



# Spatial scientometrics: Towards a cumulative research program

Koen Frenken\*, Sjoerd Hardeman, Jarno Hoekman

Urban and Regional research Centre Utrecht (URU), Faculty of Geosciences, Utrecht University, NL-3508 TC Utrecht, The Netherlands

## ARTICLE INFO

### Article history:

Received 9 September 2008

Received in revised form 18 February 2009

Accepted 17 March 2009

### Keywords:

Proximity

Spatial scientometrics

Geography

Collaboration

Citation

Mobility

## ABSTRACT

We propose a research program to analyse spatial aspects of the science system. First, we provide a review of scientometric studies that already explicitly take the spatial dimension into account. The review includes studies on (i) the spatial distribution of research and citations, (ii) the existence of spatial biases in collaboration, citations and mobility, and (iii) the citation impact of national versus international collaborations. Then, we address a number of methodological issues in dealing with space in scientometrics. Finally, to integrate spatial and non-spatial approaches, we propose an analytical framework based on the concept of proximity. A proximity approach allows for combining hypotheses from different theoretical perspectives into a single framework.

© 2009 Elsevier Ltd. All rights reserved.

## 1. Introduction

There seems to be a surge in studies that address spatial aspects of science. Though research on national differences in publication output and international collaboration go back a long time, it is only recently that the spatial analysis of science is broadened to include the regional<sup>1</sup> unit of analysis and the effects of physical distance on scientific interaction. Doing so, scientometrics follows the increased interest in science and technology studies in the globalisation of knowledge production on the one hand (Stichweh, 1996; Ziman, 1994) and the location of such activities in specific places on the other hand (Cronin, 2008; Shapin, 1998). We suggest to group these contributions under the heading of spatial scientometrics.

We first present a review of quantitative science studies that explicitly address spatial aspects of scientific research activities. Here, we limit ourselves to classic papers and recent contributions, and only to those studies that made use of information as it can be retrieved from publication data.<sup>2</sup> For analytical purposes we treat space in our framework as a Euclidian surface. This perspective makes it possible to introduce physical proximity and multiple spatial units in the analysis of scientific knowledge production and diffusion. In geography, such an approach falls under the heading of regional science or spatial analysis (for overviews and critiques see Barnes, 2001; Martin, 1999).<sup>3</sup>

\* Corresponding author.

E-mail address: [k.frenken@geo.uu.nl](mailto:k.frenken@geo.uu.nl) (K. Frenken).

<sup>1</sup> In the remainder, we mean by region a sub-national region.

<sup>2</sup> Other large datasets that have been used for spatial analysis of science include Framework Programme data for the European Union (Autant-Bernard, Billand, Franchisse, & Massard, 2007; Maggioni & Uberti, 2007; Scherngell & Barber, this issue), student mobility flows (Maggioni & Uberti, 2007), editorial boards (Gutierrez & Lopéz-Nieva, 2001) and biographies of famous scientists (Catell, 1906, 1910; Taylor, Hoyler, & Evans, 2007).

<sup>3</sup> We refrain from using the terms 'geography' and 'geographical', as most contemporary geographers no longer exclusively use these terms to refer to Euclidean space (Lefebvre, 1991; Massey, 2004). See also the discussion on the notion of space in visualisation techniques provided by Skupin (2009).

From the review it will become apparent that key activities in scientific interaction (co-publication, citation, labour mobility) display clear spatial patterns. The review also makes clear that spatial analyses of science are generally done without taking into account other dimensions in terms of which scientific interaction can be characterized. It is for this reason that we turn to an analytical framework in the second part of the paper based on the proximity concept (Boschma, 2005; Rallet, 1993; Rallet & Torre, 1999). In short, the proximity concept allows one to integrate the analysis of the spatial organisation of scientific research with cognitive, organisational, institutional and social dimensions in scientific research. Such a framework provides researchers with a platform to combine hypotheses from different theoretical perspectives into a single scientometric framework.

## 2. An overview of spatial scientometrics

The spatial scientometric literature is multifaceted in terms of the topics addressed and methodologies used. We choose to organize our review in this section under three headings: (i) spatial distribution, (ii) spatial biases, and (iii) citation impact.

### 2.1. Spatial distribution

Probably the first comprehensive studies discussing the spatial distribution of science are performed by Narin and Carpenter (1975) and Frame, Narin, and Carpenter (1977).<sup>4</sup> In the latter more extensive study, the ISI Science Citation Index is assessed for 117 countries and 2300 journals divided into 92 disciplines for the year 1973. This study found that publications are highly concentrated at the country level. The world's 10 most productive countries accounted for almost 84% of all ISI publications. More recent studies showed that the spatial concentration has remained high with OECD countries still dominant in world output (Adams, 1998; Cole & Phelan, 1999; Dosi, Llerena, & Labini, 2006; Glänzel, Schubert, & Braun, 2002; Horta & Veloso, 2007; King, 2004; May, 1997).

The US typically ranks first and the UK second with respect to their share in the world's papers and citations. It remains unclear, however, to what extent the exceptional performance of the US and the UK can be attributed to an English-language advantage, an English-language bias in ISI data, or to the alleged better functioning of Anglo-Saxon institutional structures. A recent phenomenon worth noting in these descriptive studies is the rapid increase in scientific publications coming from China, which is likely to affect the top rankings in the near future (Leydesdorff & Zhou, 2005; Zhou & Leydesdorff, 2006).

Moving from the national to the regional unit of analysis, we find only few scientometric studies. The lack of regional research is probably due to the fact that the address information in many scientific publications does not contain the postal code information, which implies that regional information must be derived indirectly from address information. An early study on regions concerns the study by Matthiessen and Schwarz (1999), who addressed aggregated publication records of 1994–1996 for European regions. They found a leading group in terms of publications of only four regions (London, Paris, Moscow, and the Amsterdam–The Hague–Rotterdam–Utrecht region) publishing more than 30,000 publications each with London as the absolute number one (64,742 publications). Normalising for population size produced a somewhat different picture boosting the rank of city regions that perform well in science and have relatively small populations (e.g. Cambridge, Oxford–Reading, and Geneva–Lausanne). Differentiating among scientific fields did not alter group constellations very much and the four regions are present in almost every list of top 10 regions per discipline.

