[This is the author post-print version of the final paper available at http://link.springer.com/article/10.1007/s11192-015-1819-8]

Please, cite as:
Fernández, A., Ferrándiz, E., & León, M. D. (2016). Proximity dimensions and scientific collaboration among academic institutions in Europe: The closer, the better?. *Scientometrics*, 106 (3), pp. 1073-1092.

Proximity Dimensions and Scientific Collaboration among Academic Institutions in Europe: The Closer, the Better?

FERNÁNDEZ, A., FERRÁNDIZ, E.*, LEÓN, M.D.

Department of Economics, Facultad de Ciencias Económicas y Empresariales de Cádiz. University of Cadiz, Cadiz, Spain.

ABSTRACT: The main objective of this paper is to examine the effect of various proximity dimensions (geographical, cognitive, institutional, organizational, social and economic) on academic scientific collaborations (SC). The data to capture SC consists of a set of co-authored articles published between 2006 and 2010 by universities located in EU-15, indexed by the Science Citation Index (SCI Expanded) of the ISI Web of Science database. We link this data to institution-level information provided by the EUMIDA dataset. Our final sample consists of 240,495 co-authored articles from 690 European universities that featured in both datasets. Additionally, we also retrieved data on regional R&D funding from Eurostat. Based on the gravital equation, we estimate several econometrics models using aggregated data from all disciplines as well as separated data for *Chemistry & Chemical Engineering*, *Life Sciences* and *Physics & Astronomy*. Our results provide evidence on the substantial role of geographical, cognitive, institutional, social and economic distance in shaping scientific collaboration, while the effect of organizational proximity seems to be weaker. Some differences on the relevance of these factors arise at discipline level.

Keywords: scientific collaboration, co-authorship, proximity dimensions, gravital equation

Introduction

In recent decades, there has been an increasing trend towards scientific collaboration (Gazni et al., 2012; Waltman et al., 2011). Scientific collaboration (SC) is assumed to enhance the quality of the research for a number of emerging benefits, widely discussed in the literature (Franceschet and Costantini, 2010; Katz and Martin, 1997; Sonnenwald, 2007): (1) it brings together complementary knowledge and expertise from different sources to solve complex problems, and to the creation of new knowledge or technologies; (2) it usually implies a higher internal quality control than single authored papers; (3) it enhances learning and the acquisition of skills from partners for future research activities; (4) it creates social networks and facilitates knowledge diffusion, not only among individuals but also in the enhancement of cross-fertilization across disciplines. From an economic viewpoint, SC also provides benefits that include access to a wide variety of resources and new foundations or instruments. These benefits, together with the well-known role of knowledge creation and diffusion as the main sources for sustainable economic growth in the long run (Foray, 2004; Romer, 1990), have shaped the European Policy. The European government initiative aimed to convert Europe into "the most competitive and dynamic knowledge-based economy in the world" (CEC, 2000a), giving priority to increased investment in knowledge and innovation and to give Europe a new "fifth liberty", the free circulation of knowledge, in order to construct a European Research Area (CEC, 2000b). In the knowledge-based economy, universities are called upon to play a role in regional economic growth due to their direct contribution to the generation of knowledge, which may, ultimately, impact on industry and economy through spillover effects (e.g. Abramovsky et al., 2007;

* Correspondance address: Esther Ferrándiz. Department of Economics, Facultad de Ciencias Económicas y Empresariales de Cádiz.University of Cadiz. Enrique Villegas Vélez, 2, 11002 Cadiz, Spain. Email: esther.ferrandiz@uca.es.

Acosta et al., 2011b; Audretsch et al., 2005).

The extant empirical literature on proximity and collaboration often has two limitations. First, although Boschma (2005) in his influential contribution identified five notions of proximity (geographical, cognitive, institutional, organizational, and social), most research has focused on one proximity dimension, namely geographical distance (for reviews, see Frenken et al., 2009; Heringa et al., 2014). However, as noted by Frenken et al. (2009), it is important to examine multiple proximity dimensions simultaneously because they are often correlated, which implies that the effect of a certain form of proximity can only be properly determined when controlling for the others. The recent study by Plotnikova and Rake (2014) provides, to the best of our knowledge, the only attempt to analyse all proximity dimensions at one time. However, their paper focused on Pharmaceutical research and was carried out at country level. Second, previous research has frequently drawn upon data at country or regional level. This paper aims to fill this gap by providing a comprehensive analysis of proximity using mainly data at university level. The objective of this paper is twofold: first, to identify those universities more prone to SC using a sample of 690 universities in Europe; and second, to jointly examine the effects of different proximity dimensions on academic SC.

In this paper, we extend the empirical literature in several ways. Firstly, following Boschma (2005), this paper mainly aims to examine the effect of geographical, cognitive, institutional, organizational and social distance on SC among academic institutions. Besides, we also analyse the role of economic distance, which has recently raised the interest of some authors (Acosta et al., 2011a; Hwang, 2008; Sonnenwald, 2007), inspired by the centre-periphery hypothesis. Secondly, we focus on SC at university-level, in which knowledge creation and diffusion is the primary mission. This data allows us to rank universities in terms of collaboration and undertake a more in-detailed analysis[†]. Thirdly, we provide a joint analysis of SC in all disciplines included in the Science Citation Index (SCI) of the WoS, and a separate analysis for *Chemistry & Chemical Engineering, Life Sciences and Physics & Astronomy* in order to examine whether there are differences across disciplines[‡]. For this purpose, we use an original dataset containing information on 240,495 scientific papers in Science and Engineering indexed in the Science Citation Index (SCI) provided by the ISI Web of Science (WoS) and co-authored among academics from different universities. Our analysis includes 690 universities from EU-15 countries except France and Denmark, for which EUMIDA does not provide data. The methodology relies on a descriptive analysis of co-authored publications and an econometric model based on the gravital equation to estimate the impact of different notions of proximity in promoting academic SC in Europe.

The remainder of the paper is organized as follows. In Section 2, we review the relevant literature. Section 3 establishes the methodology. Section 4 describes the data. Section 5 provides the results. The main conclusions and policy implications are offered at the end of the paper.

2. Literature review

The French School of Proximity Dynamics was a pioneer in the consideration of other notions of proximities beyond the geographical (Carrincazeaux et al., 2008; Rallet and Torre, 1999; Torre and Gilly, 2000). Drawing upon this line of research, Boschma (2005), from a theoretical point of view, identified five kinds of proximities: geographical, cognitive, institutional, organizational, and social. Recent research has also highlighted the relevance of economic differences as an explanatory factor of SC (Acosta et al., 2011a; Hwang, 2008; Sonnenwald, 2007).

The empirical evidence has largely shown that geographical distance among actors hinders SC because face-to-face interactions, which facilitate knowledge flows and tacit knowledge share, become costly as distances increase (e.g. Hoekman et al., 2010; Katz, 1994; Ponds et al., 2007). Although, motivated by the development of ICT, some authors claimed the death of distance as a hindrance, the empirical contribution

[†] In this paper, we use the term universities in a broad sense, to include all Higher Education Institutions.

