

Hennemann, Stefan; Rybski, Diego; Liefner, Ingo

## Conference Paper

# The Myth of Global Science

50th Congress of the European Regional Science Association: "Sustainable Regional Growth and Development in the Creative Knowledge Economy", 19-23 August 2010, Jönköping, Sweden

### **Provided in Cooperation with:**

European Regional Science Association (ERSA)

*Suggested Citation:* Hennemann, Stefan; Rybski, Diego; Liefner, Ingo (2010) : The Myth of Global Science, 50th Congress of the European Regional Science Association: "Sustainable Regional Growth and Development in the Creative Knowledge Economy", 19-23 August 2010, Jönköping, Sweden, European Regional Science Association (ERSA), Louvain-la-Neuve

This Version is available at:

<https://hdl.handle.net/10419/119076>

### **Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

### **Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

## **The Myth of Global Science**

S. Hennemann<sup>a</sup>, D. Rybski<sup>b</sup> and I. Liefner<sup>a</sup>

<sup>a</sup> Justus-Liebig-University of Giessen, Dept. of Economic Geography, Senckenbergstrasse 1, D-35390 Giessen, Germany

<sup>b</sup> Potsdam Institute for Climate Impact Research (PIK), P.O. Box 60 12 03, D-14412 Potsdam, Germany

Corresponding author:

S. Hennemann

stefan.hennemann@geogr.uni-giessen.de

Phone +49(0)641-99-36222

Fax +49(0)641-99-36229

### **Abstract**

Scientific collaboration, in most cases, is seen as a joint action on a global scale that involves researchers from not just one region or one country but instead forming an international network of researchers. This type of epistemic communities builds up especially in the case of analytical modes of knowledge production. Rationales for a global science system are needs for complementary resources in an increasingly specialized world. Further, information and communication technologies contributed to flatten the world for intense, yet spatially distant collaboration. Based on the large scale analysis of the production of scientific publications in six distinct technologies/scientific fields from 2004 to 2008 these notions have to be reconsidered. The collaboration probability is a power-law function of the distance between the participating authors: The higher the spatial proximity, the higher the chance of a jointly written paper, i.e. collaboration.

Two main effects can be isolated: an intra-country effect with sharply decreasing probability with distance, and an inter-country effect that shows features of a random distributed distance function. The former effect out-competes the latter by a factor of around 50-100, depending on the distance. These empirical results are contradicting the belief of global science collaboration. And there are good arguments that the observed collaboration network structure may offer efficient global knowledge creation.

### **Keywords:**

knowledge creation, global science networks, bibliometric network analysis, geography, epistemic communities, research collaboration

## 1. Introduction

The dyadic play of proximity and collaboration has been discussed for decades in geography, economics, and innovation research. With the advent of improving communication technologies as well as innovations in transportation including respective cost reductions, globalization of all kinds of socio-economic activity has accelerated significantly (Dicken 2007). Inevitably, there is a tendency towards a shrinking world. Some authors argue that the world flattens more and more, making local peculiarities less relevant. This debate of the "death of the distance" or "death of the nation state" (cp. Cairncross 1997, O'Brian 1992, Toffler 1980) triggered opposing notions that put the spatial dimension back into the center of interest (cp. Morgan 2004, Rodríguez-Pose and Crescenzi 2008). With respect to science and innovation, some authors completely neglect the idea of shrinking worlds through eased communication and pressures of competition (Goldenberg and Levy 2009), whereas others highlight increasingly globalized science arenas (Power and Malmberg 2008).

The socio-mathematical reasoning for the co-existence of strong local forces and global reach was given by complex networks scholars who analyzed social interaction and found evidence for "small worlds", where the average distance between all members in a social community is short, although the average number of direct connections is rather low.

What does this have to do with global science collaboration? The link is the way in which social and physical proximity is connected in specific configurations like a community of friends. Usually, information and communication technologies (ICT) do not ease the finding of new communication partners, but, once a contact is established, these technologies ease the communication between the formerly established counterparts. Hence, an amplification of the exchange between socially already connected actors may be expected (Goldenberg and Levy 2009). Most friends are made within short physical distances, so it is straight-forward to expect a network based on social communication to have a strong local component. However, professional environments are sometimes different to private ones and there is evidence that there is a reinforcement of global structures especially for science related collaboration (Wagner and Leydesdorff 2005). This situation is often present at the frontiers of science, where new knowledge is created in joint effort in international teams of excellence (Power and Malmberg 2008).

In this article, "science collaboration" is understood as joint effort and will be defined as a mutual conscious and unconscious information exchange between the participants with the purpose to create new knowledge. The rest of this article is organized as follows: Section 1

provides the theoretical embedding of the topic. Section 3 critically evaluates the used indicators, data and explains the methods. Section 4 gives detailed information about the calculation. This is followed by the presentation (section 5) and the discussion (section 6) of the results. Section 7 provides concluding remarks and suggestions for further research.

## 2. Theoretical embedding

### 2.1. Reasoning behind globalization in science

Research, and basic research in particular, has been one of the earliest internationalizing activities for centuries. Scientists were much more cosmopolitan than members of other professions. Basic scientific activity is often regarded as enhancing the world outlook in general, hence there is an inherent global component in “the sciences“. However, today, not only this global and common aim lead to strong internationalization, but also economic, competitive and political forces are influencing the rules of the game. Some of these measures seem to stretch the reach, some seem to tighten the activity reach.

#### 2.1.1. Economic explanation

A usual way to deal with socio-economic activity and space is to analyze the transactional dimension of the process of interest. The total transaction cost of the activity as well as the necessity of proximity for the transaction determines the spatial dimension (i.e. the distance) of the transaction (cp. Storper 1997, p. 39). For scientific collaboration the equation can be seen as simple as for business transactions, which is the original context of Storper’s approach. If there is need to have frequent contacts with your collaborators and costs to transcend distance are high, than the chance of having a local cooperation is high. What if, there is no reasonable local collaborator available? Then the collaborator would be the nearest available one which balances the outcome. However, in highly specialized environments, the collaboration is, first and foremost, defined by content. In addition, information transparency might be low, outside an established research community. The high degree of specialization in knowledge creating organizations and the competition-driven need for highly specialized complementary assets decreases the ability to find local partners.

