The geography of collaborative knowledge production: Entropy techniques and results for the European Union

Paper to be presented at 42nd EUROPEAN CONGRESS Regional Science Association, Dortmund, Germany, August 27-31, 2002

Koen Frenken

Department of Economic Geography
Faculty of Geographical Sciences
Utrecht University
P.O. Box 80115, 3508 TC, Utrecht, The Netherlands
phone: +3130 253 2466, fax: +3130 253 2037
e-mail: k.frenken@geog.uu.nl, homepage: http://econ.geog.uu.nl/frenken.html

Version 26 April 2002

Abstract

While the process of integration of national systems into supranational systems has been widely discussed, comprehensive indicators of such processes are largely lacking. The author proposes a new indicator of integration using the mutual information formula of frequency matrices of country-country interactions. The indicator measures the degree in which the observed frequency distribution of interactions differs from the distribution of random interaction (perfect integration). The indicator proposed here takes into account both intra-national and international interactions to control for country differences in probabilities of intra-national and international collaboration: in larger countries interaction is expected to take place relatively more often at the national level because there exist more opportunities to interact at the national level. Controlling for countries' size avoids the drawback of other indicators that typically show excessively high integration values for smaller countries.

The integration indicator is applied to data from the *Science Citation Index* on inter-institutional collaborations in scientific output of the fifteen EU countries to analyse the integration process of European science during the period 1993-2000. Evidence is found that the European science system has indeed become more integrated. Further analysis shows that the higher level of integration has resulted exclusively from a more evenly distributed pattern of European collaborations, while the strong bias towards intra-national collaborations persisted. The results also show that larger countries are typically better integrated than smaller countries, which suggests that larger countries benefit from network externalities that trigger collaboration from smaller countries. The use of indicators of integration for both future academic research and science policy is elaborated.

Key-words: globalisation, European integration, information theory, mutual information, scientometrics, research collaboration, science policy, network externalities
JEL-code: O38, O52, R11, R12, R15

1. Introduction[1]

The goal of this study is to develop a comprehensive indicator of integration among countries within a supranational system. The main novelty incorporated in the indicator proposed here holds that an integration process among countries is not analysed in terms of the *growth* in interactions among countries, but in terms of the *matrix distribution of relative frequencies* of interactions among countries. The degree of integration of a supranational system can then be indicated in terms of interaction biases among participating countries as measured by the difference between the observed distribution of interactions and the hypothetical distribution of random interactions.[2]

In this study, the indicator of integration is applied to yearly data on collaborations among European research institutions as listed in the *Science Citation Index* during the period 1993-2000. Collaborations are counted by publications containing multiple institutional addresses, and each co-occurrence of two institutional addresses counts as an interaction. In this way, a matrix can be constructed of intra-national and international collaborations. The application of the indicator provides one with a comprehensive measurement of integration of the European science system and the development hereof over time. Other social and economic systems can be analysed in the same way. We will elaborate on other applications in the final discussion section.

2. International collaboration in science

International collaboration in research is expected to generate important benefits in many ways. The rationales for collaboration can be divided in economic benefits and intellectual benefits. Collaboration provides economic opportunities to realise economies of scale, for example, with regard to costs of training and research infrastructures (Katz and Martin, 1997). European examples of such large research infrastructures that have emerged through intergovernmental collaboration, are the European Space Agency (ESA) and European Centre for Nuclear Research (CERN). Collaboration is also expected to generate intellectual benefits from the cross-fertilisation of ideas that previously were unconnected and from a better quality control through internal refereeing. More generally, collaboration is intellectually

[1] I thank Bas van Waveren and two referees for helpful comments. All errors are my own.
[2] The indicator will be used here to analyse the integration process among countries into a supranational system, but the indicator can equally be applied to lower levels (*e.g.*, integration of cities in a regional system, integration of regions in a national system). In this context, the indicator may be of relevance in urban and regional studies.

required when specialised knowledge and skills are distributed among different persons (Gibbons et al., 1994; Ziman, 1994). And given the increasing level of specialisation within disciplines and sub-disciplines resulting from an increasing division-of-labour among scientists, research would benefit from international rather than national recruitment of scientists to participate in particular projects.

The growing internationalisation in scientific research may appear, at first sight, contradictory to the recent literature on the geography of innovation. This literature tends to emphasise the increasing localised nature of knowledge production at sub-national scale (Arthur, 1990; Boschma and Lambooy, 1999; Caniëls, 2000; Feldman, 1999; Van Oort, 2002). A number of theoretical arguments support this thesis. Most importantly, localisation economies of various sorts arise when people and firms engaged in knowledge production are geographically concentrated. These externalities range from labour market pooling, informal networking and knowledge spillovers. In particular, tacit components of knowledge production typically develop and diffuse through close interaction with suppliers and clients. Furthermore, tacit knowledge is often reproduced in spin off firms that typically locate in the region of the parent company. Another important reason for localised innovation concerns the difficulties of governance of collaborative industrial R&D. As the modalities of collaboration are hard to encode in contracts, collaboration often relies on informal contacts, reciprocity and trust between partners, which is facilitated when participants share local ties and a similar institutional environment. If one accepts that the economy has become rapidly more knowledge-intensive, economic activity can be expected to have become more localised in recent times.