[‡] We selected these disciplines because they are among the disciplines most prone to collaborate according to our descriptive analysis.

of Hoekman et al. (2010) showed that physical distance does still impede research collaboration, with no evidence of a declining effect in the period 2000-2007.

Cognitive proximity, i.e. the degree of shared knowledge base between organizations, facilitates knowledge transfer by contributing to build absorptive capacity, that enables actors to identify, acquire, understand and exploit knowledge available from others (Cohen and Levinthal, 1990). Nevertheless, recent studies have shown a certain degree of cognitive distance as a potential source of complementarities in order to improve knowledge bases (Gilsing et al., 2008; Nooteboom et al., 2007), the challenge is to collaborate with actors who provide access to heterogeneous sources of knowledge in order to generate sufficiently diverse complementarities while ensuring the absorption capacity enabled by a shared knowledge base.

Institutional proximity is defined by the degree of similarity in formal institutions, such as laws and rules, and informal institutions, like cultural norms and habits, which may enable knowledge flows by facilitating trust and reducing uncertainty and risks (Boschma, 2005; Boschma and Frenken, 2009). Barjak and Robinson (2008) consider the different countries of origin and cultural diversity of academic research teams related to their research performance, concluding a moderate level of cultural diversity (which, according to our framework, would be institutional distance). Empirical literature has usually addressed this variable by using dummies to capture if collaborators belong to the same region or country. For example, Hoekman et al. (2010) found that SC is more likely to occur within the same sub-national region, within the same country and within the same linguistic area. Lakitan et al. (2012) use data on scientific productivity and co-authored papers to conclude that Indonesian researchers at public R&D institutions showed a higher degree of dependency with their foreign partners than academicians at universities. Hennemann et al. (2012) look in detail at the spatial structures of scientific activity (epistemic communities) showing that intra-country collaboration is more likely to occur than international collaboration.

Organizational proximity is related to the extent to which relations are shared in an organizational arrangement (micro-level), either within or between organizations involving the rate of autonomy and degree of control that can be exerted in organizational arrangements (Boschma, 2005). Thus, it can be understood as a variable, capturing organizations that share the same or similar regulation and routines at a micro-level. In that sense, a certain degree of organizational proximity is desirable to reduce uncertainty and opportunism in knowledge creation within and between organizations. In research collaboration literature, this dimension has often been captured by a variable noting whether partners share the same institutional arrangement, for example, belong to the same corporation (Balland, 2011). In this research, difficulties of considering organizational proximity, in Boschma´s sense, arise due to the absence of hierarchical relations among universities. However, they cannot be considered homogenous organizations because research institutions differ in their norms, structure, size and strategy (Cummings and Kiesler, 2007; Mowery and Sampat, 2004).

Social proximity, i.e. socially embedded relations based on friendship, kinship and past experience between agents at the micro-level, is expected to stimulate interactive learning due to trust and commitment (Boschma, 2005). It is commonly accepted practice to measure social proximity based on prior collaborations or previous research experiences (Hoekman et al., 2012; Hong and Su, 2013; Petruzzelli, 2011). Niedergassel and Leker (2011) analysed R&D cooperation projects of university professors and procure evidence on the relevance of factors like trust, dependency of partners and strength of ties.

Differences in economic resources among geographic areas (economic distance) may determine the spatial patterns in SC, as derived from the centre-periphery hypothesis (Schott, 1998; Schubert and Sooryamoorthy, 2010) applied to research collaboration (Acosta et al., 2011a). According to this literature, scientists in peripheral countries are willing to collaborate with core countries to gain access to resources, while core areas seeking complementarities (Hwang, 2008; Sonnenwald, 2007). Acosta et al. (2011a) using data on a sample of co-authored papers among regions in EU-15, put forward a gravity equation that includes economic distance. Their results showed that having similar levels of resources devoted to R&D play a positive role in facilitating SC. As they recognise, this finding is not in line with the centre-periphery hypothesis applied to SC, by which increasing levels of collaboration among core-periphery regions would be expected to benefit from complementarities. However, Acosta et al. (2011a) argue that this result is not strange because the greater the amount of resources, the greater the opportunities for mobility and attendance to international conferences, which encourage the establishing and reinforcing of personal contacts for future collaborations.

3. Methodology

In order to estimate the influence of different proximity dimensions on university SC, we use cross-sectional data and suggest a gravity model where SC between university *i* and *j* is a function of characteristics of the origin *i*, characteristics of the destination *j*, and some measurement of distance between both universities. The gravity equation is based on the original gravity equation defined by Newton, in which the gravitational force that attracts objects *i* and *j* is directly related to the mass of *i*, the mass of *j* and is inversely proportional to the distance separating them. It has been largely used in Economics to examine trade flows across regions and countries (e.g. see Costantini and Mazzanti, 2012; Kahouli and Maktouf, 2014). This equation has also been extensively used in the literature on research collaboration among European regions (see, for examples, Hoekman et al., 2009; Hoekman et al., 2010; Maggioni and Uberti, 2009; Ponds et al., 2007; Scherngell and Barber, 2009a; Scherngell and Barber, 2009b; Scherngell and Hu, 2010).

In the analysis of count data, such as the number of co-authored papers, estimates obtained from linear regression can be inconsistent, inefficient, and biased (Amano and Fujita, 1970; Long, 1997). Therefore, count models are preferred to estimate our gravity model. Since our dependent variable presents both overdispersion (i.e. its variance is greater than its mean), and a large number of zeros because many university pairings have no co-authored papers, a zero-inflated negative binomial regression is recommended (Cameron and Trivedi, 2009; Cameron and Trivedi, 2013; Long, 1997). The zero-inflated negative binomial regression model comprises a two stage process: first, a negative binomial regression estimates the number of co-authored papers between a pair of universities given that each of them have at least one publication; second, a logistic regression estimates the probability of a zero count (i.e. no co-publications). Two coefficients are thus obtained for each specified predictor. Zero-inflated models consider two sources of zero observations, "true zeros" that are part of the underlying sampling distribution and "excess zeros" that cannot score anything other than zero. In our sample, it may be that a pair of universities co-authored zero papers because one or both of them had no publications; other pairs of universities may own publications but report zero because they did not collaborate.

$$\Pr(F_{ij}) = \begin{cases} \lambda_{ij} + (1-\lambda_{ij})h(F_{ij}=0,\theta \mid X) & for\ F_{ij}=0 \\ (1-\lambda_{ij})h(F_{ij},\theta \mid X) & for\ F_{ij}=1,2,... \end{cases}$$

where $h(F_{ij},\theta \mid X)$ is the negative binomial density with mean $\exp(X,\beta)$, dispersion parameter α, and $\theta = (\beta'\alpha)'$. Here, λ is a zero-inflation parameter representing the proportion of observations with a strictly zero count $(0<\lambda<1)$ as determined by a logit model on all (or several) observed explanatory variables: $\lambda_{ij} = \exp(X\varphi)/1-\exp(X\varphi)$. The dependent variable F_{ij}= **Asc$_{ijt}$** represents the counts of academic scientific collaborations between university *i* and university *j* for the period 2006-2010. We use co-publications to measure collaboration, as it is often used to proxy scientific collaboration (for reviews, see Katz and Martin, 1997; Laudel, 2002; Melin and Persson, 1996). *X* is a vector including the independent variables described in Table 1 as suggested by the literature. Details on the estimation procedure of the ZINB model can be found in Cameron and Trivedi (2013).

