

# Geographical distance in bibliometric relations within epistemic communities

Per Ahlgren · Olle Persson · Robert Tijssen

Received: 25 June 2012 / Published online: 26 July 2012  
© Akadémiai Kiadó, Budapest, Hungary 2012

**Abstract** Scientists collaborate increasingly on a global scale. Does this trend also hold for other bibliometric relations such as direct citations, cocitations and shared references? This study examines citation-based relations in publications published in the journal *Scientometrics* from 1981 to 2010. Different measures of Mean Geographical Distance (MGD) are tested. If we take all citation links into consideration, there is no indication of MGD increase, but when we look at maximum distances of each relation, a weak tendency of increasing MGD could be observed. One major factor behind the lack of growth of mean distances is the form of the distribution of citation links over distances. Our data suggest that the interactions might grow simultaneously for both short and long distances.

**Keywords** Direct citations · Co-authorships · Co-citations · Geographical distance · Shared references

## Introduction

Science is becoming increasingly dispersed in terms of where research is being done. Based on this, we should expect scientists to collaborate and communicate over ever longer distances. If not, distance would interfere negatively with scientific development, making scientists less familiar with research at distant locations and perhaps less effective in utilizing the competencies of their peers.

---

P. Ahlgren (✉)  
Department of e-Resources, University Library, Stockholm University,  
SE-106 91 Stockholm, Sweden  
e-mail: per.ahlgren@sub.su.se

O. Persson  
Department of Sociology, Inforsk, Umeå University, SE 901 87 Umeå, Sweden  
e-mail: olle.persson@soc.umu.se

R. Tijssen  
Centre for Science and Technology Studies (CWTS), Leiden University,  
PO Box 905, 2300 AX Leiden, The Netherlands  
e-mail: tijssen@cwts.leidenuniv.nl

What do we know about the role of geographical distance in scientific interaction? Already in the early 1990s, Luukkonen et al. (1992) showed a doubling of international institutional co-authorships in 10 years, and this collaboration trend is still strong (Leydesdorff and Wagner 2008). In terms of geographical distance, the globalization trend in science is even more striking. Recent studies show that the average collaboration distance per publication increased nearly five times from year 1980 to year 2009: from 334 to 1,553 km (Tijssen et al 2011; Waltman et al 2011). However, in spite of the fact that the globalization trend is strong, interaction still decreases with distance. Hennemann et al (2012) observed a strongly decreasing relation between distance and the probability of collaboration between organizations. Hoekman et al (2010) concluded that distance impedes collaboration, and they did not find evidence that the importance of distance for collaboration is declining over time. The authors also concluded that territorial borders impede collaboration, but that their effect has declined over time.

The distance effect is strong also at the local level. Many years ago, Allen (1984) showed that communication between engineers in research labs dropped exponentially down to 10 m of separation, and after that the probability of communication was very low and slightly decreasing. Geographical closeness and citation impact appears to be positively related within short distances. In a study of co-authors at Harvard University, Lee et al (2010) showed that the closer the distance between the first and last author of a publication got, the more likely the publication was cited. However, the maximum distance of separation was only 12 km in this study, and when one expands the proximity beyond the local level, international collaboration publications yields significantly more citations than domestic publications (Narin et al 1991).

According to Frenken et al (2009), there are at least five proximity dimensions that can be explored in scientific interactions: physical, cognitive, social, organizational, and institutional. Social interactions in terms of co-authorships, and in relation to distance, have been studied to some extent. However, what really keeps a network of scientists together on a global scale is the common reading of research literature. Yan and Sugimoto (2011) showed that citations between information science institutions generally decreased when the geographical distance between the institutions increased. Persson and Ellegård (2012) analyzed a citation network by time and space, another example of a study that involves citations and distance. To our knowledge, though, there are not many studies on distance in relation to citations. The following three research questions remain largely unanswered:

1. Do authors cite each other in an increasingly globalized way?
2. Is the structure of the knowledge base of a research field, in which publications are linked to each other by more recent citing publications, becoming more globalized?
3. Is the common use of information, in terms of reference lists of publications, of research front publications becoming more globalized?

The first question can be answered by analyzing direct citation links and the location of citing and cited cities. The second question can be answered by analyzing co-citations among cities of publishing, and the locations of co-cited cities. The third question can be analyzed by analyzing shared references between author cities represented in different publications, and the locations of those cities.<sup>1</sup> To place these three citation aspects in broader analytical perspective, they can be compared to spatial patterns regarding

---

<sup>1</sup> The number of shared references between two publications is equal to the bibliographic coupling strength for the two publications (Kessler 1963).

collaboration among cities. It is reasonable to assume that co-authorships will be geographically more concentrated than direct citations, shared references and co-citations, since connecting people by readership is much easier than making them collaborate. But we do not know the magnitude of these differences or the rate of change over time.