Only a few studies address the spatial distribution of citations. Bonitz, Bruckner, and Scharnhorst (1997) develop a Matthew index for countries. It turns out that both for science at large and for particular scientific fields only a few countries receive more citations than expected whereas a large majority of countries receive fewer citations than expected. Though the Matthew effect for countries is said to be rather stable over time, its magnitude is said to be rather small: only 5% of all citations account for the redistribution effect of citation winning countries and citation losing countries.

More recently, Batty (2003) assessed the concentration of scientific citations at the national, regional and organisational level. His study was limited to highly cited individuals in 12 scientific fields. He found that only a few countries (especially the United States), a few regions (especially the south-west coast and the north-east coast in the United States), and a few organisations (especially Harvard University, Stanford University and University of California San Diego) are populated by highly cited researchers. At all different units of analysis there is considerable concentration of highly cited individual researchers.

Whether concentration of research also brings advantages is another relevant issue in spatial scientometrics. The notion of agglomeration advantages is useful here. These are efficiency gains for a researcher or research institute stemming from co-locating in a geographical cluster, that is, in the vicinity of many other researchers or research institutes, respectively. Advantages stem primarily from cost advantages in search costs for partners and new personnel, sharing of infrastructure, and the availability of supporting services. Furthermore, the cost of collaboration is lower as travel costs increase with physical distance.

We know of only two studies that have been specifically focused on the measurement of agglomeration advantages in scientific knowledge production. From an economic point of view, agglomeration advantages are best measured by the

<sup>4</sup> Note that most studies until then either make use of publication data without taking space into account (for instance Merton, 1973 and Price, 1963) or take space into account without making use of publication data (Catell, 1906, 1910).

effects of spatial concentration on efficiency. In the context of scientific knowledge production, Bonaccorsi and Daraio (2005) analyzed the effect of spatial concentration of research on the publications per researcher. Looking at CNR researchers in Italy and INSERM researchers in France, they found agglomeration advantages to be present indeed, though the evidence was found not to be very strong. Carvalho and Batty (2006) use a technique in which they are able to detect whether research productivity is spatially concentrated in the US after controlling for investments in Research and Development and for population. They also find agglomeration advantages to be present though again the evidence is thin.

## 2.2. Spatial biases

The extreme spatial concentration of scientific activity is quite remarkable. There may exist systematic spatial biases in the interactions of researchers favouring those located in the vicinity of many fellow researchers. At least three mechanisms may explain why interactions in science are spatially biased towards physically proximate actors. First, serendipitous encounters are more likely when two actors are in close vicinity of each other. Second, the need for face-to-face interaction when engaging in interactions comes at a cost, which increases as a function of travel time. Third, 'the rules of the game' that matter for scientific knowledge production (e.g. funding, labour market regimes, intellectual property right regimes, languages) are spatially differentiated and constrain interaction between institutional frameworks, in particular, between nation-states.

One particular bias that has received a lot of attention within scientometrics is the bias to collaborate domestically rather than internationally. It is commonly assumed that the bias to collaborate nationally has decreased over time due to globalisation of the science system. The classic study by Narin, Stevens, and Whitlow (1991) looked at publications in 28 scientific fields in the period 1977–1986 and found that the share of papers that are internationally co-authored increased from around 10% in 1977–1979 to around 13% in 1983–1985. A number of studies have used the same definition of internationalisation (Georgiou, 1998; Glänzel, 2001; Hicks & Katz, 1996) with the most recent one showing that the share of internationally co-authored papers in all ISI publications increased from 10% in 1990 to 23% in 2005 (Leydesdorff & Wagner, 2008). Liang, Zhang, Kretschmer, and Scharnhorst (2006) show in this respect that internationalisation has also occurred at a higher level of spatial aggregation, as the share of collaborations within the EU has decreased, while the share of papers that list both a EU-country and a non-EU country has increased.

Yet, the conclusion of 'internationalisation' or 'globalisation' in research collaboration is not undisputed. If one does not measure internationalisation by the share of internationally co-authored papers, but by the share of international collaborations over national collaborations counting each co-occurrence of two addresses as one research collaboration, other results have been obtained. Using this alternative methodology, Frenken (2002) observed for EU collaborations in the period 1993–2000 that the strong bias toward domestic collaboration persists over time. Similarly, looking at eight disciplines, Ponds (2009) found for papers involving at least one address from the Netherlands that the internationalisation of research seemed to have come to a halt in the Dutch case.

The tendency to collaborate domestically may also be related to the number of researchers in a country. An early study by Frame and Carpenter (1979) found that larger countries are less prone to collaborate internationally. Their analysis, however, did not control for the fact that if researchers would choose their partner randomly, researchers in countries with many researchers will automatically have a stronger domestic bias than researchers in smaller countries. Controlling for this effect, Frenken (2002) found that the countries with most researchers actually display the weakest bias to collaborate domestically.

Apart from analysing biases towards domestic collaboration, studies have also focused on biases in the country of origin of the collaboration partner. Frame and Carpenter (1979) observed that the strongest collaboration patterns exist among nearby countries sharing socio-political characteristics. Frenken's (2002) study on the EU also found a bias toward collaboration with neighbouring countries, while Frenken et al. (in press), in a study on the 36 most productive countries in the world, showed that the propensity to collaborate was negatively affected by the flight distance between capitals. Liang et al. (2006) drew similar conclusions for the countries of the European Union.

Studies on collaboration patterns among regions or cities are rare. The first study has been Katz's (1994) analysis of the effect of physical distance for university–university collaboration in the UK, Canada and Australia. Distance was computed from the physical distance ('as the crow flies') between the cities in which two universities are located. The main conclusion was that an increase in distance significantly decreases the frequency of research collaboration pointing toward the importance of face-to-face communication for collaboration. A subsequent study by Liang and Zhu (2002) for Chinese regions confirmed Katz's earlier study. At the city-level, Havemann, Heinz, and Kretschmer (2006) came to a different conclusion. They found that German immunological institutes are more likely to collaborate with partners in the same city, but in collaborations with institutes outside the city distance was found not to affect the probability of collaboration.

