−0.1205** [0.0353]		−0.1605** [0.0478]	−0.1357** [0.0471]
Nuniv			−0.0206 [0.0571]		−0.0499 [0.0607]	−0.1259* [0.0602]
Hybrid			0.0158 [0.0347]		−0.0061 [0.0431]	−0.0735+ [0.0427]
Nation				−0.0585+ [0.0332]	0.0131 [0.0458]	0.0055 [0.0456]
EU				0.0290 [0.0433]	0.0933+ [0.0496]	0.1863** [0.0523]
Inter				0.0217 [0.0375]	0.1009* [0.0480]	0.0550 [0.0475]
AUS						−0.1990** [0.0580]
BEL						−0.2297** [0.0521]
DEN						0.0593 [0.0564]
FIN						0.0130 [0.0592]
FRA						−0.4396** [0.0286]
GER						0.0907** [0.0278]
GRE						−0.4790** [0.0786]
IRE						−0.1060 [0.0774]
ITA						−0.3142** [0.0391]

Table 2 (*Continued*)

Specification	1	1′	2	3	4	5
LUX						−22.4708 [21,874.9230]
NET						0.0995** [0.0378]
POR						−0.3315** [0.0579]
SPA						−0.3629** [0.0343]
SWE						0.1529** [0.0427]
y1989	0.0930 [0.0620]	0.0924 [0.0620]	0.0972 [0.0620]	0.0919 [0.0620]	0.0964 [0.0620]	0.0590 [0.0612]
y1990	0.0929 [0.0608]	0.0928 [0.0608]	0.0927 [0.0607]	0.0916 [0.0608]	0.0925 [0.0607]	0.0837 [0.0599]
y1991	0.0469 [0.0571]	0.0468 [0.0571]	0.0493 [0.0570]	0.0457 [0.0571]	0.0480 [0.0570]	0.0276 [0.0562]
y1992	0.0651 [0.0571]	0.0649 [0.0570]	0.0665 [0.0570]	0.0636 [0.0570]	0.0643 [0.0570]	0.0555 [0.0562]
y1993	−0.0648 [0.0569]	−0.0645 [0.0569]	−0.0624 [0.0568]	−0.0660 [0.0569]	−0.0631 [0.0568]	−0.0606 [0.0561]
y1994	0.0710 [0.0556]	0.0719 [0.0556]	0.0756 [0.0556]	0.0687 [0.0556]	0.0740 [0.0556]	0.0615 [0.0549]
y1995	−0.1168* [0.0557]	−0.1162* [0.0557]	−0.1109* [0.0557]	−0.1181* [0.0557]	−0.1138* [0.0557]	−0.1280* [0.0550]
y1996	−0.2333** [0.0550]	−0.2329** [0.0550]	−0.2287** [0.0550]	−0.2363** [0.0550]	−0.2323** [0.0550]	−0.2484** [0.0543]
y1997	−0.4727** [0.0576]	−0.4726** [0.0576]	−0.4575** [0.0578]	−0.4732** [0.0579]	−0.4655** [0.0579]	−0.4889** [0.0571]
y1998	−0.6321** [0.0568]	−0.6321** [0.0568]	−0.6224** [0.0568]	−0.6359** [0.0568]	−0.6269** [0.0568]	−0.6397** [0.0561]
y1999	−0.5991** [0.0548]	−0.5973** [0.0548]	−0.5955** [0.0547]	−0.6041** [0.0548]	−0.6018** [0.0548]	−0.6303** [0.0541]
y2000	−1.0044** [0.0553]	−1.0038** [0.0553]	−1.0000** [0.0553]	−1.0058** [0.0553]	−1.0041** [0.0553]	−1.0145** [0.0547]
y2001	−1.8104** [0.0549]	−1.8088** [0.0550]	−1.8048** [0.0549]	−1.8147** [0.0550]	−1.8105** [0.0550]	−1.8330** [0.0544]
y2002	−2.9214** [0.0613]	−2.9195** [0.0613]	−2.9183** [0.0613]	−2.9250** [0.0613]	−2.9251** [0.0613]	−2.9469** [0.0609]
Constant	2.1509** [0.0489]	2.1335** [0.0643]	2.1132** [0.0521]	2.1446** [0.0515]	2.1275** [0.0524]	2.1900** [0.0525]
α	1.43**	1.43**	1.43**	1.43**	1.38**	1.38**

Table 2 (*Continued*)

Specification	1	1′	2	3	4	5
LR-test α	167873.80	167658.89	167791.41	167560.90	161278.40	161169.61
Observations	19769	19769	19769	19769	19769	19769
Condition Index	6.31	7.60	7.48	8.81	8.92	8.09
Log-likelihood	−58666.78	−58657.11	−58663.02	−58652.11	−58361.02	−58354.57
Log-likelihood (intercept only)	−61372.36	−61372.36	−61372.36	−61372.36	−61372.36	−61372.36

Notes: Results from negative binomial regression; dependent variable is citations, independent variables see text; the reference for country dummies is the UK, for year dummies the year 1988. α is the overdispersion parameter. Standard errors in brackets.