We will try to shed light on the three questions in a case study of publications published in the journal *Scientometrics*. This journal captures fairly well the citation patterns and cognitive interactions within an epistemic community that focuses on mostly bibliometric studies of research. The advantage of looking at one fairly coherent research network is that we can study distance effects within a common framework that enables us to follow changes over time more easily.

The remainder of this paper is organized as follows. Data and methods are treated in the next section, while the following section reports the results of the study. The final section contains a discussion, as well as conclusions.

## Data and methods

We retrieved from the Web of Science database all publications, of the document types *article* and *proceedings paper*, which were published in the journal *Scientometrics* between 1978, the starting year of *Scientometrics*, and 2010.<sup>2</sup> A total of 2,504 publications were retrieved. Since we need author affiliate addresses in this study, and cited references, we extracted each publication such that its Web of Science record has (a) either a C1 (Author Address) field or an RP (Reprint Address) field, and (b) a CR (Cited References) field. 2,424 of the 2,504 publications satisfied the conditions (a) and (b), and thereby a set of 2,424 publications is used in the study. Let  $P$  denote this set.

For each publication in  $P$ , we reduced all addresses occurring in the C1 and RP field of its record to city + country expressions (like “Leuven, Belgium”). These expressions were then standardized: variant expressions standing for the same city were mapped to a standard expression. For instance, “Amsterdam, The Netherlands” and “Wx Amsterdam, The Netherlands” were mapped to “Amsterdam, The Netherlands”. After standardization, duplicate city + country expressions within a publication were deleted. In order to obtain geographical distances between cities, a program was written that collected, for each standard city + country expression, values on latitude and longitude. The program collected the coordinates from two sources, *Google Maps API Web Services* and *Yahoo! PlaceFinder*. The total number of city + country expressions was 639. Yahoo! returned coordinates for all expressions, while Google failed to return coordinates for 4 expressions. For 38 of 635 (639 – 4) expressions, the distance between the Google and the Yahoo! coordinates was >50 km. We manually controlled the cases with distances >200 km, 18 cases. The outcome was that Google was more accurate in 14 of the cases, Yahoo! in 2, whereas two cases were such that we could not determine the accuracy order of the two tools. Based on these observations, and the observation made by Waltman et al. (2011) that Google seemed to be more accurate than Yahoo! regarding geocoding, we decided to use the coordinates provided by Google in all cases except the 4 + 2 cases mentioned above, where the Yahoo! coordinates were used. With values on latitude and longitude at our disposal, the geographical distance between two cities were calculated with the Haversine formula (Sinnott 1984), multiplied by 6,371, the mean radius of the earth in km.

<sup>2</sup> The following query was used:  $SO = (SCIENTOMETRICS) AND PY = 1978-2010 AND Document Types = (Article OR Proceedings Paper)$ .

In this work we consider four bibliometric relations: direct citations, co-citations, shared references, and co-authorships. The first three of these relations are citation-based, while the last one does not involve citations and concerns a single publication. We experimented with two approaches, **max** and **all**, with respect to the definition of geographical distance. Further, we use two indicators: Mean Geographical Distance (MGD) and Share of Links between Distinct Countries (SLDC). With regard to the latter indicator, a link with two identical (distinct) country expressions was assigned the score 0 (1). We calculate indicator values for several 5 year time periods: 1981–1985, 1986–1990, 1991–1995, 1996–2000, 2001–2005, and 2006–2010. We also use two 15 year periods, 1981–1995 and 1996–2010, to somewhat reduce the effect of lower number of publications from the first 10 years.

As for the MGD measure, there are several ways of defining geographical distance. The **max** approach means that one focuses only on the maximum distance between cities within a publication for co-authorships (Waltman et al. 2011), or between two publications for the citation-based relations. In addition to the **max** approach, the **avg** approach would be to calculate the mean distance within or between two publications, based on exactly the same distances as the **max** calculation. We will return to the **avg** approach in the end of the results section.

### Direct citations

We extracted cited references from the publications in  $P$  in order to generate direct citations or direct citation links, within  $P$ . A total of 7,545 citation links were generated, where the citing publication is published in the interval 1981–2010. Each link is from a  $P$  publication  $x$  to another  $P$  publication  $y$ , where  $y$  is represented in the reference list of  $x$ .

In the **max** approach, we define the maximum geographical distance between a citing publication  $x$  and a cited publication  $y$  as the maximum geographical distance between  $\gamma$  and  $\delta$ , where  $\gamma$  is a city + country expression obtained from the record of  $x$ , and  $\delta$  a city + country expression obtained from the record of  $y$ . For each of the 7,545 citation links, we recorded the maximum geographical distance and the pair of countries associated with the distance.<sup>3</sup> In that way, 7,545 links between a *citing* city + country expression and a *cited* city + country expression were obtained.