Despite our theoretical equation collaboration between a pair of universities depending on the "mass" of publications of each of them, reverse causality is also possible due to the potential effect of collaboration on scientific productivity (see Abramo et al., 2009; Lee and Bozeman, 2005 for reviews). Thus, to avoid endogeneity, the mass of publications of each university refers to the period 2001-2005, while collaborations refer to the period 2006-2010 (see Table 1). Since the variable Specij is based on publication data, it has also been lagged and captures information for the period 2001-2005. Finally, R&D expenditures have been included with a two-year lag because it is to be expected that economic resources take time to be reflected in scientific output.

Note that the description of the variables refers to those in the model using data for all 12-disciplines. For separated regressions by disciplines, the dependent variable, the mass of publications and previous collaborations refer to the respective counts for that specific discipline. At the discipline level, $Specdist_{ij}$ represents the dissimilarity in specialization in a certain discipline. Since it is not possible to calculate it as a correlation coefficient, it was calculated by following a different procedure for models by disciplines: first, for each university we calculated the share of publications in each discipline over its total number of publications; second, we obtained the absolute difference in this indicator for each pair of universities.

Table 1. Description of the explanatory variables in Model 1

Variable	Description
MASS OF PUBLICATIONS	
Pub_i	No. of publications in university *i* in the period 2001-2005 (variable in logarithms)
Pub_j	No. of publications in university *j* in the period 2001-2005 (variable in logarithms)
GEOGRAPHICAL DISTANCE[(1)]	
$Geodist_{ij}$	Geographical distance between universities *i* and *j*, in kilometres.
COGNITIVE PROXIMITY[(2)]	
$Specdist_{ij}$	Correlation index calculated as Paci and Usai (2009) for the 12 discipline composition of scientific papers in university *i* and university *j* for the period 2001-2005. This coefficient ranges between 0 (minimum distance, identical specialization) and 1 (maximum distance).
INSTITUTIONAL PROXIMITY	
Reg_{ij}	Dummy variable, which takes value 1 when universities *i* and *j* are in the same region, 0 otherwise.
Nat_{ij}	Dummy variable, which takes value 1 when universities *i* and *j* are in the same country, 0 otherwise.
ORGANIZATIONAL PROXIMITY:	
$Eduprox_{ij}$	Correlation coefficient between the 9 education fields, as identified in EUMIDA, corresponding to university *i* and university *j*
$Staffdist_{ij}$	Absolute difference in total staff of universities *i* and *j* (variable in logarithms)
SOCIAL PROXIMITY	
$Prior_{ij}$	Dummy variable which takes value 1 if universities "i" and "j" have collaborated for the five-years previous period 2001-2005.
ECONOMIC PROXIMITY [(3)]	
$R\&Ddist_{ij}$	Absolute difference in the average Higher education R&D expenditures as % of the GDP in 2004-2008 between regions where universities *i* and *j* are located

[(1)] Geographical distance was alternatively calculated as the euclidean distance between regions *i* and *j* where universities are located, obtaining similar estimation results. [(2)] Cognitive proximity was also calculated as the correlation coefficient between the 12-field composition of scientific papers in university *i* and university *j* for the period 2001-2005 but the results of our estimations were not significantly different. [(3)]R&Ddist also captures differences in the Higher Education funding structure.

It is worth pointing out that organizational proximity measures are not completely perfect because they attempt to capture a complex phenomenon and its measurement is a difficult task. A possible alternative could have been to identify any collaboration between the same departments in each university. However, as we are excluding intra-university collaboration, this alternative was not possible here. Then we chose the differences in size and educational profiles as factors proxying organizational characteristics that may their culture or orientation. Additionally, due to data availability limitations, it has not been possible to include R&D funding information at the level of institutions: to the best of our knowledge, Data collection 2, which includes information on funding for universities, has not been made public. Then, to complete our analysis, we have included the amount of R&D expenditure in the region in which the university is located.

4. Data

The empirical data used in this paper consists of a set of 240,495 articles co-authored by scientists affiliated to different universities and published in journals indexed by the Science Citation Index Expanded (SCI Expanded) provided by the Thomson Reuters Web of Science (WoS). Our period of analysis has been 2006-2010. This dataset was built following a similar procedure to Acosta et al. (2011a) and Acosta et al. (2014). Since our focus is at the university level, we had to harmonize the name variations of universities, mainly stemming from the use of the native versus the English name or the use of different acronyms. Then, papers were assigned to universities following the full-counting process (crediting 1 publication to each co-author institution). Next, data on academic collaboration was placed into a symmetrical matrix containing all co-publications between university *i* and university *j* and therefore excluding intra-university collaboration. Publications were classified into 12 scientific disciplines following the CWTS classification (Tijssen and van Leeuwen, 2003) and Torres-Salinas et al. (2011), again using the full counting method for those publications included in journals related to more than one discipline. In a further step, we matched this dataset with the EUMIDA dataset (Data Collection 1) in order to get information about organizational characteristics of the universities. EUMIDA data is the result of an initiative of the European Commission to provide a complete census of European universities and provides information at the university level including organizational details such as education offered and staff employed[§]. Our final sample includes 690 universities that were present in both datasets. Consequently, there are potentially (690*689) ÷ 2 = 237,705 collaboration links (observations). However, only 34,095 pairs have at least one paper in collaboration over the period of analysis. As mentioned in the previous section, this makes appropiate the use of zero-inflated negative binomial regression. Additional information about regional R&D expenditures was extracted from Eurostat.

Descriptive analysis

For the purpose of contextualization, we present a brief description of academic scientific publications and co-publications that enable us to provide a picture of the temporal evolution of scientific output in our sample and its distribution across universities and disciplines. A summary of the main statistics is reported in Table 2. Following the full-counting procedure, the total number of academic papers during 2006-2010 was 1,283,632, of which 543,446 have been published in collaboration with other universities. From the 690 universities that comprised the original sample, 56 do not have any co-publication in the period. On average, each university has 1,860.34 academic publications, of which 787.60 are co-authored with researchers affiliated to other universities.

Regarding the temporal evolution, academic publications have increased by 24.92% during the period of analysis, while academic collaborations have increased by 38.10% over the same period. This is also naturally reflected in an increasing share of collaborations, which rose from 40.47% of total publications in 2006 to 44.75% in 2010. These results confirm the positive trend towards collaboration in scientific research. The Gini coefficients included in Table 2 also reveal a high level of concentration of scientific output in a few universities. The remaining concentration indexes in Table 2 lead to the same conclusion. For example, the value of the C25 index suggests that just twenty-five universities account for 25.54% of papers and 25.30% of co-publications with academics in other universities.

Table 3 ranks the top-15 universities in terms of intensity of scientific collaborations (i.e. co-publications). As can be observed in this table, the top 15 collaborative universities accounted for 17.08% of total number of co-publications, which confirms, similar to the case of publications, that the exchange of scientific knowledge is highly concentrated among a few universities. By normalizing the data according to the size of the staff, there are notable changes in the rank.

