NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS

Stanislav Avdeev

INTERNATIONAL COLLABORATION IN HIGHER EDUCATION RESEARCH: A GRAVITY MODEL APPROACH

BASIC RESEARCH PROGRAM

WORKING PAPERS

SERIES: EDUCATION
WP BRP 54/EDU/2019

***SERIES**: EDUCATION*

INTERNATIONAL COLLABORATION IN HIGHER EDUCATION RESEARCH: A GRAVITY MODEL APPROACH *

Stanislav Avdeev †

HSE University

Abstract

Geographical distance has become less relevant in co-authorship for monodisciplinary research fields such as economics, mathematics, physics. Multidisciplinary fields are influenced by the norms and traditions of other fields. Higher education is a multidisciplinary field of research in which multiple communities operate under different norms and paradigms. We study collaboration patterns in higher education research across different world regions using the Scopus database with the application of the gravity model. Our results show that international collaboration has intensified and increased rapidly in the last two decades. We confirm that the intensity of collaboration is dependent on geographical distance and linguistic commonality. The importance of geographical proximity differs significantly between various world regions. EU scholars appear to give preference to linguistic proximate partners over geographical neighbors. Despite the fact that English is the lingua franca in science, language is not a significant factor for the formation of collaboration for North American researchers. This study contributes to the current discussion on the importance of international collaboration in science, paying special attention to the growing public interest in multidisciplinary research.

Keywords: gravity model, higher education, international collaboration, multidisciplinarity, spatial scientometrics

JEL Codes: C21, I23, N30

*The article was prepared within the framework of the HSE University Basic Research Program and funded by the Russian Academic Excellence Project "5-100". I gratefully acknowledge valuable comments and suggestions from Maria Yudkevich, colleagues at the Center for Institutional Studies, and conference and workshop participants in Amsterdam, Jyväskylä, Kassel, Moscow, and Saint Petersburg.

†Center for Institutional Studies, HSE University, Moscow, Russia; e-mail: stnavdeev@gmail.com

1 Introduction

Geographical distance has become less relevant while linguistic ties are still a strong factor in collaboration between countries (Hoekman et al., 2010). Countries located in close physical proximity profit from each other through knowledge spillovers (Almeida et al., 2009). When authors speak the same language, they are more likely to collaborate. However, these factors are found to be relevant for monodisciplinary research fields such as economics, mathematics, physics, biology (Newman, 2004). Increased multidisciplinary research requires the involvement of researchers from different disciplinary backgrounds (Gates et al., 2019). Scholars from multidisciplinary fields have different patterns of collaboration (Leahey, 2016). An example of a multidisciplinary field is higher education which is defined by its object of research. Scholars from different fields come to research higher education with their own methods and theories (Tight, 2004; Yokoyama, 2016). The field of higher education is influenced by the norms and traditions of other fields, particularly sociology, psychology and economics which are the roots of the field (Lovakov and Yudkevich, forthcoming). Nevertheless, higher education researchers remain unreflective about the structure of their own collaboration.

Waltman et al. (2011) found that social scientists are less likely to collaborate with people from different countries. Economists are more likely to cooperate with international scholars than sociologists and psychologists (Rosenblat and Mobius, 2004; Kliegl and Bates, 2011; Kuld and O'Hagan, 2018). There are substantial differences in collaboration patterns of higher education researchers who come from different fields (usually social sciences). Scholars choices of international collaboration partners are influenced by the norms, habits, and routines of specific disciplines (Henriksen, 2016). Almeida et al. (2009) found that countries located in close geographical proximity to each other show similar specialization patterns. Policy-oriented papers involve more researchers from different countries, whereas individual scholars focus more on student-specific topics (Altbach, 2014). Therefore, the extent to which different disciplines have various collaboration patterns is understudied.

Collaboration patterns are affected not only by discipline but by country of origin. European, American and Asian scholars have different collaboration patterns. For instance, European researchers benefit from the EU and the variety of funding agencies where research programs often require cross-country teams (Wagner et al., 2015). Asian researchers prefer to choose scholars from other Asian countries due to the similarities of higher education systems and academic traditions. Likewise, if scholars in Asia are cited by authors within the region, they do not need recognition from European and North American scholars (Glänzel, 2001). Geographical partners might be preferred not only due to their close proximity but due to the similarities and benefits of collaboration with neighboring partners.

We study collaboration patterns in the multidisciplinary field of higher education across different world regions. The question of international collaboration is critical in science since the national level of organization is challenged by researchers focusing on the globalization process (Frenken et al., 2009). In order to measure the research collaboration of different countries, the most robust approach is to use spatial scientometric techniques analyzing the institutional affiliations contained in bibliometric data. Our results show that the number of international publications has grown exponentially during last two decades. The main research producers are English-speaking countries, Western Europe and China. EU countries are the major international partners for non-EU countries: the share of cooperation within the EU is decreasing while the share of international collaboration of EU countries with non-EU countries is increasing. The larger a country's research output, the smaller the proportion of the international co-authorship of the country. The mean number of internationally co-authored papers is 1.46 per year and the mean distance is 6,472 km. The regression results show that the number of international collaborations is highly dependent on the number of scholars and the distance between co-authors for most of the countries, varying considerably by region. The official language is not a significant factor for the formation of collaboration for North American and Asian countries but it plays an important role for EU countries. These encouraging results contribute to the current discussion of the

importance of international collaboration in science, paying special attention to the growing public interest in multidisciplinary research.

The paper is structured as follows. In Section 2 we discuss the role of geographical proximity and linguistic commonality in collaboration as well as international collaboration in the field of higher education. Section 3 describes the data, methodology and specifications. The results of descriptive and regression analysis are described and summarized in Section 4. In Section 5 we interpret the empirical results and discuss possible mechanisms for the formation of international collaboration in the multidisciplinary field of higher education.

2 Literature Review

2.1 International collaboration and the role of geographical proximity and linguistic commonality

Katz and Martin (1997) noted that co-authorship of scientific papers is only one of the possible outcomes of research collaboration, and it is only one of the forms in which collaboration can be expressed. Nevertheless, co-authorship has been adopted as a proxy for measuring research collaboration, i.e. this is a classical outcome of research activities. Therefore, in this study international research collaboration is a co-authorship relation between researchers from two or more countries.