In the **all** approach, all links between citing city + country expressions and cited city + country expressions, together with their associated geographical distances, are taken into account. Obviously, this yields more links compared to the **max** approach, and the number of links is here 16,391. For each link, the distance was recorded, as well as its pair of countries.

For the assignment of links between citing city + country expressions and cited city + country expressions to publication years, and thereby to time periods, a given link was assigned to the publication year of the corresponding *citing* publication. For instance, a link from “Amsterdam, The Netherlands” to “Leiden, The Netherlands”, where the citing publication is published 2010 and the cited 1978, is assigned to 2010.

For a given time period, and for both **max** and **all**, we calculated MGD and SLDC over the city + country links assigned to the period.

<sup>3</sup> For the **max** approach, regardless of bibliometric relation, if there is more than one maximum distance, it is possible that the distances are associated with different pairs of countries. Therefore, in cases with more than one maximum distance, one of the distances was randomly selected, together with its associated pair of countries.

## Co-citations

A full scale co-citation study is not possible with our data set, since we do not have all publications that cite the publications in *Scientometrics*. However, we are able to generate a substantial number of co-citations from the direct citation links within the set  $P$  of *Scientometrics* publications. These co-citation links would be the same as shared inlinks (Björneborn and Ingwersen 2004).

We extracted each triple  $(z, x, y)$  of publications in  $P$  such that  $x$  and  $y$  are co-cited by  $z$  and such that  $z$  is published in the interval 1981–2010. This generated 24,399 triples, or the same number of co-citation links. For each triple, the maximum geographical distance between  $x$  and  $y$  (the two co-cited publications), defined in the same way as for direct citations, and the pair of countries associated with the distance, were recorded. Thereby, 24,399 triples  $(z, \gamma, \delta)$  such that  $\gamma$  is a city + country expression obtained from the record of  $x$  and  $\delta$  a city + country expression obtained from the record of  $y$  were obtained. For the **all** approach, triples (49,895) were obtained by taken into account, for  $z, x$  and  $y$ , all combinations of a city + country expression in  $x$  and a city + country expression in  $y$ , with associated distances and country pairs. Each considered triple, regardless of approach, was assigned to the publication year of its citing publication,  $z$ .

For a given time period, and for both **max** and **all**, MGD and SLDC were calculated over the triples assigned to the period.

## Shared references

We extracted each pair  $(x, y)$  of publications in  $P$  such that the number of shared cited references for  $x$  and  $y$  is at least 3, and such that  $x$  and  $y$  are published the same year, during the interval 1981–2010. Such a pair is bibliographically coupled, and the coupling strength for the publications is  $\geq 3$ . In this way, 1,182 pairs were generated. For each pair, the maximum geographical distance, defined in the same way as for direct citations, and the pair of countries associated with the distance, were recorded. Thereby, 1,182 links between two city + country expressions were obtained. Regarding the **all** approach, links (3,229) between city + country expressions, with distances and country pairs, were obtained analogously to how they were obtained for direct citations. Each considered link, regardless of approach, was assigned to the publication year of its two corresponding publications.

For both **max** and **all**, and for a given time period, MGD and SLDC were calculated over the city + country links assigned to the period.

## Co-authorships

For co-authorships and the **max** approach, we define the *maximum geographical distance* in a publication  $x$  as the maximum geographical distance between  $\gamma$  and  $\delta$ , where both  $\gamma$  and  $\delta$  are city + country expressions obtained from the record of  $x$ . For both **max** and **all**, if exactly one city + country expression,  $\gamma$ , is obtained from the record of  $x$ , we define the maximum geographical distance of  $x$  as 0. Further, we let  $(\gamma, \gamma)$  be the only city + country link in  $x$ .

For each publication in  $P$  published in the interval 1981–2010 (2,373 publications), the maximum distance of the publication was recorded, together with the pair of countries associated with the distance. Thereby, 2,373 links between city + country expressions were obtained. With regard to the **all** approach, links (2,852) were obtained by taken into

account all combinations of city + country expressions in  $x$ , with associated distances and country pairs. Each considered link, regardless of approach, was assigned to the publication year of its corresponding publication.

For both **max** and **all**, and for a given time period, MGD and SLDC over the links assigned to the period were calculated.

## Results

The first subsection describes the results for the four bibliometric relations, under both distance approaches. In the second subsection, we briefly exemplify how geographical distances can be used as link weights in a citation network.

The four bibliometric relations under **max** and **all**

Table 1 reports the number of observations and publications for the four bibliometric relations over eight time periods and across all studied years, for both the **max** and the **all** distance approach. It shows that the number of observations is low for shared references for the earliest 5 year period (1981–1985). This should be taken into account when the indicator values are interpreted for shared references within this period.