#### 2.1.2. Competitive forces and spatial division of labor

More and more scientific endeavors go beyond one single discipline and require experts from more than one scientific field. Enabling technologies are sometimes the result of a single research team, but the usage in subsequent projects requires far more than simply mixing

these ideas with those from the own discipline. As a result, interaction is simply necessary because of the increasing complexity of (boundary crossing) technologies. A single research team or even a single researcher is no longer capable of expanding the frontiers of science, i.e. to find something unique (Subramanyam 1983). This implicates a higher specialization to master the full process and cut it down into smaller fractions. In effect, not only business activity is shaped by an increasing division of labor, but also scientific activity is more and more characterized by vertical and horizontal division of labor to be able to quickly expand the frontiers of science. This development efficiently collapses the distinction between science and products (Slaughter and Leslie 1997), adding an additional blurring of boundaries between public and private research.

Recent years have seen a fierce competition for research grants, publications, citations. This leads to increasing pressures on the organization level with increasing global competition. Research teams are forced to combine efforts to work ahead of other teams. To be the center of excellence in a science field does frequently mean that industry as well as national governments or supra-national initiatives are more selective. Scientific experiments (e. g. in natural sciences, medical sciences) are becoming increasingly expensive to be conducted by only one researcher or one organization. The research costs for a blockbuster drug may easily exceed 1 billion US-dollar until it reaches the market (cp. Berenson 2005).

The division of labor can be largely expected in almost all variants of knowledge creation. Within a discipline, there may be a tendency that the more experimental and the less theory-oriented the research is, the higher the vertical specialization level of the whole process (e. g. group leader, experimenters, laboratory assistants). Vertically specialized research processes may enjoy advantages from strong group organization, which is often present in formalized research teams. Theory-driven research within the same discipline, in contrast, can also be group independent and spatially less confined. In cross-disciplinary types of research, horizontal and vertical specialization processes may have different characteristics with respect to division of labor and the consequences for the spatial organization. The increase in competition as well as stronger cross-disciplinary content of the research is contributing to a spatial division of research with internationalization processes. Wagner (2002) reports that about 20 to 30 % of all co-authored scientific papers involve researchers from more than one country. In research collaborations between organizations in industrialized and developing countries, Hwang (2008) found a strong division of labor based on complementary assets. A prior argument of the inverse relation between the size of the science-base of an economy and the necessity to seek complementing resources abroad (cp. Davidson Frame and Carpenter

1979) may not hold today, as the international collaboration with scientists from the US is most frequent (Wagner 2002).

In summary, the general observation of higher vertical and horizontal specialization in science (e. g. due to strategic positioning also of national science systems) may contribute to a higher internationalization of scientific collaboration activity.

### 2.1.3. Research Policy

Scientific knowledge creation is a socio-economically desired process to improve the pool of available technologies, techniques and methods that eventually lead to innovation and create economic progress, i. e. welfare. Policy is highly motivated to integrate research teams to reduce costs, to pool ideas from heterogeneous research environments. One source of motivation is to be found in research policy itself, another one stems from equalization policy ambitions. While the former can be found in almost all countries, the latter is unique to supranational organizations like the European Union. Policies to promote science and technology (S&T) can potentially lead into opposing directions in terms of collaboration distance. Geopolitical macro-changes in the last decades have fueled an increasingly international science arena. Today, up to 20 % of national S&T budgets are attributed to international science activities (Wagner 2002). At the same time national funds are trying to out-compete research organizations from other countries to preserve or improve the competitive advantage and prestige of their own organizations. Hence, the way research grants are announced, may affect the spatial distribution of the collaborators. Some may lead to more (mandatory) international collaboration, some may have strong localizing effects. It is far beyond the scope of this article to reflect on the impacts of available funding initiatives. However, some examples should be given to understand different effects on spatial collaboration. Especially the European Union (EU) pays very much attention on collaboration and on international joint projects at the same time. Two third of the total budget of the framework budget is devoted to cooperation and the mobility of researchers. Further, the EU is always concerned with the reduction of barriers to international collaboration. However, the collaboration pattern is not based on "scholarly ground" (Hoekman et al. 2009) but is still bounded within national barriers, although most measures favor international research proposals (Wagner and Leydesdorff 2005).

National research policies seem to interfere with supra-national measures, at least in case of the EU. The intrinsic motivation of international collaboration in the scientific arena is opposing national policies that remain within the countries' borders (Wagner 2002). More and

more countries, including Germany, are pursuing the establishment of clusters of excellence in recent years to increase the competitive state of their science-base. This is mirrored with regional initiatives in many cases. Whereas the former demand for strong international embedding, the latter usually seek to complement regional imbalances with local collaboration (e. g. LOEWE in Hesse).

		Research Complexity	
		low	high
Territoriality	low	<p>little specialized, well-defined problems, informal collaboration, absence of locally specific assets.</p> <p><i>ubiquitous local collaboration</i></p>	<p>complex research, joint and formal, specialized and high degree of cross-disciplinary knowledge.</p> <p><i>global networked science</i></p>
	high	<p>niche research in well-structured science fields.</p> <p><i>sparse global spots of local collaboration</i></p>	<p>formal research collaboration, complex and little structured problems with highly specialized knowledge pools, which are organized centrally.</p> <p><i>globally linked spots</i></p>
		synthetic	analytical
		Mode of Knowledge Production	

Table 1: Classification scheme for research typologies and the potential collaboration structure; Source: based on Wagner (2005).

Examples from the US National Science Foundation (NSF) show that international research collaboration is also encouraged by funding opportunities through a large pool of counterpart science funding agencies (cp. NSF Office of International Science & Engineering OISE).

Today, international collaborations with US research organizations are steadily increasing and most common for researchers world-wide (Wagner 2002).