Table 4 shows the top 15 universities, which are more prone to collaboration (threshold: 1,000

[§] A description of data and the collection procedure is provided in EUMIDA. 2010. Feasibility Study for Creating a European University Data Collection [Contract No. RTD/C/C4/2009/0233402].
Data collection 1 is available at http://ec.europa.eu/research/era/areas/universities/universities_en.htm. (Accessed at 18/10/2012). Data Collection 2, which contains more detailed data, was not available to us at the time of this research.

publications). The data shows that Italian and British universities are the most likely to collaborate, with 7 and 4 universities respectively within the top 15. This ranking also includes universities from Belgium, Finland, Germany and Spain. Altogether, the results from Table 2, 3 and 4 confirms the importance of analysing the factors shaping collaboration at an institutional level, since there are strong differences across universities in scientific output and its propensity to collaborate.

Table 2. Descriptive statistics of academic scientific publications and collaborations

		2006	**2007**	**2008**	**2009**	**2010**	**06-10**
Publications (Pub)	No. Pub.	226,940	245,821	257,321	270,066	283,484	1,283,632
	Mean	328.90	356.26	372.93	391.40	410.85	1,860.34
	Max.	3,651	3,887	4,202	4,305	4,756	20,530
	Min.	0	0	0	0	0	0
	Desv. Est	557.05	600.34	625.88	653.31	688.50	3115.71
	C. Var(1)	1.69	1.69	1.68	1.67	1.68	1.68
	Coef. Gini(2)	0 .74	0.74	0.73	0.73	0 .73	0.73
	C25(3)	25.77	25.69	25.54	25.32	25.42	25.54
	C50(4)	41.77	41.81	41.77	41.56	41.57	41.69
	C100(5)	64.33	64.04	63.57	63.48	63.52	63.77
Collaborations (Col)	No. Col.	91,847	100,175	108,067	116,512	126,845	543,446
	Mean	133.11	145.18	156.62	168.86	183.83	787.60
	Max.	1,396	1,524	1,652	1,802	2,040	8,386
	Min.	0	0	0	0	0	0
	Desv. Est	223.33	242.90	261.13	279.53	307.18	1,309.07
	C. Var(1)	1.67	1.67	1.67	1.66	1.67	1.66
	Coef. Gini(2)	0.74	0.73	0.73	0.72	0.73	0.73
	C25(3)	25.33	25.39	25.24	25.19	25.34	25.30
	C50(4)	41.35	41.70	41.44	41.17	41.48	41.43
	C100(5)	63.94	63.77	63.39	63.19	63.56	63.55
	Col/pub	40.47	40.75	42.00	43.14	44.75	42.34

(1)Coefficient of variation = Std Dev. /Mean; (2) The Gini coefficient ranges between 0 and 1; the larger the value the higher the level of concentration in publications or collaborations. (3)(4)(5) Concentration indexes of publications and collaborations for the top 25, 50 and 100 universities with the largest number of scientific papers and collaborations, respectively

Source: ISI. Own elaboration.

Table 3. Top 15 universities by collaborations (average value 2006-2010)

	No. Col. (A)	**A/Total A *100 (%)**	**Acum. (%)**	**Col/ Staff (in thousand)**	
Univ. College London (UK)	8,386	1.54	1.54	Univ. of Applied Sciences Schmalkalden (DE)	13.35
The Univ. of Oxford (UK)	8,199	1.51	3.05	Paracelsus Univ. Salzburg (AT)	9.79
Imperial College of Science, Technology and Medicine London (UK)	7,808	1.44	4.49	Swansea Univ. (UK)	4.01
The Univ. of Cambridge (UK)	7,702	1.42	5.91	Univ. of Applied Sciences Karlsruhe (DE)	2.72
Univ. "La Sapienza" Rome (IT)	6,639	1.22	7.13	Cardiff Univ. (UK)	2.65
Ludwig-Maximilians Univ. Munich (DE)	6,265	1.15	9.40	Univ. of Lübeck (DE)	2.50
Univ. of Helsinki (FI)	6,078	1.12	9.40	Univ. of Groningen (NL)	2.43
Univ. of Milan (IT)	5,658	1.04	10.44	London School of Hygiene and Tropical Medicine (UK)	2.06
Univ. of Manchester (UK)	5,397	0.99	11.43	Univ. Vita-Salute San Raffaele (IT)	1.95
Univ. "Federico II" Naples (IT)	5,234	0.96	12.40	Karolinska Institute (SE)	1.66
Karolinska Institute (SE)	5,216	0.96	13.36	Univ. College London (UK)	1.65
Univ. of Bologna (IT)	5,216	0.96	14.32	Univ. of Milano (IT)	1.54
Heidelberg Univ. (DE)	5,055	0.93	15.25	Univ. of Piemonte orientale "A. Avogadro" (IT)	1.48

K.U. Leuven (BE)	4,982	0.92	16.16	International School for Advanced Studies Trieste (IT)	1.42
Univ. of Amsterdam (UVA) (NL)	4,965	0.91	17.08	Leiden Univ. (NL)	1.41
Others	427,230	78.61			
Total	543,446	100			

Source: ISI. Own elaboration

Table 4. Ranking of universities with the highest collaboration intensity (2006-2010)

	No. Pub (A)	**No. Col (B)**	**Collaboration intensity (%) (B/A*100)**
Univ. of Piemonte orientale "A. Avogadro" (IT)	1,656	1,032	62.32
Hasselt Univ. (BE)	1,095	657	60.00
Hannover Medical School (DE)	3,704	2,207	59.58
Politechnic Univ. of Milan (IT)	2,716	1,590	58.54
Univ. Gabriele D'Annunzio (IT)	2,035	1,186	58.28
Univ. of L'Aquila (IT)	2,302	1,333	57.91
Univ. of Jaen (ES)	1,435	830	57.84
Univ. of Hertfordshire (UK)	1,035	597	57.68
Univ. of Tampere (FI)	2,328	1,323	56.83
Univ. of Foggia (IT)	1,184	671	56.67
Univ. of Kent Canterbury (UK)	2,062	1,168	56.64
Univ. of Ferrara (IT)	3,383	1,899	56.13
Univ. College London (UK)	15,226	8,386	55.08
The Open University (UK)	1,747	962	55.07
Univ. of Verona (IT)	2,906	1,600	55.06
Total	1,283,143	543,071	42.32

Source: ISI. Own elaboration.