Many studies have shown the inexorable growth of international collaboration across different fields (Adams, 2012, 2013). Marginson (forthcoming) observed the rapid growth in the number and proportion of papers co-authored from more than one country. Wagner et al. (2015) found that the proportion of internationally co-authored papers rose from 10.1% in 1990 to 19.5% in 2000 and to 24.6% in 2011. After the year 2000 the worldwide collaboration rate rose in all disciplines, including hard sciences and social sciences (Wagner and Leydesdorff, 2005; Wagner et al., 2017). In addition, Adams (2013) found that the rise in the total annual output for each country is due to international collaboration. Almost all

countries have similar patterns in the growing proportion of articles that have international co-authors. However, the world average figures of international collaboration do not explain patterns of variation in international co-authorship on a smaller scale: by regions and disciplines. This study focuses not only on the world scale but distinguishes regional patterns of international collaboration in the multidisciplinary field of higher education.

To find out how co-authors work together, previous papers explored some characteristics of such collaboration. Geographical proximity and linguistic commonality play an important role in fostering high intensity of collaboration. Researchers are biased towards international collaboration with partners who are proximate with respect to geography and language (Boschma, 2005; Hoekman et al., 2010). Typically, geographical proximate partners are preferred (Waltman et al., 2011; Hoekman et al., 2010). Waltman et al. (2011) revealed that research has globalized in recent decades: the average collaboration distance per publication has increased from 334 km in 1980 to 1553 km in 2009. Increasingly, funding agencies and public policies have encouraged collaboration by prioritizing research in partnership. The establishment of the European Research Area has stimulated international collaboration (Hoekman et al., 2009). Hoekman et al. (2010) using spatial scientometrics found a diminishing effect of geographical proximity on co-publishing, with territorial borders becoming less relevant. Overall, the research design of spatial scientometrics papers is more elaborated than previous descriptive papers on international collaboration.

When authors speak the same language, they are more likely to collaborate with each other rather than with researchers with different characteristics (Hoekman et al., 2009). Adams (2012) found that Nigeria collaborates not with its neighbors in West Africa but with co-linguists in East Africa. Hoekman et al. (2010) showed that co-publication intensity is higher within regional, national and linguistic areas after controlling for the size of regions and their research specialization profiles. Although researchers speaking a common language are not necessarily located within a single geographical area, we expect co-authors who speak shared language to have a higher intensity of international collaboration.

Geographical clusters tend to influence the intensity of international collaboration regardless of the distance and linguistic proximity. The rapid growth of regional links in Asia and in established regions such as North America and Europe reveal the importance of collaboration within world regions (Glänzel, 2001). Leydesdorff and Wagner (2008) revealed that the number of internationally co-authored publications has grown linearly while the number of addresses of internationally collaborating authors grew exponentially, suggesting that the growth of networks extends to many more places around the globe. Adams (2012) showed growing research networks of countries in Asia, the Middle East and the Latin America. China's dramatic growth of international collaboration with neighboring countries is one example of regional development (Jung and Horta, 2013). Therefore, despite language differences, geographical proximate countries within a certain region are preferred. Despite the vast body of empirical evidence, the effects of geography and language on international collaboration in a multidisciplinary field have yet to be studied.

2.2 International collaboration in multidisciplinary fields

To explore international collaboration in a multidisciplinary field, there should be a clear definition of mono- and multi- disciplinary fields. Monodisciplinary research is concerned with the study of a research topic within a single discipline, and with methods from this discipline. Multidisciplinary research is concerned with the study of a research topic across multiple disciplines, and with the transfer of methods from one discipline to another.

Different disciplines operate under different norms and paradigms (Henriksen, 2016). We exclude studies of international collaboration in hard sciences from our analysis due to their distinctive collaboration patterns. To the best of our knowledge, international collaboration in a multidisciplinary field have never been investigated with the application of spatial scientometrics.

Several studies have investigated patterns of international collaboration in social sciences: Rosenblat and Mobius (2004); Kuld and O'Hagan (2018) in economics; Leahey and

Reikowsky (2008) in sociology; Kliegl and Bates (2011) in psychology. Preliminary work on international collaboration in social sciences focused primarily on a descriptive analysis of the co-occurrence of countries in joint papers. Few studies have been published about the geographical barriers of social scientists (Waltman et al., 2011). Different barriers exist when researchers in social disciplines having differing norms and traditions are collaborating. For instance, Waltman et al. (2011) explored the differences in average geographical distance per publication among the fields of science: economics and business (1939 km) and psychology (1478 km) are more globalized fields than sociology and anthropology (1063 km) and educational sciences (969 km).

Some studies focused on the influence of international collaboration on scientific production in multidisciplinary fields. Bartneck and Hu (2010) found that there are no significant differences between domestic and international collaboration in terms of citations in the Computer–Human Interaction community. They found that North America and Europe collaborate most within the Computer–Human Interaction community, particularly, the USA, UK and Canada co-author many papers. Correia et al. (2018) analyzing the Computer Supported Cooperative Work community demonstrated that distance is no longer the barrier it was in the past, despite the heterogeneity between some regions in their propensity to collaborate. Wang et al. (2015) found that sport science researchers showed a strong tendency to collaborate, especially among European countries. They highlighted that the share of international collaboration in Asian countries is below 40%, and the growth rate is lower than that of these countries' overall output, while the trend is reversed in many western countries: the share is above 50% and the growth rate is higher. Therefore, it is not straightforward what the collaboration patterns are between scholars from various disciplines, as in higher education research, as they could be less or more inclined to collaborate with each other than two researchers in monodisciplinary fields.

Overall, there is a large body of research that primarily focuses on the growth of international collaboration in multidisciplinary fields. We take a different approach with the

application of spatial scientometrics and study geographical and linguistic proximity, which influence the formation of international collaboration in higher education. We hypothesize that geographical proximity still plays an important role for international research collaboration because of the language they speak and the region in which they operate. By focusing on international collaboration in a multidisciplinary field of research, this study contributes to the broader literature that explores geographical proximity and linguistic commonality across various fields.

2.3 Higher education as a multidisciplinary field

Tight (2004) defined higher education as an "interdisciplinary field of research in which multiple communities of practice operate". The field of higher education research is a fragmented community of researchers from various disciplines with different types of degrees, theoretical approaches and methodologies. It is even not a sub-field of educational studies (Yokoyama, 2016). Lovakov and Yudkevich (forthcoming) covered the disciplinary foundations and roots of higher education research and revealed the different influences of psychology, economics, and sociology on the field. Psychology has the highest citation rate by higher education researchers, followed by sociology and economics.

Altbach (2014) said "higher education is not a scholarly or scientific discipline; it has no central and accepted methodology nor does it have a set of concerns for research and study. Rather, it is a field that uses the disciplinary insights of other fields, mainly in the social sciences, to inform research themes that often require interdisciplinary insights". Yokoyama (2016) defined higher education as a "multi-disciplinary and loosely coupled community [which] suggest diversity in the field and in its identity rather than coherence and consolidation". Internationalization in the form of increased student and academic mobility have broadened the scope of higher education, particularly for scholars of the economics of education, the sociology of education, public administration and psychology (Altbach, 2014). The multidisciplinary field of higher education has been formed by interaction with other

fields and disciplines, therefore, is influenced by them and their collaboration patterns.

