



# Characterizing research leadership on geographically weighted collaboration network

Chaocheng He<sup>1,2</sup> · Jiang Wu<sup>1</sup>  · Qingpeng Zhang<sup>2</sup>

Received: 24 July 2020 / Accepted: 4 March 2021 / Published online: 20 March 2021  
© Akadémiai Kiadó, Budapest, Hungary 2021

## Abstract

Research collaborations, especially long-distance and international collaborations, have become increasingly prevalent worldwide. Recent studies highlighted the significant role of research leadership in collaborations. However, existing measures of the research leadership do not take into account the intensity of leadership in the co-authorship network. More importantly, the spatial features, which influence the collaboration patterns and research outcomes, have not been incorporated in measuring the research leadership. To fill the gap, we construct an institution-level weighted co-authorship network that integrates two types of weight on the edges: the intensity of collaborations and the spatial score (the geographical distance adjusted by the cross-linguistic-border nature). Based on this network, we propose a novel metric, namely the spatial research leadership rank, to identify the leading institutions while considering both the collaboration intensity and the spatial features. The leadership of an institution is measured by the following three criteria: (a) the institution frequently plays the corresponding rule in papers with other institutions; (b) the institution frequently plays the corresponding rule in longer distance and even cross-linguistic-border collaborations; (c) the participating institutions led by the institution have high leadership status themselves. Harnessing a dataset of 323,146 journal publications in pharmaceutical sciences during 2010–2018, we perform a comprehensive analysis of the geographical distribution and dynamic patterns of research leadership flows at the institution level. The results demonstrate that the SpatialLeaderRank outperforms baseline metrics in predicting the scholarly impact of institutions. And the result remains robust in the field of Information Science and Library Science.

**Keywords** Social network analysis · Spatial scientometrics · Research leadership

---

✉ Jiang Wu  
jiangw@whu.edu.cn

✉ Qingpeng Zhang  
qingpeng.zhang@cityu.edu.hk

<sup>1</sup> School of Information Management, Wuhan University, Wuhan, Hubei, China

<sup>2</sup> School of Data Science, City University of Hong Kong, Kowloon, Hong Kong, China

## Introduction

In the “Big Science” era, research collaboration plays an important role in knowledge creation, iteration, and dissemination (Gazni et al., 2012; Katz & Martin, 1997). Thanks to the increasing scale and complexity of scientific projects, we have witnessed a rapid growth in the frequency and influence of collaborative research (Chen et al., 2019; Wang et al., 2013). This trend could be evidenced by the increase in research collaborations at a distance (Bercovitz & Feldman, 2011; Hoekman et al., 2010).

Research collaborations involve not only explicit but also tacit knowledge, which cannot be shared viacommunication technologies (Holste & Fields, 2010). The coordination activities in collaboration projects also require frequent mobility of researchers, such as sharing lab facilities and exchanging data/samples (Hoekman et al., 2010). Geographical distance remains significant in research collaboration, especially those larger-scaled projects, which often lead to more impactful results (Larivière et al., 2015). It is essential to gain an in-depth understanding of the trend towards long-distance and even international collaborations for the following reasons. First, it enhances the quality of research by combining expertise and resources and sharing the cost (Hoekman et al., 2010). Compared with single-authored publications, collaborative work often leads to more novel and higher internal quality control (Fernandez et al., 2016). Collaborative work does go with increasing self-citation, but the increase is much weaker than that of foreign citations (Glänzel & Thijs, 2004). Furthermore, these benefits increase as the distances between collaborators increase, as potential collaboration partners are more likely to be found within a larger radius (Hoekman et al., 2010). Indeed, international collaborative research, particularly those with elite partners (Tang, 2013) from high-performing countries (Glänzel & Schubert, 2001), tends to be of higher quality and more significant impact than local collaborators (Guan et al., 2016; Jiang et al., 2018). Second, significant public policies and expenditures have been established and allocated to facilitate long-distance research collaborations (Hoekman et al., 2010). For example, China’s 12th Five Year (2011–2015) Plan for Science and Technology Development stated that China would actively take part in international science and technology organizations and international research collaborations (Wang & Wang, 2017). Similarly, the European government attempts to construct a *European Research Area* by coordinating regional, national, and EU research activities (Hoekman et al., 2010) to improve European states’ internal consistency and to break down barriers to innovation. Third, many countries and regions are making great efforts to attract overseas talents, which further facilitate international collaborations (Baruffaldi & Landoni, 2012). For example, the 1000-talent program of China has attracted more than 7000 full-time and part-time senior scholars by 2018. Many are the leader in the field, and 22.5% are non-Chinese researchers (Jia, 2018). Within the first five years of the 1000-talent program, China has attracted more full professors than the past 30 years combined.<sup>1</sup> These overseas talents bridged the gap between the research communities in two countries and significantly improved both the intensity and quality of international collaborations.

Existing literature on spatial research collaboration has two main limitations. First, most studies assume that the research collaborations are homogeneous while ignoring the existence of leadership. It is sensible that the first and corresponding author(s) play a leading role in the collaboration. Thus their relationship with other authors should be more potent

<sup>1</sup> <http://kfq.job1001.com/>.

than those among others (Wang et al., 2014). Several studies examine the research leadership (González-Alcaide et al., 2017; Wang et al., 2014). However, they either do not consider the social network structure of the co-authorship network (Chaocheng et al., 2019) or ignores the intensity of collaborations while measuring the leadership (Zhou et al., 2018). Second, a collaboration relationship is assumed to be evenly distributed to authors of the same article without considering spatial features between the institutions. Existing algorithms for identifying key research entities in collaborations mainly focus on the topological features of the entities in the collaboration network (Zhou et al., 2018), without incorporating the spatial features into the network. Collaboration at a longer distance has a higher impact than that at a short distance (Hoekman et al., 2010). Identifying important research entities in research collaboration from a spatial perspective becomes necessary and essential to knowledge creation and dissemination (Forman & van Zeebroeck, 2019; Wu, 2013).

To fill the above research gaps, we extend the literature as follows. First, we model the collaboration relationships as a directed network at the institution level, where the direction of an edge indicates the leadership flow between the two institutions. Second, we incorporate the spatial features between institutions as a weight on edge in the collaboration network. Third, based on the constructed geographically weighted collaboration network, we propose a novel metric, namely the spatial research leadership rank (SpatialLeaderRank), to identify the leading institutions while considering both the collaboration intensity and the spatial features. More specifically, an institution is considered with higher leadership status according to the following three criteria: (a) the institution frequently plays the corresponding role in papers with other institutions; (b) the institution frequently plays the corresponding role in longer distance and even cross-linguistic-border collaborations; (c) the participating institutions led by the institution have high leadership status themselves. We exemplify and validate the proposed SpatialLeaderRank metric using the journal publications in the pharmaceutical sciences, a field that has witnessed a dramatic increase in collaborations between multinational scientists in both academia and the biotechnology sector because of the need for very diverse expertise from various disciplines and the rise of R&D outsourcing (Herrling, 1998; McKelvey et al., 2003; Plotnikova & Rake, 2014). We also examine the SpatialLeaderRank in Information Science and Library Science as the robustness check.