From these theoretical rationales for localised industrial innovation, however, one can not conclude that one should expect scientific research to develop into a more localised activity as well. Scientific research is qualitatively different from industrial innovation. Though in some disciplines the distinction between science and innovation has become less relevant such as in biotechnology and informatics, scientific knowledge production generally differs from industrial knowledge production in a number of ways. First, the tacit component is expected to be much smaller in scientific knowledge production, which renders communication and collaboration at a distance much easier. 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 determined by the local context, but emerge from a global *discours*. 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 than industrial innovation.

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). For example, a U.S. patent citation study has 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. Concluding, both theory and evidence suggests that scientific knowledge production differs from industrial knowledge production in that the latter may primarily be accumulating at a regional scale, while the latter is expected to internationalise over time.

The recent rise in international collaborations in scientific research can relatively easily be indicated by computing the share of international co-authorships in all publications. It has been estimated that the share of international collaborations has doubled during the period 1987-1997 to account for 15 percent of world publications (Wagner, 2002). To assess the benefits of international collaboration is a somewhat more difficult exercise. Empirical studies that addressed the benefits of international collaboration have focused on scientific impact and productivity (Katz and Martin, 1997). The impact of scientific output resulting from collaboration as measured by citation rates is substantially higher than average. The difference in citation impact is even higher for international collaboration. Furthermore, it has been found that the productivity of scientists is positively dependent on the frequency of collaboration. Collaboration tends to increase the level of personal productivity as measured by the number of publications produced per year.

A large part of European science policy can be considered as an attempt to capitalise on the potential of scientific collaboration among member states. Not surprisingly, research collaboration and mobility of researchers is at the core of its policies. Given the evidence on the positive effects of research collaboration, and the policy importance attached to it, an important question holds whether empirical data show that the European science system has indeed become more integrated or not. Within the European context, the number of European collaborations has undeniably increased over the last few decades. However, the number of collaborations is in itself no indication of integration (Leydesdorff, 1992). For example, the number of collaborations can double in a period, but at the same time the distribution of collaborations may fragment in a number of islands of collaborating countries. The question is whether the increase in collaborations, as a general phenomenon, also contributed to a more integrated pattern of collaborations.

The hypothesis holds that European science indeed evolved towards a more integrated system. However, in an empirical research design, the hypothesis of increasing integration requires further specification. What does it mean that a set of countries becomes more integrated? Below, I address this hypothesis using data on both national and European collaborations as indicated by publications with multiple addresses. European integration, then, can then be analysed by comparing the propensities of countries to collaborate nationally with the propensity to collaborate with other European member states. It is crucial to distinguish intra-national from international collaboration, because, other things equal, larger countries are expected to collaborate relatively more often nationally than internationally, simply because there exist more potential national partners in larger countries than in smaller countries. Controlling for differences in size of countries leads us to specify two hypothesis:

1. European integration increases over time as indicated by a declining bias to collaborate nationally controlling for differences in size of countries.

2. European integration increases over time as indicated by a convergence in the bias of each pair of countries to collaborate controlling for differences in size of each of the countries.

3. A new measure of integration

An inter-institutional collaboration is here defined as a pair of different institutional addresses occurring in a publication record contained in the *Science Citation Index*. Note that this definition is not restricted to co-authorship as one and the same person can be associated with more than one institution.[3]

The number of inter-institutional collaboration between two European member states i ($i=1,..,15$) and j ($j=1,..,15$) as a share of the total number of collaborations is denoted as q_{ij}, which results in a 15x15 matrix of 225 q_{ij}-values. National collaborations are present on the

[3] More on this, see Katz and Martin (1997: 11-13).

diagonal for which holds $i=j$, while all other cells refer to country-country collaborations for which holds $i \neq j$. A co-occurrence of two addresses in different countries is attributed to both cells that refer to a pair of countries so we get a symmetric matrix ($q_{ij} = q_{ji}$).[4] The share of each country in the total number of collaborations is then given by:

$$q_{i.} = \sum_{j=1}^{15} q_{ij} \tag{1}$$

and, because of symmetry in the matrix, $q_{.j}$ is equal to $q_{i.}$ for $i=j$. In other words, to derive the marginal totals for each country one can either sum over the rows or over the columns.

3.1 Mutual information

The degree of integration of country i with respect to country j is measured here as the difference between the observed share of collaborations q_{ij} and what would be expected from the product of the individual shares $q_{i.}$ and $q_{.j}$. The difference between the observed share and the expected share is measured by the natural logarithm of the division of q_{ij} by the products of $q_{i.}$ and $q_{.j}$:

$$T_{ij} = \ln \frac{q_{ij}}{q_{i.} \cdot q_{.j}} \tag{2}$$

The T_{ij}–value is a measure o*f bias*. The value is positive when country i is collaborating with country j more than what is expected from the product of the individual country shares in all output. The T_{ij}–measure takes on a negative sign when country i is collaborating with country j less than what was expected from their shares. Put another way, a positive value indicates a positive bias in the propensity of country i to collaborate with country j and *vice versa* while a negative value indicates a negative bias in the propensity of country i to collaborate with country j and *vice versa*.