Higher education studies differ not only because of the several disciplines that comprise it, but because of its substantial geographical variation. The American higher education community is more practice-oriented, engaging in meso- and micro- level research, whereas European scholars are more policy-focused, emphasizing analysis at the macro level (Yokoyama, 2016). Although very few studies have explored international collaboration in higher education research, several studies have analyzed the field of higher education research and some collaboration patterns (Kosmützky and Krücken, 2014; Kuzhabekova et al., 2015). Kuzhabekova et al. (2015) found a steady increase of the share of international higher education research for 2002–2011: from 2.9% to 20.5%. Kosmützky and Krücken (2014) focused specifically on comparative research and found a much higher share of international collaboration in comparative research compared to non-comparative. They showed that 46% of comparative papers are international co-authored publications compared to about 20% of non-comparative papers. On the other hand, Kuzhabekova et al. (2015) found just 11.3% of papers are authored by researchers coming from at least two countries. International comparative research appears to be the most globalized topic in the field of higher education and many international publications have been co-authored by scholars from different countries.

3 Methodology

3.1 Data

In this section, we describe the data and the main variables we use in the analysis, paying particular attention to the construction of the proximity measures. We take into account only scientific papers without accounting for books and other possible ways of cooperating and publishing research as they are under-represented and, therefore, further analysis would be inhibited by a lack of data. Our analysis is based on publication types "article" and "review" from Scopus. Scopus has a wider coverage of papers in social sciences than other databases, so it seems meaningful to exploit this database.

We examine articles and reviews published in 24 journals considered key in the field of higher education (Lovakov et al., forthcoming). Data was retrieved from the Scopus database in January 2019. The overall data set covers 17,413 publications from the period 1978–2017 excluding papers with no affiliation information.

First, we choose papers that are internationally co-authored: papers that have been published with the cooperation of at least two different countries. International co-publications, international co-authored papers, international collaboration will be used as synonyms. The determination of the country of origin of the authors is based on their institutional affiliation. This method of identifying the current location of authors is not foolproof. For example, a scholar might work temporarily in a foreign university and choose to list their home institute. We would consider such a collaboration as an international one. Previous studies have found that the impact of these misassignments is relatively small. In sum, there are 1,414 papers which have at least two coauthors from different countries.

Second, we find the total number of authors in each country counting the number of authors with at least one paper in the overall data set. In addition, if an author has multiple affiliations in different countries, the paper is considered as an international paper and this author is counted multiple times: scholars with multiple affiliations gain access to addi-

tional research resources or networks and form stronger ties between institutions in different countries, therefore, these authors and papers are included and analyzed.

Third, all author affiliations of the selected publications are reduced to a country, giving 96 countries. For each country we find the capital and its latitude and longitude. Since we are interested in international collaboration between two countries rather than cities, we take into account only one geographical point per country. Then, for each pair of countries, the number of times it occurs in the selected publications as an international co-authorship is counted. For all capitals, coordinates are obtained using the R package *geosphere*. Next, distance is calculated for all pairs as the Euclidean distance in kilometers between the capitals of two countries. For each country the official or de facto official language is identified using R package *lingtypology*: if a country has several languages, all of them are counted.

3.2 Empirical model

We follow the spatial scientometrics framework of Frenken et al. (2009) in exploring the geographical patterns of international collaboration in higher education research. We start estimating a base model which is modelled by analogy to Newton's law of universal gravitation (Tinbergen, 1962; Hoekman et al., 2010). The gravity model states that the gravitational force between two entities is dependent on their masses and the distance between them. Collaboration frequency between two countries is assumed to be dependent on the number of authors and the distance between them:

$$I_{ij} = a_0 \frac{MASS_i^{a_1} MASS_j^{a_2}}{DISTANCE_{ij}^{a_3}} \tag{1}$$

Empirically, taking natural logarithms on both sides of the gravity model and adding a random error term, Model 1 can be converted into a testable Model 2:

$$\ln I_{ijt} = a_0 + a_1 \ln MASS_{it} + a_2 \ln MASS_{jt} + a_3 \ln DISTANCE_{ij} + \varepsilon_{ijt} \tag{2}$$

It is important to take into consideration the total number of authors in a country, because collaboration intensity is highly dependent on size.

Dummy variables can be added to the model to account for factors that are not included in the basic model. To take into account linguistic proximity, we use a dummy variable which equals to 1 if two countries share a common official language/languages, and 0 otherwise.

$$\ln I_{ijt} = a_0 + a_1 \ln MASS_{it} + a_2 \ln MASS_{jt} + a_3 \ln DISTANCE_{ij} + a_4 LANG_{ij} + \varepsilon_{ijt} \quad (3)$$

I_{ijt} is the total number of co-authored papers between country i and country j in year t;

$MASS_{it}, MASS_{jt}$ is the number of authors in country i and country j in year t;

$DISTANCE_{ij}$ is the Euclidean distance between capitals of country i and country j;

$LANG_{ij}$ is a dummy variable equals to 1 if country i and country j have a common official language/languages;

ε_{ijt} is the error term.

3.3 Model specifications and estimation

The dependent variable is the number of internationally co-authored papers between two countries, i.e. collaboration frequency. The standard model for count data (non-negative integers) is the Poisson regression model. A Poisson distribution assumes that the mean and variance are equal. Our data show that the mean is less than the variance, respectively, 3.68 and 7.83, suggesting some overdispersion. This occurs when, for a random variable $Y \sim Pois(\lambda)$:

$$\mathbb{E}(Y) < Var(Y) \quad (4)$$

There may be an issue of unobserved heterogeneity in our data. Unobserved heterogeneity leads to overdispersion. The standard parametric model to account for overdispersion is the negative binomial. The negative binomial model assumes a particular form of dependence for the underlying stochastic process, with the occurrence of an event increasing the probability

of further occurrences.

Zero event counts are often observed in our data, leading to a skewed distribution. Zero events count is a situation when two countries do not have any papers for a given year, which leads to inconsistency with the Poisson model. A zero-inflated count model provides another way to model excess zeros, therefore, we include this model with the negative binomial model.

In sum, we conduct our analysis using negative binomial and zero-inflated count models to partially solve the problem of the spatial gravity model. The approach that we use provides measurements of the geographical proximity and linguistic commonality in higher education research within and across various regional groupings.