More recent studies estimated gravity models, which determine the collaboration frequency between two regions from their physical distance and their respective sizes, where size is measured by the total number of publications in a region. For Dutch NUTS3<sup>5</sup> regions, a gravity analysis by Ponds, van Oort, and Frenken (2007) used travel time as a distance indi-

<sup>5</sup> The Nomenclature of Territorial Units for Statistics (NUTS) is a uniform breakdown of spatial units in the European Union which follows a four-level hierarchy that ranges from NUTS0 to NUTS3. The NUTS0 level corresponds to the territory of individual member states, whereas NUTS3 roughly corresponds to labour market regions in most countries.

cator and concluded that a longer travel time decreases collaboration frequency, but less so for university–university collaborations than for university–industry–government collaborations. In a study on NUTS3 regions in the EU27, a gravity analysis by Hoekman, Frenken, and Van Oort (2009) found that both physical distance (as measured ‘as the crow flies’) and country borders render collaboration between regions less likely. This study also found that the most producing regions and the regions hosting a capital city are collaborating more than proportionally. Such observations suggest that there are ‘elite structures’ consisting of well-off regions with excessive number of collaborations among them.

Concerning the analysis of spatial bias in citations, only few examples can be mentioned. Matthiessen, Schwarz, and Find (2002) identified the 40 most publishing regions in the world in terms of publication output in 1997–1999 and compared collaboration with citation patterns among these regions. Though not using any systematic statistical methodology, they concluded from their data that (i) both citation and collaboration relations occur most frequently domestically, (ii) citations are much less affected by distance than collaboration, and (iii) that the domestic bias in citation and collaboration increase with the size of the country.<sup>6</sup> Recently, Börner, Penumathy, Meiss, and Ke (2006) assessed the distance decay of the 500 most cited research institutions in the United States between 1982 and 2001 statistically. The results suggest that there is a distance-decay in citation relations between research organisations: articles from nearby research organisations are more likely to be cited than articles from research organisations further away. Over time, the effect of physical distance on the probability of citation relations to occur did not decrease.

Spatial biases may also exist in the labour mobility<sup>7</sup> patterns of researchers. Stephan and Levin (2001) selected a group of researchers based on highly cited papers in ISI and complemented the data with biographical information on their countries of birth. The main result held that the international mobility balances of researchers are highly skewed across countries, resulting in brain drain of researchers in some countries and exceptionally high contributions of the foreign-born to the science system in others, in particular, the United States. This result qualifies the finding of U.S. dominance in science as being partially an effect of mobility. A more comprehensive scientometric framework to study movements of researchers is provided by Laudel (2003). The author asserts that longitudinal bibliometric data open up the possibility of tracking down changes in affiliation addresses of individuals over time. An application to elite researchers in two small research fields confirms the magnetic forces that draw researchers to the United States, but also points towards disciplinary specificities of the phenomenon (Laudel, 2005).

A final topic has been addressed by Van Dijk and Maier (2006) and concerns conference attendance patterns. Looking at papers presented at the annual European conferences of the Regional Science Association (ERSA), the authors analyzed the effect of physical distance on conference attendance. Using conference dummies to control for city-specific effects, they found that the physical distance to a conference venue indeed affects participation.

### 2.3. Citation impact

A large number of studies in scientometrics deal with the citation impact of different types of collaboration. In particular, several studies have compared the citation impact of internationally co-authored papers with domestically co-authored papers in the light of investigating the rationale for internationalisation policies. Such an effect is expected since more resources are invested in international collaboration and more diffusion channels are activated once results are published.

Among the first to assess this issue has been the study by Narin et al. (1991) on a subset of papers covering biomedical research published in 1977. They found that international co-publications are cited on average more often than domestic co-publications, while no differences were found for international co-publications among EU countries, between EU and non-EU countries and among non-EU countries. Katz and Hicks (1997) posed the same question for all publications from the UK in 1981–1991. Controlling for the number of authors, organisations and countries, they also found that the average citation rate of papers increases more by adding an author from a foreign organisation as compared to adding an author from a domestic organisation. Frenken, Hözl, and De Vor (2005) assessed the citation impact of international research collaboration in European biotechnology for the period 1988–2002, while controlling for number of authors and organisations as well as country dummies. They found only evidence of a higher citation impact of international co-publications for EU publications, while publications between an EU country and a country outside the EU did not receive more citations than a domestic collaboration. Using a similar methodology, but adding author-fixed effects to control for the observation that more successful authors tend to collaborate more internationally than less successful authors, Singh (2007) and He (2008) still found a significant and positive effect of international collaboration on citation impact compared to domestic co-publications. In all, these studies suggest that international research has more impact, on average, than domestic research.

<sup>6</sup> Note that Matthiessen et al. (2002) do not control for the observation that researchers in countries with many researchers will automatically have a stronger domestic bias than researchers in smaller countries.

<sup>7</sup> We prefer to speak of labour mobility instead of migration as we cannot retrieve authors' intentions concerning their length of stay in a particular place from publication data. Thus, we only observe mobility and not migration per se.

## 2.4. Summary

In quantitative science studies that explicitly take into account space, we distinguished between three major topics. First, there are descriptive studies on differences between countries and between regions in terms of their publication output and citations. Second, a range of studies found systematic evidence on spatial biases in collaboration, citation, labour mobility and conference attendance. Third, the citation impact for international co-publications is higher than for national co-publications. [Table 1](#) provides a summary table.

## 3. Methodological issues

Spatial scientometrics relies on the address information on publications to locate the places where knowledge is created or diffused to. Using such data is not without problems and a number of methodological issues are recurrent. Here, only those methodological issues are discussed that are of particular interest to spatial scientometrics rather than of scientometrics at large.<sup>8</sup> First, the use of address information is based on the assumption that addresses listed on publications tell something about the location where the actual research was conducted. In general, there are no reasons to believe that this assumption is unreasonable. Yet, we know that publication data exhibit noise and we do not yet have clear estimates of the amount and nature of this noise. For example, researchers on temporary visit may choose to list their home institute and researchers may sometimes list the grant organisation rather than the institute where the research has been done. Concerning research institutes and companies, headquarters are sometimes listed instead of the subsidiary where the research was carried out.