The remainder of this paper is organized as follows. Firstly, we review related literature. Secondly, we describe the data and methodologies. Thirdly, we present the results. Fourthly, we conclude the paper. Finally, we discuss the limitations and future directions.

## Related literature

### Spatial research collaboration

The spatial co-presence remains significant in carrying out the research collaboration tasks, though the hindering effect of spatial distance to research collaboration has declined due to the advancement in communication technologies (Hoekman et al., 2010). However, research collaboration involves not only explicit but also tacit knowledge. Explicit knowledge can be easily articulated or exchanged with communication technologies. Tacit knowledge, such as the abilities, developed skills, experience, and undocumented processes, is embedded in one's experience and values (Holste & Fields, 2010). In research

collaboration, tacit knowledge can be transferred between researchers by personal contact. However, it cannot be passed via communication techniques in formulas, diagrams, verbal descriptions, or instructions (Collins, 2001). Face-to-face interaction enables researchers to conduct intense and complex forms of communication, which greatly facilitate the transfer of tacit knowledge (Collins, 2001). What's more, the coordination activities in large-scale collaboration projects involve frequent mobility of researchers, such as sharing lab facilities, meetings, seminars. This imposes frequent and considerable travel time and economic costs. As the scale, complexity, and period of collaboration projects increase, the tacit knowledge sharing tends to be more essential, and the coordination activities tend to be more costly. The ongoing COVID-19 pandemic poses new challenges to research collaborations (Callaway, 2020), particularly in the pharmaceutical field (Yu et al., 2020), due to the travel bans, quarantine and disruption of the supply chain system. Therefore, spatial distance remains significant in research collaboration, especially those with large scale, high complexity, and long period.

The existing research on spatial research collaboration mainly focuses on the spatial pattern of research collaboration and the role of geographical proximity on research collaboration. The spatial pattern of research collaboration has been systematically studied at multiple levels. A recent study illustrated the spatial patterns of international knowledge flows and evaluated the effect of various factors, including the geographical factor (Gui et al., 2018). The establishment of research alliances with more developed countries constituted a critical mechanism, which integrated developing countries into the global research community (González-Alcaide et al., 2017). Researchers in small European states were found to be less homogeneously collaborating with both domestic and foreign partners (Ukrainski et al., 2014). Research activities in Brazil were found to be spatially heterogeneous. A geographical decentralization process of scientific research activities across the country was needed to stimulate the development of those privileged areas (Sidone et al., 2017). The role of geographical proximity has been explored in various fields. Plotnikova and Rake (2014) examined the country-level determinants in pharmaceutical research and found that the geographical distance was negatively associated with the research collaboration. In humanities, arts, and social sciences, the geographical distance was found to be critical to the collaborative activities (Luo et al., 2018). Similar conclusions could be drawn from the field of ecology (Parreira et al., 2017) and immunology (Lander, 2015).

There are three main types of drivers of long-distance and cross-border collaboration (Jeong et al., 2014): *input factor*, *motivation*, and *informal communication*. A typical driver of *input factor* is academic excellence. Academic excellence may not only lead to superior synergy through cross-border collaboration but also lower the cost of searching cross-border partners due to the high reputation of researchers with academic excellence (Jones et al., 2008). As for *motivation*, researchers at low academic levels tend to undertake long-distance and cross-border collaboration for higher academic productivity and quality (Laband & Tollison, 2000). As for *informal communication*, it's widely acknowledged that most research collaborations begin informally and are triggered by informal communication. Particularly, informal face-to-face communication plays an important role in the formulation of long-distance and cross-border collaborations (Jeong et al., 2011). Under some conditions, the geographical distance may not as significant as usual to research activities. For instance, once social relationships are established, patent inventors can remain socially close in citing patents even though they are geographically separated (Agrawal et al., 2006). Collaboration with local elite partners may lead to higher research quality than long-distance or international collaboration (Tang, 2013). However, most studies show that research quality is positively

**Table 1** Selected empirical studies about collaboration at a distance and research performance

	Data source	Country	Research scope	Method	Results
Wagner et al. (2017)	Web of Science	Global	Six fields	Descriptive and social network analysis	+Quality
Wagner et al. (2019)	Web of Science and Scopus	Global	All fields	Modeling	+Quality
Kwiek (2020a)	EUROAC	European Countries	Eight fields	Modeling	+Productivity
Kwiek (2020b)	Scopus	European Countries	Six fields	Descriptive and social network analysis	+Quality
Persson et al. (2004)	Web of Science	Global	All fields	Descriptive and social network analysis	+Quality
Fang et al. (2020)	UTD24 Journals	China	Business fields	Modeling	+Quality
Barjak and Robinson (2008)	Survey	European Countries	Biology	Modeling	+Quality; +Productivity
Larivière et al. (2015)	SCIE, SSCI, AHCI	Global	All fields	Modeling	+Quality

*EUROAC* European Academic Profession: Responses to Societal Challenges, *UTD24* 24 top journals proposed by the University of Texas at Dallas, *SCIE* SSCI, *AHCI*: Science Citation Index Expanded, Social Sciences Citation Index, Arts & Humanities Citation Index

correlated with international collaboration. Table 1 summarizes the relevant literature reporting the correlation between collaboration at a distance and research performance. Research collaboration at a distance has become increasingly prevalent (Bercovitz & Feldman, 2011; Hoekman et al., 2010). Given the abundant evidence that the spatial features influence the research performance, existing research on collaboration networks did not account for the spatial patterns and mainly assumed that the edges are homogeneous in the collaboration network.