The use of a logarithm renders this measure symmetric regarding to whether a country collaborates x times more than expected or x times less than expected with another country. For example, when two countries collaborates two times more than expected, the T_{ij}–measure

[4] Consequently, a co-occurrence of two addresses in the same country is counted twice. The complete procedure is also illustrated in the example in Table 1.

equals *ln 2 = 0.693* and when two countries collaborates two times less than expected, the T_{ij} –measure equals $ln\ ½ = –0.693$.

The degree of integration of the network of fifteen member states as a whole is measured by T, which is the sum of the values for T_{ij} weighted for the share in the total number of collaborations q_{ij}. In information theory, the measure *T* is known as the “mutual information” value, which measures dependence in a frequency matrix (Frenken, 2000, 2001; Langton, 1990; Leydesdorff, 1991; Theil, 1967, 1972):

$$T = \sum_{i=1}^{15} \sum_{j=1}^{15} q_{ij} \cdot \ln \frac{q_{ij}}{q_{i.} \cdot q_{.j}} \qquad (3)^{5}$$

It has been shown that mutual information is non-negative for any frequency distribution (Theil, 1972). When all pairs of countries would collaborate exactly as much as expected from their individual shares, we have $q_{ij} = q_{i.} \cdot q_{.j}$. In this case, all pair wise bias values T_{ij} equal zero, and the T-value consequently equals zero too (total independence). In the context of research collaboration, a zero T-value indicates perfect integration of all fifteen member states within the European science system. When any bias exists in the propensity to collaborate, mutual information will be positive. The higher the T-value, the less countries are integrated in a system (higher dependence).

Theil (1967, chapter 9) initially used the mutual information measure to characterise the amount of information contained in input-output tables. In this application, the values of q_{ij} stand for the inter-industry flows as fractions of the aggregate output. Total independence of the matrix (T=0) would mean that the input-output table would not contain any information at all, since the inter-industry flows q_{ij} can readily be derived from the product of the marginal totals $q_{i.}$ and $q_{.j}$. Any other input-output table would yield a positive mutual information. The higher the value of the mutual information, the more structure is present in the input-output-table, and the higher its information content.[6] [7]

[5] For x=0 we have $x \cdot \ln x$ 0. In information theory, one usually uses base two logarithm instead of the natural logarithm to express the value of mutual information in bits. When the natural logarithm is used, as in this study, one speaks of “nits” (Theil, 1972).

[6] Theil (1967) also showed why the mutual information decreases when sectors in an input-output table are aggregated. In this context, he showed that minimisation of input heterogeneity of sectors that are aggregated minimises the loss of information due to aggregation. A similar aggregation procedure, though not followed below, could be applied to the matrices of research collaborations.

[7] More recent applications of mutual information in social sciences can be divided in two groups: applications to empirical data and application to simulation data. Empirical applications include the dependence between different donors and different recipients of grants in the United States (Theil, 1972), the dependence between journals as reflected in matrices of journal-journal citation matrices

In the context of research collaboration addressed here, it is important to note that the indicator takes into account both intra-national ($i=j$) and international ($i \neq j$) interactions. In this way, the degree of integration is *adjusted for differences in size of countries* as measured by the number of collaborations in which a country participates, as a fraction of the total number of collaborations. Collaboration patterns are assessed by means of comparing the observed frequency of collaboration (q_{ij}) to what is expected from the individual shares of countries ($q_{i.} \cdot q_{.j}$). What follows is that a large country should be expected to collaborate more intensively at the national level than a small country, because there are more researchers available in larger countries to interact with at the national level. In this, the indicator proposed above differs from more other measures that indicate internationalisation either by looking at international collaboration only (Katz, 2000) or by taking the ratio between national and international activity (Kearney, 2001). The latter types of indicators typically show high integration values for smaller countries compared to larger countries as these indicators do not control for the size of countries.

3.2 Analysing subsets

As explained above, the integration measure is a weighted sum of all intranational and international T_{ij}–values weighted for their share in the population. For the European Union, there are $15^2 = 225$ T_{ij}–values. By summing non-overlapping subsets of the 225 T_{ij}–values, and dividing the sum by the share of the subset in the population, one can focus on the degree of integration of a subset of the matrix.