First, we analyze a model for all 96 countries. Second, we estimate several models that include various regional groupings of countries. For the purpose of the regression analysis, we divide and restrict our database into three regions: Europe, North America, and Asia (top-3 contributors): each co-authored paper is included to European, North American or Asian databases, if one of the authors is from this region.

4 Results

4.1 Descriptive statistics

The number of internationally co-authored publications has grown exponentially, as demonstrated in Figure 1 and Figure 2, which indicate the increased interest in higher education research. The share of international collaboration rose annually by 5.5% before 2000, but by 13.8% after 2000. In the following analysis, we restrict the sample to 2000–2017 since there are few publications before 2000. There are 1,262 publications from 2000 to 2017.

Figure 1: The share of international papers in 1978-2017

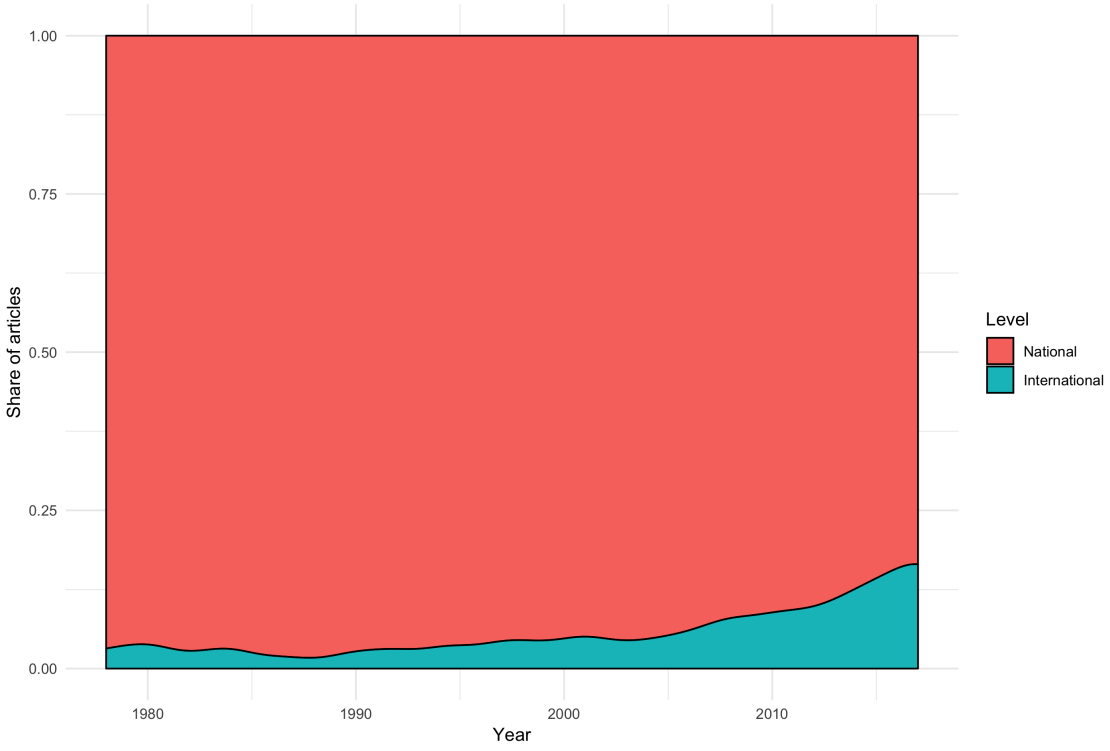
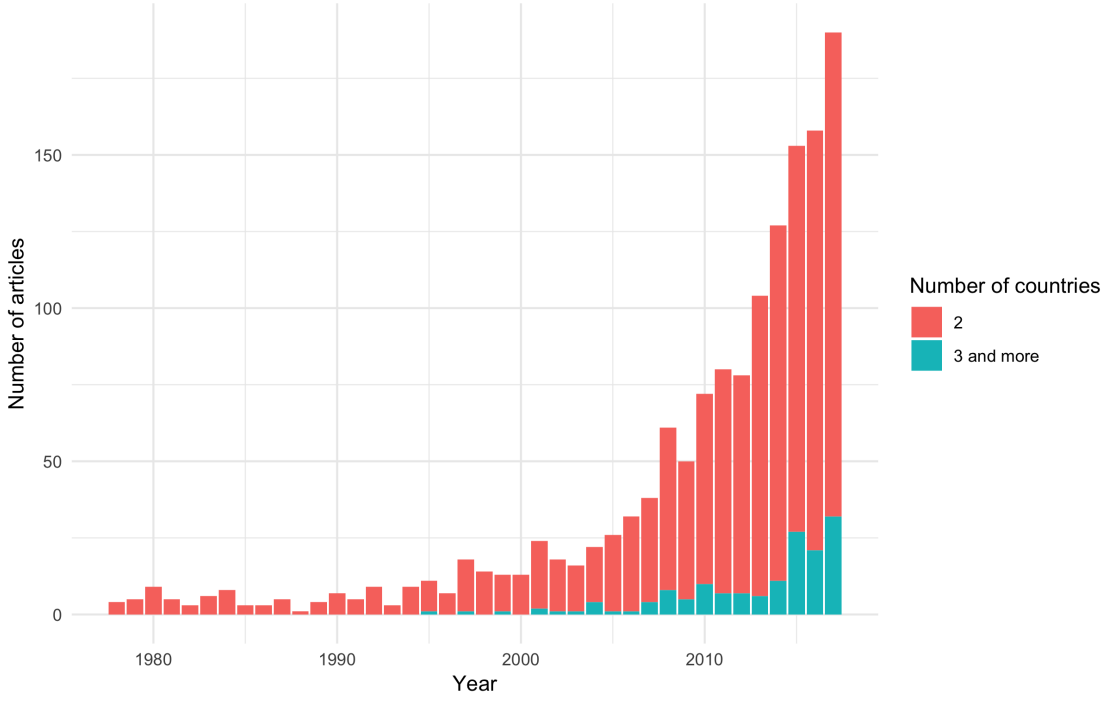


Table 1 presents the total national publication output and the share of international papers by country for the observed period. The most active 40 countries are ranked in descending order by their total number and the share of international papers. Most publications are from English-speaking countries: UK, US, Australia, Canada and New Zealand. Some European and Asian countries are also among the top publishing countries: China,

Figure 2: The number of international papers in 1978-2017



Netherlands, Germany, Norway, Hong Kong.

The share of international publications in a country's output can be found in Table 2. The most internationally-oriented countries are Vietnam, Brazil, Romania, United Arab Emirates, Cyprus, China, South Korea, Switzerland, Czech Republic and Austria. There is at least one country in each region that accumulates the largest amount of international collaboration: Europe – UK and the Netherlands; North America – US and Canada; Asia – China and Hong Kong; Oceania – Australia and New Zealand; Africa – South Africa; South America – Brazil.