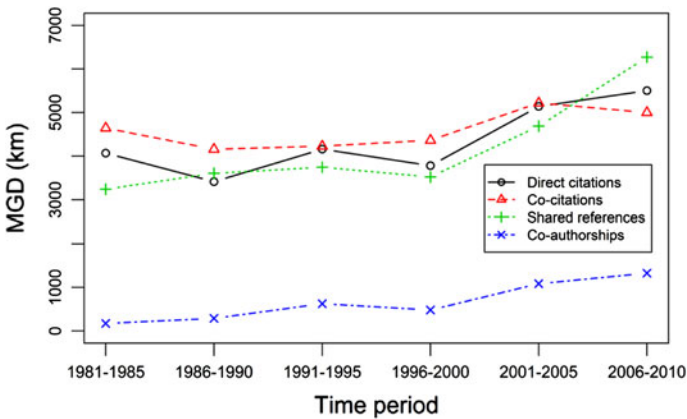
In Table 2, values on the indicators MGD and SLDC are given for the relations over time periods, for the **max** approach. For the six 5 year periods, visualizations of the outcome can be found in Figs. 1 (MGD) and 2 (SLDC). The MGD values for direct citations and shared references exhibit an increasing trend regarding the last three 5 year periods, an outcome that also holds for co-authorships (Table 2; Fig. 1). For all four relations, the MDG value is considerably higher for the period 1996–2010 than for the period 1981–1995, with the most pronounced increase, about 60 %, for shared references (Table 2). For SLDC, and the importance of national borders, the indicator values generally show an increasing trend across the six periods, regardless of relation (Table 2; Fig. 2). This means that the importance of national borders tends to decrease over time.

**Table 1** Frequency counts of observations and publications for the four bibliometric relations over eight time periods, and across all studied years (1981–2010). Numbers are given for both the **max** and the **all** distance approach

Time period	# Observations								# Publications
	Direct citations		Co-citations		Shared references		Co-authorships		
	Max	All	Max	All	Max	All	Max	All	
1981–1985	135	149	637	712	19	24	143	143	143
1986–1990	399	483	723	907	56	87	250	265	250
1991–1995	728	1,107	1,531	2,098	71	112	323	355	323
1996–2000	1,210	2,128	3,494	5,643	84	156	409	449	409
2001–2005	1,588	3,544	5,480	10,961	158	481	434	540	434
2006–2010	3,485	8,980	12,534	29,574	794	2,369	814	1,100	814
1981–1995	1,262	1,739	2,891	3,717	146	223	716	763	716
1996–2010	6,283	14,652	21,508	46,178	1,036	3,006	1,657	2,089	1,657
1981–2010	7,545	16,391	24,399	49,895	1,182	3,229	2,373	2,852	2,373

**Table 2** MGD and SLDC values for the four bibliometric relations over eight time periods, and across all studied years (1981–2010). Distance approach: **max**

Time period	Direct citations		Co-citations		Shared references		Co-authorships	
	MGD	SLDC	MGD	SLDC	MGD	SLDC	MGD	SLDC
1981–1985	4,068	0.69	4,643	0.69	3,241	0.58	174	0.02
1986–1990	3,412	0.61	4,157	0.67	3,605	0.61	289	0.04
1991–1995	4,159	0.73	4,229	0.76	3,744	0.76	627	0.11
1996–2000	3,781	0.73	4,364	0.83	3,521	0.73	481	0.10
2001–2005	5,139	0.81	5,219	0.88	4,688	0.81	1,085	0.18
2006–2010	5,503	0.87	5,000	0.91	6,271	0.89	1,328	0.22
1981–1995	3,913	0.69	4,302	0.72	3,625	0.68	418	0.07
1996–2010	5,079	0.82	4,953	0.89	5,807	0.86	1,055	0.18
1981–2010	4,884	0.80	4,875	0.87	5,537	0.84	863	0.15

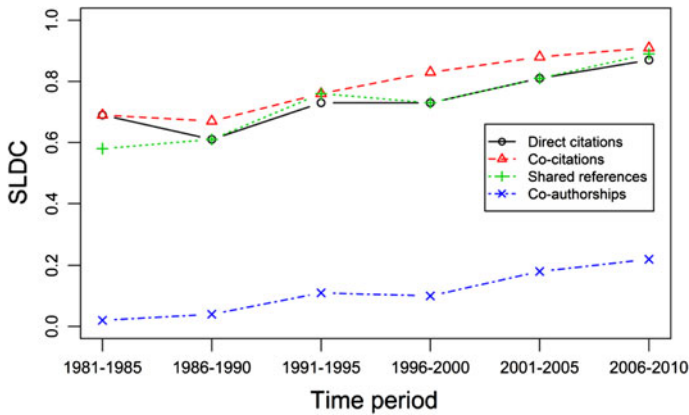


**Fig. 1** MGD values for the four bibliometric relations over six time periods. Distance approach: **max**

Also in this case, and for all four relations, are the SLDC values for 1996–2010 considerably higher than for 1981–1995, with 18 and 17 % units as the maximum increases (shared references and co-citations, respectively).