## Research leadership

Research leadership is associated with the capability to integrate the research resources, such as knowledge, expertise, and equipment, to launch, advance, and sustain research projects (Wagner et al., 2001), and is associated with high production as well as quality (Klavans & Boyack, 2010). Research leadership has been a topic of perennial interest in clinical research (Campbell et al., 2001), entrepreneurial organizing (Hansson & Mønsted, 2008), social network analysis (Babbar et al., 2019), etc. Research leadership can be classified into three sub-classes (Klavans & Boyack, 2010): publication leadership, reference leadership (Calero-Medina et al., 2008), and thought leadership (Klavans & Boyack, 2008). From the bibliometrics perspective, a typical way to measure research leadership is to examine the authorship position (Chinchilla-Rodríguez et al., 2019; Quan et al., 2019). The authorship positions in the byline reflect each author's relative contribution (Liu & Fang, 2014). Despite the disciplinary bias, it is commonly acknowledged that the first author and the corresponding author often lead the research collaboration and make a major contribution (Wang & Wang, 2017). In the biomedical field, the first author is often an early-career researcher assigned to carry out the research and write the research paper. Simultaneously, other non-leading co-authors act in more specialized roles (Sekara et al., 2018). To limit irresponsible authorship listing, the International Committee of Medical Journal Editors (2018) establishes a series of authorship criteria in medical journals. Notably, it elaborates on the corresponding author's role that the corresponding author is responsible for communication with the journal during the submission, peer review, and publication process. The corresponding author should be available during the whole process to timely respond to any editorial queries. Therefore, corresponding authors should be responsible for both scientific and non-scientific contributions. Moreover, the corresponding author's responsibility is more prominent with the increase of collaboration scales, the growing complexity and depth of the research (Hemlin et al., 2013).

Recently, the term “research guarantor” has been widely adopted as an indicator of research leadership (Moya-Anegón et al., 2013). The concept of “research guarantor” was initially introduced by Rennie et al. (1997). Literature does not claim that an individual corresponding author is the research guarantor, but rather the research group/institution to which the corresponding author belongs is the research guarantor. In collaboration, the research guarantor contributes substantially and makes sure the integration of the entire project. Thus, the research guarantor has a special role in collaboration, including negotiating the authorship within the team, obtaining the corresponding authorship, and receiving more credit (Moya-Anegón et al., 2013; Sánchez-Jiménez et al., 2017). Building on the above studies, since we are studying research leadership at the institution level, we employ the corresponding author's institution as an indicator of research leadership.

## PageRank and LeaderRank

PageRank was initially proposed by Google to rank the importance of webpages (Brin & Page, 1998). There is a boom in its variation and application in a broad set of fields in the following decades. PageRank has also been widely applied to analyzing the research collaboration network. Liu et al. (2005) transformed each undirected edge in the research collaboration network into a set of bi-directional, symmetrical edges and defined modification of PageRank, namely the AuthorRank. Fiala et al. (2008) modified the PageRank by incorporating both citation and co-authorship graph property.

Successful as it is, PageRank has several drawbacks (Gleich, 2015). The stability of ranking and the robustness to noise and manipulation vary given different parameters (Lu et al., 2016). Moreover, if there are disconnected components in the network, the ranking result is not unique. To this end, Lu and Zhou (2011) proposed the LeaderRank, an adaptive and nonparametric algorithm, by adding a ground node that bi-directionally connected to every other node and then performing random walks. As a result, LeaderRank has a faster convergence rate, higher stability for noisy data, and more robustness to manipulations (Li et al., 2014). LeaderRank is further applied to identify the influential nodes in complex products and systems (Li et al., 2019); in power grids (Zhou et al., 2019); in manufacturing services (Wu et al., 2019). Notably, in the field of “Library & Information Science”, Xu et al. (2020) found that LeaderRank had the best performance in ranking science and technology citation data, compared with other 17 network-based metrics. However, it is noteworthy that all the previous variants of PageRank, including LeaderRank, only consider the topological features of nodes while ignoring other non-topological features, especially the spatial features that are very important to the academic performance and impact.

## Data and methodology

### Data collection

We leverage the Web of Science Core Citation (WOS) Database to perform a data collection. Specifically, following (Plotnikova & Rake, 2014) we collect publications in categories related to pharmaceutical study during 2010–2018. To check the robustness and have a comprehensive understanding, we also collect the publications in the category of “Information Science & Library Science (ISLS)” and compare the results in these two very different fields. More precisely, the data was retrieved using the search term “WC=A AND PY=B”, where A is the sub-categories in pharmaceutical sciences, and B is 2010–2018. We restrict publications of journal articles in English and exclude other non-journal publications such as meeting abstracts, letters, editorial materials, or reviews. We further restrict the data and sample 323,146 publications in pharmaceutical sciences and 28,158 publications in ISLS with at least two co-institutions. It is noteworthy that in the dataset of WOS, the problem of missing information from the address field remains significant even for recent publications (Liu et al., 2018). This problem may affect the validity of the address-based metric, SpatialLeaderRank. Following (Liu et al., 2018), we collect the publications with missing address information using the search term (Query #1) NOT (Query #2), where (Query #1) is the same as the former search term, “WC=A AND PY=B”, and (Query #2) is  $AD=(A^* OR B^* OR C^* OR D^* OR E^* OR F^* OR G^* OR H^* OR I^* OR J^* OR K^* OR L^* OR M^* OR N^* OR O^* OR P^* OR Q^* OR$

R\* OR S\* OR T\* OR U\* OR V\* OR W\* OR X\* OR Y\* OR Z\* OR 0\* OR 1\* OR 2\* OR 3\* OR 4\* OR 5\* OR 6\* OR 7\* OR 8\* OR 9\*). We obtain 6074 and 2613 address-missing publications in English in pharmaceutical sciences and ISLS, respectively. Most of the address-missing publications are either single-authored or without abstract and full text and thus are excluded from the data. We further concentrate on 540 and 77 multi-authored publications with abstracts in pharmaceutical sciences and ISLS, respectively. We manually obtained the full address information of the authors of these papers by searching them on Google. Finally, we extract 2459 institutions in pharmaceutical sciences and 841 institutions in ISLS, which have been the primary affiliation of the corresponding author for at least one paper (with multiple institutions) in each year. To disambiguate institutions' names, since we are studying at the institution level and WOS provides a complete list of authors and their institutions, we first remove the duplicates, next sort them alphabetically, then manually disambiguate them. We utilize Google Map to obtain the latitude, longitude, and country of each institution. For an institution with multiple addresses, we only adopt the Google Map's first search result. The linguistic data is from Melitz and Toubal (2014). Note that the Chinese Academy of Sciences (CAS) has over 100 branches distributed nationwide, covering six academic divisions: Earth Sciences, Chemistry, Information Technological Sciences, Life Sciences & Medical Sciences, Mathematics & Physics, and Technological Sciences. In this paper, we only focus on pharmaceutical sciences and ISLS. And the majority of the pharmaceutical (63%) and ISLS (82%) publications are published by the branches in Beijing, where CAS's headquarter is located. Following Fu et al. (2011) and Li et al. (2009), we pooled the different branches as one heading in Beijing. We do the same for other mega organizations, including Russian Academy of Sciences, French National Centre for Scientific Research (CNRS).