These two tables demonstrate that the larger a country's research effort, the smaller the proportion of its international co-authorship: scholars in smaller (resource-poor) countries are forced to look outside for co-authors, while researchers in the larger countries tend to collaborate more frequently with domestic partners.

The groups of countries clearly reflect the effects of geography on international co-authorship. The countries of Africa, Asia, Europe, North America, and Oceania are brought

Table 1: The number and the share of international papers in 2000-2017

№	Country	N of int. papers	Share (total)	№	Country	N of int. papers	Share (total)
1	UK	391	0.31	21	Japan	30	0.02
2	US	390	0.31	22	Singapore	29	0.02
3	Australia	276	0.22	23	Taiwan	28	0.02
4	Canada	142	0.11	24	Switzerland	26	0.02
5	China	113	0.09	25	France	25	0.02
6	Netherlands	110	0.09	26	Ireland	23	0.02
7	Germany	90	0.07	27	UAE	21	0.02
8	New Zealand	78	0.06	28	Turkey	21	0.02
9	Norway	74	0.06	29	Brazil	19	0.02
10	Hong Kong	70	0.06	30	Austria	19	0.02
11	Sweden	60	0.05	31	Greece	16	0.01
12	Spain	50	0.04	32	Czech Republic	12	0.01
13	Portugal	49	0.04	33	Israel	12	0.01
14	Finland	49	0.04	34	Mexico	11	0.01
15	South Africa	48	0.04	35	Vietnam	10	0.01
16	Belgium	45	0.04	36	Romania	10	0.01
17	South Korea	39	0.03	37	Cyprus	10	0.01
18	Malaysia	37	0.03	38	Chile	10	0.01
19	Denmark	34	0.03	39	Russia	10	0.01
20	Italy	31	0.02	40	India	9	0.01

Only countries with 9 papers and more are shown in the table.

together for further analysis. Figure 3 examines trends in international collaboration by region of the world. If a country has a joint paper with a country that is not in the country's region, it counts as an international co-authored paper. Europe is the dominant region for collaboration, reflecting the total output of European countries (529 papers). The countries of North America (417) are second, followed by Asian countries (361), Oceania (294), and Africa (75). This reveals the differences in growth by regions: collaboration with Asia increased most rapidly, followed by Africa, Oceania, Europe, and North America.

EU countries extended and collaborated more with researchers from non-EU countries. The proportion of the collaboration within EU countries is decreasing, while the proportion of the collaboration of EU countries with co-authors from non-EU countries is increasing. This illustrates that in international scientific collaboration the EU has become more open

Table 2: The share of international papers by country in 2000-2017

№	Country	Share	№	Country	Share
1	Vietnam	0.71	21	Netherlands	0.33
2	Brazil	0.56	22	Russia	0.31
3	Romania	0.56	23	Turkey	0.30
4	UAE	0.54	24	Hong Kong	0.30
5	Cyprus	0.53	25	Japan	0.29
6	China	0.53	26	Portugal	0.28
7	South Korea	0.47	27	Canada	0.28
8	Switzerland	0.46	28	Taiwan	0.26
9	Czech Republic	0.44	29	Sweden	0.26
10	Austria	0.44	30	Italy	0.25
11	Belgium	0.41	31	Spain	0.24
12	Malaysia	0.40	32	India	0.24
13	Greece	0.39	33	New Zealand	0.23
14	Singapore	0.39	34	Finland	0.21
15	France	0.38	35	Ireland	0.21
16	Denmark	0.38	36	Australia	0.14
17	Germany	0.38	37	South Africa	0.14
18	Chile	0.37	38	UK	0.14
19	Norway	0.36	39	Israel	0.12
20	Mexico	0.35	40	US	0.09

Only countries with 9 papers and more are shown in the table.

towards non-EU countries.

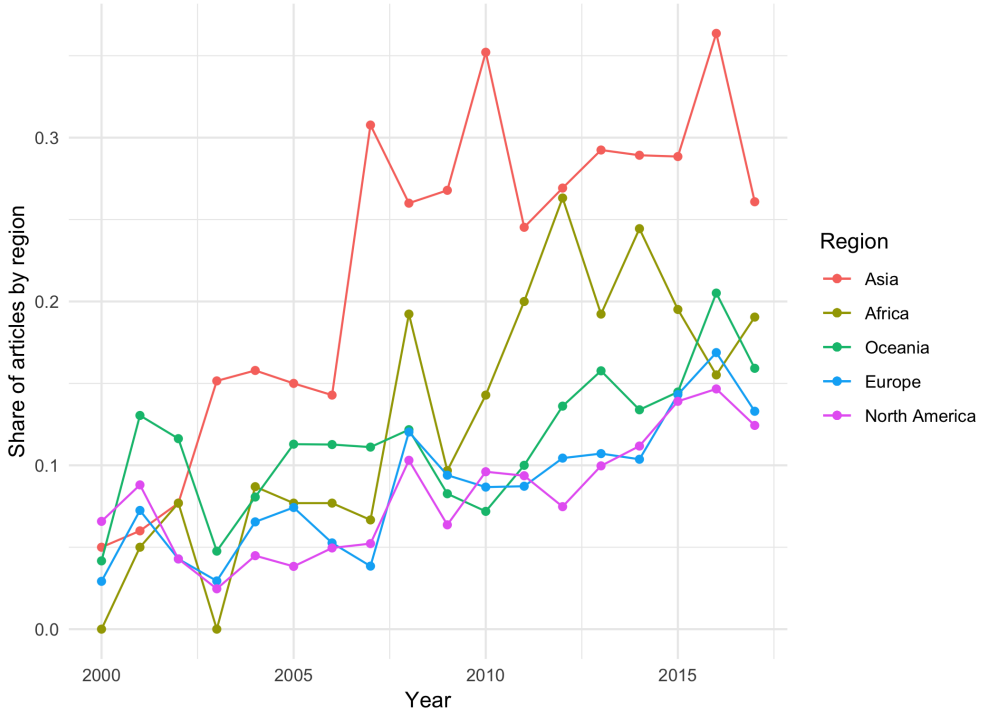
Table 8 in the appendix shows the joint papers of the top 25 countries. There are no clear geographical preferences between countries: there are close geographical partners such as the pairs of Canada-US and Australia-New Zealand, and there are distant pairs of countries such as Australia-UK, UK-US, and Australia-US. These are the most productive countries, suggesting that the English language is the current lingua franca of science.

4.2 Regression analysis

The summary statistics for the variables included in the model are presented in Table 3. The average number of co-authored papers between countries is 1.46 per year. The average distance per paper is 6,472 km. About 22% of countries share the official language/languages.