In the case of the European Union, one can think of two ways of splitting the matrix into subsets. First, one can compare the T_{ij}–values for national ($i=j$) and international ($i \neq j$) collaborations to analyse to what extent integration is due to intra-national biases versus international biases. We get, respectively:

$$T_{i=j} = \frac{1}{\sum_{i=1}^{15} \sum_{j=1}^{15} q_{ij}} \cdot \left(\sum_{i=1}^{15} \sum_{j=1}^{15} q_{ij} \cdot \ln \frac{q_{ij}}{q_{i.} \cdot q_{.j}} \right) \qquad (i = j) \qquad (4)$$

(Leydesdorff, 1991), and the dependence of countries on technologies and markets (Frenken, 2000). The application of mutual information to simulation data concern the measurement of dependency and

$$T_{i\neq j} = \frac{1}{\sum_{i=1}^{15}\sum_{j=1}^{15} q_{ij}} \cdot \left(\sum_{i=1}^{15}\sum_{j=1}^{15} q_{ij} \cdot \ln \frac{q_{ij}}{q_{i.} \cdot q_{.j}} \right) \qquad (i \neq j) \tag{5}$$

A second way to split the matrix into subsets is to sum the subset of 15 T_{ij}–values belonging to a country *i*. In this way, one obtains the level of integration for each individual country labelled as T_i. We get:

$$T_i = \frac{1}{q_{i.}} \cdot \sum_{j=1}^{15} q_{ij} \cdot \ln \frac{q_{ij}}{q_{i.} \cdot q_{.j}} \tag{6}$$

for each of the fifteen member states (*i=1,…,15*). Table 1 provides a numerical example of the application of all integration measures described above, using imaginary data on collaboration patterns of three countries.

TABLE 1 AROUND HERE

Note that the application of this measure is by no means restricted to the analysis of country-country collaborations. The indicator can be applied to the level of regions in a country (or in the European union) and to the level of cities in a region (or in a country or in the European union). Similarly, the indicator can be applied to supranational blocks within the world system.

4. Results

Data were collected from the *Science Citation Index* for the period 1993-2000 covering the the large majority of publications of natural and life sciences. I first selected for each year all

relations between the states of cells in cellular automata (Langton, 1990) and the characterisation of multi-dimensional NK fitness landscapes in terms of the distribution of local optima (Frenken, 2001).

records containing at least one address located in an EU member state.[8] I further reduced the size of the dataset by excluding records containing the most common inter-institutional collaborations with the major countries outside the EU.[9] The resulting number of records amounts to over 200.000 on average each year, with the number increasing from 183.020 in 1993 to 230.561 in 2000.

To sample the number of inter-institutional collaborations within and between European member states, I used only the first three listings of addresses. Each first and second address, each first and third address, and each second and third address were counted as one collaboration. Thus a single-address record yields no collaboration, a double-address record yields at most one collaboration between a pair European countries, and a record containing three (or more) addresses yields at most three collaborations between pairs of European countries.[10]

4.1 Collaboration in the European Union as a whole

Fig. 1 shows the values of mutual information *T* as computed from formula (3) for each year during the period 1993-2000. As explained above, a lower *T*–value indicates lower levels of biases in the choice of partners, and thus a higher level of integration between the EU member states. The trend of *T*–values indicates a gradual integration process suggesting that EU member states indeed, on average, have become less biased with regard to the country of origin of their research collaboration partners.

FIGURE 1 AROUND HERE

The integration process, however, is very slow as the integration indicator decreased only from 1.526 to 1.461. Put another way, the level of integration in 2000 is 95.7 percent of the level in 1993. This seemingly slow process of European integration, however, should not be

[8] Member states are Austria (AU), Belgium (BE), Denmark (DE), Finland (FI), France (FR), Germany (GE), Greece (GR), Ireland (IR), Italy (IT), Luxembourg (LU), The Netherlands (NE), Portugal (PO), Spain (SO), Sweden (SW), and United Kingdom (UK). Note that the United Kingdom refers to records in the *Science Citation Index* containing addresses from England, Northern Ireland, Scotland, or Wales.
[9] Being collaborations between an EU country and either Canada, Japan, Russia, Switzerland, or United States.
[10] The use of only the first three addresses provides, apart from computational advantages, a way to circumvent the fact that in some disciplines collaboration is much more common than in others.

judged from the reference level of full integration (no bias at all as indicated by T=0), but from a (unknown) reference level of bias that would occur when institutional and language barriers between countries were to be fully removed.[11] This reference level of bias that is expected to remain in homogeneous geographical territories is expected to be substantial following from the theory of geography of innovation as discussed in the second section. Even if scientific knowledge does not share the characteristics of industrial innovation for what concerns its degree of specificity, tacitness, and appropriability, some degree of spatial concentration is expected to remain.

4.2 Intra-national versus international collaboration

The next question regarding the integration process at the EU level holds whether the integration process is an effect of a decreasing bias of countries to collaborate nationally or an effect of a decreasing bias with regard to the choice of EU partners, or a combination between the two. Fig. 2 plots the $T_{i=j}$-values of intranational collaborations within the EU countries for each year (formula 4). The results indicate that there is a stable positive bias to collaborate nationally. Over the years, the average value is about +2.04 with means that the probability of national collaboration is $e^{2.04}$ = 7.7 times higher than when partner selection would have been at random. There is no real trend in the national bias over the years. The results at least suggest that the bias to collaborate nationally has not decreased during the period 1993-2000.