## Geographically weighted and directed network

### Measurement of research leadership flow

There are mainly three approaches to share the credit of a co-authored paper to each contributor (Gauffriau et al., 2007). From an institution-level perspective, the first approach is all counting approach, which attributes one full credit to each unique collaboration institution. The second approach is fractional counting, where one credit is shared by all institutions. The third approach is straight counting, where only the most prominent institution receives one full credit, while others receive none. Considering we are studying research leadership, we adopt straight counting and attribute one full credit to the leading institution (the corresponding author's institution). For those papers with multiple corresponding authors and multiple leading institutions, we share the one full credit to the leading institutions in each co-authored paper. In this paper, the notion and measurement of research leadership flow are based on (Chaocheng et al., 2019). Specifically, for a paper  $p$  with leading institution number  $LN_p$  and total institution number  $TN_p$ , each leading institution receives  $\frac{1}{LN_p}$  credit. Therefore, the research leadership flow intensity between one leading institution  $a$  and one participating institution  $b$  is defined as,

$$RLF_{p,ab} = \frac{1}{LN_p} \times \frac{1}{TN_p} = \frac{1}{LN_p \times TN_p}, \quad (1)$$

And the aggregated research leadership intensity between institution  $a$  and  $b$  is defined as,



$$RLF_{ab} = \sum_{p=1}^{P_{ab}} RLF_{p,ab}, \tag{2}$$

where  $P_{ab}$  is the number of co-authored papers by  $a$  and  $b$ , where  $a$  is the leading institution. And the research leadership mass of  $a$ ,  $RLM_a$  (the total research leadership intensity between leading institution  $a$  and all other collaboration institutions) is defined as

$$RLM_a = \sum_{b=1}^{B_a} RLF_{ab}, \tag{3}$$

where  $B_a$  is the number of participating institutions that  $a$  has led.

### Measurement of spatial score

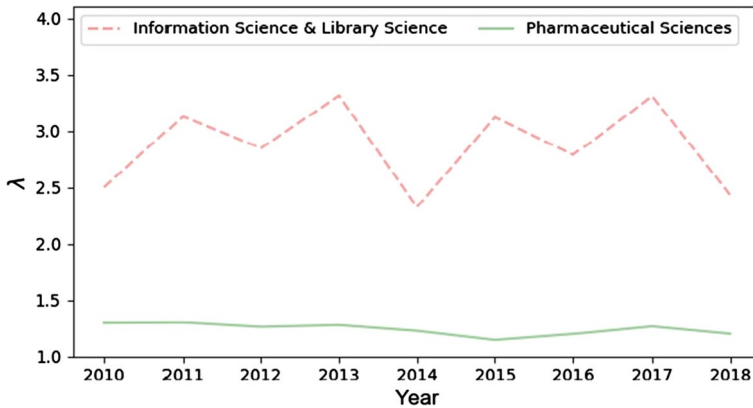
Boschma (2005) identifies five notions of proximity (geographical, cognitive, institutional, organizational, and social) on research collaboration and interactive learning. Following Boschma (2005), a broad set of literature employs the gravity model and finds that geographical proximity and the other proximities are hinder determinants to the formation of research collaboration (Fernandez et al., 2016; Zhang & Guo, 2017). In the gravity model, the estimation equation is as follows

$$I_{ab} = \beta_0 + \beta_1 \lg \text{Pubmass}_a + \beta_2 \lg \text{Pubmass}_b + \beta_3 \lg \text{Geographical}_{ab} + \beta_4 \text{Institutional}_{ab} + \sum_{k=5}^K \beta_k \lg s_k + \epsilon_{ab}, \tag{4}$$

where  $I_{ab}$  denotes the collaboration intensity, measured by the number of co-publications,  $\text{Pubmass}_a$  and  $\text{Pubmass}_b$  denote the number of previous publications of institutions  $a$  and  $b$ , respectively.  $\text{Geographical}_{ab}$  is geographical proximity, which is measured by the spatial distance between two institutions.  $\text{Institutional}_{ab}$  is the institutional proximity. Institution refers to common habits, cultural norms. Institutional proximity can be measured by a dummy variable, which takes 1 if two institutions are from the same country (Fernandez et al., 2016) or linguistic area (Gui et al., 2018; Zhang & Guo, 2017), and 0 otherwise.  $s_k$  denotes other dimensions of proximity, including cognitive, social and economic proximity.

From a geographical point of view, geographical and institutional proximities are spatial features. On the one hand, geographical distance remains significant in research collaboration. On the other hand, linguistic borders render the cross-linguistic-border collaboration less likely due to national funding schemes, institutional framework, norms, or value (Hoekman et al., 2010). Collaborators are expected to use their native languages if a native language is shared among them. Particularly, knowledge sharing and coordinating activities will become easier when collaborators speak the same language (Liang et al., 2006). The cross-linguistic-border nature can capture the institutional proximity more comprehensively. Therefore we adopt cross-linguistic-border nature as the institutional proximity. Long-distance and even cross-linguistic-border collaborations bring unique benefits, such as better access to international data (Jonsen et al., 2013), a higher tendency to stimulate new ideas (Ellis & Zhan, 2011), and increased international visibility and impact (Kwiek, 2015).

Therefore, based on the gravity model, we propose to measure the spatial proximity using the following spatial score, which takes an additive form of both the geographical distance



**Fig. 1** The values of  $\lambda$  for pharmaceutical sciences and ISLS during 2010–2018

and the cross-linguistic-border nature. For a publication  $i$ , in line with the form of the gravity model, the spatial score of leading institution  $a$  and collaborator institution  $b$  pair is defined as:

$$\text{SPS}_{i,ab} = \lg(\text{Geographical}_{ab}) + \lambda \times \text{Institutional}_{ab}, \quad (5)$$