Second, address information refers to research institutes and not to authors. Only recently, information is available linking authors to addresses. In the absence of information on author–address links, multiple address publications may actually refer to a single author with multiple affiliations or an author who conducted research at one institute and subsequently moved to another institute. Even though someone with multiple affiliations can be said to establish collaboration between multiple organisations, its meaning is clearly different from a project where multiple researchers from different organisations are involved. It was estimated that within the UK circa 2.5% of all authors the number of organisations exceeds the number of articles (Hicks & Katz, 1996).

Third, constructing aggregated data at the level of spatial units also poses some methodological problems. The fundamental problem in spatial research is to decide what spatial unit is a relevant unit of analysis. Most studies aggregate addresses to predefined administrative areas (e.g. countries, states, provinces, municipalities, etc.). Yet, there is generally no reason to believe that such administrative boundaries coincide with relevant boundaries as perceived by researchers. In geography this problem is known as the Modifiable Areal Unit Problem (Openshaw, 1984). More in general this problem occurs when data are aggregated into discrete units. Ideally, one would therefore like to analyse individual authors or institutes with a specific address and the effects of physical distance between each pair of authors or institutes. Effects of higher order spatial units can then be addressed in a multi-level framework. Yet, collecting and treating such detailed data is very time-consuming. Aggregating authors or organisations to a spatial unit is often more practical. In that case, labour market areas consisting of one major city and its commuting area can be considered the most relevant unit of analysis as such areas facilitate face-to-face interaction at a daily bases (Hoekman et al., 2009).<sup>9</sup>

Finally, from the relational nature of the data on research collaboration, citations, labour mobility and conference attendance raises the problem of measurement of physical distance. Distance can be measured in a categorical way (domestic vs. foreign) or in a continuous way. In the latter case, most studies rely on physical distance ('as the crow flies') as such information is readily available from Geographical Information System software. However, more appropriate measures of distance would be travel time or travel cost as these measures indicate more directly the real burden people have to overcome while travelling from A to B.

## 4. Towards a proximity approach

From our review it has become clear that, thus far, research in spatial scientometrics is rather fragmented. Studies differ in methodologies, units of analysis, notions of distance, and explanatory frameworks. What is more, spatial analyses of science are generally done without taking into account other dimensions in terms of which scientific interaction can be characterized. Following the analytical notion of proximity as introduced in the field of economic geography, we propose a framework that can function as one possible conceptual core allowing to link research in spatial scientometrics with other endeavours in scientometrics.

The proximity concept has been developed by a group of French economists interested in the spatial evolution of economic activities (Carrincazeaux, Lung, & Vicente, 2008; Rallet, 1993; Rallet & Torre, 1999). The main contribution of their approach

<sup>8</sup> For an account on methodological issues in scientometrics at large see for example Schubert, Glänzel, and Braun (1989).

<sup>9</sup> Labour market regions are best covered by the NUTS3 level of analysis, whereas in the US context labour market regions are best covered by Metropolitan Statistical Areas.



**Table 1**

Notions of space in spatial scientometrics.

Topic	Author	Geographical scope	Spatial unit	Notion of distance
Spatial distribution of publications	Frame et al. (1977)	World (117 countries, 7 sub-continental regions)	Country and sub-continental region	–
Spatial distribution of publications and citations	May (1997)	World (15 countries)	Country	–
Spatial distribution of publications and citations	Bonitz et al. (1997)	World (50 countries)	Country	–
Spatial distribution of publications and citations	Adams (1998)	World (7 countries)	Country	–
Spatial distribution of citations	Cole and Phelan (1999)	World (95 countries)	Country	–
Spatial distribution of publications	Matthiessen and Schwarz (1999)	Europe (39 regions)	Region	–
Spatial distribution of publications and citations	Glänzel et al. (2002)	World (32 countries)	Country	–
Spatial distribution of citations	Batty (2003)	World (multiple countries and regions)	Country and region	–
Spatial distribution of publications and citations	King (2004)	World (31 countries)	Country	–
Spatial distribution of publications and citations	Leydesdorff and Zhou (2005)	World (31 countries)	Country	–
Spatial distribution of publications and citations	Dosi et al. (2006)	World (multiple countries)	Country and sub-national region	–
Spatial distribution of publications and citations	Zhou and Leydesdorff (2006)	World (6 countries, EU 15 and EU 25)	Country and sub-continental region	–
Spatial distribution of publications	Carvalho and Batty (2006)	USA	Region	Continuous
Spatial distribution of publications	Bonaccorsi and Daraio (2005)	France and Italy	Region	Categorical
Spatial distribution of publications and citations	Horta and Veloso (2007)	World (EU 15 and USA)	Country and sub-continental region	–
Spatial bias in research collaborations	Frame and Carpenter (1979)	World (15 countries, 8 sub-continental regions)	Country and region	Categorical
Spatial bias in research collaborations	Narin et al. (1991)	World (5 countries and EU 9)	Country	Categorical
Spatial bias in research collaborations	Katz (1994)	UK, Canada, Australia	Region	Continuous
Spatial bias in research collaborations	Hicks and Katz (1996)	UK	Country	Categorical
Spatial bias in research collaborations	Georghiou (1998)	World (EU – Economic zone, North-America, Japan, the Republic of Korea and Australasia)	Sub-continental regions	Categorical
Spatial bias in research collaborations	Glänzel (2001)	World (multiple countries)	Country	Categorical
Spatial bias in labour mobility of scientists	Stephan and Levin (2001)	USA	Country	Categorical
Spatial bias in research collaborations	Frenken (2002)	Europe (15 countries)	Country	Categorical
Spatial bias in research collaborations and citations	Matthiessen et al. (2002)	World (40 regions)	Region	Categorical
Spatial bias in labour mobility of scientists	Laudel (2005)	USA	Country	Categorical
Spatial bias in research collaborations	Wagner and Leydesdorff (2005)	World	Country	Categorical
Spatial bias in research collaborations	Havemann et al. (2006)	Germany (multiple cities)	City	Continuous
Spatial bias in citations	Börner et al. (2006)	USA	Region	Continuous
Spatial biases in research collaborations	Liang et al. (2006)	Europe (EU 15)	Region	Continuous and categorical

Table 1 (Continued)