The regression results for all countries are in Table 4. We start estimating the basic

Figure 3: The share of international papers by region



model without a dummy variable: Model 1 (negative binomial), Model 3 (zero-inflated count data). The coefficients for the number of authors in countries are 0.605 and 0.486 for negative binomial model, and 0.491 and 0.434 for zero-inflated count data model, statistically significant at the 0.01 level. This shows that the higher the total number of scholars in a country, the higher the intensity of international collaboration. This is in line with previous research on collaboration.

Table 3: Statistics for variables

Variables	N	Mean	St. Dev.	Min	Max
N of co-authored papers	1,157	1.46	1.30	1	21
N of authors in country i	1,157	30.92	50.79	0	233
N of authors in country j	1,157	99.08	111.56	0	336
Distance, km	1,157	6,472.46	5,143.10	59.78	19,575.90
Common language, 1 = Yes	1,157	0.22	0.41	0	1

Geographical proximity shows a negative statistically significant coefficient at the 0.01

level. The coefficients are -0.262 and -0.208 for negative binomial and zero-inflated count data models, respectively. This indicates a lower intensity of collaboration over longer distances.

Table 4: Regressions results for international collaboration between all countries

	Dependent variable:			
	N of co-authored papers			
	negative binomial		*zero-inflated count data*	
	(1)	(2)	(3)	(4)
N of authors in country i	0.605***	0.541***	0.491***	0.416***
	(0.050)	(0.057)	(0.051)	(0.055)
N of authors in country j	0.486***	0.463***	0.434***	0.413***
	(0.046)	(0.047)	(0.042)	(0.040)
Distance, km	−0.262***	−0.262***	−0.208***	−0.205***
	(0.053)	(0.053)	(0.050)	(0.042)
Common language, 1 = Yes		0.325**		0.364***
		(0.149)		(0.131)
Constant	−3.074***	−2.889***	−2.739***	−2.560***
	(0.453)	(0.453)	(0.418)	(0.282)
Observations	1,157	1,157	1,157	1,157
Log Likelihood	−527.147	−524.780	−584.310	−580.905
Akaike Inf. Crit.	1,062.293	1,059.561		

Standard errors in parentheses clustered at the country level.
*p<0.1; **p<0.05; ***p<0.01

Model 2 (negative binomial), Model 4 (zero-inflated count data) include an additional regressor: common official language. The coefficients are 0.325 and 0.364 and statistically significant at the 0.05 and 0.01 level, respectively. This indicates that collaboration between countries that share a common official language occur more often than collaboration between other countries in the world.

Table 9 and Table 10 in the appendix show the regression results for all countries in

two separated periods: 2000-2008 and 2009-2017. In line with previous results we observe similar coefficients for the number of authors in countries. Remarkably, the coefficients for geographical distance and linguistic commonality show a slight decline over time, suggesting that international collaboration is less dependent on geography and language in the period 2009-2017 than in the period 2000-2008. All coefficients are statistically significant.

Table 5: Regressions results for international collaboration for European researchers

	Dependent variable:			
	N of co-authored papers			
	negative binomial		*zero-inflated count data*	
	(1)	(2)	(3)	(4)
N of authors in country i	0.636^{***}	0.558^{***}	0.645^{***}	0.435^{***}
	(0.069)	(0.075)	(0.066)	(0.151)
N of authors in country j	0.432^{***}	0.403^{***}	0.245^{***}	0.680^{***}
	(0.059)	(0.060)	(0.068)	(0.117)
Distance, km	-0.233^{***}	-0.248^{***}	-0.177^{***}	-0.317^{***}
	(0.069)	(0.069)	(0.066)	(0.109)
Common language, 1 = Yes		0.445^{**}		1.031^{***}
		(0.200)		(0.349)
Constant	-3.216^{***}	-2.836^{***}	-2.642^{***}	-4.050^{***}
	(0.549)	(0.557)	(0.528)	(0.951)
Observations	776	776	776	776
Log Likelihood	−350.015	−347.626	−389.954	−178.125
Akaike Inf. Crit.	708.030	705.252		

Standard errors in parentheses clustered at the country level.
*p<0.1; **p<0.05; ***p<0.01

Regression results for EU countries are presented in Table 5. We observe similar collaborative trends among European countries compared to the world. The total number of authors in a country is positive and statistically significant in both models. Geographical

patterns of collaboration for the EU are similar to the world average, suggesting that EU collaboration does not occur more often over longer distances than the collaboration of other countries.

Table 6: Regressions results for international collaboration for North American researchers

	Dependent variable:			
	N of co-authored papers			
	negative binomial		*zero-inflated count data*	
	(1)	(2)	(3)	(4)
N of authors in country i	0.559^{***}	0.481^{***}	0.514^{***}	0.428^{**}
	(0.072)	(0.095)	(0.120)	(0.194)
N of authors in country j	0.390^{***}	0.395^{***}	0.629^{**}	0.678^{**}
	(0.114)	(0.117)	(0.304)	(0.276)
Distance, km	-0.334^{***}	-0.280^{**}	-0.363^{***}	-0.304^{*}
	(0.106)	(0.113)	(0.137)	(0.169)
Common language, 1 = Yes		0.326		0.287
		(0.281)		(0.437)
Constant	-1.846	-2.261^{*}	-3.257	-3.887^{*}
	(1.176)	(1.226)	(2.140)	(2.088)
Observations	350	350	350	350
Log Likelihood	-191.206	-190.540	-114.720	-114.795
Akaike Inf. Crit.	390.411	391.081		

Standard errors in parentheses clustered at the country level.
*p<0.1; **p<0.05; ***p<0.01

The coefficient for linguistic commonality is positive and statistically significant. The coefficient for the common official language is higher for EU countries compared to the world which shows the greater importance of linguistic commonality for EU scholars and their co-authors. EU researchers collaborate significantly more often with co-authors who speak the same language.

The results for North American researchers are presented in Table 6. As in previous models, the coefficients for the total number of authors are positive and statistically significant and the geographical distance is negative and statistically significant in both specifications. However, the coefficient for a common official language shows no statistically significant results for this region. Despite the fact that English is a lingua franca in higher education research and English is the most widely spoken language in the region, this does not correlate with the intensity of international collaboration of North American scholars with other countries.