+ Significant at the 10% level.

* Significant at the 5% level.

** Significant at the 1% level.

our results are due to a dummy variable trap. Second, we checked the appropriateness of the NEGBIN II specification against the Poisson specification by using two tests. First, we tested the significance of the overdispersion parameter α by means of a likelihood ratio test. This is necessary as the usual Gaussian test of $\alpha = 0$ occurs on the boundary. In such cases, the limiting distribution of the maximum-likelihood estimate of the parameter is bounded from below and the asymptotic normality of the estimate does not hold under the null hypothesis that the Poisson model is appropriate. However, it can be shown that the likelihood ratio statistic has a distribution with probability mass of 0.5 at 0 and a chi-square with 1 d.f. for positive values (see Winkelmann, 2000, p. 105). This modified likelihood-ratio test is valid on the boundary, and is reported in Table 2, beneath the rows with heading LR-test α. α reports the estimated value of the overdispersion parameter. In order to check this result, we used also a regression-based approach devised by Cameron and Trivedi (1986). This test confirmed the results of the LR-test reported in the regression table.

4.2. Results

We run six specifications with an increasing number of independent variables. In all specifications, we included year dummies in order to control for the age of a paper, as older papers simply have more chance of being cited, other things being equal, and whether a top institute participated. Specification 1, then, explains the number of citations by two additional variables being the number of authors and the number of organizations. These variables can be considered core variables in our analysis testing hypothesis 1 and hypothesis 2. Model 1' specifies the same model, but includes the square term of author and address to test for an inverted-U shape as there may be an optimal size of collaboration. The square terms were not significant in model 1' neither when added to models 2–5. Therefore, we report specifications 2–5 without the squared terms for author and address. Specification 2 is specification 1 extended by the institutional type of collaboration, while specification 3 adds the geographical type of collaboration to specification 1. Specification 4 combines specifications 2 and 3, while the final specification 5 also includes country fixed effects.

In all specifications, the number of authors (author) and the number of organizations (address) contributed positively and significantly to the number of citations a paper received. Hypothesis 1 and hypothesis 2 can, therefore, not be rejected. Hypothesis 3, which stated that collaboration within academia (univ) increases a paper's citation impact compared to collaboration involving non-academic organizations, is rejected in all specifications. In fact, the results show a significant negative effect of academic collaborations on the citation impact of a paper. Also note that in the final specification in specification 5, the other types of institutional collaborations (nuniv and hybrid) also contribute negatively to citation impact. Finally, hypothesis 4, which stated that international collaboration increases a paper's citation impact compared to national collaboration, can largely be confirmed. In particular, the European level of collaboration (eu) contributes to citation impact in all specifications. International collaborations (inter) are significant in specification 4, but no longer in specification 5 when country dummies are included. This latter result reflects the fact that the most successful countries are also most internationalised, which explains that when country dummies are included, the variable for international collaboration is no longer significant.

Specification 5 further suggests that there are significant differences between countries in their citation impact. The coefficients for the country dummies compare to the number of citations from articles from UK, which has been used as a benchmark. This can be interpreted not only as a statistical benchmark, but also carries empirical meaning as the UK is the largest producer of articles in general in the EU (Batty, 2003), including in the field of biotechnology and applied microbiology (see Table 1). The results on country dummies show that small countries (den, fin, ire and lux) do not display significant effects, which probably reflects the relative low number of observations.[6] The larger countries in biotechnology research all show significant country-fixed effects. The results also point to a geographical divide. Nordic countries including Sweden, The Netherlands and Germany do better than the UK (in this order). By contrast, Latin countries including France, Greece, Italy, Portugal and Spain all receive much less citations than the benchmark case of UK. This result confirms the profile of European biotechnology sketched earlier as consisting of two country clubs.[7]

4.3. Limitations

Though our estimations provide a number of robust results, it should be borne in mind that our research design has some limitations. First, the use of citations

[6] The extreme case is Luxembourg with only two observations. This also explains the extreme results for the Luxembourg dummy.