Clearly, one expects citation-based relations to be less dependent on distance than co-authorships, and for both MGD and SLDC, the indicator values are consistently much smaller for co-authorships compared to the values for the other three relations.

Turning our attention to the **all** distance approach and MGD, we observe interesting differences compared to the **max** case. The MGD values for direct citations, co-citations and shared references are more or less constant across the five first periods (Table 3; Fig. 3), whereas there is a clear increase in the corresponding **max** cases. Comparing the two 15 year periods 1981–1995 and 1996–2010, we note that MGD for co-citations decreases slightly, while the increase for direct citations is relatively small (Table 3). For shared references and co-authorships, however, more substantial increases are observed. Regarding SLDC, the general pattern is similar to the general pattern for the **max** approach (Table 3; Fig. 4): the importance of national borders tends to decrease over time.



**Fig. 2** SLDC values for the four bibliometric relations over six time periods. Distance approach: **max**

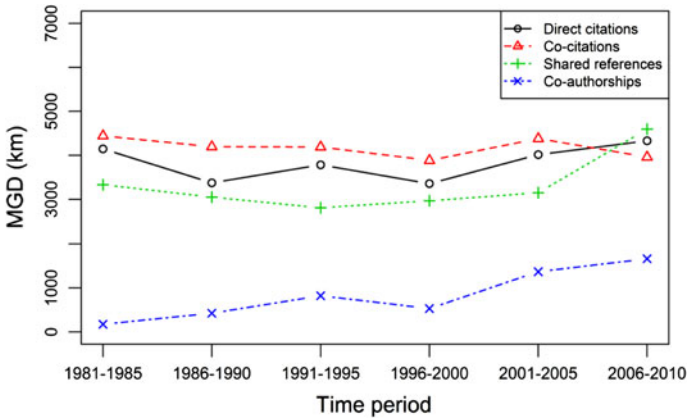
**Table 3** MGD and SLDC values for the four bibliometric relations over eight time periods, and across all studied years (1981–2010). Distance approach: **all**

Time period	Direct citations		Co-citations		Shared references		Co-authorships	
	MGD	SLDC	MGD	SLDC	MGD	SLDC	MGD	SLDC
1981–1985	4,145	0.69	4,444	0.67	3,331	0.58	174	0.02
1986–1990	3,374	0.61	4,196	0.67	3,050	0.54	425	0.06
1991–1995	3,784	0.70	4,190	0.78	2,807	0.70	823	0.15
1996–2000	3,356	0.71	3,885	0.82	2,967	0.69	532	0.11
2001–2005	4,016	0.75	4,381	0.85	3,149	0.73	1,370	0.25
2006–2010	4,330	0.82	3,961	0.86	4,596	0.84	1,664	0.31
1981–1995	3,701	0.68	4,240	0.73	2,958	0.62	563	0.10
1996–2010	4,113	0.78	4,052	0.86	4,279	0.82	1,345	0.25
1981–2010	4,069	0.77	4,066	0.85	4,188	0.80	1,136	0.21

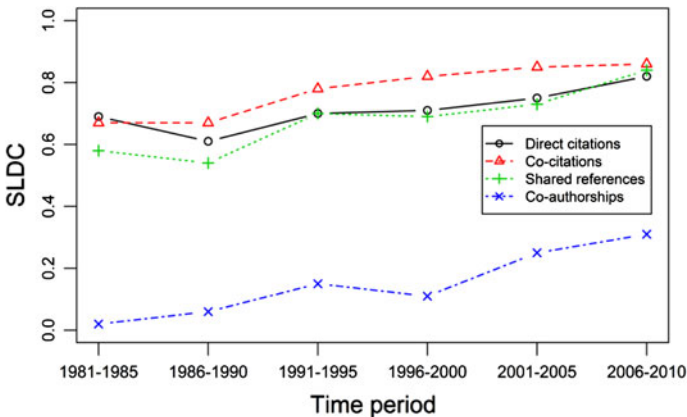
Also for the **all** approach, and not surprisingly, the indicator values for both MGD and SLDC are consistently much smaller for co-authorships compared to the values for the other three relations.