Topic	Author	Geographical scope	Spatial unit	Notion of distance
Spatial bias in conference attendance	Van Dijk and Maier (2006)	Europe (6 cities)	City	Continuous
Spatial bias in research collaboration	Ponds et al. (2007)	Netherlands	Region	Continuous
Spatial bias in research collaboration	Tijssen (2008)	Europe (EU 15 and EU 27)	Region	Categorical
Spatial bias in research collaboration	Frenken et al. (in press)	World (36 countries), Europe (1316 regions), the Netherlands (40 regions)	Country and region	Continuous and categorical
Spatial bias in research collaboration	Hoekman et al. (2009)	Europe (27 countries)	Region	Continuous and categorical
Spatial bias in research collaboration	Leydesdorff and Wagner (2008)	World (multiple countries)	Country	Categorical
Spatial bias in research collaboration	Ponds (2009)	Netherlands	Region	Categorical
Citation impact	Narin et al. (1991)	World	Country	Categorical
Citation impact	Katz and Hicks (1997)	UK	Country	Categorical
Citation impact	Frenken et al. (2005)	Europe (15 countries)	Country	Categorical
Citation impact	Singh (2007)	Europe (15 countries)	Country	Categorical
Citation impact	He (2008)	New Zealand	Country	Categorical

lies in disentangling physical proximity from other forms of proximity. Like physical proximity, other forms of proximity structure interactions. To mention one prominent example, the presence of social relationships among employees of two firms may create trust and lower transaction costs (Uzzi, 1997).

In a review of the various literatures on the role of proximity on collaborative innovation, Boschma (2005) distinguished between five forms of proximity: physical,<sup>10</sup> cognitive, social, organisational and institutional. All five dimensions can be expected to play a role in scientific interaction.

- *Physical proximity*, as already discussed, is generally taken as the kilometric distance (Hoekman et al., 2009; Liang et al., 2006), travel time (Ponds, 2009) or in a binary fashion contrasting domestic versus foreign relations (Frenken et al., 2005; He, 2008; Hoekman et al., 2009; Liang et al., 2006; Ponds et al., 2007; Singh, 2007). From the review provided above, we conclude that physical proximity indeed affects scientific interaction patterns.
- *Cognitive proximity* can be defined as the extent to which two researchers share the same knowledge base. Cognitive proximity among researchers is fundamental as to engage in meaningful interaction. As sociologists of scientific knowledge have been arguing, the establishment of a knowledge claim as a scientific fact is not exclusively determined by observation, but also requires that researchers have a shared understanding of the meaning of an observation (Collins, 1985; Shapin, 1984). Previous statistical studies have shown that cognitive proximity has a strong impact on citation (Baldi, 1998; White, Wellman, & Nazer, 2004). On the one hand, one thus expects collaboration patterns to be primarily present among researchers who already share disciplinary knowledge. On the other hand however, the bridging of cognitively distant research collaborators might lead to relatively radical contributions (Lambiotte & Panzarasa, this issue) or even new scientific fields (Lucio-Arias & Leydesdorff, 2009).
- *Social proximity* can be defined as the extent to which researchers have friendly relationships. Social relationships facilitate interaction by creating trust among researchers. Trust is important in carrying out complex research projects, but also plays a role in judging the validity of claims written down in papers (Shapin, 1984). Social relations can stem from private life or professional life and have been measured as such in statistical studies addressing citation patterns between researchers (Baldi, 1998; White et al., 2004).<sup>11</sup>
- *Organisational proximity* can be defined as the extent to which two researchers are under common hierarchical control, which is important to coordinate research activities. Such a variable can be constructed as a dichotomous variables indicating whether two people work for the same or for different organisation as done in studies on the citation impact of research collaboration (He, 2008; Singh, 2007).
- *Institutional proximity* can be defined as the extent to which researchers operate under the same incentive structures, which aligns the objectives of researchers. In the context of spatial scientometrics, the main institutional spheres that are generally distinguished are universities, industry and government (Frenken et al., 2005; Ponds et al., 2007; Singh, 2007). University researchers primarily aim at transforming research into publication, corporate researchers at transforming research into commercial knowledge and government researchers at transforming knowledge into policies. This conflict of interest has been a central topic in the economics of science literature (Dasgupta & David, 1994; Stephan, 1996). The lack of institutional

<sup>10</sup> Boschma (2005) calls this form of proximity 'geographical proximity' but for reasons of consistency we use the term physical proximity.

<sup>11</sup> White et al. (2004) also speak of socio-cognitive proximity structures, which they defined as whether two researchers have collaborated in the past. In our framework, this form of proximity would be treated as social proximity following Singh (2007).

**Table 2**  
Spatial scientometric papers using the proximity framework.

Author	Topic	Unit of analysis	Proximity dimensions				
			Geographical	Cognitive	Social	Organisational	Institutional
Frenken et al. (2005)	Citation impact of collaboration	Papers	Categorical: domestic vs. foreign	–	–	–	Categorical: university vs. other
Ponds et al. (2007)	Collaboration bias	Organisations	Continuous: travel time	–	–	–	Categorical: university–industry–government Categorical: firms vs. other
Singh (2007)	Citation impact of collaboration	Papers	Categorical: domestic vs. foreign	–	Categorical: past collaboration	–	–
He (2008)	Citation impact of collaboration	Papers	Categorical: domestic vs. foreign	–	–	Categorical: intra-university vs. extra-university	–
Ponds (2009)	Collaboration bias	Organisations	Categorical: domestic vs. foreign	–	–	–	Categorical: university vs. other

proximity in university–industry–government relationships reflects the complexity of such projects and the difficulty to design policies aimed at fostering projects (Etzkowitz & Leydesdorff, 2000; Gibbons et al., 1994).<sup>12</sup>

The list of proximity dimensions presented here is by no means exhaustive. Analytically, the proximity approach as further outlined below allows for the inclusion of any number of proximity dimensions. For example, lingual proximity (Liang et al., 2006), ethnic proximity, ideological proximity or proximity in terms of age may also play a role in scientific interaction. Moreover, to develop a coherent framework incorporating additional theories is always necessary.