[7] In this context, it is worth noting that the language barrier does not provide an explanation for this difference. Most journals considered in our database are indeed English, but this should not taken to mean that research organizations from Latin countries should have less impact, even though one would expect them to send relatively fewer articles to these journals. One could even argue that these organizations typically send only the best papers to the widely read English journals, and this should receive relatively more citations, while sending the other papers to journals in other languages (given a more specialised readership).

as a measure of impact of research collaboration only addresses one of all potential benefits from collaboration. As discussed in Section 2, benefits from collaboration are diverse ranging from cost reduction to commercialisation of research output. Output indicators other than citation should be included in future research. A second limitation is the focus on a single discipline. By doing so, we fail to see what role collaboration plays in interdisciplinary research, which is known to be an important driver of research collaboration (Gibbons et al., 1994; Rosenkopf and Almeida, 2003). Finally, our results are valid only for biotechnology; the extent to which they are representative for other fields of scientific research is left to further research.

5. Summary and discussion

Our study started from the premise that research collaboration enhances the quality of research, and hereby, the extent knowledge is diffused as indicated by the citation impact of a paper. Different aspects of research collaboration are taken into account simultaneously to assess their relative importance. We distinguished between the number of contributing authors and the number of addresses as to differentiate between the effect of the collaboration between individuals and between organizations. We further distinguish different geographical and institutional scales of collaboration. We found evidence that the diffusion of scientific knowledge, as measured by citation rate, is positively dependent on the number of authors and organizations. Another important finding has been that the differences in citation impact relate to the geographical scale of collaboration with the European collaboration being most successful as well as to country fixed effects reflecting the national systems being still in place.

Our analysis has a number of theoretical and practical implications. The inclusion of collaboration and its characteristics significantly contributes to the explanation of differences in citation impact of scientific papers. Collaboration is thus to be considered a fundamental and integral part of scientific research, both with regard to its production and its diffusion. The theoretical understanding of citation impact of scientific publications, however, is still limited. In this context, much is to be learned from the (rapidly increasing) literature on patent citations, which has shown important advances in theoretical and methodological respects. However, as argued earlier, one should always bear in mind the differences between scientific publications and patents. Conclusions from patent studies are thus not expected to hold automatically for scientific publications as well.

Practically, our analysis serves as an example how to assess the (citation) impact of collaboration, for example, in a research evaluation exercise. As science policy, both at national levels and within the EU framework programmes, devotes much of its resources to collaborative projects, one is in need of systematic ex post evaluation of the differential impact of research collaborations. Our analysis suggests that the organisation of scientific knowledge production in biotechnology is organised differently in different European countries, and is organised along a North–South divide. The findings suggest that the European Union, though successful as a scale of collaboration, still harbours national differences in knowledge production. An important question becomes to what extent these differences can be related to national varieties in science policy and institutions.

Acknowledgements

We would like to thank Hans van Amsterdam, Frank van Oort, Tim Schwanen, Olav Sorenson, David Waguespack and two anonymous referees for useful suggestions and comments. The usual caveats apply.

Appendix A

A.1. Data source

Source: Web of Science, Institute of Scientific Information, www.isinet.com

Date: 24-05-03

Category Name: *Biotechnology and Applied Microbiology*

Category Description: The biotechnology and applied microbiology category includes resources on a number of subjects that relate to the exploitation of living organisms or their components. In CC/AB&ES, the emphasis is on applied biology, including industrial microbiology. Applications include industrial chemicals and enzymes, biosensors, bioelectronics, pesticide development, food, flavor and fragrance industry applications, waste treatment, and pollution bioremediation.

Journals: *Acta Biotechnologica*, *Applied Microbiology and Biotechnology*, *Biochemical Engineering Journal*, *Biodegradation*, *Biofutur*, *Biomass & Node Count*, *Bioenergy*, *Bioprocess and Biosystems Engineering*, *Bioresource Technology*, *Biosensors & Bioelectronics*, *Biotechnology Advances*, *Biotechnology and Applied Biochemistry*, *Biotechnology and Bioengineering*, *Biotechnology Letters*, *Biotechnology Progress*, *Canadian Journal of Microbiology*, *Critical Reviews in Biotechnology*, *Current Opinion in Microbiology*, *Cytotechnology*, *Enzyme and Microbial Technology*, *Extremophiles*, *Folia Microbiologica*, *Food Technology and Biotechnology*, *Journal of Bioscience and Bioengineering*, *Journal of Biotechnology*, *Journal of Chemical Technology and Biotechnology*, *Journal of Industrial Microbiology & Biotechnology*, *Letters in Applied Microbiology*, *Metabolic Engineering*, *Nature Biotechnology*, *Process Biochemistry*, *Seibutsu-Kogaku Kaishi*, *Trends in Biotechnology*, *World Journal of Microbiology & Biotechnology*, *Yeast*.