The fact that the mean distances do not increase for the citation-based relations (Fig. 3), while the share of cross country links, i.e., SLDC, do (Fig. 4), appears to be contradictory at a first glance, and calls for a closer analysis. In Fig. 5, we plot the number of direct citation links by different distance intervals for the two 15 year time periods, 1981–1995 and 1996–2010, and with respect to the **all** distance approach. When we compare the two curves, we observe a strong increase of the number of citation links between the periods up to 2,000 km (the rightmost endpoint of the interval 1,001–2,000). Between 2,000 and 5,000 km there is but a few citation links, because researchers are now separated by oceans and wastelands. A substantial increase of the number of citation links, in all intervals with a leftmost endpoint  $>5,000$  km, can also be observed. Thus, there is a simultaneous growth of short and long distances, which means that scientific networks will grow both locally and globally. Furthermore, this simultaneous growth explains why the MGD of direct





**Fig. 3** MGD values for the four bibliometric relations over six time periods. Distance approach: **all**

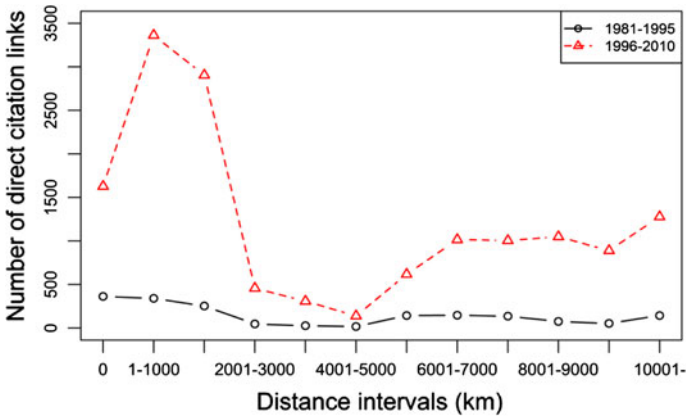


**Fig. 4** SLDC values for the four bibliometric relations over six time periods. Distance approach: **all**

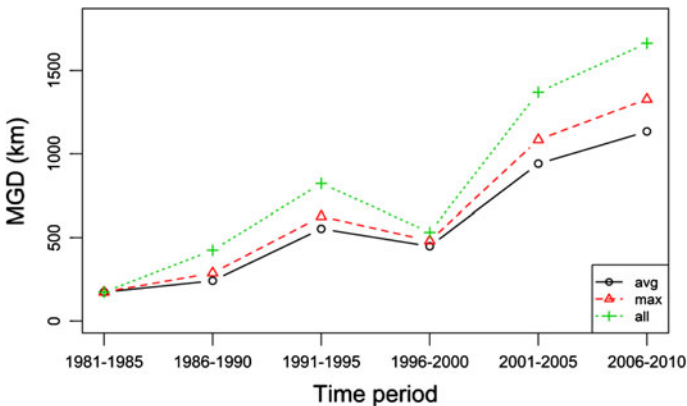
citations is almost constant over the two periods (Table 3): the relatively stronger increase of short distance direct citation links will even out the effect of the increase of long distance links. We call this the “local–global distance effect”, and we might assume that it will occur in other fields of science as well.

The strong increase of direct citation links between cities separated by 1 up to 1,000 km partly reflects a relative growth of citations within Europe. In the early period, 78 % of all direct citation links are between European cities in the distance class corresponding to the interval 1–1,000, and the corresponding share for the late period is 86 %.

We conclude this section by introducing a third geographical distance approach, **avg**, mentioned above in the section on data and methods. In a footnote, Waltman et al. (2011) mention the possibility of studying the mean collaboration distance within a publication in addition to the **max** approach. For co-authorships, we calculated the mean distance within each publication, and then calculated the average over all publications. In Fig. 6, the **avg** distance values, together with the corresponding values for **all** and **max**, over the six time