Examples from China show the implementation of measures to support international joint projects with EU partners in the context of the 973 Program<sup>1</sup> and by the Ministry of Science and Technology. In effect, the number of international joint projects involving foreign and Chinese counterparts has risen by more than 50% since 2000 according to the Chinese

<sup>1</sup> <http://www.973.gov.cn/English/AreaCoop.aspx>

Statistical Yearbook on Science and Technology 2008. Similar policy initiatives were integrated into the S&T policy strategies in most developed and many developing countries.

## 2.2. Science typology and differences in spatial organization

The motivation to collaborate internationally may vary not only due to external factors, but also with the underlying technology and can be understood as science-specific culture. "As a result, the expectation is that the extent of growth in international linkages will differ across subfields" (Wagner 2005). Wagner, thus, suggests distinguishing four principal categories: 1) data driven scientific fields (e. g. biomedical, virology), 2) resource driven fields (e. g. seismology, zoology), 3) equipment driven fields (e. g. polymers, manufacturing), and 4) theory driven sciences (e. g. mathematics, economics). The subject culture has, therefore, effects on the principle mode of knowledge production and the way of collaboration.

According to Moodysson et al. (2008) one type of knowledge creation is analytical, i.e. based on natural laws. In contrast to synthetic modes of knowledge creation, analytical modes aim to create new knowledge, independent of the immediate use for business, i.e. companies. However, the authors expect both modes to be relevant for most technologies and sciences at the same time, but varying with the underlying disciplines as well as with the stage of the knowledge production process.

Well-defined research problems that can be tackled efficiently in a non-centralized and less specialized way are usually organized in rather informal local research collaborations (e. g. short-term consulting style projects) (ubiquitous local collaboration). In contrast, complex and less structured research problems are best solved within and across highly specialized knowledge pools that may be more likely organized centrally with a global reach. This type of globally linked spots is often characterized by high funding budgets. A third general type of research problem (global networked science) can be captured with highly complex and completely unstructured problems. In terms of spatial collaboration distribution this type shows tendencies to assembly formal research collaborations worldwide (Raymond 1997). A fourth type of niche research can be categorized by high local specialization in well-structured sciences (e. g. specialized engineering departments with relation to adjacent niche industries). The global collaboration, even in rather tacit knowledge exchanges, was recently explained by the concept of cognitive proximity (cp. Boschma 2005, Nooteboom et al. 2007). Communities of practice can induce cognitive proximity even beyond spatial and organizational boundaries and may substitute for spatial proximity (Gertler 2003, p. 86). Usually, these communities are seen as formalized through institutions (e. g. scientific associations). "Epistemic



communities” can thereby be described as more loosely organized networks, less framed by formal structures than communities of practice (i.e., transcending the boundaries of organizations and inter-organizational alliances). Typical examples of epistemic communities are international networks of academic researchers who specialize in similar or related fields” (Moodysson 2008). For the establishment of these communities, homophily is a strong explaining factor, i.e. groups with a common socio-cultural background tend to attach to each other in social networks (McPherson et al. 2001). According to Wagner such communities are beneficial for the research outcome. Thus, distance and crossing boundaries, no matter whether spatial or cognitive, are more likely to provide complementary assets (e. g. ideas, data, equipment), resulting in more innovative outcomes of the joint research compared to regionally or intellectually locked-in collaborations. Therefore, the global level is frequently seen as "providing increasingly a system of reference other than national systems" (Wagner and Leydesdorff 2005).

In a situation of low territoriality and high research complexity, there is the highest chance to find global science networks. In this case the social embeddedness in a ‘community of practice’ / ‘epistemic community’ is able to substitute for spatial proximity in highly tacit and interactive knowledge creation processes (cp. Moodysson 2008). All other things being equal, researchers can find partners all over the world, as long as they are member of that community.

The crucial question then is, do we have barriers to the worldwide knowledge sourcing (e. g. transaction costs in terms of travel expenses, policy induced barriers) or not? The answer is certainly dependent on the type of research field. Considering a military technology, political barriers for international joint research might be restricted to allies. In a non-sensitive field there may be less restricted or even internationally encouraged collaboration. Considering a world-wide epidemic, there is a higher chance that there is geopolitical consensus to jointly solve the complex problem (e. g. vaccine research for pandemic diseases). Consumer electronics related technologies may also be relatively unrestricted.

In summary, the underlying field of science should arrange for different collaboration distance patterns.

### 2.3. Global knowledge networks and flat worlds in science - mixed empirical evidence

Rational economic behavior, competition and the spatial division of labor tend to shape a global science system that comes closer to the theoretic ideal of optimal sourcing of collaboration partners. However, social network aspects as well as policy initiatives seem to

either support this direction or to oppose it. Empirical evidence can be found for both extremes as well as in between.

"The production of scientific endeavors is in many respects one of the most globalized of human activities. [...] Most well-developed scientific fields are strongly globalized in terms of publication outlets" (Power and Malmberg 2008). The number of cross-country co-authored papers increased dramatically, showing that "the processes whereby excellence in research is being developed are more often than not 'extra-regional' in scope" (Power and Malmberg 2008). Wagner (2005) points into the same direction. She found an increase in international collaboration activity for a diversity of fields, including polymers, physics, but also theory-driven fields like mathematics. With the advancement of countries, Wagner and Leydesdorff (Wagner and Leydesdorff 2005), moreover, argue that they are enabled to compete for collaborators from other advanced and also in more peripheral countries. Ergo, this inevitable progression seems to lead automatically towards a greater internationalization of science. But there were also mixed patterns to be found. For example, Almeida et al. (2009) report strong neighboring effects in research patterns in Europe with similar overall research profiles of bordering countries. The authors found that the similarity of the research patterns affect the propensity for the organizations to collaborate. However, these effects were classified small and proximity is treated as no more than an underlying factor. Similarly, Moodysson et al. (2008) found that scientific collaboration in the life-sciences tend to be non-local. Rather, it focuses on global centers of knowledge production and their immediate neighborhood in the very first stage of idea generation. In this stage, there is a strong focus on interaction within research groups, because it is hard to articulate and define the actual problem. To work together with external partners is, therefore, hardly possible, due to a highly unstructured type of tacit knowledge fragments.