FIGURE 2 AROUND HERE

Fig. 3 plots the $T_{i\ j}$-values of international collaborations in the EU for each year (formula 5). From the results two important observations can be made. First, the bias to collaborate with other EU countries has been negative over the whole period considered. Over the years, the average value is about –1.62 with means that the probability to collaborate with another member state is on average $e^{-1.62}$ = 0.20 times lower than when partner selection would have been at random. Second, the negative bias towards European collaboration has become less and less over the years. The integration process as indicated by the trend in Fig. 1 can thus be understood as the result of a decreasing bias in the selection of a European partner while the

[11] The application of the mutual information indicators to regions within a homogeneous territory such as the United States would give an indication of the degree of spatial concentration that is expected to remain. This analysis falls outside the scope of this study.

bias towards national collaboration persisted.[12] From this, we can conclude that only the second hypothesis regarding the convergence of inter-national biases is confirmed, while no evidence is found that the bias to collaborate nationally, has declined.

FIGURE 3 AROUND HERE

4.3 Country comparison

The integration values for each of the fifteen member states are plotted in Fig. 4 following formula (6). Differences between countries are quite pronounced. In particular, that the degree of integration is closely related to the size of a country. The three largest countries (UK, Germany, France) have the lowest T_i –values indicating the highest degrees of integration while smaller countries (Greece, Finland, Portugal, Ireland) have the highest T_i –values indicating the lowest degrees of integration.[13]

FIGURE 4 AROUND HERE

The strong correlation between the size of a country and its level of integration can be further analysed by plotting the yearly q_i –values with the corresponding yearly T_i –values of each country (Fig. 5). Note again here that 'size' does not refer to more conventional measures such as the number of researchers in a country, but to a country's share in collaborations (q_i) as it is directly derived from the marginal totals of the collaboration matrix. The shape of the function that best explains the scatter plot in Fig. 5 could is asymptotic rather than linear suggesting that scale effects are marginally decreasing. The propensity to collaborate internationally thus tends to rise with country size, but decreasingly so. Importantly, the

[12] Given the largely unchanged positive $T_{i=j}$-values and the trend of the negative $T_{i \neq j}$-values towards zero, the falling trend in T-values Fig. 1 must be understood as resulting from a rising share of international collaboration as a percentage of all collaborations. The share of international collaborations $\Sigma\Sigma q_{ij}$ (i≠j) has indeed risen fom 0.137 to 0.161, while the share of intranational collaboration $\Sigma\Sigma q_{ij}$ (i=j) has fallen from 0.863 to 0.839.

[13] The result on scale effects is empirically not conflicting with Katz's (2000) result that smaller countries tend to engage more often in international collaboration, because Katz (2000) made use of an indicator that did not relate the amount of international collaborations to the amount of national collaborations.

correlation between degree of integration of a country and its share in the research output is not perfect. For example, the share of Sweden is much higher than that of Belgium while the latter country is better integrated in the European system. Similarly, the share of Greece is much higher than the share of Portugal, but the latter is better integrated. The results can thus be used to benchmark individual countries in terms of their level of integration and what would be expected when their size is taken into account. In this respect, Belgium and Portugal score relatively well compared to countries of similar size.

FIGURE 5 AROUND HERE

The positive relationship between country size and degree of integration suggests that larger countries benefit from scale effects that trigger European collaboration. Though comprehensive theories on scale advantages on a national scale are lacking, one can expect that the degree of diversification in scientific research in a country is strongly related to its size. Specialised research institutes require a critical mass regarding investments in training programmes and research infrastructures. If one accepts that more specialised knowledge is available in the larger and supposedly more diversified countries, it will be generally more attractive for researchers from any other country to collaborate with researchers from larger countries.

A special kind of positive network externality present in larger countries concerns the number of people that are able to communicate in the national language. Clearly, people from smaller countries typically invest in learning languages that are widely spoken throughout the (academic) world. In this sense, language can be considered a network standard, the adoption of which is characterised by network externalities (Arthur, 1989). A related notion is language as a 'hypercollective' good: the more people that are able to communicate in a language, the higher the benefit for each single able to communicate in this language (De Swaan, 2001). Following this reasoning, the largest European countries (UK, Germany, France) will enjoy the largest network externalities in that their languages are more widely spoken within the European Union, in particular English that has become the global language in academia.[14]

[14] Note that in the *Science Citation Index* from which the empirical data have been extracted, includes journals in other European languages, too, though a bias exists towards the inclusion of journals in English.

4.4 Country-country comparison

The 225 individual T_{ij} –values for each pair of countries following formula (2) are given in Table 2. Each value in the table is the average over the eight yearly T_{ij} –values in the period 1993-2000. This static representation still gives a fairly good idea of the yearly T_{ij} –values since no single of the 225 time-series shows a consistent falling or rising trend over time. Put another way, though a clear integration pattern emerges from the collective of countries, the country-county dynamics tend to fluctuate over time.

In Table 2 the highest values for T_{ij} are indicated by bold values using an otherwise arbitrary threshold of –1.00. As expected, using this threshold the strongest collaboration is found for all intra-national collaborations ($i=j$) reflecting that all European countries strongly favour national over European collaboration. The scale effect can now also be observed in greater detail. Larger countries like France, Germany, Italy, and the UK have a smaller positive bias values to collaborate nationally (ranging from 1.46 to 1.84) while smaller countries like Greece, Ireland, Luxembourg, and Portugal, have higher positive bias values to collaborate nationally (ranging from 4.22 to 6.16).