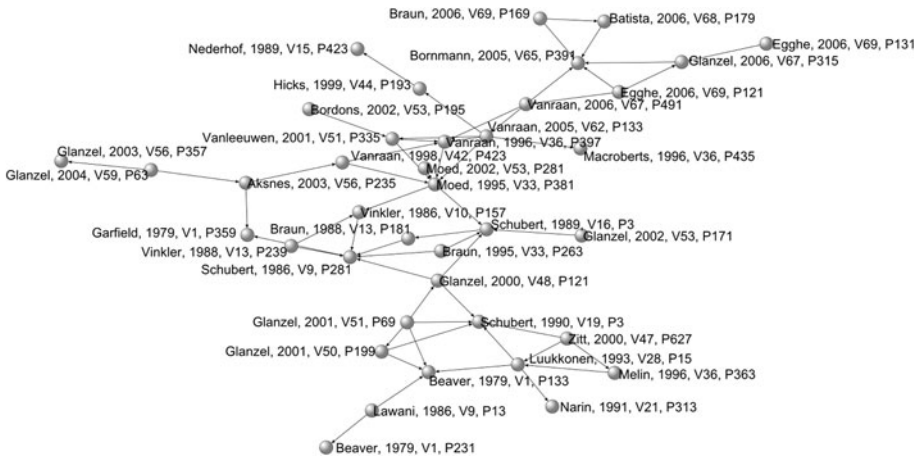


**Fig. 5** Frequency count of direct citations links by distance intervals (km). Distance approach: **all**



**Fig. 6** Three distance approaches with respect to co-authorships. MGD (km) values over six time periods

periods are indicated. We find that the **avg** approach yields shorter distances than the **max** approach. Clearly, this is what we should expect, since the two approaches are based on the same distances for each publication. However, the **all** approach is not calculated on the publication level, but by summing all distances over all publications, and then dividing the resulting sum by the number of distances. As can be seen from Fig. 6, **all** generally gives longer distances compared to **avg** and **max**, and at the end **all** appears to grow faster. The main reason for **all** to exhibit longer distances is that all distances is at work when the mean is calculated, but for **avg** and **max** they are not. Still, all three approaches co-vary strongly, and it is more of a matter of perspective which of them we choose. If we talk about the network rather than the publications, we would recommend using **all**, since this approach includes all relations of the network. We further note that the **avg** approach does not take into account that the number of observations (distances) within a publication varies across the publications. It might be more reasonable to take this variation into account, and calculate a weighted mean, as is done in the **all** approach.



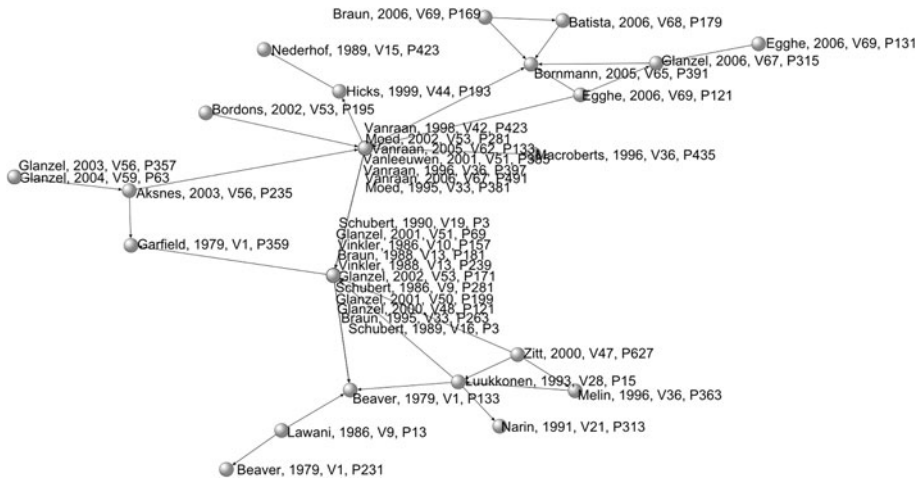
**Fig. 7** Direct citation network with 39 publications, published in *Scientometrics*, as nodes. Values of lines set to the Pajek option “Forget”

### Mapping direct citation networks

Traditionally, mapping bibliometric relations is not based on geographical distances, but rather on similarities in terms of co-citations or shared references. However, in a directed network consisting of citation links between publications, the citation links have the same weight at the outset. What will happen if geographical distance is introduced as weights in such a direct citation network? Consider the maps in Figs. 7 and 8, visualizations of network structures that were obtained by means of the Pajek software tool (the Kamada-Kawai algorithm). These maps show direct citation links among 39 publications (nodes) in our set, where each publication is cited by at least 20 other publications within this same set (indegree  $\geq 20$ , within the full network of publications). Both maps use distances as link weights. The distances are calculated on the basis of the citing and cited reprint (author) address (RP field). In Fig. 8, values of lines are set to dissimilarities (the Pajek option “Dissimilarities”). We observe two major changes between the maps: (a) some publications associated with Leiden and with citation links between them, including publications by Moed, Nederhof, van Leuwen and van Raan have in Fig. 8 been collapsed into one place, and (b) some publications associated with Budapest and with citation links between them, including publications by Braun, Glänzel, Schubert and Vinkler have in Fig. 8 been collapsed into one place. Thus, when distances are used as link weights in a citation network, and values of lines set are to dissimilarities, some interconnected nodes are superimposed and their citation links will not show. As a result, there is a link to the far left in the map of Fig. 7 between the two publications by Glänzel that is missing in Fig. 8.

### Discussion and conclusions

In this study, we focus on the role of geographical distance within a specific epistemic community, as represented by publications in the journal *Scientometrics*. We have discussed four types of bibliometric relations (direct citations, co-citations, shared references



**Fig. 8** Direct citation network of 39 publications published in *Scientometrics*. Values of lines set to the Pajek option “Dissimilarities”

and co-authorships) relative to geographical distance, while emphasizing the importance of national borders. We applied two approaches to the definition of geographical distance: **max** and **all**. We employed the two indicators MGD and SLDC. We found that citation-based relations are several times more globalized in terms of MGD compared to co-authorships.