Table 5 shows the distribution of scientific publications by scientific fields. This analysis is useful to identify the disciplines in which collaboration is more prominent (i.e. plays a more relevant role). Taking base on this, in a further section we estimate a separate econometric model to test for the specific effect of distance on collaboration in these disciplines. The top four scientific fields in terms of publications accounted for 65.16% of the total number of publications. These scientific fields are *Medicine & Biomedicine* (26.14%), *Physics & Astronomy* (14.04%), *Life Science* (13.93%) and *Chemistry & Chemical Engineering* (11.05%). Similarly, these scientific disciplines are those with highest share of collaborations. It is also noticeable that the percentage of collaborations differs among scientific fields, ranging from 28.87%% in *Medicine & Biomedicine* and 14.93 in Life Science, to just a 0.98% in Multidisciplinary and 2.38% in *Pharmacology*. As shown by last column of Table 5, *Multidisciplinary* (49.45%), *Medicine & Biomedicine* (46.11%), *Life Science* (44.75%), *Earth & Environmental Science* (44.42%), *Physics & Astronomy (*42.89%) and *Pharmacology* (42.25%) are scientific fields with collaborations higher than average.

Table 5. Publications and collaborations by scientific discipline (2006-2010)

	Publications		**Collaborations**		
	No. Pub. (A)	**Pub. Share (%)** (A/Total A)	**No. Col.** (B)	**Col. Share (%)** (B/Total B)	**Col. Intensity (%)** (B /A)
Agriculture & Food Sci.	64,456	3.97	22,868	3.38	35.48
Chemistry & Chemical Eng.	179,237	11.05	68,006	10.04	37.94
Earth & Environmental Sci.	118,560	7.31	52,666	7.77	44.42
Enginnering Science	127,982	7.89	41,656	6.15	32.55
Information & Communic. Sci.	51,803	3.19	18,699	2.76	36.10
Life Sci.	225,996	13.93	101,131	14.93	44.75
Material Sci.	73,999	4.56	27,520	4.06	37.19
Mathematics	76,753	4.73	28,960	4.28	37.73
Medicine & Biomedicine	424,076	26.14	195,544	28.87	46.11
Multidisciplinary	13,367	0.82	6,610	0.98	49.45
Pharmacology	38,142	2.35	16,116	2.38	42.25

Physics & Astronomy	227,678	14.04	97,640	14.41	42.89
Total	1,622,049	100	677,416	100	41.76

Source: ISI. Own elaboration.

5. Results

Table 6 shows the descriptive statistics of the variables included in Model 1, i.e. the model including 12 disciplines in our sample and full counting data. Then, it must be noted that the values displayed for these variables may differ when focusing on only one discipline and/or using fractional counting data.

Table 6. Descriptive Statistics

Variable	Mean	Std. Dev	Min	Max
ASC_{ij}	1.67	13.39	0	941
Pub_i	1589.2	2677.3	0	18306
Pub_j	1290.6	2331.5	0	18306
$Geodist_{ij}$	14.03	9.22	0	101.60
$Specdist_{ij}$	0.83	0.09	0.29	1
Reg_{ij}	0.007	0.087	0	1
Nat_{ij}	0.153	0.360	0	1
$Eduprox_{ij}$	0.196	0.376	-1	1
$Staffdist_{ij}$	1909.7	2140.3	0	13938
$Prior_{ij}$	0.109	0.311	0	1
$R\&Ddist_{ij}$	0.279	0.233	0	1

Table 7 shows the results of our estimations using zero inflated negative binomial models. Model 1 displays the results including data for all disciplines. Models 2, 3 and 4 show the estimation results for *Chemistry & Chemical Engineering*, *Life Sciences* and *Physics & Astronomy*, respectively. We have chosen these disciplines because, jointly with *Medicine & Biomedicine*, they have the highest publication and collaboration share[**].

The following conclusions are drawn from the results of Model 1:

- The positive and significant coefficients of the variables capturing the mass of publications of each university indicate that SC increases as the number of publications in each pair of universities rises.
- The negative and significant coefficient of the variable accounting for geographical distance shows that academic SC decays with physical distance.
- Scientific specialization displays a negative and significant coefficient, which suggests that cognitive proximity promotes SC.
- Dummy variables controlling for universities located in the same region and in the same country (institutional proximity) present positive and significant coefficients, suggesting that factors like language, culture and policies foster SC.
- Correlation among education fields does not seem to have a significant effect on the number of scientific collaborations. This may be explained because teaching does not have to be related to research output. The variable accounting for the differences in size of two universities has a negative and significant coefficient. However, the value of this coefficient is rather small. Then, it seems that the two variables related to organizational proximity do not significantly affect SC. This result might be motivated because organizational diversity among academic institutions may not be wide enough. Additionally, as we already mentioned, the indicator we proposed has some limitations.
- Having previous collaborations enhance SC, as shown by the positive and significant coefficient of this variable. As already discussed, this may enable mutual trust and confidence.

** Note that we do not give a detailed analysis for *Medicine & Biomedicine* because some of the publications may be associated with university hospitals, which may or may not have been co-authored by academics. Publications for which it has not been possible to establish a clear link with an academic institution have been excluded from our sample. Thus, our study may underestimate the scientific output in this discipline.

- Negative signs of the variable capturing economic distance indicate that SC is stronger among universities located in regions with similar levels of resources devoted to R&D. Thus, it seems that the centre-periphery hipothesis is not accepted to explain collaboration among universities in Europe, confirming the results obtained by Acosta et al. (2011a) for SC across regions.

Next, we check if these results hold for *Chemistry & Chemical Engineering*, *Life Science* and *Physics & Astronomy* (Model 2, 3 and 4). For these disciplines, the number of publications of each university, the binary variables capturing if universities *i* and *j* belong to the same region or country (*reg, nat*) and if they have previous collaborations in the period 2001-2005 (*prior*), remain with a positive and significant coefficient. These results confirm that the mass of publications, institutional and social proximity promotes SC. Differences in staff are also relevant in these disciplines, displaying a negative sign. For the rest of variables, results differ across disciplines. Geographical distance is relevant for *Chemistry & Chemical Engineering* and *Life Science*, but not for *Physics & Astronomy*. Conversely, the difference in scientific specialization only has a significant effect on *Physics & Astronomy*. Proximity in educational profiles is only relevant for explaining collaboration in *Chemistry & Chemical Engineering*, while distance in R&D resources only becomes relevant for *Life Science*. From these results, it is clear that the influence of the mass of publications, institutional and social proximity enhances SC independently of the disciplines under examination. However, the effect on SC of geographical, cognitive, organizational and economic distance differs across scientific fields.