Any two entities (researchers, research institutes, regions, countries) can now be conceptualized as having a relational distance in at least five dimensions. Thus, for a study comprising of  $n$  entities, we have  $1/2(n^2 - n)$  pair wise relationships, which constitute the observations in a proximity analysis. For each observation, the distance between the two entities involved is then described in all proximity dimensions. Once the data are constructed in the manner, one can start to combine proximity dimensions in a single research design.

Table 2 summarizes the analyses done so far in quantitative science studies that both take the spatial dimension into account and at the same time make use of one other proximity dimension. Looking at biases in scientific research collaboration Ponds et al. (2007) and Ponds (2009) assessed whether the role of geographical proximity is mediated by institutional proximity. Both studies conclude that scientific research collaboration between institutionally different organisations is more geographically localised than scientific research collaboration between institutionally similar organisations. With respect to the citation impact of scientific research collaborations Frenken et al. (2005) found that international research collaboration has a higher citation impact than national research collaboration. Singh (2007) and He (2008) concluded that once controlled for social and organisational proximity, respectively, the difference in citation impact of international and national research collaboration is much smaller.

The importance of explicitly including a proximity dimension varies across research designs. Often, by focusing the study on a specific subset of publications one can already control for proximity dimensions in an implicit manner, albeit imperfectly. For example, in the analysis of the effect of physical proximity on collaboration between universities, Katz (1994) implicitly controls for institutional proximity. Similarly, focusing the analysis on a particular discipline (Frenken et al., 2005; Hoekman et al., 2009; Ponds, 2009; Ponds et al., 2007; Singh, 2007) renders the inclusion of an explicit cognitive proximity dimension less important.

Future research can make use of multiple proximity dimensions to address various kinds of research questions. There are at least four types of analyses that can be done within the proposed proximity framework. In such analyses, several hypotheses can be developed and tested, and separate studies can be compared more easily in terms of their mutual consistency.

First, taking these proximity dimensions as independent variables, the strength of scientific interaction as the dependent variables can be predicted. As a dependent variable one can take collaboration frequency, citation frequency or labour mobility flows.<sup>13</sup> This is essentially what is done when applying the gravity model (Frenken et al., in press; Hoekman et al., 2009; Ponds et al., 2007), whether addressing collaboration, citations or labour mobility flows. The main advantage of the proximity approach as described here is that scientific interaction is assessed in a multivariate framework. Taking into

<sup>12</sup> Note that Hoekman et al. (2009) defined institutional proximity as a collaboration that takes place under the same set of territorially bounded institutions as in domestic collaboration. However, for reasons of consistency, we define domestic relations here under physical proximity.

<sup>13</sup> Another dependent variable can be the similarities among science systems. Using co-structure cluster analysis Bonitz, Bruckner, and Scharnhorst (1993) argue that some countries are similar in their publication patterns to other countries. They suggest that these similarities might be due to physical proximity as in the case of the Scandinavian countries or institutional proximity as in the case of the commonwealth countries.



account multiple proximity dimensions simultaneously is important since proximity dimensions are generally significantly correlated.<sup>14</sup> That means that the effect of a particular form of proximity (e.g. physical proximity) can only be properly assessed when controlling for other proximity dimensions. For example, in general, one expects cognitive proximity to be much more important in structuring scientific interaction than physical proximity, since at least some minimum of cognitive proximity is required in any meaningful interaction among scientists.

Second, within a proximity framework one can analyse whether proximities are substitutes. The proximity concept includes the idea that being proximate in one dimension allows distance in another dimension. For example, physical proximity is helpful in many forms of scientific interaction, but it is expected to be less important if two researchers are proximate in, say, the cognitive dimension. In the latter case, interaction through the Internet is expected to be very effective (Amin & Cohendet, 2004). Another example concerns the relation between physical and institutional proximity. In university–industry–government collaboration institutional proximity is lacking, which might explain why such collaborations are relatively more often realised within the boundaries of a region (Ponds et al., 2007).

Third, in the case of research collaboration, one can relate the proximity in each dimension to citation impact. For example, one can ask the question whether physical distance between collaborating actors contribute to the citation impact of the joint paper, while controlling for cognitive, organisational, social and institutional distance. Controlling for all proximity dimensions other than physical proximity is important to analyse whether physical proximity, or its absence, truly affects citation impact.

Finally, the notion of temporary geographical (i.e. physical) proximity is useful as to extend the proximity framework from mere static to dynamic analysis (Rallet & Torre, 1999; Rychen & Zimmermann, 2008; Torre, 2008). Using the notion of temporary proximity one can analyse how different proximity dimensions interact over time. For example, a new field may emerge from a single research institute where researchers develop cognitive proximity over time. When these researchers subsequently move to other organisations, physical and organisational proximity vanishes, but the cognitive and social proximity built up in the past allows them to continue to collaborate effectively.

## 5. Concluding remarks

Our objective was to provide an analytic framework for spatial scientometric research. In our review of the literature, we extracted collaboration, citation and mobility as the main research topics. To analyse these forms of scientific interaction we then proposed to use the concept of proximity, which distinguishes physical proximity from other forms of proximity as determinants of scientific interaction. The framework provides researchers in scientometrics with an analytical starting point to combine hypotheses from different theoretical perspectives.