Later, collaboration distance will increase (Cooke 2009). When the problem definition is clearer and interfaces can be identified, the collaboration strongly relies on communities of practice. "Actually, the interpersonal networks tend to be globally distributed [...] among academic research groups" (Moodysson 2008). Interestingly enough, local concentrations of research units do not correlate with the global collaboration pattern that has been observed by the author. So, those, who found global-local patterns, usually argue with time dependent spatial effects of research activities. As this stage is likely to be the most relevant for scientific publishing, the collaboration pattern should emphasize the global, rather than the local scale. And this is even more so in epistemic communities that are less formally organized than communities of practice.

## 2.4. Research framework and hypotheses

With the theoretical discussion and the mixed empirical results presented in the preceding sections, the following hypotheses can be derived. They address and account for the expected variations related to the type of research and the underlying scientific field.

- H1a In state-of-the-art scientific knowledge creation processes there is high probability of social embeddedness, which involves territorially unbound research communities. As a result, collaboration occurs independent of the distance.
- H1b State-of-the-art science and research are highly internationalized activities. Hence, national borders are of ancillary importance for the collaboration propensity.
- H2 The probability of collaboration in dependence to distance varies with the investigated technology. In scientific fields where low territoriality and high research complexity coincide, research cooperation is organized on the global scale.

## 3. Material and methods

### 3.1. Organizing and selecting the data

Six datasets for distinctive techniques and research fields have been collected for testing the hypotheses. The raw data was downloaded from ISI Web of Knowledge databases in August 2009. The query to match one of the given techniques was restricted to the years 2004 to 2008. Using the typology from section 2.2, the scientific fields can be exemplarily characterized as follows in terms of organization, space, and content:

- Bluetooth: The wireless communication standard has been defined by a company-driven research consortium. The underlying technology is based on electrical engineering and has little cross-disciplinary and enabling character. This field is rather engineering oriented (synthetic mode of knowledge production) and therefore subject to lower global research integration.
- Image Compression: Software algorithms to reduce the amount of required storage capacity are rather cross-disciplinary in use, ranging from medicine-related imaging procedures to cineastic usage. As a theory-driven field its potential global integration is high.
- Heart Valves: The related technologies combine animal research, human medicine, technical aspects and may be subject to influence by stakeholders (e. g. companies, public health systems and policies, ethic commissions). These features make this field set in between synthetic and analytical modes of knowledge production with the effect of both global and local potential for scientific collaboration.

- H5N1: This is a rather spontaneously emerging research field where ad hoc solutions were needed to prevent from pandemic spreading of the virus. The location origin was Asia, but effects are highly relevant world-wide. Interaction between vaccine producing companies, government agencies and global research hot-spots joint forces to solve this issue. Although being an analytical type of biotechnology related knowledge, this field has high potential for global collaboration due to the external pressures.
- Tissue Engineering: Generally, a medicine-related field, but also interdisciplinarily linked to material research and biology. The general issue is condensed in ethical debates, however, the field is of high relevance in most developed as well as developing countries, where political influence regularly provide positive location factors. As more analytical mode of knowledge production, this field can be expected to global reach in collaboration, which may be hindered again by political influence.
- Carbon Nanotubes: Highly cross-disciplinary and enabling for other science and technology fields. Heterogeneous in terms of place and content. This field is equipment-intensive and, therefore, a mix of synthetic and analytic knowledge content with the consequence of potential of showing global and local tendencies at the same time.

The fields have been selected to represent various modes of knowledge production. They can be assumed to be sufficiently distinct with respect to the expected spatial organization. A summary of the used data is given in table 2.

Table 2: Description of the article raw data

Technology, method, research field	Preponderant mode of knowledge production following typology in Table 1	Articles		No. of individual organizations	Search string <sup>a</sup>
		Raw no.	Included into analysis		
bluetooth	synthetic	2,171	274	455	TS=(bluetooth)
image compression	synthetic	2,399	403	614	TS=(image compression algorithm)
heart-valve	synthetic/analytic	1,483	603	1,034	TS=(heart-valve)
H5N1	analytic	1,787	934	1,271	TS=(h5n1)
tissue-engineering	synthetic/analytic	8,821	4,240	2,721	TS=(tissue-engineering)
carbon nanotubes	analytic	29,076	12,656	4,483	TS=(nanotub* SAME carbon)

<sup>a</sup> Databases=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH Timespan=2004-2008

### 3.2. Collection and pre-processing of the data

To allow for misleading effects that may arise from unsuited data representation, the raw data was edited based on several assumptions/restrictions:

- Articles with up to 10 contributors have been included. With more than 10 people an intensive knowledge sharing is questionable. Sometimes the contributors even do not know each other (cp. Laudel 2002).
- The production of new knowledge is a strong social construct as is learning. The publication outcome is the result of a negotiation between the collaborators and a leveling between the respective knowledge pools.
- The lowest level of aggregation is the organizational level rather than the individual author. Therefore, publications that have been co-authored only by individuals from one organization are not taken into account. This can be interpreted as optimal intra-organizational knowledge-sharing.

In using organization aggregated data, all problems arising from team-strategic listing (e. g. supervisor has to be on the list, when a PhD student is writing) are leveled out, because the organization is in the list anyway and only counted once.

Algorithm for building the empirical network:

1. Select all publications with between 2 and 10 individual co-authors AND at least two different organizations from the pool of all publications in the given scientific field in the given period of time.
2. Take the information on the organization of a selected paper and completely connect all contributing organizations with each other.
3. Do this with all selected papers and add a weight attribute to the edge in case of multiple publications between two organizations.
4. Calculate the physical distance of each edge with Eq. 1.

### 3.3. Geo-coding

The location information of all contributors was hand-cleaned and automatically geocoded, using the GOOGLE-Maps API<sup>2</sup>.