Table 7. Estimation results from ZINB regressions (full counting of publications)

	MODEL 1			**MODEL 2**			**MODEL 3**			**MODEL 4**		
	12 disciplines			**Chemistry & Chemical Eng.**			**Life Sciences**			**Physics & Astronomy**		
	Coeff.	Std. Err	Sig.	Coeff.	Std. Err	Sig.	Coeff.	Std. Err	Sig.	Coeff.	Std. Err	Sig.
Constant	-3.059	0.143	***	-5.372	0.140	***	-5.494	0.138	***	-4.837	0.160	***
Pub$_i$	0.462	0.006	***	0.523	0.012	***	0.476	0.012	***	0.411	0.014	***
Pub$_j$	0.406	0.007	***	0.533	0.013	***	0.461	0.012	***	0.416	0.014	***
Geodist$_{ij}$	-0.015	0.001	***	-0.019	0.002	***	-0.023	0.002	***	0.004	0.002	*
Specdist$_{ij}$	-3.648	0.133	***	-0.063	0.175		-0.098	0.174		-0.960	0.146	***
Reg$_{ij}$	1.313	0.045	***	1.592	0.066	***	1.158	0.057	***	1.250	0.078	***
Nat$_{ij}$	1.406	0.022	***	1.338	0.036	***	1.170	0.034	***	0.886	0.041	***
Eduprox$_{ij}$	0.030	0.020		0.143	0.033	***	-0.002	0.030		-0.021	0.036	
Staffdist$_{ij}$	-0.013	0.006	**	-0.061	0.011	***	-0.044	0.010	***	-0.063	0.012	***
Prior$_{ij}$ (1)	0.867	0.020	***	-	-	-	0.801	0.029	***	1.052	0.032	***
R&Ddist$_{ij}$	-0.129	0.031	***	-0.077	0.055		-0.224	0.050	***	0.016	0.058	
Inflated (Logit)												
Constant	6.424	0.307	***	9.367	0.222	***	7.234	0.262	***	8.927	0.277	***
Pub$_i$	-0.547	0.013	***	-0.851	0.018	***	-0.665	0.022	***	-0.731	0.022	***
Pub$_j$	-0.530	0.012	***	-0.786	0.020	***	-0.571	0.022	***	-0.651	0.023	***
Geodist$_{ij}$	0.009	0.003	***	0.029	0.004	***	0.018	0.004	***	0.016	0.004	***
Specdist$_{ij}$	1.401	0.299	***	1.967	0.257	***	0.502	0.347		0.952	0.282	***
Reg$_{ij}$	-1.385	0.170	***	-1.663	0.175	***	-0.951	0.193	***	-1.623	0.271	***
Nat$_{ij}$	-1.351	0.063	***	-1.909	0.074	***	-1.130	0.086	***	-1.397	0.093	***
Eduprox$_{ij}$	-1.136	0.052	***	-0.073	0.061		0.073	0.075		-0.175	0.076	**
Staffdist$_{ij}$	0.004	0.016		-0.005	0.019		-0.007	0.023		-0.050	0.023	**
Prior$_{ij}$ (1)	-1.928	0.093	***	-	-	-	-1.749	0.137	***	-1.946	0.117	***
R&Ddist$_{ij}$	-0.107	0.082		-0.302	0.098	***	-0.190	0.121		0.283	0.115	**
Lnalpha	0.025	0.015	*	0.303	0.023	***	-0.199	0.029	***	0.365	0.024	***
Alpha	1.025	0.015		1.355	0.031		0.819	0.024		1.440	0.035	
LR-test	29239.1		***	7998.91		***	10867.3		***	6146.7		***
Likelihood-ratio test alpha	1.5e+05		***	5.9e+04		***	2.0e+04		***	5.0e+04		***
Vuong test	18.21		***	14.64		***	14.87		***	15.11		***
N. obs. (2)	132045		89463				85671			81709		

**p<0.10, **p<0.05, ***p<0.01*

(1) Model 2 does not include prior collaboration because ZINB models could not be estimated due to perfect correlation of the zero values in this variable and the zero values in the dependent variable, i.e. universities that did not collaborate in 2001-05 did not collaborate in the 2006-10 period either (2) Regressions were estimated with a different number of observations due to the inclusion of variables presenting missing data for some observations, especially for Higher Education R&D expenditures.

Robustness checks
To test the robustness of our results, we first apply the fractional counting method to publications and then estimate Tobit regressions. In fractional counting, a paper co-authored by researchers affiliated to different universities is weighted by the number of universities in the paper. For example, in an article co-authored by researchers from two universities, each university is credited 1/2. To take into consideration that the new dependant variable is a fractional count, we estimate Tobit models where we consider zero collaborations as left censoring of the distribution (see e.g. Wooldridge, 2002 for details about the Tobit model). Apart from this change in the dependant variable, we also apply fractional count to the mass of publications. The rest of explanatory variables are the same as described above.

The results of Model 1 and Model 5 (Table 7 and 8) lead to the same conclusions. Thus, confirming the robustness of our estimations for the Models using data from 12 disciplines. Some differences arise in the analysis per discipline. Geographical distance and specialization proximity turn to play a relevant role for the three disciplines. Proximity in education fields and R&D distance becomes significant to explain collaboration in *Physics & Astronomy*. The significance of the rest of the coefficients remains unchanged.

Table 8. Estimation results from Tobit regressions (fractional counting of publications)

	Model 5			**Model 6**			**Model 7**			**Model 8**		
	12 disciplines			**Chemistry & Chemical Eng.**			**Life Sciences**			**Physics & Astronomy**		
	Coeff.	Std. Err	Sig.	Coeff.	Std. Err	Sig.	Coeff.	Std. Err	Sig.	Coeff.	Std. Err	Sig.
Constant	-39.370	0.946	***	-26.372	0.404	***	-18.511	0.257	***	-21.346	0.324	***
Pub_i	3.794	0.041	***	1.827	0.031	***	1.459	0.021	***	1.535	0.025	***
Pub_j	3.539	0.038	***	1.836	0.033	***	1.374	0.021	***	1.544	0.025	***
Geodistij	-0.083	0.009	***	-0.069	0.007	***	-0.052	0.004	***	-0.016	0.005	***
$Specdist_{ij}$	-24.381	0.912	***	-2.696	0.480	***	-0.858	0.314	***	-2.706	0.299	***
Reg_{ij}	18.013	0.407	***	9.458	0.270	***	5.140	0.168	***	5.587	0.208	***
Nat_{ij}	11.14	0.182	***	4.003	0.131	***	3.289	0.080	***	3.107	0.097	***
$Eduprox_{ij}$	0.262	0.154	*	0.464	0.112	***	-0.078	0.068		0.205	0.082	**
$Staffdist_{ij}$	-0.203	0.046	***	-0.090	0.034	***	-0.107	0.021	***	0.059	0.025	**
$Prior_{ij}$ [(1)]	4.581	0.147	***	7.604	0.106	***	2.225	0.065	***	3.114	0.076	***
$R\&Ddist_{ij}$	-0.771	0.243	***	0.054	0.183		-0.343	0.109	***	-0.392	0.134	***
Log likelihood	-109011			-42703			-37765			-35037		
LR-test	61310.2		***	35246.7		***	31586.5		***	28416.7		***
Pseudo R^2	0.2195			0.2921			0.2949			0.2885		
Sigma	11.394	0.050		5.887	0.039		3.63	0.025		4.16	0.030	
Left-censored observations	106982			78175			79729			71977		
Uncensored observations	25063			11288			10942			9732		
N. obs.	132045			89463			85671			81709		

**p<0.10, **p<0.05, ***p<0.01*

6. Conclusions

The objective of this paper was twofold: first, to identify those universities more prone to scientific collaboration using a sample of 690 universities in Europe; and second, to jointly examine the effects of several proximity dimensions on academic SC. For this purpose, the methodology of this paper is based on a descriptive analysis of co-authored papers among academics affiliated to different universities and an econometric model. The results from the descriptive analysis suggest a strong concentration of scientific output within few universities, with 15 universities accounting for 17.08% of total number of co-publications. There are also differences in the intensity of collaborations (calculated as the share of co-publications divided by publications) across universities and scientific disciplines. For example, *Chemistry & Chemical Engineering*, *Life Sciences, Medicine & Biomedicine* and *Physics & Astronomy* show the greatest collaboration intensity.