## References

- Adams, J. (1998). Benchmarking international research. *Nature*, 396, 615–618.
- Amin, A., & Cohendet, P. (2004). *Architectures of knowledge: Firms, capabilities, and communities*. Oxford: Oxford University Press.
- Autant-Bernard, C., Billand, P., Franchisse, D., & Massard, N. (2007). Social distance versus spatial distance in R&D cooperation: Empirical evidence from European collaboration choices in micro and nanotechnologies. *Papers in Regional Science*, 86(3), 495–519.
- Baldi, S. (1998). Normative versus social-constructivist processes in the allocation of citations: A network-analytic model. *American Sociological Review*, 63(6), 829–846.
- Barnes, T. J. (2001). Retheorizing economic geography: From the quantitative revolution to the “cultural turn”. *Annals of the Association of American Geographers*, 91(3), 546–565.
- Batty, M. (2003). The geography of scientific citation. *Environment and Planning A*, 35, 761–765.
- Bonaccorsi, A., & Daraio, C. (2005). Exploring size and agglomeration effects on public research productivity. *Scientometrics*, 63(1), 87–120.
- Bonitz, M., Bruckner, E., & Scharnhorst, A. (1993). The science strategy index. *Scientometrics*, 26(1), 37–50.
- Bonitz, M., Bruckner, E., & Scharnhorst, A. (1997). Characteristics and impact of the Matthew effect for countries. *Scientometrics*, 40(3), 407–422.
- Börner, K., Penumarth, S., Meiss, M., & Ke, W. (2006). Mapping the diffusion of scholarly knowledge among major US research institutions. *Scientometrics*, 68(3), 415–426.
- Boschma, R. A. (2005). Proximity and innovation: A critical assessment. *Regional Studies*, 39(1), 61–74.
- Carrincazeaux, C., Lung, Y., & Vicente, J. (2008). The scientific trajectory of the French school of proximity: Interaction- and institution-based approaches to regional innovation systems. *European Planning Studies*, 16(5), 617–628.
- Carvalho, R., & Batty, M. (2006). The geography of scientific productivity: Scaling in US computer science. *Journal of Statistical Mechanics*, 10(P10012), 1–11.
- Catell, J. M. (1906). A statistical study of American men of science. III: The distribution of American men of science. *Science*, 24(623), 732–742.
- Catell, J. M. (1910). A further statistical study of American men of science. *Science*, 32(827), 633–648.
- Cole, S., & Phelan, T. J. (1999). The scientific productivity of nations. *Minerva*, 37(1), 1–23.
- Collins, H. M. (1985). *Changing order: Replication and induction in scientific practice*. London: Sage.
- Cronin, B. (2008). On the epistemic significance of place. *Journal of the American Society for Information Science and Technology*, 59(6), 1002–1006.
- Dasgupta, P., & David, P. A. (1994). Toward a new economics of science. *Research Policy*, 23, 487–521.
- Dosi, G., Llerena, P., & Labini, M. S. (2006). The relationships between science, technologies and their industrial exploitation: An illustration through the myths and realities of the so-called ‘European Paradox’. *Research Policy*, 35(10), 1450–1464.
- Etzkowitz, H., & Leydesdorff, L. (2000). The dynamics of innovation: From National Systems and “Mode 2” to a Triple Helix of university–industry–government relations. *Research Policy*, 29(2), 109–123.
- Frame, J. D., & Carpenter, M. P. (1979). International research collaboration. *Social Studies of Science*, 9(4), 481–497.
- Frame, J. D., Narin, F., & Carpenter, M. P. (1977). The distribution of world science. *Social Studies of Science*, 7(4), 501–516.

<sup>14</sup> This correlation can be understood dynamically from the mutually reinforcing nature of proximities. For example, social and cognitive proximity often stems from physical proximity as many social and cognitive processes take place through face-to-face interaction.