where  $\text{Geographical}_{ab}$  denotes the geographical distance between institutions  $a$  and  $b$  and is calculated as the Great-circle distance according to institutions' latitudes and longitudes (Sinnott, 1984).  $\text{Institutional}_{ab}$  is a dummy variable indicating whether  $a$  and  $b$  are from different linguistic areas.  $\lambda$  represents the relative importance of the cross-linguistic-border nature and geographical distance of the collaboration. The effect of long-distance and cross-border nature on collaborative work varies in different research fields (Adams et al., 2005).  $\lambda$  is thus a field-specific parameter. The value of  $\lambda$  is obtained by calculating the ratio between the coefficient of  $\text{Institutional}_{ab}$  and the coefficient of  $\text{Geographical}_{ab}$ ,  $\lambda = \left| \frac{\beta_2}{\beta_3} \right|$ , in the gravity model (Gui et al., 2018; Zhang & Guo, 2017). Following Chaocheng et al. (2019), we perform regression based on the gravity model to obtain the coefficients of different dimensions of proximities. The results showed that, during 2010–2018,  $\lambda$  was 1.29 and 2.86 in pharmaceutical sciences and ISLS, respectively, indicating that the hindering effect of cross-linguistic-border nature in pharmaceutical sciences was lower than that in ISLS. As a field with ever-growing and widely distributed knowledge bases, pharmaceutical sciences have a particularly pronounced cross-border dimension of collaboration (Cantner & Rake, 2014). Regionally clustered pharmaceutical organizations locate R&D facilities to extend their collaboration beyond borders, to connect research partners, and source knowledge on a global scale (Cooke, 2006). The rapid increase of codified knowledge and its dispersion among a variety of research actors induced a pronounced trend of cross-border collaboration. However, for social sciences and humanities, research collaboration relies heavily on language and cultural proximity (Pajić, 2015). And this conclusion could also apply to ISLS (Jokić, 2020), a typical social science field. Indeed, in ISLS, mutual collaboration is intensive between authors with a similar historical and linguistic background (Jokić, 2020). We can also obtain the evolution of  $\lambda$  by running the gravity model for each year. Figure 1 shows the yearly evolution of  $\lambda$  during 2010–2018. The value of  $\lambda$  for pharmaceutical sciences was stable, with a slightly decreasing trend, indicating that the hindering effect of cross-linguistic-border nature has declined slightly

over time. This is in line with the important cross-border dimension of collaboration in pharmaceutical sciences. Interestingly, the value of  $\lambda$  for ISLS is relatively high and oscillated periodically. The cross-border collaboration in ISLS relies more heavily on linguistic, cultural, and historical background, compared with that in pharmaceutical sciences.

Therefore, the longer the geographical distance is, the smaller the probability of collaboration is, and the higher the spatial score is. Meanwhile, cross-linguistic-border nature positively influences the spatial score. The spatial score of a publication  $p$  is the average spatial score of the leading institution and participating institution pairs, which can be expressed as

$$SPS_p = \frac{1}{LN_p} \times \frac{1}{TN_p} \times \sum_a^{LN_p} \sum_b^{TN_p} SPS_{p,ab}, \tag{6}$$

where  $TN_p$  is the number of total institutions and  $LN_p$  is the number of leading institutions. Similar to Eq. (1), leading institutions share the spatial score of paper  $p$ . Each leading institution receives a spatial score  $\frac{SPS_p}{LN_p}$ . Therefore, the spatial research leadership flow intensity between a leading institution  $a$  and one participating institution  $b$  is as follows,

$$SL_{p,ba} = \frac{SPS_p}{LN_p \times TN_p}, \tag{7}$$

The total spatial research leadership flow intensity between leading institution  $a$  and participating institution  $b$  is as

$$SL_{ba} = \sum_{p=1}^{P_{ba}} SL_{p,ba}, \tag{8}$$

where  $P_{ba}$  is the number of papers where  $a$  is the leading institution, and  $b$  is a participating institution. We define the direction of spatial research leadership flow from participating institution  $b$  to leading institution  $a$ , indicating that  $b$  obtains (cites) the research leadership from  $a$ .

### Construction of the network

Now, we can construct a geographically weighted network where institutions are nodes and research leadership flows are directed edges. The weight on edge represents the spatial research leadership flows' intensities between the two institutions. Figure 2 illustrates the construction of the network in pharmaceutical sciences. For more details about network construction, please refer to the supplementary information.

### Spatial research leadership rank

LeaderRank is a simple variant of PageRank but has been widely proved to outperform PageRank regarding ranking effectiveness with good robustness (Li et al., 2014). In LeaderRank, a ground node, which bi-directionally connects to every other node, is added to the existing network with  $N$  nodes and  $M$  weighted directed edges. Figure 3 is an illustration of adding the ground node to the weighted and directed spatial research leadership network. Thus, the network consists of  $N + 1$  nodes and  $M + 2N$  edges, and forms

Publication 1	Publication 2	Publication 3
Author 1, Institution <i>a</i> , Language <i>A</i>	Author 3, Institution <i>c</i> , Language <i>B</i>	Author 1, Institution <i>a</i> , Language <i>A</i>
Author 2, Institution <i>b</i> , Language <i>A</i>	Author 4, Institution <i>a</i> , Language <i>A</i>	Author 3, Institution <i>c</i> , Language <i>B</i>
Author 3, Institution <i>c</i> , Language <i>B</i>	Author 5, Institution <i>d</i> , Language <i>C</i>	
Author 4, Institution <i>a</i> , Language <i>A</i>	Author 2, Institution <i>b</i> , Language <i>A</i>	
Author 5, Institution <i>d</i> , Language <i>C</i>		
Corresponding Author	Corresponding Author	Corresponding Author
Author 4, Institution <i>a</i> , Language <i>A</i>	Author 3, Institution <i>c</i> , Language <i>B</i>	Author 1, Institution <i>a</i> , Language <i>A</i>

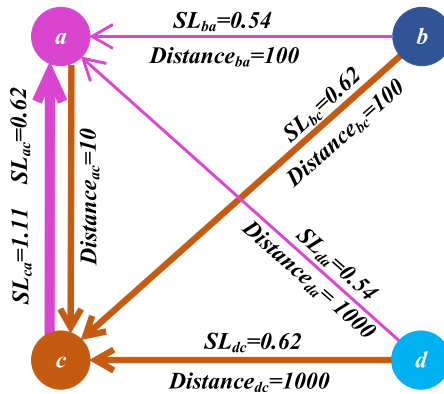


Fig. 2 An illustration of a geographically weighted network in pharmaceutical sciences

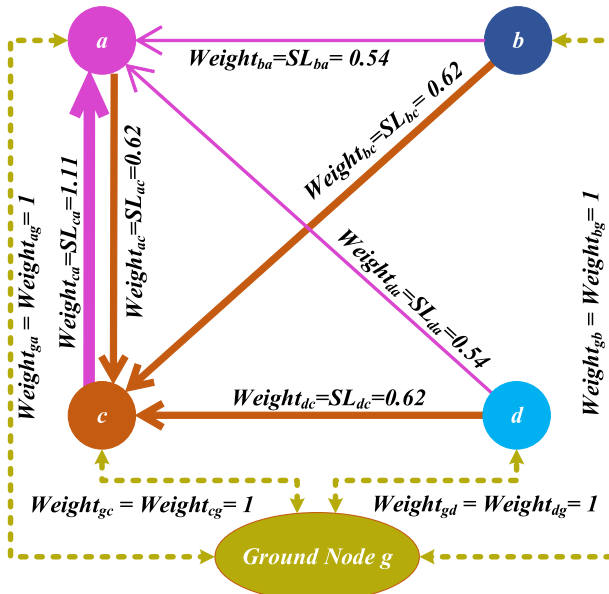


Fig. 3 An illustration of adding the ground node to the geographically weighted network to calculate SpatialLeaderRank