In a further step, we put forward a gravital equation to explain collaboration as a function of the mass of each university and different measures proximity: geographical, cognitive, institutional, organizational, social and economic distance. The results from our ZINB regressions confirm that, as expected, the larger the mass of publications, the greater its propensity to subsequently collaborate. In line with previous literature at other scale of analysis and other contexts reviewed in this paper, we found that geographical distance hinders SC; while similarities in scientific specialization (cognitive proximity) and institutional proximity encourage SC. With respect to organizational proximity, our results are less clear. Similarities in the educational fields of universities do not seem to have a significant effect on SC, while differences in staff are significant, although displaying a rather small coefficient. This paper also shows that having collaborated previously may enhance future collaboration, confirming the crucial role of social proximity. Finally, our model shows the intensity of SC to be stronger among universities located in regions with similar level of R&D resources (economic proximity). Thus, we conclude that the centre-periphery hypothesis applied to collaboration among universities does not hold in Europe. When comparing these results to those obtained from estimating separate models for *Chemistry & Chemical Engineering*, *Life Sciences* and *Physics & Astronomy*, the relevant role of institutional and social proximity is confirmed. The effect of the rest of proximity dimensions differs across fields, which may be explained by their specific nature, such as their level of interdisciplinarity or equipment requirements. Thus, this should be borne in mind when explaining the factors and dynamics of SC. The robustness check of our 12-discipline model, using fractional counting of publications and Tobit regressions, supports our results. However, the robustness checks by disciplines show slight differences compared to the ZINB models: a relevant role of geographical and cognitive proximity is additionally confirmed for the three disciplines.

The results shown in this paper allow us to draw some policy implications. Given the hindering effect of geographical distance and regional and national borders, it seems that the establishment of these collaboration links may require a special incentive, which may be not be monetary only. For example, incorporating new-to-the-group members in proposals for research projects could be positively evaluated. Science policy devoted to foster scientific collaboration in Europe could also focus on social proximity. As already mentioned above, those universities that collaborated in the past, are also more likely to collaborate in the future. This could be explained because social proximity reduces uncertainty and risk and increases trust. Then it seems that the first collaboration is the most crucial in the process of repeated collaborations. Therefore, public policy should focus on promoting these first-time relationships among scientists. Another alternative measure is to facilitate researcher mobility across universities, since in-person contact may enhance the share and finding of mutual research interests and potential co-publications. In this regard, several authors have indicated that scientific mobility flows could be a factor influencing international collaboration patterns (e.g. Furukawa et al., 2011; Jonkers and Cruz-Castro, 2013; Jonkers and Tijssen, 2008; Murakami, 2014; Scellato et al., 2015).

Future research could be aimed at including data for funding at university level and could analyse the substitute effects of proximity dimensions in fostering SC (evidence on the complementary role of proximity dimensions has been only provided on collaboration for innovation by Paci et al., 2014; and university-industry collaboration by D'Este et al., 2012).

Acknoweledgement

The authors would like to thank reviewers for their valuable comments.

References

Abramo, G., D’Angelo, C. & Di Costa, F. (2009). Research collaboration and productivity: is there correlation? *Higher Education* 57, 155-171.

Abramovsky, L., Harrison, R. & Simpson, H.(2007). University Research and the Location of Business R&D. *The Economic Journal* 117, 114-141.

Acosta, M., Coronado, D., Ferrándiz, E. & León, M.D., (2011a). Factors affecting inter-regional academic scientific

collaboration within Europe: the role of economic distance. *Scientometrics* 87, 63-74.
Acosta, M., Coronado, D., Ferrándiz, E. & León, M.D.(2014). Regional Scientific Production and Specialization in Europe: The Role of HERD. *European Planning Studies* 22, 949-974.
Acosta, M., Coronado, D. & Flores, E.(2011b). University spillovers and new business location in high-technology sectors: Spanish evidence. *Small Business Economics* 36, 365-376.
Amano, K. & Fujita, M., 1970. Long run economic effect analysis of alternative transportation facility plans - regional and national. *Journal of Regional Science* 10, 297-323.
Audretsch, D.B., Lehmann, E.E. & Warning, S.(2005). University spillovers and new firm location. *Research Policy* 34, 1113-1122.
Balland, P.A.(2011). Proximity and the Evolution of Collaboration Networks: Evidence from Research and Development Projects within the Global Navigation Satellite System (GNSS) Industry. *Regional Studies* 46, 741-756.
Barjak, F. & Robinson, S.(2008). International collaboration, mobility and team diversity in the life sciences: impact on research performance. *Soc. Geogr.* 3, 23-36.
Boschma, R.(2005). Proximity and Innovation: A Critical Assessment. *Regional Studies* 39, 61-74.
Boschma, R. & Frenken, K., 2009. Some Notes on Institutions in Evolutionary Economic Geography. *Economic Geography* 85, 151-158.
Cameron, A. & Trivedi, P.(2009). *Microeconometrics using stata*. Stata Press, College Station, Texas.
Cameron, A. & Trivedi, P.(2013). *Regression analysis of count data*. Cambridge University Press.
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, 617-628.
CEC (2000a) *Presidency conclusions of the Lisbon European council*, 23-24 March. Brussels: Commission of the European Communities.
CEC. (2000b). *Towards a European research area.* Brussels: Commission of the European Communities.
Cohen, W.M. & Levinthal, D.A., 1990. Absorptive Capacity: A New Perspective on Learning and Innovation. *Administrative Science Quarterly* 35, 128.
Costantini, V. & Mazzanti, M., 2012. On the green and innovative side of trade competitiveness? The impact of environmental policies and innovation on EU exports. *Research Policy* 41, 132-153.
Cummings, J.N. & Kiesler, S.(2007). Coordination costs and project outcomes in multi-university collaborations. *Research Policy* 36, 1620-1634.
D'Este, P., Guy, F. & Iammarino, S.(2012). Shaping the formation of university–industry research collaborations: what type of proximity does really matter? *Journal of Economic Geography*.
Foray, D.(2004). *Economics of knowledge*. MIT press, Cambridge, MA/London.
Franceschet, M. & Costantini, A.(2010). The effect of scholar collaboration on impact and quality of academic papers. *Journal of Informetrics* 4, 540-553.
Frenken, K., Hardeman, S. & Hoekman, J.(2009). Spatial scientometrics: Towards a cumulative research program. *Journal of Informetrics* 3, 222-232.
Furukawa, T., Shirakawa, N. & Okuwada, K.(2011). Quantitative analysis of collaborative and mobility networks. *Scientometrics* 87, 451-466.
Gazni, A., Sugimoto, C.R. & Didegah, F.(2012). Mapping world scientific collaboration: Authors, institutions, and countries. *Journal of the American Society for Information Science and Technology* 63, 323-335.
Gilsing, V., Nooteboom, B., Vanhaverbeke, W., Duysters, G. & van den Oord, A.(2008). Network embeddedness and the exploration of novel technologies: Technological distance, betweenness centrality and density. *Research Policy* 37, 1717-1731.
Hennemann, S., Rybski, D. & Liefner, I.(2012). The myth of global science collaboration—Collaboration patterns in epistemic communities. *Journal of Informetrics* 6, 217-225.
Heringa, P.W., Horlings, E., van der Zouwen, M., van den Besselaar, P. & van Vierssen, W.(2014). How do dimensions of proximity relate to the outcomes of collaboration? A survey of knowledge-intensive networks in the Dutch water sector. *Economics of Innovation and New Technology* 23, 689-716.
Hoekman, J., Frenken, K. & Oort, F.(2009). The geography of collaborative knowledge production in Europe. *Annals of Regional Science* 43, 721-738.
Hoekman, J., Frenken, K. & Tijssen, R.J.W.(2010). Research collaboration at a distance: Changing spatial patterns of scientific collaboration within Europe. *Research Policy* 39, 662-673.
Hoekman, J., Scherngell, T., Frenken, K. & Tijssen, R.(2012). Acquisition of European research funds and its effect on international scientific collaboration. *Journal of Economic Geography*.
Hong, W. & Su, Y.S.(2013). The effect of institutional proximity in non-local university–industry collaborations: An analysis based on Chinese patent data. *Research Policy* 42, 454-464.
Hwang, K.(2008). International Collaboration in Multilayered Center-Periphery in the Globalization of Science and Technology. *Science Technology Human Values* 33, 101-133.
Jonkers, K. & Cruz-Castro, L.(2013). Research upon return: The effect of international mobility on scientific ties, production and impact. *Research Policy* 42, 1366-1377.
Jonkers, K. & Tijssen, R.(2008). Chinese researchers returning home: Impacts of international mobility on research collaboration and scientific productivity. *Scientometrics* 77, 309-333.
Kahouli, B. & Maktouf, S.(2014). Trade creation and diversion effects in the Mediterranean area: Econometric analysis by gravity model. *The Journal of International Trade & Economic Development* 24, 76-104.