Instead of looking up the organizations' full address, only the city names have been used. This has been proven to be much less fault-prone, but still sufficient in terms of location accuracy for distances above approx. 20 km. Accordingly, results below this distance cannot be interpreted, because the distance between two organizations from the same city is set to zero. Incomplete or erroneous locations were identified and manually recoded.

---

<sup>2</sup>Information about the GClientGeocoder class can be found at <http://code.google.com/intl/en-EN/apis/maps/documentation/reference.html#GClientGeocoder>

The latitude and longitude values have been used to calculate geodesic distances by using

$$d_{a,b} = \arccos [\sin(c * lng_a * \sin(c * lng_b + \cos(c * lng_a) * \cos(c * lng_b)) * \cos(c * lat_b - c * lat_a)] * r_{earth}, \text{ where } c=2\pi/360.$$

Moreover, lng and lat are the longitude and latitude of the sites a and b.

#### 4. Calculation

The data was organized in the form of an undirected network containing nodes that represent the actors (e.g. research organizations, public authorities, universities and companies) and edges, representing each single act of collaboration (i.e. co-authoring a paper). Since it is possible to have more than one publication between any two organizations, the network contains multiple edges. Such a network was build for each of the six selected research fields and serves as empirical model. Further, the spatial distance was calculated for each connection between any two nodes, based on the latitude and the longitude value for the node.

##### 4.1. Theoretical Null-model

The uncorrelated "null"-model was build adopting a randomizing approach, where the degree distribution of the network is preserved (the degree of a node is the number of edges it has to other ones), but any other feature is destroyed by randomizing (cp. Maslov and Sneppen 2002). Rewiring algorithm for randomization preserving the degree distribution (Null-model):

1. Take the empirical network as a starting point.
2. Randomly pick two edges,  $na-nb$  and  $ma-mb$ , and swap the connections:  $na-ma$  and  $nb-mb$ .
3. Repeat step 2 at least as often as the number of edges the network consists of (we use 5 times the number of edges).

##### 4.2. Calculating probabilities for the empirical and the random network

We consider the conditional relative frequency that nodes have an edge at euclidean distance  $d$ . Therefore, we perform the following steps:

- 1a Calculate the distances  $d_{i,j}$  between the two nodes of all edges (with at least one common publication).
- b Logarithmically bin the distances:  $b_{i,j} = \text{int}(\ln d_{i,j})$ .
- c Count the number of edges in each bin,  $c_1(d)$ .
- 2a Calculate the distances between all pair of nodes.
- b Logarithmically bin these distances.
- c Count the number of distances in each bin,  $ca(d)$
- 3 The relative frequency is given by  $p(d) = c_1(d)/ca(d)$ .

The relative frequency  $p(d)$  represents an estimate for the probability to have an edge at euclidean distance  $d$ . We repeat this procedure for 100 configurations of the rewired network

(Null-model) and calculate the corresponding averages and standard deviations as a function of the distance.

## 5. Results

### 5.1. Visual representation of the global science networks

The visualizations of the six large-scale global science networks reveal first insight into the spatial organization at the macro perspective. One common aspect for all analyzed fields is the low integration of Africa, Australia and South America, which can be attributed to low overall development levels and low experience in frontier science in the cases of Africa and South-America (cp. Duque et al. 2005). The small overall population with effects on R&D in Australia may serve as reason for the low absolute integration into the global science network. Figure 1a shows the lowest overall activity of all investigated fields. Bluetooth research seems to be a collaborative effort between Europe and North-America. Asia is less integrated. The research on image compression algorithms (Fig. 1b) is marginally concentrated on North-America, but generally globally distributed. Both fields, bluetooth and image compression algorithms, are related to information and communication technologies, one being hardware-based, one being software-related. The activity magnitude reflects the type of research, which is more synthetic than analytic, which may result in lower publication activity.

The research field of heart-valves (Fig. 1c) represents both a basic-research analytical driven subject as well as synthetic mode of knowledge production in close cooperation with applying and producing organizations (e. g. hospitals, clinical trials, and medical equipment manufacturers). The global activity is rather concentrated on Europe and North-America, both intra-continentially and inter-continentially. The research on the bird-flu virus H5N1 shows a high activity in the years under investigation and involves most regions of the world, including Africa. Asia, as the core of origin, is heavily integrated into the global network and collaboration with European and North-American research partners compare (Fig. 1d).

[FIGURE 1 about here]

Figure 1: World maps of collaborations between organizations that are involved in (a) bluetooth research, (b) image compression algorithm, (c) heart valve, (d) H5N1 research, (e) tissue engineering, and (f) nanotube carbon research.

Tissue engineering is, again, a rather North-America centered research field (Fig. 1e). Collaboration between North-America and Europe is more frequent than between North-

America and Asia. The magnitude of the collaboration activity reflects the nature of the underlying research fields and the relevance of the technology. This is even more so in nanotube research, which is the network with the highest density and a reasonable involvement of all macro-regions except Africa (Fig. 1f).

So far, it is hard to tell whether or not local connectivity outweighs the global connections. Graphical representations of large-scale networks often result in overlapping and very dense areas.

## 5.2. Probability estimation of the collaboration over distance

The six plots that are discussed in this section are showing the empirical probability of a co-authored publication as function of the distance between the collaboration partners in a double-log binned plot (cp. Fig. 2a-f). The circles represent the obtained probability for a common publication as a function of the distance. The squares represent the corresponding probabilities for the rewired networks. A constant probability (i.e. parallel to the distance axis) indicates independence of the distance. Hence, the more the collaboration activity tends to follow no spatial preference, the higher is the congruence of the empirical line (circles) with random line (squares), which will be taken as theoretical distribution model. To be able to assess the influence of national effects in the co-authorship activity, the empirical data has been decomposed into national collaboration partners (labeled same country in the plots) and exclusively international (different country).