- Frenken, K. (2002). A new indicator of European integration and an application to collaboration in scientific research. *Economic Systems Research*, 14(4), 345–361.
- Frenken, K., Hoekman, J., Kok, S., Ponds, R., Van Oort, F., & Van Vliet, J. (in press). Death of distance in science? A gravity approach to research collaboration. In: Pyka, A., & Scharnhorst, A. (Eds.), *Innovation networks—New approaches in modelling and analysing*. Berlin: Springer.
- Frenken, K., Hözl, W., & De Vor, F. (2005). The citation impact of research collaborations: The case of European biotechnology & applied microbiology (1988–2002). *Journal of Engineering Management and Technology*, 22(1–2), 9–30.
- Georgiou, L. (1998). Global cooperation in research. *Research Policy*, 27, 611–626.
- Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., & Trow, P. (1994). *The new production of knowledge: The dynamics of science and research in contemporary societies*. London: Sage.
- Glänzel, W. (2001). National characteristics in international scientific co-authorship relations. *Scientometrics*, 51(1), 69–115.
- Glänzel, W., Schubert, A., & Braun, T. (2002). A relational charting approach to the world of basic research in twelve science fields at the end of the second millennium. *Scientometrics*, 55(3), 335–348.
- Gutierrez, J., & López-Nieva, P. (2001). Are international journals of human geography really international? *Progress in Human Geography*, 25(1), 53–69.
- Havemann, F., Heinz, M., & Kretschmer, H. (2006). Collaboration and distances between German immunological institutes—a trend analysis. *Journal of Biomedical Discovery and Collaboration*, 1–6.
- He, Z.-L. (2008). On the relationship between research collaboration and research output: A longitudinal study. In *Paper presented at the 25th Celebration Conference 2008 on Entrepreneurship and Innovation—Organisations, Institutions, Systems and Regions* Copenhagen, CBS, Denmark, June 17–20.
- Hicks, D. M., & Katz, J. S. (1996). Where is science going? *Science, Technology & Human Values*, 21(4), 379–406.
- Hoekman, J., Frenken, K., & Van Oort, F. (2009). The geography of collaborative knowledge production in Europe. *Annals of Regional Science*, doi:10.1007/s00168.008.0252.9
- Horta, H., & Veloso, F. M. (2007). Opening the box: Comparing EU and US scientific output by scientific field. *Technological Forecasting and Social Change*, 74, 1334–1356.
- Katz, J. S. (1994). Geographical proximity and scientific collaboration. *Scientometrics*, 31(1), 31–43.
- Katz, J. S., & Hicks, D. (1997). How much is a collaboration worth: A calibrated bibliometric model. *Scientometrics*, 40(3), 541–554.
- King, D. A. (2004). The scientific impact of nations. *Nature*, 430, 311–316.
- Lambiotte, R., & Panzarasa, P. (2009). Communities, knowledge creation and information diffusion. *Journal of Informetrics*, 3(3), 180–190.
- Laudel, G. (2003). Studying the brain drain: Can bibliometric methods help? *Scientometrics*, 57(2), 215–237.
- Laudel, G. (2005). Migration currents among the scientific elite. *Minerva*, 43, 377–395.
- Lefebvre, H. (1991). *The production of space*. D. Nicholson-Smith trans. Oxford: Basil Blackwell. (Originally published 1974)
- Leydesdorff, L., & Wagner, C. S. (2008). International collaboration in science and the formation of a core group. *Journal of Informetrics*, 2(4), 317–325.
- Leydesdorff, L., & Zhou, P. (2005). Are the contributions of China and Korea upsetting the world system of science? *Scientometrics*, 63(3), 617–630.
- Liang, L. M., & Zhu, L. (2002). Major factors affecting China's inter-regional research collaboration: Regional scientific productivity and geographical proximity. *Scientometrics*, 55(2), 287–316.
- Liang, L. M., Zhang, L., Kretschmer, H., & Scharnhorst, A. (2006). Geographical and lingual preferences in scientific collaboration of the European Union (1994–2003). In *Proceedings International Workshop on Webometrics, Informetrics and Scientometrics & Seventh COLLNET Meeting* Nancy (France).
- Lucio-Arias, D., & Leydesdorff, L. (2009). The dynamics of exchanges and references among scientific texts, and the autopoiesis of discursive knowledge. *Journal of Informetrics*, 3(3), 261–271.
- Maggioni, M. A., & Uberti, T. E. (2007). International networks of knowledge flows: An econometric analysis. In K. Frenken (Ed.), *Applied evolutionary economics and economic geography* (pp. 230–255). Cheltenham, UK: Edward Elgar.
- Matthiessen, C. W., & Schwarz, A. W. (1999). Scientific centres in Europe: An analysis of research strength and patterns of specialization based on bibliometric indicators. *Urban Studies*, 36(3), 453–477.
- Matthiessen, C. W., Schwarz, A. W., & Find, S. (2002). The top-level global research system, 1997–99: Centres, networks and nodality. An analysis based on bibliometric indicators. *Urban Studies*, 39(5–6), 903–927.
- May, R. M. (1997). The scientific wealth of nations. *Science*, 275(7), 793–796.
- Martin, R. (1999). The new 'geographical turn' in economics: Some critical reflections. *Cambridge Journal of Economics*, 23, 65–91.
- Massey, D. (2004). *For space*. London: Sage.
- Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigations*. Chicago: University of Chicago Press., pp. 267–278
- Narin, F., & Carpenter, M. P. (1975). National publication and citation comparisons. *Journal of American Society for Information Science*, 26, 80–93.
- Narin, F., Stevens, K., & Whitlow, E. S. (1991). Scientific cooperation in Europe and the citation of multinationally authored papers. *Scientometrics*, 21(3), 313–323.
- Openshaw, S. (1984). *The modifiable areal unit problem*. Norwich: Geo Books.
- Ponds, R. (2009). The limits to internationalisation of scientific research collaboration. *Journal of Technology Transfer*, 34, 76–94.
- Ponds, R., van Oort, F. G., & Frenken, K. (2007). The geographical and institutional proximity of research collaboration. *Papers in Regional Science*, 86, 423–443.
- Price, D. de Solla. (1963). *Little science, big science*. New York and London: Columbia University Press.
- Rallet, A. (1993). Choix de proximité et processus d'innovation technologique, Revue d'Economie Régionale et Urbaine, numéro special. *Economie de Proximités*, 3, 365–386.
- Rallet, A., & Torre, A. (1999). Is geographical proximity necessary in the innovation networks in the era of global economy? *GeoJournal*, 49(4), 373–380.
- Rychen, F., & Zimmermann, J.-B. (2008). Clusters in the global knowledge-based economy: Knowledge gatekeepers and temporary proximity. *Regional Studies*, 42(6), 767–776.
- Scherngell, T., & Barber, M. (this issue). Spatial interaction modelling of cross-region R&D collaborations. Empirical evidence from the 5th EU Framework Programme. *Papers in Regional Science*.
- Schubert, A., Glänzel, W., & Braun, T. (1989). Scientometric datafiles. A comprehensive set of indicators on 2649 journals and 96 countries in all major science fields and subfields 1981–1985. *Scientometrics*, 16(1–6), 3–478.
- Shapin, S. (1984). Pump and circumstance: Robert Boyle's literary technology. *Social Studies of Science*, 14(4), 481–520.
- Shapin, S. (1998). Placing the view from nowhere: Historical and sociology problems in the location of science. *Transactions of the Institute of British Geographers NS*, 23, 5–12.
- Singh, J. (2007). *External collaboration, social networks and knowledge creation: Evidence from scientific publications*. INSEAD, mimeo.
- Skupin, K. (2009). Conceptualizing science: Implications for knowledge domain visualization. *Journal of Informetrics*, 3(3), 233–245.
- Stephan, P. E. (1996). The economics of science. *Journal of Economic Literature*, 34(3), 1199–1235.
- Stephan, P. E., & Levin, S. G. (2001). Exceptional contributions to US science by the foreign-born and foreign-educated. *Population Research and Policy Review*, 20, 59–79.
- Stichweh, R. (1996). Science in the system of world society. *Social Science Information*, 35(2), 327–340.
- Taylor, P., Hoyler, M., & Evans, D. M. (2007). A geohistorical study of the rise of modern science: Career paths of leading scientists 1500–1900. *GAWC Research Bulletin* 233.
- Tijssen, R. J. W. (2008). Are we moving towards an integrated European Research Area? Some macro-level bibliometric perspectives. *COLLNET Journal of Scientometrics and Information Management*, 2, 19–25.
- Torre, A. (2008). On the role played by temporary geographical proximity in knowledge transmission. *Regional Studies*, 42(6), 869–889.
- Uzzi, B. (1997). Social structure and competition in interfirm networks: The paradox of embeddedness. *Administrative Science Quarterly*, 42(1), 35–67.
- Van Dijk, J., & Maier, G. (2006). ERS conference participation: Does location matter? *Papers in Regional Science*, 85(4), 483–504.

- Wagner, C., & Leydesdorff, L. (2005). Mapping global science using international co-authorships: A comparison of 1990 and 2000. *International Journal of Technology and Globalization*, 1(2), 185–208.
- White, H. D., Wellman, B., & Nazer, N. (2004). Does citation reflect social structure? Longitudinal evidence from the “Globenet” interdisciplinary research group. *Journal of the American Society for Information Science and Technology*, 55(2), 111–126.
- Zhou, P., & Leydesdorff, L. (2006). The emergence of China as a leading nation in science. *Research Policy*, 35, 83–104.
- Ziman, J. M. (1994). *Prometheus bound. Science in a dynamic steady state*. Cambridge and New York: Cambridge University Press.