Appendix B

B.1. Description of variables

Citation	number of citations a paper received at the day of retrieval (24-05-2003)
Author	number of co-authors of a paper
Address	number of participating organizations as indicated by the number of addresses
Top (Y/N)	whether one or more top institutes participated (see Appendix C)
Nation (Y/N)	whether two or more organizations originate from the same country
EU (Y/N)	whether two or more organizations originate from different EU countries
Inter (Y/N)	whether a country from outside the EU is involved
Univ (Y/N)	whether two or more academic institutes contributed to the paper
Nuniv (Y/N)	whether two or more non- academic institutes contributed to the paper
Hybrid (Y/N)	whether at least one academic institute and at least one non- academic institute contributed to the paper
AUS (Y/N)	whether one or more organizations originate from Austria
BEL (Y/N)	whether one or more organizations originate from Belgium
DEN (Y/N)	whether one or more organizations originate from Denmark
FIN (Y/N)	whether one or more organizations originate from Finland
FRA (Y/N)	whether one or more organizations originate from France
GER (Y/N)	whether one or more organizations originate from Germany
GRE (Y/N)	whether one or more organizations originate from Greece
IRE (Y/N)	whether one or more organizations originate from Ireland
ITA (Y/N)	whether one or more organizations originate from Italy
LUX (Y/N)	whether one or more organizations originate from Luxembourg
NET (Y/N)	whether one or more organizations originate from The Netherlands
POR (Y/N)	whether one or more organizations originate from Portugal
SPA (Y/N)	whether one or more organizations originate from Spain
SWE (Y/N)	whether one or more organizations originate from Sweden
UK (Y/N)	whether one or more organizations originate from UK
y1988 (Y/N)	paper published in 1988
y1989 (Y/N)	paper published in 1989
y1990 (Y/N)	paper published in 1990
y1991 (Y/N)	paper published in 1991
y1992 (Y/N)	paper published in 1992
y1993 (Y/N)	paper published in 1993
y1994 (Y/N)	paper published in 1994
y1995 (Y/N)	paper published in 1995
y1996 (Y/N)	paper published in 1996
y1997 (Y/N)	paper published in 1997
y1998 (Y/N)	paper published in 1998
y1999 (Y/N)	paper published in 1999
y2000 (Y/N)	paper published in 2000
y2001 (Y/N)	paper published in 2001
y2002 (Y/N)	paper published in 2002

Appendix C

C.1. Definition of a top institute

Organizations producing more than ten articles during the period 1988–2002 in the field of biotechnology and applied microbiology.

List of top institutes so defined

Austria
Graz University of Technology
Vienna University

Belgium
Leuven University

Canada
Alberta Research Institute
Alberta University

Denmark
Carlsberg Lab
Technical University of Denmark, Lungby

Finland
Helsinki University
VTT, Espoo

France
INRA
Institute Luminy, Marseille
Pasteur Institute
Technological University Compeigne

Germany
GBF
KFA Julich
Max Planck Institute
Hannover University
Stuttgart University

Greece
Athens University

Ireland
National University Ireland, Dublin/ Cork

India
Cent. Food Technology Research Institute, Mysore
Indian Technological Institute, New Delhi

Italy
CNR
Bologna University
Genoa University
Milan University

Japan
Toyama Prefectual University

New Zealand
New Zealand Research Institute, Wellington

The Netherlands
DSM
Gist Brocades
Radboud University Nijmegen
Wageningen University and Research Institute
Groningen University
Leiden University
Technological University Delft
Unilever
Amsterdam University

Portugal
Instituto Superior Técnico, Lisbon
Nova Lisbon University

Spain
CSIC
Barcelona University

Sweden
Chalmers Technological University, Gothenburg
Royal Technological Institute, Stockholm
Lund University

Switzerland
Swiss Federal Technological Institute

Taiwan
Taiwan National University

United Kingdom
AFRC
Unilever
Birmingham University
London University
Manchester University

United States of America

Cornell University, Ithaca
MIT
Oregon State University
Maryland University
Massachusetts University
Wisconsin University

References

Adams, J., Black, G.C., Clemmons, J.R., Stephan, P.E., 2004. Scientific teams and institutional collaborations: evidence from US universities, 1981–1999, NBER Working Paper 10640.

Banchoff, T., 2002. Institutions, inertia and European Union research policy. Journal of Common Market Studies 40 (1), 1–21.