[FIGURE 2 about here]

Figure 2: Probability for the occurrence of collaborations as a function of the distance. The probability is plotted vs. the distance in double-logarithmic representation, for the empirical data (black circles), the rewired networks (green squares), as well as the former separated into collaborations within countries (red triangle down), and between countries (blue triangle up). The panels show the results for (a) bluetooth research, (b) image compression, (c) heart valve, (d) H5N1 research, (e) tissue engineering, and (f) nanotube carbon research. The green dotted lines correspond to  $\pm 1$  standard deviation as obtained from the rewired networks.

Two general observations can be made for all presented fields. First, the probability for having national co-authors is about 10 to 100 times higher than for having an international collaborator. The shorter the distance, the larger is the difference of the probability in comparison. Second, there is a strong decay of co-authorship activity with distance, in most cases strictly monotonic decreasing in the double-log plot, i.e. the observed data function is following approx. a power law, rather than an exponential function as mainly claimed by



other authors (cp. Katz and Martin 1997). This behavior can be attributed to the national effect, since the collaboration with a partner from abroad is of equal chance no matter how distant the foreign co-author is located. One additional common feature is a hump at around 100 km indicating that collaborations at this distance are to some extent favored. This leads to smaller slopes at smaller distances. Further, there is no obvious and general difference in terms of network size or science field. However, there are some distinct features of the power law functions that have to be discussed.

The bluetooth-network (Fig. 2a) is the smallest of all presented fields and therefore potential subject to noise that is related to the small absolute number of included observations.

Interestingly, the general features as described above apply to this network as well. Up to a certain distance (here slightly above 100 km at around 250 km) there is a lower decay of the probability followed by a steeper decay in probabilities. Up to 1000 km the line is still above the random network function, meaning that the collaboration activity is higher than expected. As a engineering related subject the mode of knowledge production is rather synthetic. Hence a focus on smaller distances may be expected from theory.

For joint research on image compression algorithms the dependence on the distance is very pronounced. Although one might expect only small effects of distance in a classic field of ICT that results can be transferred through digital networks, the mode of knowledge production seems to be dominated by a cooperation pattern that relies on spatial proximity. Up to 100 km the likelihood of a collaboration is much higher than for spatially remote collaborations. Spatial proximate collaboration with partners from the same country is approximately 100 times more likely than collaboration with a partner from other countries.

Compared to the first two fields, the publication activity in heart-valve related research does almost not show any disruption, but a rather smooth distance-related decay. This has the effect that very proximate collaboration is of even higher relevance compared to the ICT-related fields. The foreign decomposition part shows a small tendency towards closer proximities, i.e. co-authors in other countries are preferably from spatial proximate countries, an effect that is frequently acknowledged in the literature (cp. Almeida et al. 2009), but not so prominent in our results.

H5N1 research is one of the most evenly distributed global sciences (cp. the visual inspection). This result is confirmed by the comparably high probability for foreign collaboration partners, which may be related to time-pressures to find vaccines that help preventing from expected pandemic developments. This short-term relevance may have had an effect on the willingness to cooperate more or less only on the basis of scientific expertise,

rather than on territorial preferences or on the participation in existing (local) collaboration networks. In the distance-classes around 200 to 1000 km the probability is very close to the expected value from the random network. However, although being rather globally organized, there is still a dominant intra-national component with a slightly smaller negative slope than for the other fields.

The field of tissue engineering shows the smoothest intra-country decay combined with an inter-country function which shows features for the rewired network. With increasing network size, the probability for the smaller distances in international collaboration, i.e. collaborators are in close proximity but divided through a national border, can be better compared to the matching distance without borders. This is of special interest as spatial and non-spatial influence can be compared directly.

For tissue engineering, this minimal distance between international collaborators is approx. 40 km. The probability for co-authoring a paper at this distance is  $p(d) = 0.002$ , which is around 200 times lower than for collaboration between authors within the same country at this distance.

Research dealing with nanotubes build from macro-molecular carbon connections is very similar in the national structure, but different in the international behavior to collaborate. First, intra-country collaboration is characterized by a slower decay up to 100 km. As this example represents the largest network with 10 times more nodes than in the bluetooth network, this empirical finding of a intra-country differentiation in collaboration propensity is fairly robust and independent from network-/community-size. The international collaboration component, however, shows distinct features. There are some very adjacent cross-border collaborations, very few collaborations within a radius up to 100 km, and then a sharp increase in the distance classes from above 100 km to approx. 800 km. Then the collaboration probability drops again rather sharply.

The result section has shown that the size of the network does not affect the clarity of the results. Moreover, the underlying subject, either being more analytical or synthetic, does not seem to affect the probability of the collaboration activity.

Two main effects can be isolated: an intra-country effect following a power law with a negative exponent, and an inter-country effect that shows features of a random distributed distance function. The former effect out-competes the latter by a factor of around 50-100 for usual intra-country distances up to 1,000 km compared to global (international) distances. The intra-country effect can be further decomposed into an effect of the near surrounding (<100

km) for most of the examples, where the decay is less steep than for collaboration with farther partners but still in the same country.

## 6. Discussion

The notion of a global science needs to be readjusted. All raised hypotheses in this paper have to be rejected. There is no evidence of greater account of global collaboration activities compared to local collaboration in the scientific world, but this should not be confused with the importance of collaboration with respect to the quality aspects of knowledge creation. However, in terms of quantity, local linkages are overwhelmingly relevant [H1a] especially *inside* national borders [H1b]. This mirrors the importance of spatial proximity in any collaboration as has been proven in other studies for pure business relations and for industry-to-university relations. Further, there is no sign of differentiation between subjects with respect to collaboration distance [H2]. How can the observed spatial pattern of collaborative scientific knowledge creation be explained? Melin (2000) suggested to consider individual scientists as globally integrated, but not research teams. Our results tell a slightly different story. As pure co-publishing teams of the same organization were not included into the analysis and only true organization spanning co-authorships were acknowledged, both individuals and cross-organizational teams are to be characterized by non-global interactions. Obviously, homophily effects (i.e. social selection effects) may contribute to a spatial selection strategy that prefers shorter over longer distances and amplifies already personally established contacts with the use of ICT systems. Therefore, communication in these scientific networks reflect structures of simple neighborhood effects. Communities of specialists are based on interpersonal ties that are, indeed, frequently territorialized (Goldenberg and Levy 2009, Storper 1985, 1997). Other empirical analyses suggest an additional force at play that fits well to the results presented here: cooperation is increasingly global, but this effect is over-compensated by even stronger local collaboration propensities. This effect seems to gain importance with the overall increase of collaboration and co-authorship in science (Moodysson et al. 2008, Wagner and Leydesdorff 2005). It would also help to explain the strong country effect, since many of these social selection effects are influenced by the socio-economic and cultural background.