Katz, J.(1994). Geographical proximity and scientific collaboration. *Scientometrics* 31, 31-43.
Katz, J.S. & Martin, B.R.(1997). What is research collaboration? *Research Policy* 26, 1-18.
Lakitan, B., Hidayat, D. & Herlinda, S.(2012). Scientific productivity and the collaboration intensity of Indonesian universities and public R&D institutions: Are there dependencies on collaborative R&D with foreign institutions? *Technology in Society* 34, 227-238.
Laudel, G.(2002). What do we measure by co-authorships? *Research Evaluation* 11, 3-15.
Lee, S. & Bozeman, B.(2005). The Impact of Research Collaboration on Scientific Productivity. *Social Studies of Science* 35, 673-702.
Long, J.S.(1997). *Regression Models for Categorical and Limited Dependent Variables*. Sage, Thousand Oaks.
Maggioni, M. & Uberti, T.(2009). Knowledge networks across Europe: which distance matters? *The Annals of Regional Science* 43, 691-720.
Melin, G. & Persson, O.(1996). Studying research collaboration using co-authorships. *Scientometrics* 36, 363-377.
Mowery, D. & Sampat, B.(2004). Universities in National Innovation Systems, in: Fagerberg, J., Mowery, D., Nelson, R. (Eds.), *The Oxford Handbook of Innovation*. Oxford University Press Oxford, 209-239.
Murakami, Y.(2014). Influences of return migration on international collaborative research networks: Cases of Japanese scientists returning from the US. *Journal of Technology Transfer* 39, 616-634.
Niedergassel, B. & Leker, J.(2011). Different dimensions of knowledge in cooperative R&D projects of university scientists. *Technovation* 31, 142-150.
Nooteboom, B., Van Haverbeke, W., Duysters, G., Gilsing, V. & van den Oord, A.(2007). Optimal cognitive distance and absorptive capacity. *Research Policy* 36, 1016-1034.
Paci, R., Marrocu, E. & Usai, S.(2014). The Complementary Effects of Proximity Dimensions on Knowledge Spillovers. *Spatial Economic Analysis* 9, 9-30.
Paci, R. & Usai, S.(2009). Knowledge flows across European regions. *The Annals of Regional Science* 43, 669-690.
Petruzzelli, A.M.(2011). The impact of technological relatedness, prior ties, and geographical distance on university–industry collaborations: A joint-patent analysis. *Technovation* 31, 309-319.
Plotnikova, T. & Rake, B.(2014). Collaboration in pharmaceutical research: exploration of country-level determinants. *Scientometrics* 98, 1173-1202.
Ponds, R., van Oort, F. & Frenken, K.(2007). The geographical and institutional proximity of research collaboration. *Papers in Regional Science* 86, 423-443.
Rallet, A. & Torre, A.(1999). Is geographical proximity necessary in the innovation networks in the era of global economy? *GeoJournal* 49, 373-380.
Romer, P., 1990. Endogenous technological change. *Journal of Political Economy* 98, S71-S102.
Scellato, G., Franzoni, C. & Stephan, P.(2015). Migrant scientists and international networks. *Research Policy* 44, 108-120.
Scherngell, T. & Barber, M.(2009a). Distinct spatial characteristics of industrial and public research collaborations: evidence from the fifth EU Framework Programme. *The Annals of Regional Science*.
Scherngell, T. & Barber, M.J.(2009b). Spatial interaction modelling of cross-region RD collaborations: empirical evidence from the 5th EU framework programme. *Papers in Regional Science* 88, 531-546.
Scherngell, T. & Hu, Y.(2010). Collaborative Knowledge Production in China: Regional Evidence from a Gravity Model Approach. *Regional Studies* 45, 755-772.
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, 112 - 144.
Schubert, T. & Sooryamoorthy, R.(2010). Can the centre–periphery model explain patterns of international scientific collaboration among threshold and industrialised countries? The case of South Africa and Germany. *Scientometrics* 83, 181-203.
Sonnenwald, D.H.(2007). Scientific collaboration: a synthesis of challenges and strategies, in: Cronin, B. (Ed.), *Annual Review of Information Science and Technology, vol. 41.* Medford NJ: Information Today, Inc.
Tijssen, R., J.W & van Leeuwen, T.N.(2003). Bibliometric Analyses of World Science, Extended Technical Annex to Chapter 5 of the Third European Report on Science & Technology Indicators.
Torre, A. & Gilly, J.P.(2000). On the Analytical Dimension of Proximity Dynamics. *Regional Studies* 34, 169-180.
Torres-Salinas, D., Delgado-López-Cozar, E., García-Moreno, J. & Herrera, F.(2011). Rankings ISI de las universidades españolas según campos científicos: descripción y resultados (2ª Ed. 2011). Available at http://rankinguniversidades.es/ (Last accessed 1/11/2011).
Waltman, L., Tijssen, R.J.W. & Eck, N.J.v.(2011). Globalisation of science in kilometres. *Journal of Informetrics* 5, 574-582.
Wooldridge, J. (2002). Econometric Analysis of cross section and panel data. Cambridge: MIT Press.