a strongly connected network. LeaderRank performs a standard random walk process to rank every node. However, LeaderRank only considers the topological features of the collaboration network while ignoring other non-topological features, especially the spatial features, which have been recognized as important factors of the academic performance and research impact. Therefore, we propose a new metric, namely the SpatialLeaderRank, to incorporate the spatial features into the measure of research leadership. Specifically, the SpatialLeaderRank of node  $a$  at the time step  $t$  denoted as  $\text{SpatialLeaderRank}_a(t)$ . Thus, the dynamics of SpatialLeaderRank is described by the following iterative process,

$$\text{SpatialLeaderRank}_a(t + 1) = \sum_{b=1, b \neq a}^{N+1} \left[ \frac{\text{SL}_{ba}}{\sum_{c=1, c \neq b}^{N+1} \text{SL}_{bc}} \times \text{SpatialLeaderRank}_b(t) \right], \tag{9}$$

where  $\text{SL}_{ba}$  is the spatial research leadership flow intensity from institution  $b$  to  $a$ . The spatial research leadership flow intensity from other institutions to the ground node and from the ground node to other institutions is set to 1 (Lu & Zhou, 2011). The process starts with the initialization where all institutions' SpatialLeaderRank being equal to 1. According to the iterative process described by Eq. (9), the SpatialLeaderRank value will converge to a unique and steady-state  $\text{SpatialLeaderRank}(\infty)$ , ( $a = 1, 2, \dots, N, N + 1$ ). We rank all institutions according to  $\text{SpatialLeaderRank}(\infty)$ .

## Result and analysis

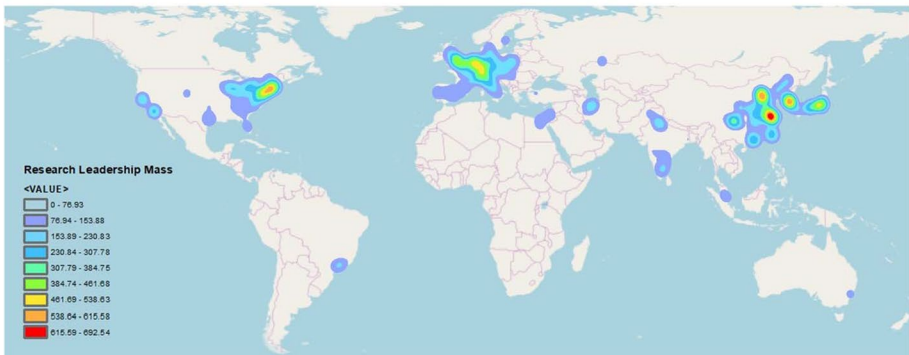
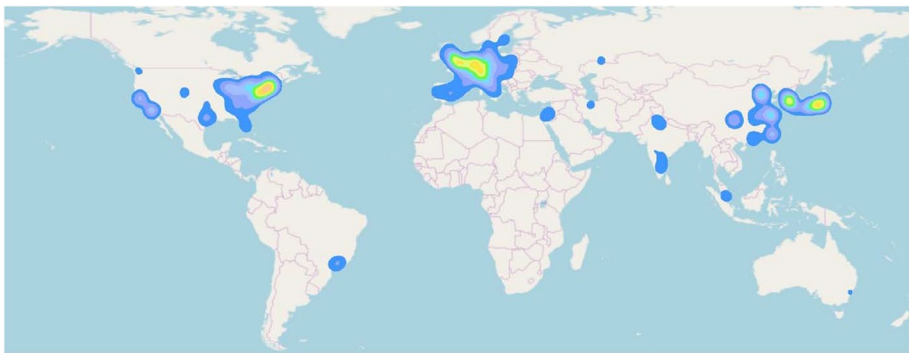
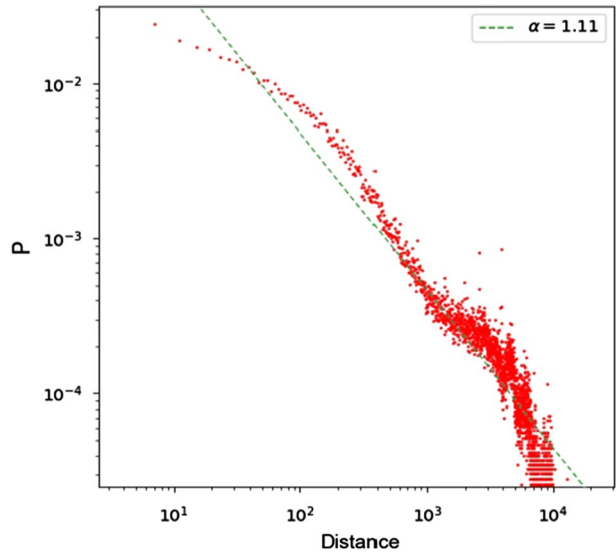
For simplicity, we mainly introduce the results for pharmaceutical sciences and then summarize the results for ISLS in “[Results in Information Science & Library Science](#)” section.

### Spatial patterns and evolution of research leadership flows in pharmaceutical sciences

Figure 4 illustrates the distance distribution of research leadership flows during 2010–2018. The data can well fit a power-law distribution  $P(k) \propto k^{-\alpha}$ . The majority of research leadership flows within a small radius, indicating that most research leaders integrate research resources within a short radius.

To present the changes in the research leadership mass in pharmaceutical sciences, we visualize the geographical distribution of research leadership mass in Fig. 5, the kernel density heat map of global research leadership mass in two split periods, 2010–2014 and 2015–2018. The kernel density estimation smooths the spatial coordinates to generate a probability density surface of a set of point locations (Downs & Horner, 2012). We can identify three main research leadership mass clusters, Northeastern United States, the European Union, and Northeastern Asia. More specifically, during 2010–2014, the cluster of European Union covered the largest area, and the Northeastern United States had the highest density. In the Northeastern Asia cluster, Japan and the Republic of Korea led the region. In the Greater China Region, the research leadership mass was mainly distributed in the eastern part. In general, during 2010–2014, pharmaceutical sciences research was dominated by the most economically developed countries. Differently, during 2015–2018, although these developed countries were still playing a prominent role in leading pharmaceutical research, Eastern Asia has emerged as a key leader, with multiple significant

**Fig. 4** The distance distribution of research leadership flow



**Fig. 5** The kernel density heat map of the distribution of research leadership mass at the regional level

clusters in China, Japan, and the Republic of Korea. Remarkably, the Yangtze River Delta Region (the densest region in the figure, including Shanghai, Jiangsu, Anhui and Zhejiang Provinces), the Jing-Jin-Ji Metropolitan Region (including Beijing, Tianjin and Hebei Provinces), and Sichuan Province have emerged as prominent leaders in pharmaceutical research, because of the finest universities and medical schools located in these regions. Other developing countries also had significant improvements, especially India and Iran. In general, the distribution of research leadership mass has become more balanced between western and eastern countries.