There are a number of other country-county values that exceed the threshold of –1.00: Austria-Germany, France-Luxembourg, Belgium-Luxembourg, Finland-Sweden, Belgium-The Netherlands, Germany-Luxembourg, Belgium-France, Ireland-UK, Belgium-Portugal, Portugal-Spain, Denmark-Sweden, and Portugal-UK. These results indicate that relative high propensities to collaborate with another country are very much organised along geographical lines: high values are typically found for neighbouring states. Also note that in many cases the countries that collaborate relatively often share a common or similar language. However, from the data of the *Science Citation Index* it is not possible to analyse in detail the propensity of researchers to collaborate with researchers that speak the same language. From the information contained in publications records, the working language cannot be derived.

TABLE 2 AROUND HERE

5. Discussion

In this study, it has been proposed to understand integration as the degree in which interaction patterns among countries are biased. Integration has been measured by the difference between the observed frequency matrix of interactions and the matrix that would have resulted from random interactions. The main novelty of the approach holds that it takes into account both the interactions between countries and the interactions within each country. In this way, the measurement of integration among countries is adjusted for differences in the size of the national systems.

The indicator has been applied to data on multiple address publications in the EU to analyse the process of integration of the European science system. Using data on institutional addresses in records of the *Science Citation Index* for the period 1993-2000, results show that the process of European integration has indeed occurred. It is also found that the integration process has not been the result of a falling bias of countries to collaborate nationally, but solely the result of a falling bias in the choice of partner in European collaboration. Furthermore, the size of a country correlates with the degree of integration of a country, which indicates that larger countries have contributed most to the process of integration. The latter result has been related to scale advantages arising from diversification in large countries and network externalities stemming from language. These explanations are largely suggestive and merit further theoretical and empirical elaboration.

From a policy evaluation perspective, the results suggest that European science policy has led to a more evenly distribution of European partnerships, but has not led to a "substitution" of national for European partnerships. This is not to say that European science policy has not succeeded. On the contrary, for what concerns European collaboration the bias in partner selection has steadily decreased. European funding, including the equability conditions attached to it, can be expected to have a substantial effect on a more evenly collaboration. The fact that the bias to collaborate nationally has not decreased may well reflect the effect of national science policies aiming to increase national collaboration[15] rather than the ineffectiveness of European science policy as such. This is in line with the observation that from a budgetary point of view, European science policy has not been successful: until now, member states still account for 95 percent of expenditures on public civil R&D in the European Union (Banchoff, 2002).

[15] For example, the Dutch government has promoted the creation of national research schools in all disciplines during this period.

The results obtained in this study offer us a macroscopic picture of the integration process since the sample from the *Science Citation Index* includes all disciplines in natural and life sciences. The conclusion that the European union is integrating can by no means be generalised for all scientific disciplines in which research communities are predominantly organised. It may well be the case that the application of integration measures at the level of scientific disciplines would show disintegration for some disciplines.[16] To look for more detailed explanations of patterns of collaboration, future research could extend the analysis presented here by decomposing the collaboration matrix into the lower level of scientific disciplines. Methodologies based on journal-journal citation reports are readily available do delineate scientific disciplines using clustering techniques (Leydesdorff and Cozzens, 1993; Van den Besselaar and Leydesdorff, 1996). Alternatively, one can use exiting classifications that are available from ISI.[17] Having delineated scientific disciplines, one can apply the proposed indicators of integration in the same manner as we applied these to the science systems as a whole. One can then attempt to explain the level of integration of scientific disciplines as a dependent variable from independent variable that characterise disciplines (tacitness, specificity, appropriability, fixed costs). From earlier patent studies it has become clear that differences in spatial concentration of innovative activities in Europe are highly sector-specific (Breschi, 2000). Similarly, one can expects important differences to exist in the degree of European integration of different scientific disciplines.

The research agenda outlined above is also expected to contribute to (European) policy design and policy evaluation. Understanding the determinants of research collaboration from characteristics of scientific disciplines can be helpful in designing policies that promote collaboration and mobility at the level of particular disciplines. For example, the current emphasis of EU science policy on applied knowledge production in Framework Programmes and the lack of EU funding of basic science, has been criticised at various occasions (Banchoff, 2002; Pavitt, 2000). This criticism calls for empirical studies that test whether basic science benefits from supranational networks, and whether applied science more often emerges from local networks. If so, there may be reasons to re-adjust the orientation of collaboration and mobility programs of the European Union towards basic science.

[16] It is even theoretically possible that all disciplines are disintegrating in islands of collaborating clubs of countries, while the macroscopic system as a whole is integrating. If specific pairs of countries would specialise in specific scientific disciplines, but different pairs of countries would specialise in different scientific disciplines, the integration values of disciplines would show disintegration, while the aggregated science system could still show a macro process of integration.
[17] http://www.isinet.com

A final note concerns the application of the integration measure to other domains. In principle, the integration measure can be applied to any data that can be summarised in frequency matrices of interaction. Within the context of European integration, valuable information can be generated through labour market analyses using intra-national and European migration data and through commodity market using intra-national and European trade data. One can also think of studies that analyse the frequencies of intra-national and European collaborations, mergers and acquisitions among firms. Analyses of these kinds would provide us with important empirical and policy-relevant information on the level, structure, and dynamics of integration in the European Union. After all, lacking any real precedent, the process of European integration is still both theoretically and empirically poorly understood.