Table 7: Regressions results for international collaboration for Asian researchers

	Dependent variable:			
	N of co-authored papers			
	negative binomial		*zero-inflated count data*	
	(1)	(2)	(3)	(4)
N of authors in country i	0.680***	0.670***	0.454***	0.442***
	(0.122)	(0.122)	(0.115)	(0.116)
N of authors in country j	0.745***	0.760***	0.619***	0.627***
	(0.119)	(0.123)	(0.092)	(0.096)
Distance, km	−0.570**	−0.540**	−0.457***	−0.427***
	(0.227)	(0.233)	(0.045)	(0.034)
Common language, 1 = Yes		0.312		0.274
		(0.351)		(0.321)
Constant	−1.435	−1.789	−1.017	−1.329
	(1.752)	(1.829)		
Observations	379	379	379	379
Log Likelihood	−150.860	−150.490	−166.353	−166.025
Akaike Inf. Crit.	309.721	310.980		

Standard errors in parentheses clustered at the country level.
*p<0.1; **p<0.05; ***p<0.01

The results for Asian scholars are presented in Table 7. Negative binomial and zero-inflated count models show similar results consistent with previous regions: the coefficients for the total number of authors are positive and statistically significant, those for distance are negative and statistically significant. The higher coefficients for distance for Asian authors compared to European and North American ones suggest that geographical proximity is more important for this region. Asian scholars have significantly less collaboration with distant partners suggesting that they prefer to choose co-authors from Asia. The coefficients for common language show no statistical results in both specifications. Asian countries share common languages with few countries around the world, thus, it is not significant in this subset for the Asian region.

5 Conclusions and discussions

This study contributes to the literature on the spatial analysis of geographical proximity by analyzing international collaboration in a multidisciplinary field. First, our results show that international collaboration in higher education research has intensified and increased rapidly in the last two decades. A greater share of the articles are being created in international collaboration, but the majority of the articles are still produced by co-authors from the same country. Second, in line with previous papers we confirm that the intensity of collaboration is dependent on geographical distance. Our most important findings are that the geographical proximity differs to a great extent between various world regions: the dense clustering of European countries makes the intensity of international collaboration far less dependent on geography and more on language commonality. Third, linguistic proximity plays a significant role mostly for European researchers leaving North American and Asian scholars less connected with international co-authors. The results are robust to various models. The statistical significance is not sensitive to the specifications used: the negative binomial and zero-inflated count model estimates lead to similar conclusions. The

various estimators and clustered standard errors of the regression coefficients are consistent, although they differ for some variables by over 15% across the different models.

Several contemporary trends in higher education are highlighted. First, geographical distance has become less relevant in some fields, but our findings are that it remains relevant in the field of higher education. However, we observe a slight decline of the role of geographical distance and language commonality in the formation of international collaboration between 2000-2008 and 2009-2017. We identify that the multidisciplinarity of the field determines the structure and the intensity of international collaboration. In contrast with previous research on international collaboration in Europe, we find that despite the encouragement of multidisciplinary research by increasing funding opportunities, the substantial influence of geography remains. Nonetheless, European higher education researchers are more inclined to collaborate with regional neighbors: the UK is an example of a country that has less research capacity but outperforms the US. Second, the share of internationally co-authored higher education publications is on the rise as is the mean distance between collaborating scholars. This implies that the field of higher education is becoming internationally oriented. We suggest that the American higher education community is more engaging in meso- and micro- level research, suggesting a higher rate of national rather than international collaboration, whereas European scholars emphasize analysis at the macro level and are more inclined to collaborate on the EU level. Unsurprisingly, the main centers in terms of knowledge production and research in higher education are in the major English-speaking developed countries, with significant strength in Western Europe and China. Overall, it is difficult to talk about coherence and solidarity in higher education community. The scholars are still geographically divided and form relatively isolated scientific communities.

References

Adams, J. (2012). Collaborations: The rise of research networks. *Nature*, 490:335–336.

Adams, J. (2013). International collaboration research. *Nature*, 497:557–560.

Almeida, J. A. S., Pais, A. A. C. C., and Formosinho, S. J. (2009). Science indicators and science patterns in Europe. *Journal of Informetrics*, 3(2):134–142.

Altbach, P. G. (2014). The emergence of a field: research and training in higher education. *Studies in Higher Education*, 39(8):1306–1320.

Bartneck, C. and Hu, J. (2010). The fruits of collaboration in a multidisciplinary field. *Scientometrics*, 85(1):41–52.

Boschma, R. A. (2005). Proximity and innovation: A critical assessment. *Regional Studies*, 39(1):61–74.

Correia, A., Paredes, H., and Fonseca, B. (2018). Scientometric analysis of scientific publications in CSCW. *Scientometrics*, 114(1):31–89.

Frenken, K., Hoekman, J., Kok, S., Ponds, R., van Oort, F., and van Vliet, J. (2009). Death of distance in science? A gravity approach to research collaboration. In *Innovation Networks*, chapter 3, pages 43–57. Springer Berlin Heidelberg.

Gates, A. J., Ke, Q., Varol, O., and Barabási, A. L. (2019). Nature's reach: Narrow work has broad impact. *Nature*, 575:32–34.

Glänzel, W. (2001). National characteristics in international scientific co-authorship relations. *Scientometrics*, 51(1):69–115.

Henriksen, D. (2016). The rise in co-authorship in the social sciences (1980-2013). *Scientometrics*, 107(2):455–476.

Hoekman, J., Frenken, K., and Tijssen, R. J. (2010). Research collaboration at a distance: Changing spatial patterns of scientific collaboration within Europe. *Research Policy*, 39(5):662–673.

Hoekman, J., Frenken, K., and van Oort, F. (2009). The geography of collaborative knowledge production in Europe. *Annals of Regional Science*, 43(3):721–738.

Jung, J. and Horta, H. (2013). Higher education research in Asia: A publication and co-publication analysis. *Higher Education Quarterly*, 67(4):398–419.

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

Kliegl, R. and Bates, D. (2011). International collaboration in psychology is on the rise. *Scientometrics*, 87(1):149–158.

Kosmützky, A. and Krücken, G. (2014). Growth or steady state? A bibliometric focus on international comparative higher education research. *Higher Education*, 67(4):457–472.

Kuld, L. and O'Hagan, J. (2018). Rise of multi-authored papers in economics: Demise of the 'lone star' and why? *Scientometrics*, 114(3):1207–1225.

Kuzhabekova, A., Hendel, D. D., and Chapman, D. W. (2015). Mapping global research on international higher education. *Research in Higher Education*, 56(8):861–882.

Leahey, E. (2016). From sole investigator to team scientist: Trends in the practice and study of research collaboration. *Annual Review of Sociology*, 42(1):81–100.

Leahey, E. and Reikowsky, R. C. (2008). Research specialization and collaboration patterns in sociology. *Social Studies of Science*, 38(3):425–440.

Leydesdorff, L. and Wagner, C. S. (2008). International collaboration in science and the formation of a core group. *Journal of Informetrics*, 2(4):317–325.