Figure 6 intuitively presents the detailed evolution of research leadership mass over the year with a bump graph. The width of the bump is proportional to the research leadership mass of the corresponding country/region. We present the top 15 countries/regions in terms of their research leadership mass. At first glance, the bump lines of all countries/regions were widening, indicating that the research leadership masses of all were growing during 2010–2018. In line with Fig. 5, some significant changes were underway. Mainland China had taken #1 from the United States in 2018. Japan retained the place at #3. India and Iran have risen to #4 and #5, respectively. In a nutshell, given these changes during 2010–2018, it is apparent that the distribution of research leadership mass is rebalancing between western and eastern countries.

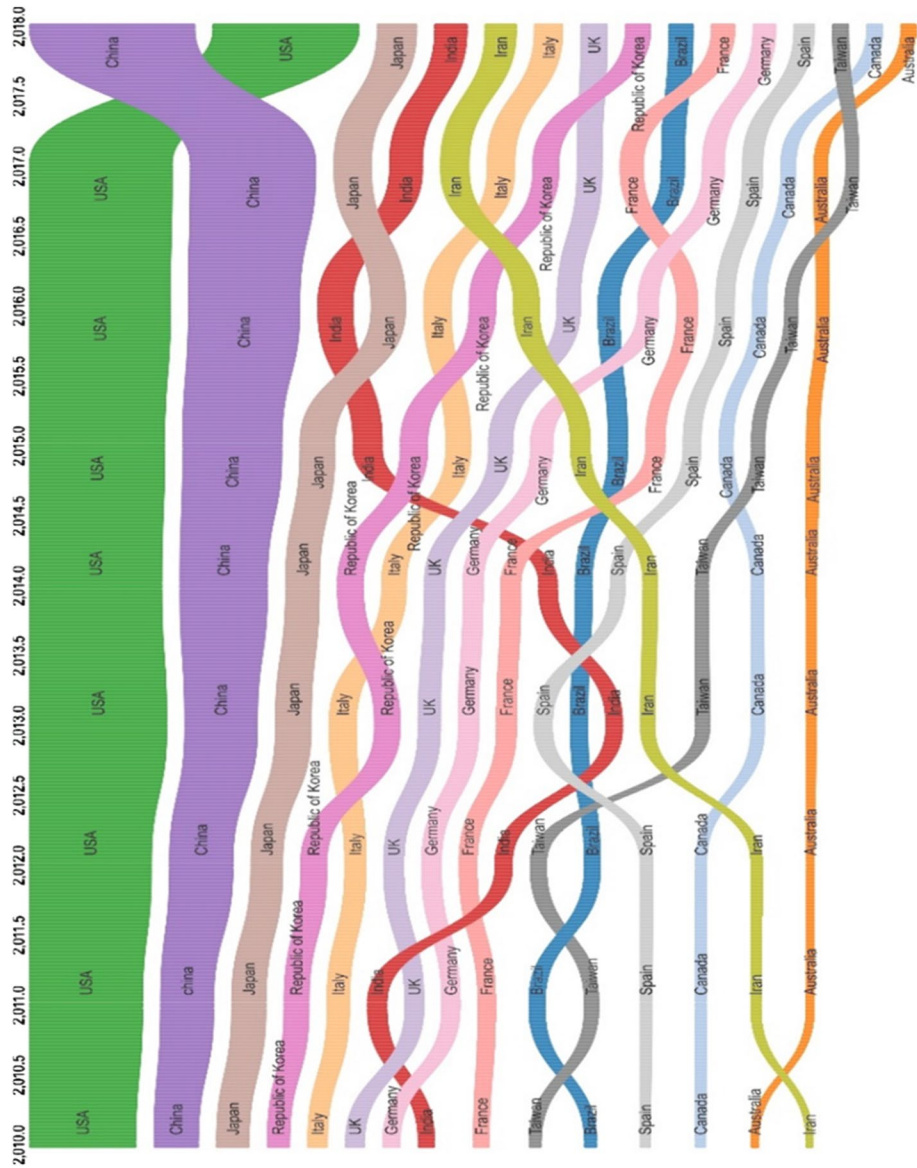
Figure 7 details the cross-linguistic-border research leadership flows at the country level. There are three chord charts, uncovering the cross-linguistic-border research leadership flow pattern in 2010, 2014, and 2018. Also, we present the top 15 countries/regions in terms of cross-linguistic-border research leadership flows. In general, the cross-linguistic-border research leadership pattern is still dominated by western countries. Several developing countries are improving rapidly, such as India, Brazil, etc. Although in Fig. 6, Mainland China has taken the #1 in terms of research leadership mass in 2018, the USA remains the first and far ahead in terms of cross-linguistic-border research leadership flows. This reflects that the increase in research leadership mass of China is mainly domestic.

## Effectiveness analysis of SpatialLeaderRank

To evaluate the effectiveness of the proposed SpatialLeaderRank, we then perform the receiver operating characteristic (ROC) curve analysis to evaluate the ability of SpatialLeaderRank and other conventional indices to identify the top 5% institutions with high academic impact. We further use *Ksim* (Haveliwala, 2003) to measure the similarity between each indices' ranking results and the academic impact rank. Last, we compare the detailed rank of institutions in terms of SpatialLeaderRank, other conventional indices, and academic impact indices. The conventional indices include PageRank, betweenness centrality, closeness centrality, indegree centrality, and publication number (Kim & Diesner, 2015; Wu, 2013). And the academic impact indices include citation count, citation-based h-index, altmetrics count, altmetrics-based h-index.