For inter-country collaboration Maggioni and Uberti (2009) reported that while controlling for important country specific biases (e. g. cultural, geographical, and institutional) there is still a relevant influence of distance on knowledge flows. This is partly in contrast to our findings, where the intra-country effect of distance on collaboration is, indeed, clearly visible,

whereas all collaborations that cross national borders are more or less random in terms of distance. This becomes explicitly clear in the cases of proximate cross-border collaborations compared to intra-national collaboration of the same distance. This phenomenon of domestic seeking first and, if not available, abroad, no matter where, was earlier found in the context of science-based industries (Moodysson and Jonsson 2007).

The analyzed networks are varying in size (i.e. number of nodes) by the factor 10, although even the small ones (e. g. bluetooth) are truly international as they contain nodes that are very distant from each other. But still, all networks show the same strong relations between spatial proximity and collaboration probability. This finding may point to universality in scale independence of science relations in space.

Obviously, strategically/intentionally built global networks, like the ones that can be expected in science communities of practice, can fail quickly even with small network size and small numbers of potential collaborators. This can mean that not all options for collaboration are known and, even if they are known, not all potential connection can be established due to resource limits on either side of the collaboration partnership.

An alternative interpretation in this respect is based upon the argument of Davidson Frame and Carpenter (1979). If the national science-base is sufficiently large, there is no need for international cooperation. Further, our results suggest strong local components in collaboration, which can be considered as either to reflect a high redundancy in the science systems, where partners are always adjacent, or as a tendency towards a second-best choice (probably better suited partners available, but too distant to contact).

National policies, language and national specifics in the science culture seem to be driving forces behind the revealed collaboration activities. Long term development and manifestation of these structures is amplified through the means of increased communication. Otherwise, language may not be relevant enough as explanatory factor, because international publications are usually written in English. The pre-negotiation and idea-generation part of the knowledge creation process may, however, be strongly influenced by a similar cultural background (including language) of the collaborators. This relates to the tacit knowledge discussion. The absorption of knowledge may be global "across long distances, transcending national and geopolitical boundaries" (Schott 1998), but the production of state-of-the-art knowledge has a strong local component.

## 7. Conclusion

In this article, the global scientific collaboration structure in six selected fields has been evaluated. The proxy indicator “co-authorships” revealed a strong local collaboration propensity. It is up to 100 times more likely to have a national coauthor than an international one. The mixed empirical evidence about the relevance of global/local research interactions can be stream-lined by the systematic results from six distinct fields in terms of size, scientific community, dynamic, or analytic/synthetic style of knowledge production.

There are several mechanisms related to the empirical evidence presented here.

- The *tacit knowledge effect*, i.e. the pre-negotiation activity in collaborative research is of much higher relevance. Hence, scientific collaboration is not different from innovation-related business-prone parts of knowledge creation.
- The *competition effect*, i.e. groups of researchers separate from each other in order to ensure being the first, who publishes.
- The *grant allocation effect* and public funding, i.e. taxpayers money has to be spend at home.
- The *temporary cluster* effect may be exaggerated, i.e. conferences may be solely used to assess the capability of competing research teams.
- The *social effect*, i.e. homophily, social embedding, and language is comparably strong.
- The *technical amplification effect*, i.e. the real world social contacts are strengthened by information and communication technologies.

However, there are strong arguments for an efficient science system without global collaboration. The fierce competition for national grants may guaranty the selection of the national leader. For reputation purposes of scientific nations, this selection makes perfect sense and is amplified by the individual or group-based desire for the first encounter of pioneering research results. On a global scale, research teams are obviously better off with keeping their national partners as long as there is a critical mass. This country-based competitive system is likely to push the global knowledge pool straighter forward than a large earth-spanning network of collaborators.

Today, most of the tax-payers money for research grants goes into national funding programs (e. g. NSF) and has to be spent primarily within the country of origin. A second funding stream (e. g. the Fulbright program) is used to induce individual mobility. The latter one probably has an effect on international co-publication. However, much less money goes into this second funding stream. Thus, if international scientific collaboration should be promoted, a third funding stream seems necessary, one that is designed to give research grants to

international teams. The relevant funding body had to be multilateral (e. g. the World Bank), or based on bi-lateral or multi-lateral agreements. EU-research funding could serve as a model. Such a funding stream could define international collaborative structures that inevitably result in co-publications, and are not blocked by competition between national research teams. However, above-mentioned barriers with respect to culture and language could diminish the productivity and cost-effectiveness of international teams when compared with national teams. It seems by no means clear that science collaboration can be truly international, but it is obvious that the interplay of local and global interactions need to be accounted for in science-related national policies even in large science-based economies. Still unknown is the dynamic development and time-dependency of new sciences with respect to distance. The evolution of a field may reveal further aspects of the global science arena, the main drivers and the spatial diffusion of knowledge over time. Related to this dynamical perspective, further research may be also directed to the investigation of hierarchical structures of the presumed multi-layered science network, where the social network structure is much more reflecting spatial patterns than discussed in the upcoming literature on cognitive proximity and communities of practice. A hierarchical structure is surprisingly efficient in the complex network sense (cp. small-worlds), but obviously unwanted in the way science policies are steering towards internationalization and open science (cp. Cooke 2009). The science world of today is at least as curved and mountainous as other globalized socio-economic activities (cp. McCann 2008, Rodríguez-Pose and Crescenzi 2008).