Barabasi, A.L., Jeong, H., Neda, Z., Ravasz, E., Schubert, A., Vicsek, T., 2002. Evolution of the social network of scientific collaborations. Physica A 311 (3–4), 590–614.

Batty, M., 2003. The geography of scientific citation. Environment and Planning 35 (5), 761–770.

Belsley, D.A., Kuh, E., Welsch, R.E., 1980. Regression Diagnostics. John Wiley and Sons, New York.

Breschi, S., Lissoni, F., 2002. Mobility and social networks: localised knowledge spillovers revisited. Paper presented to the Workshop ‘‘Clusters in high-technology: Aerospace biotechnology and software compared’’. UQAM, Montreal, Canada, November.

Cameron, A.C., Trivedi, P.K., 1986. Econometric models based on count data: comparison and application of some estimators and tests. Journal of Applied Econometrics 1, 29–53.

Cooke, P., Uranga, M.G., Etxebarria, G., 1997. Regional innovation systems: institutional and organisational dimensions. Research Policy 26 (4–5), 475–491.

Danell, R., Persson, O., 2003. Regional R&D activities and interactions in the Swedish Triple Helix. Scientometrics 58 (2), 205–218.

Dasgupta, P., David, P.A., 1994. Toward a new economics of science. Research Policy 23, 487–521.

Etzkowitz, H., Leydesdorff, L., 2000. The dynamics of innovation: from National Systems and ‘‘Mode 2’’ to a Triple Helix of university–industry–government relations. Research Policy 29, 109–123.

Feldman, M.P., 1999. The new economics of innovation, spillovers and agglomeration: a review of empirical studies. Economics of Innovation and New Technology 8, 5–25.

Fornahl, D., Brenner, T. (Eds.), 2003. Cooperation, Networks and Institutions in Regional Innovation Systems. Edward Elgar, Cheltenham & Northampton MA.

Frenken, K., 2002. A new indicator of European integration and an application to collaboration in scientific research. Economic Systems Research 14 (4), 345–361.

Frenken, K., Van Oort, F.G., 2004. The geography of research collaboration: theoretical considerations and stylised facts in biotechnology in Europe and the United States. In: Cooke, P., Piccaluga, A. (Eds.), Regional Development in the Knowledge Economy. Edward Elgar Publishing Ltd., Cheltenham.

Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scot, P., 1994. The New Production of Knowledge. Sage, London.

Jaffe, A.B., Trajtenberg, M., Henderson, R., 1993. Geographic localization of knowledge spillovers as evidenced by patent citations. Quarterly Journal of Economics 108 (3), 577–598.

Katz, J.S., 1994. Geographical proximity and scientific collaboration. Scientometrics 31, 31–43.

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

Liang, L.M., Zhu, L., 2002. Major factors affecting China’s inter-regional research collaboration: regional scientific productivity and geographical proximity. Scientometrics 55 (2), 287–316.

Lotka, A.J., 1926. The frequency distribution of scientific productivity. Journal of the Washington Academy of Science 16, 317–323.

Narin, F., Stevens, K., Whitlow, E.S., 1991. Scientific cooperation in Europe and the citation of multinationally authored papers. Scientometrics 21, 313–323.

Nelson, R.R. (Ed.), 1993. National Innovation Systems. A Comparative Analysis. Oxford University Press, Oxford & New York.

Newman, M.E.J., 2004. Coauthorship networks and patterns of scientific collaboration. Proceedings of the National Academy of Science of the USA 101 (Suppl. 1), 5200–5205.

Rosenberg, N., 1990. Why do firms do basic research (with their own money)? Research Policy 19, 165–174.

Rosenkopf, L., Almeida, P., 2003. Overcoming local search through alliances and mobility. Management Science 49 (6), 751–766.

Singh, J., 2004. Collaborative networks as determinants of knowledge diffusion patterns. Mimeo, Harvard Business School, May.

Sorenson, O., Fleming, L., in press. Science and the diffusion of knowledge, Research Policy.

Verbeek, A., Debackere, K., Luwel, M., 2003. Science cited in patents: a geographic "flow" analysis of bibliographic citation patterns in patents. Scientometrics 58 (2), 241–263.

Wagner, C.S., 2002. International linkages: is collaboration creating a new dynamic for knowledge creation in science? Manuscript. Amsterdam School of Communications Research, University of Amsterdam, The Netherlands.

Wagner-Doebler, R., 2001. Continuity and discontinuity of collaboration behaviour since 1800—from a bibliometric point of view. Scientometrics 52, 503–517.

Winkelmann, R., 2000. Econometric Analysis of Count Data, third ed. Springer, Berlin.