Newman, M. E. (2004). Coauthorship networks and patterns of scientific collaboration. *PNAS*, 101(1):5200–5205.

Rosenblat, T. and Mobius, M. (2004). Getting closer or drifting apart? *The Quarterly Journal of Economics*, 119(3):971–1009.

Tight, M. (2004). Research into higher education: An a-theoretical community of practice? *Higher Education Research and Development*, 23(4):395–411.

Tinbergen, J. (1962). Shaping the world economy. Suggestions for an international economic policy. *The 20th Century Fund: New York*.

Wagner, C. and Leydesdorff, L. (2005). Network structure, self-organization and the growth of international collaboration in science. *Research Policy*, 34(10):1608–1618.

Wagner, C. S., Park, H. W., and Leydesdorff, L. (2015). The continuing growth of global cooperation networks in research: A conundrum for national governments. *PLoS ONE*, 10(7):1–15.

Wagner, C. S., Whetsell, T. A., and Leydesdorff, L. (2017). Growth of international collaboration in science: revisiting six specialties. *Scientometrics*, 110(3):1633–1652.

Waltman, L., Tijssen, R. J., and van Eck, N. J. (2011). Globalisation of science in kilometres. *Journal of Informetrics*, 5(4):574–582.

Wang, L., Thijs, B., and Glänzel, W. (2015). Characteristics of international collaboration in sport sciences publications and its influence on citation impact. *Scientometrics*, 105(2):843–862.

Yokoyama, K. (2016). Reflections on the field of higher education: Time, space and subfields. *European Journal of Education*, 51(4):550–563.

6 Appendix

Table 8: The number of joint publications between top 25 countries in 2000-2017

	AU	BE	CA	CH	CN	DE	DK	ES	FI	FR	HK	IE	IT	JP	KR	MY	NL	NO	NZ	PT	SE	SG	UK	US	ZA
AU	0																								
BE	1	0																							
CA	16	1	0																						
CH	0	2	4	0																					
CN	11	3	4	0	0																				
DE	9	2	3	8	1	0																			
DK	2	0	1	0	2	6	0																		
ES	2	1	2	1	0	6	1	0																	
FI	3	5	3	0	7	5	2	3	0																
FR	0	2	2	1	0	1	0	3	0	0															
HK	12	1	5	0	18	1	0	0	0	0	0														
IE	4	0	0	2	0	1	0	1	0	1	0	0													
IT	2	2	0	3	1	1	1	3	1	3	0	0	0												
JP	3	0	3	0	1	1	0	1	0	0	2	0	0	0											
KR	1	0	0	0	2	0	0	0	0	0	2	0	0	0	0										
MY	12	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0									
NL	8	14	3	1	2	8	4	5	1	1	3	0	2	0	0	0	0								
NO	6	5	2	2	3	10	7	0	6	2	0	3	0	0	0	0	5	0							
NZ	33	0	3	0	3	0	1	0	3	0	4	1	0	2	1	3	3	0	0						
PT	6	1	1	0	4	2	1	3	3	0	3	3	2	2	1	0	2	2	1	0					
SE	6	0	2	1	3	1	5	3	9	1	1	0	0	2	0	0	4	12	0	2	0				
SG	6	0	1	0	3	0	0	0	0	1	5	0	0	2	0	1	4	0	1	0	0	0			
UK	96	10	31	6	29	14	6	15	6	8	10	9	5	6	3	7	24	15	19	9	16	1	0		
US	42	4	58	3	25	17	2	7	5	4	7	4	8	7	31	4	10	6	3	5	7	2	51	0	
ZA	8	0	3	0	0	1	1	1	0	0	0	1	0	0	2	0	3	3	3	2	5	0	17	5	0

AU - Australia; BE - Belgium; CA - Canada; CH - Switzerland; CN - China; DE - Germany; DK - Denmark; ES - Spain; FI - Finland; FR - France; HK - Hong Kong; IE - Ireland; IT - Italy; JP - Japan; KR - South Korea; MY - Malaysia; NL - Netherlands; NO - Norway; NZ - New Zealand; PT - Portugal; SE - Sweden; SG - Singapore; UK - United Kingdom; US - United States; ZA - South Africa

Table 9: Regressions results for all countries in 2000-2008 and 2009-2017

	Dependent variable:			
	N of co-authored papers			
	negative binomial			
	2000-2008		2009-2017	
	(1)	(2)	(3)	(4)
N of authors in country i	0.641^{***}	0.528^{***}	0.632^{***}	0.580^{***}
	(0.141)	(0.172)	(0.055)	(0.065)
N of authors in country j	0.635^{***}	0.604^{***}	0.514^{***}	0.495^{***}
	(0.169)	(0.171)	(0.050)	(0.051)
Distance, km	-0.359^{***}	-0.353^{***}	-0.265^{***}	-0.264^{***}
	(0.135)	(0.134)	(0.058)	(0.058)
Common language, 1 = Yes		0.470^{***}		0.242^{**}
		(0.138)		(0.068)
Constant	-3.369^{***}	-3.186^{**}	-3.277^{***}	-3.112^{***}
	(1.305)	(1.290)	(0.492)	(0.500)
Observations	252	252	905	905
Log Likelihood	−79.572	−79.001	−442.221	−441.200
Akaike Inf. Crit.	167.144	168.001	892.442	892.400

Standard errors in parentheses clustered at the country level.
*p<0.1; **p<0.05; ***p<0.01

Table 10: Regressions results for all countries in 2000-2008 and 2009-2017

	Dependent variable:			
	N of co-authored papers			
	zero-inflated count data			
	2000-2008		2009-2017	
	(1)	(2)	(3)	(4)
N of authors in country i	0.624***	0.495***	0.606***	0.541***
	(0.127)	(0.155)	(0.052)	(0.059)
N of authors in country j	0.263***	0.314**	0.404***	0.361***
	(0.073)	(0.145)	(0.048)	(0.050)
Distance, km	−0.299**	−0.302**	−0.208***	−0.220***
	(0.121)	(0.122)	(0.057)	(0.055)
Common language, 1 = Yes		0.411***		0.300*
		(0.105)		(0.158)
Constant	−1.740	−1.819*	−3.007***	−2.582***
	(1.109)	(1.083)	(0.458)	(0.463)
Observations	252	252	905	905
Log Likelihood	−89.711	−90.308	−490.300	−482.428

Standard errors in parentheses clustered at the country level.
*p<0.1; **p<0.05; ***p<0.01

Stanislav Avdeev
National Research University Higher School of Economics (Moscow, Russia)
Research Assistant, Center for Institutional Studies
E-mail: stnavdeev@gmail.com