There are some, but not uniform, indications of increased globalization in the field of scientometrics. If we take all citation links into consideration, there is no indication of MGD increase, but when we look at maximum distances of each relation, a weak tendency of increasing MGD could be noted. One major factor behind the lack of growth of mean distances is the form of the distribution of citation links over distances. Our data suggests that the interactions might grow simultaneously on both short and long distances in this particular field, which is dominated by researchers from Europe and the USA with local collaboration and trans-Atlantic ties. Thus, we have a local–global effect. For co-authorships in scientometrics we observed, like Waltman et al. (2011), a clear increase in MGD over time. The increase in MGD is indicative of further internationalization of knowledge production, where researchers outside Europe and the US will become more integrated into the epistemic community and its collaboration networks. The local effect is likely to remain a powerful determinant of proximity structures in contemporary science, as argued by Hennemann et al. (2012). Nonetheless, the global effect will become stronger if world science continues to evolve into a more uniform geographically distributed system where joint knowledge production across borders become common practice. With quantitative performance indicators becoming accepted tools of practice in science policy and research evaluation in more and more countries, the geographical composition of the scientometrics community will broaden its reach accordingly. The production of scientometric research papers, of citations of relevant publications, will reflect this cognitive expansion. Based on this scenario we should expect all proximity measures to increase over the foreseeable future, especially the MGD measures.

Clearly, there are limitations inherent in bibliometric data (Bornmann et al. 2011). For instance, it is not necessarily so that the cities listed in a publication reflect the locations where the research was carried out. One should take this into account when considering the validity of an indicator like MGD, even if we have no reason to believe that the indicated problem is extensive.

The various indicators applied in this paper reflect different features of cognitive and geographical proximity. The joint analysis across a suite of related indicators provides a proximity profile with obviously analytical value compared to a single leading indicator like MGD **all**. There is room for further development of these profiles. Notably, the binary nature of the SLDC (i.e., the same country or not) is a crude measure, which could be supplemented by analogous measures of other geographical zones (e.g., within or outside Europe).

The agenda for future research on this topic should also include a closer look at both the dynamics of globalization processes and its key determinants. Obviously one would like to ascertain the robustness of the findings in this journal-based case study by tracking global trends, as well as those in related academic journals and neighboring subfields of science such as library and information science, research policy management, and social studies of science. From a sociological point of view, it would be interesting to identify general patterns and trends within the publication of newcomers, as well as those multidisciplinary researchers that are also active in other subfields. The results would shed light on the intricate co-evolving relationship between knowledge influx and globalization processes within an epistemic community.

## References

- Allen, T. J. (1984). *Managing the flow of technology: Technology transfer and the dissemination of technological information within the R&D organization*. Cambridge: MIT Press.
- Björneborn, L., & Ingwersen, P. (2004). Toward a basic framework for webometrics. *Journal of the American Society for Information Science and Technology*, 55(14), 1216–1227.
- Bornmann, L., Leydesdorff, L., Walch-Solimena, C., & Ettl, C. (2011). Mapping excellence in the geography of science: An approach based on Scopus data. *Journal of Informetrics*, 5(4), 537–546.
- Frenken, K., Hardeman, S., & Hoekman, J. (2009). Spatial scientometrics: Towards a cumulative research program. *Journal of Informetrics*, 3(3), 222–232.
- Hennemann, S., Rybski, D., & Liefner, I. (2012). The myth of global science collaboration—Collaboration patterns inepistemic communities. *Journal of Informetrics*, 6(2), 217–225.
- 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(5), 662–673.
- Kessler, M. M. (1963). Bibliographic coupling between scientific papers. *American Documentation*, 14(1), 10–25.
- Lee, K., Brownstein, J.S., Mills, R.G., & Kohane, I.S. (2010). Does collocation inform the impact of collaboration? *Plos One*, 5(12).
- Leydesdorff, L., & Wagner, C. S. (2008). International collaboration in science and the formation of a core group. *Journal of Informetrics*, 2(4), 317–325.
- Luukkonen, T., Persson, O., & Sivertsen, G. (1992). Understanding patterns of international scientific collaboration. *Science, Technology, & Human Values*, 17(1), 101–126.
- Narin, F., Stevens, K., & Whitlow, E. S. (1991). Scientific cooperation in Europe and the citation of multinationally authored papers. *Scientometrics*, 21(3), 313–323.
- Persson, O., & Ellegård, K. (2012). Torsten Hägerstrand in the citation time web. *The Professional Geographer*, 64(X), 1–12.
- Sinnott, R. W. (1984). Virtues of the haversine. *Sky and Telescope*, 68(2), 159.
- Tijssen, R. J. W., Waltman, L., & van Eck, N. J. (2011). Collaborations span 1,553 kilometres. *Nature*, 473(7346), 154.

- Waltman, L., Tijssen, R. J. W., & van Eck, N. J. (2011). Globalisation of science in kilometres. *Journal of Informetrics*, 5(4), 574–582.
- Yan, E. J., & Sugimoto, C. R. (2011). Institutional interactions: Exploring social, cognitive, and geographic relationships between institutions as demonstrated through citation networks. *Journal of the American Society for Information Science and Technology*, 62(8), 1498–1514.