## References

- Almeida, J. A. S., Pais, A. A. C. C., Formosinho, S. J., 2009. Science indicators and science patterns in Europe. *Journal of Informetrics* 3(2), 134–142.
- Berenson, A., 2005. Blockbuster drugs are so last century. *The NY Times* July 3.
- Boschma, R., 2005. Proximity and innovation: A critical assessment. *Regional Studies* 39(1), 61–74.
- Cairncross, F., 1997. *The Death of Distance: How Communications Revolution Will Change Our Lives*. Orion Business Books, London.
- Cooke, P., 2009. The economic geography of knowledge flow hierarchies among internationally networked medical bioclusters: A scientometric analysis. *Tijdschrift voor Economische en Sociale Geografie* 100(3), 332–347.
- Davidson Frame, J., Carpenter, M. P., 1979. International research collaboration. *Social Studies of Science* 90(4), 481–497.
- Dicken, P., 2007. *Global shift: Mapping the Changing Contours of the World Economy*, fifth ed. Sage Publications Ltd., New York.
- Duque, R. B., Ynalvez, M., Sooryamoorthy, R., Mbatia, P., Dzorgbo, D.-B. S., Shrum, W., 2005. Collaboration paradox. scientific productivity, the internet, and problems of research in developing areas. *Social Studies of Science* 35(5), 755–785.
- Gertler, M. S., 2003. Tacit knowledge and the economic geography of context, or the undefinable tacitness of being (there). *Journal of Economic Geography* 3(1), 75–99.
- Goldenberg, J., Levy, M., 2009. Distance is not dead: Social interaction and geographical distance in the internet era. Arxiv preprint arXiv:0906.3202.
- Hoekman, J., Frenken, K., van Oort, F., 2009. The geography of collaborative knowledge production in Europe. *The Annals of Regional Science* 43(3), 721–738.
- Hwang, K., 2008. International collaboration in multilayered center-periphery in the globalization of science and technology. *Science, Technology, and Human Values* 33(1), 101–133.
- Katz, J. S., Martin, B. R., 1997. What is research collaboration? *Research Policy* 26(1), 1–18.
- Laudel, G., 2002. What do we measure by co-authorships? *Research Evaluation* 11(1), 3–15.

- Lundberg, J., Tomson, G., Lundkvist, I., Skor, J., Brommels, M., 2006. Collaboration uncovered: Exploring the adequacy of measuring university-industry collaboration through co-authorship and funding. *Scientometrics* 69(3), 575–589.
- Maggioni, M. A., Uberti, T. E., 2009. Knowledge networks across Europe: which distance matters? *The Annals of Regional Science* 43(3), 691–720.
- Maslov, S., Sneppen, K., 2002. Specificity and stability in topology of protein networks. *Science* 296(5569), 910–913.
- McCann, P., 2008. Globalization and economic geography: the world is curved, not flat. *Cambridge Journal of Regions, Economy and Society* 1(1), 351–370.
- McPherson, M., Smith-Lovin, L., Cook, J. M., 2001. Birds of a feather: Homophily in social networks. *Annual Review of Sociology* 27(8), 415–444.
- Melin, G., 2000. Pragmatism and self-organization: Research collaboration on the individual level. *Research Policy* 29(1), 31–40.
- Moodysson, J., 2008. Principles and practices of knowledge creation: On the organization of buzz and pipelines in life science communities. *Economic Geography* 84(4), 449–469.
- Moodysson, J., Coenen, L., Asheim, B., 2008. Explaining spatial patterns of innovation: analytical and synthetic modes of knowledge creation in the medicon valley life-science cluster. *Environment and Planning A* 40(5), 1040–1056.
- Moodysson, J., Jonsson, O., 2007. Knowledge collaboration and proximity. *European Urban and Regional Studies* 14(2), 115–131.
- Morgan, K., 2004. The exaggerated death of geography: learning, proximity and territorial innovation systems. *Journal of Economic Geography* 4, 3–21.
- Nooteboom, B., Vanhaverbeke, W., Duysters, G. M., Gilsing, V. A., van den Oord, A. J., 2007. Optimal cognitive distance and absorptive capacity. *Research Policy* 36(7), 1016–1035.
- OBrian, R., 1992. *Global Financial Integration: The End of Geography*. Pinter, London.
- Power, D., Malmberg, A., 2008. The contribution of universities to innovation and economic development: in what sense a regional problem? *Cambridge Journal of Regions, Economy and Society* 1(2), 233–245.



- Raymond, S. U., 1997. Global cooperation in science, engineering, and medicine: an overview of the issues. *Technology in Society* 19(1), 7–16.
- Rodríguez-Pose, A., Crescenzi, R., 2008. Mountains in a flat world: why proximity still matters for the location of economic activity. *Cambridge Journal of Regions, Economy and Society* 1(3), 371–388.
- Schott, T., 1998. Ties between center and periphery in the scientific world-system: Accumulation of rewards, dominance and self-reliance in the center. *Journal of World-Systems Research* 4(2), 112–144.
- Slaughter, S., Leslie, L. L., 1997. *Academic Capitalism. Politics, Policies, and the Entrepreneurial University*. The John Hopkins University Press, Baltimore and London.
- Storper, M., 1985. Oligopoly and the product cycle: Essentialism in economic geography. *Economic Geography* 61(3), 260–282.
- Storper, M., 1997. *The Regional World: Territorial Development in a Global Economy*. Guilford Press, New York.
- Subramanyam, K., 1983. Bibliometric studies of research collaboration: A review. *Journal of Information Science* 6(1), 33–38.
- Toffler, A., 1980. *The Third Wave*. William Morrow, New York.
- Wagner, C., 2002. The elusive partnership: science and foreign policy. *Science and Public Policy* 29(6), 409–417.
- Wagner, C., 2005. Six case studies of international collaboration in science. *Scientometrics* 62(1), 3–26.
- Wagner, C., Leydesdorff, L., 2005. Mapping the network of global science: comparing international co-authorships from 1990 to 2000. *International Journal of Technology and Globalisation* 1(2), 185–208.