Example (for three countries)

Co-occurrences in the *Science Citation Index*:

FRANCE – FRANCE:	250	UK – FRANCE:	120	GERMANY – FRANCE:	140
FRANCE – UK:	80	UK – UK:	800	GERMANY – UK:	90
FRANCE – GERMANY:	160	UK – GERMANY:	110	GERMANY – GERMANY:	750

Collaboration matrix:[18]

	FRANCE	UK	GERMANY	SUM
FRANCE	250	100	150	500
UK	100	800	100	1000
GERMANY	150	100	750	1000
SUM	500	1000	1000	2500

Frequency matrix:

	FRANCE	UK	GERMANY	
FRANCE	$q_{11} = 0.10$	$q_{21} = 0.04$	$q_{31} = 0.06$	$(q_{.1} = 0.20)$
UK	$q_{12} = 0.04$	$q_{22} = 0.32$	$q_{32} = 0.04$	$(q_{.2} = 0.40)$
GERMANY	$q_{13} = 0.06$	$q_{23} = 0.04$	$q_{33} = 0.30$	$(q_{.3} = 0.40)$
	$(q_{1.} = 0.20)$	$(q_{2.} = 0.40)$	$(q_{3.} = 0.40)$	$(q_{..} = 1.00)$

$\mathbf{T_{ij}}$-values:

	FRANCE	UK	GERMANY
FRANCE	$T_{11}= \ln(0.10/0.04)= 0.92$	$T_{21}= \ln(0.04/0.08)= -0.69$	$T_{31}= \ln(0.06/0.08)= -0.29$
UK	$T_{12}= \ln(0.04/0.08)= -0.69$	$T_{22}= \ln(0.32/0.16)= 0.69$	$T_{32}= \ln(0.04/0.16)= -1.39$
GERMANY	$T_{13}= \ln(0.06/0.08)= -0.29$	$T_{23}= \ln(0.04/0.16)= -1.39$	$T_{33}= \ln(0.30/0.16)= 0.63$

Integration indicators :

$\mathbf{T} = (0.10 \cdot 0.92) + (0.04 \cdot -0.69) + (0.06 \cdot -0.29) + (0.04 \cdot -0.69) + (0.32 \cdot 0.69) + (0.04 \cdot -1.39) + (0.06 \cdot -0.29) + (0.04 \cdot -1.39) + (0.30 \cdot 0.63) = 0.30$

$\mathbf{T_{i=j}} = (1/0.72) \cdot ((0.10 \cdot 0.92) + (0.32 \cdot 0.69) + (0.30 \cdot 0.63)) = 0.70$

$\mathbf{T_{i\ \ j}} = (1/0.28) \cdot ((0.04 \cdot -0.69) + (0.06 \cdot -0.29) + (0.04 \cdot -0.69) + (0.04 \cdot -1.39) + (0.06 \cdot -0.29) + (0.04 \cdot -1.39)) = -0.72$

$\mathbf{T_1} = (1/0.2) \cdot ((0.10 \cdot 0.92) + (0.04 \cdot -0.69) + (0.06 \cdot -0.29)) = 0.23$
$\mathbf{T_2} = (1/0.4) \cdot ((0.04 \cdot -0.69) + (0.32 \cdot 0.69) + (0.04 \cdot -1.39)) = 0.35$
$\mathbf{T_3} = (1/0.4) \cdot ((0.06 \cdot -0.29) + (0.04 \cdot -1.39) + (0.30 \cdot 0.63)) = 0.29$

Table 1. Example of application of integration indicators

[18] An address that is listed before or after another address is treated in the same way. The share of collaborations between two different countries is therefore computed as half the mean to obtain $q_{ij}=q_{ji}$.

	AU	BE	DE	FI	FR	GE	GR	IR	IT	LU	NE	PO	SP	SW	UK
AU	**3.63**	-1.92	-2.14	-2.40	-2.20	**-0.34**	-2.24	-2.08	-2.12		-1.81		-2.19	-2.13	-2.27
BE	-1.92	**3.18**	-1.63	-2.54	-0.77	-1.40	-1.14	-1.24	-1.99	**2.15**	**-0.33**	-0.90	-1.45	-1.64	-1.58
DE	-2.14	-1.63	**3.40**	-1.28	-2.28	-1.35	-1.83	-1.56	-2.29		-1.53	-1.52	-2.00	**-0.16**	-1.48
FI	-2.40	-2.54	-1.28	**3.39**	-2.91	-2.04	-3.71	-2.30	-3.28		-2.30	-3.08	-2.55	**-0.47**	-2.34
FR	-2.20	-0.77	-2.28	-2.91	**1.60**	-1.75	-1.34	-2.00	-1.94	**0.04**	-2.24	-1.02	-1.41	-2.40	-2.06
GE	**-0.34**	-1.40	-1.35	-2.04	-1.75	**1.68**	-1.19	-1.64	-2.07	**0.06**	-1.47	-1.49	-1.86	-1.74	-1.77
GR	-2.24	-1.14	-1.83	-3.71	-1.34	-1.19	**4.22**		-2.10		-2.26	-1.70	-2.36	-2.17	-1.19
IR	-2.08	-1.24	-1.56	-2.30	-2.00	-1.64		**4.69**	-2.19		-1.43	-1.13	-1.81	-1.95	**-0.26**
IT	-2.12	-1.99	-2.29	-3.28	-1.94	-2.07	-2.10	-2.19	**1.84**	-1.66	-2.25	-1.93	-1.81	-2.45	-2.07
LU		**2.15**			**0.04**	**0.06**			-1.66	**6.16**					
NE	-1.81	**-0.33**	-1.53	-2.30	-2.24	-1.47	-2.26	-1.43	-2.25		**2.44**	-1.37	-2.19	-2.04	-1.68
PO		-0.90	-1.52	-3.08	-1.02	-1.49	-1.70	-1.13	-1.93		-1.37	**4.55**	**-0.32**	-1.69	-0.91
SP	-2.19	-1.45	-2.00	-2.55	-1.41	-1.86	-2.36	-1.81	-1.81		-2.19	**-0.32**	**2.57**	-2.36	-1.66
SW	-2.13	-1.64	**-0.16**	**-0.47**	-2.40	-1.74	-2.17	-1.95	-2.45		-2.04	-1.69	-2.36	**2.75**	-1.88
UK	-2.27	-1.58	-1.48	-2.34	-2.06	-1.77	-1.19	**-0.26**	-2.07		-1.68	-0.91	-1.66	-1.88	**1.46**

Table 2. Country-country T_{ij}-values averaged over the period 1993-2000 (empty cells refer to pairs of countries that have not collaborated in each year)

Fig. 1. T-values indicating the level of integration of all EU countries for both intra-national and international collaborations

Fig. 2. $T_{i=j}$-values indicating the level of integration of all EU countries for intra-national collaborations only

Fig. 3. $T_{i\ j}$-values indicating the level of integration of all EU countries for international collaborations only

Fig. 4. T_i-values for EU countries for intra-national and international collaborations

Fig. 5. T_i-values for EU countries plotted against their share in output q_i

Bibliography

Arthur, W. B., 1989. Competing technologies, increasing returns, and lock-in by historical events. Economic Journal, 99, 116-131.

Arthur, W. B., 1990. Silicon Valley locational clusters - when do increasing returns imply monopoly. Mathematical Social Sciences, 19(3), 235-251.

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

Boschma, R. A. and Lambooy, J. G., 1999. Evolutionary economics and economic geography. Journal of Evolutionary Economics, 9(4), 411-429.

Breschi, S., 2000. The geography of innovation: A cross-sector analysis. Regional Studies, 34(3), 213-229.

Caniëls, M.C.J., 2000, Knowledge spillovers and economic growth : regional growth differentials across Europe. Edward Elgar, Cheltenham.

De Swaan, A., 2001, Words of the World. Polity Press, Oxford.

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.

Frenken, K., 2000. A complexity approach to innovation networks. The case of the aircraft industry (1909-1997). Research Policy, 29(2), 257-272.

Frenken, K., 2001, Understanding Product Innovation using Complex Systems Theory. University of Amsterdam and University of Grenoble, PhD Thesis.

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

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

Katz, J.S., 2000. Scale-independent indicators and research evaluation. Science and Public Policy, 27, 23-36.

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

Kearney, Inc., 2001. Measuring globalization. Foreign Policy, January-February, 56-65.

Langton, C. G., 1990. Computation at the edge of chaos. Phase-transitions and emergent computation. Physica D, 42(1-3), 12-37.

Leydesdorff, L., 1991. The static and dynamic analysis of network data using information theory. Social Networks, 13(4), 301-345.

Leydesdorff, L., 1992. The impact of EC science policies on the transnational publication system. Technology Analysis & Strategic Management, 4, 279-298.

Leydesdorff, L. and Cozzens, S. E., 1993. The delineation of specialities in terms of journals using the dynamic journal set of the SCI. Scientometrics, 26(1), 135-156.

Pavitt, K., 2000. Why European Union funding of academic research should be increased: a radical proposal. Science and Public Policy, 27(6), 455-460.

Theil, H., 1967, Economics and Information Theory. North-Holland, Amsterdam.

Theil, H., 1972, Statistical Decomposition Analysis. North-Holland, Amsterdam.

Van den Besselaar, P. and Leydesdorff, L., 1996. Mapping change in scientific specialties: A scientometric reconstruction of the development of artificial intelligence. Journal of the American Society for Information Science, 47(6), 415-436.

Van Oort, F., 2002, Agglomeration, Economic Growth and Innovation. Tinbergen Institute Research Series, Rotterdam.

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.

Ziman, J., 1994, Prometheus Bound. Science in a Dynamic Steady State. Cambridge University Press, Cambridge.