## Evaluation

We evaluate the effectiveness of the SpatialLeaderRank by using it to predict institutions' academic impact, which is measured by four common indices. We also compare its performance with that of conventional indices. Citation count is a widely recognized measure of academic impact (Yan & Ding, 2011). However, simple citation count is not robust against manipulations (Hirsch, 2005). To this end, Hirsch (2005) proposed h-index, the

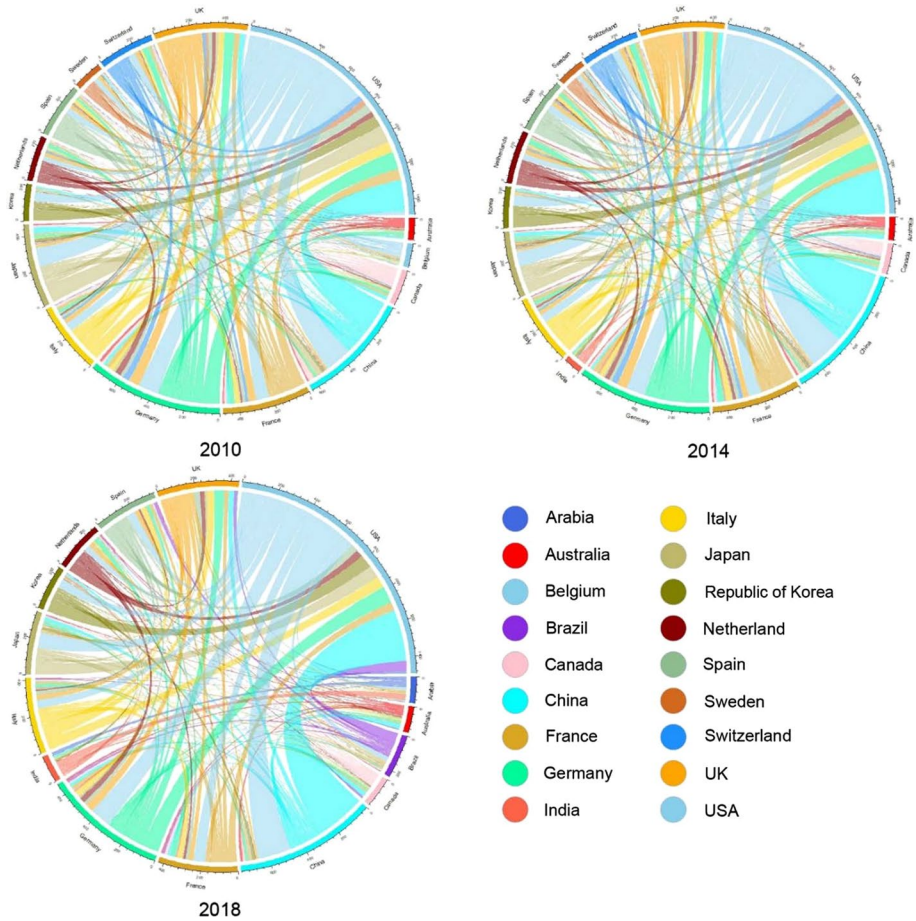


**Fig. 6** The bump graph of research leadership mass at the country level

maximum value of  $h$  papers being cited at least  $h$  times for each entity (author, journal, institution, etc.). H-index combines the quantity and quality of publications and has been widely adopted by the scientific community (Lund, 2019). Recently, Altmetrics indices<sup>2</sup> have emerged as popular tools to measure academic impact because they are less subject to

<sup>2</sup> We derive the altmetrics data from <https://www.altmetric.com>.

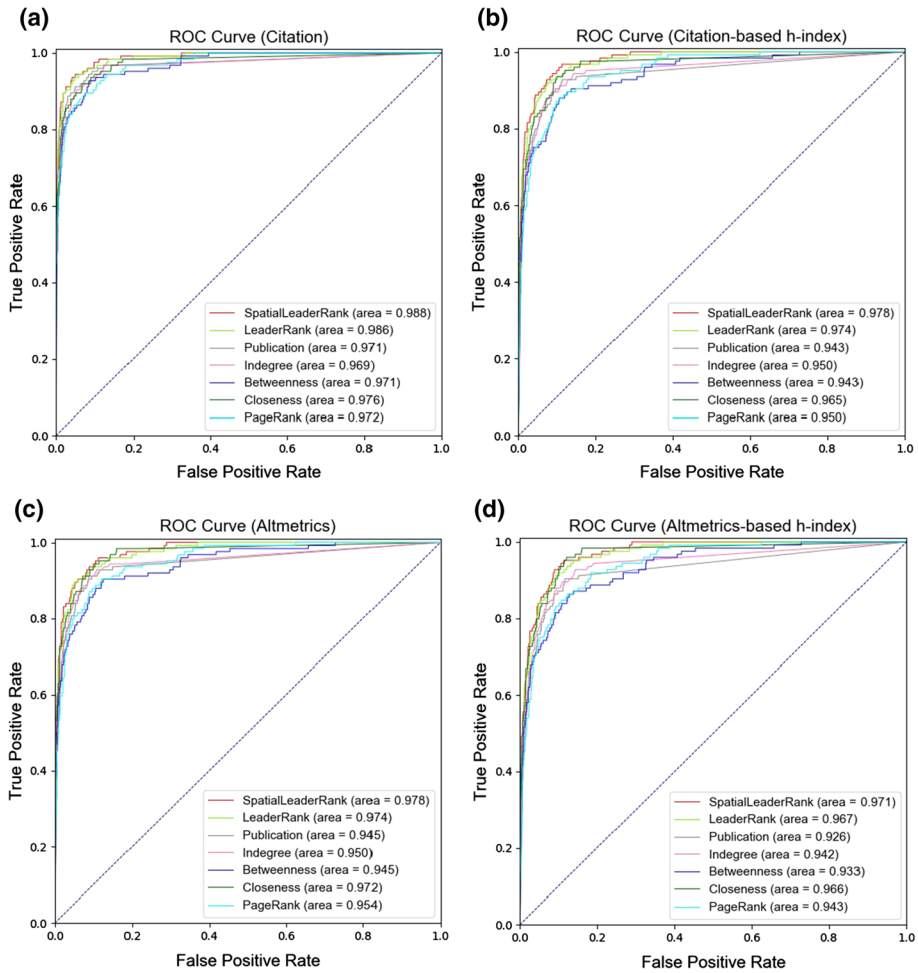




**Fig. 7** The chord graph of cross-linguistic-border research leadership flows

publication delays than citation count. Similarly, the simple Altmetrics count is not robust against manipulations either. To this end, Askeridis (2018) proposed an Altmetrics-based h-index, namely, Mendeley-based h-index, which replaces the citation count with the view count of Mendeley readers. Mendeley is a free reference manager and academic social network. The view count of Mendeley users is associated with the later citation count (Aduku et al., 2017) because the view count is crowdsourced from the research community and thus reflects the academic impact earlier than the citation count. In this study, we have two variants of the h-index, the citation-based h-index, and the Altmetrics-based h-index. Both of them will not be large if an entity publishes either many publications with low citation/Altmetrics or very few publications with high citation/Altmetrics. In total, four evaluation metrics (citation, citation-based h-index, Altmetrics, and Altmetrics-based h-index) are employed as the model outcome to evaluate the effectiveness of SpatialLeaderRank. And These evaluation metrics cover the all journal publications of an institution.

We use the ROC curve analysis to examine the capability of SpatialLeaderRank in identifying top institutions with high academic impact (top 5% in terms of citation,



**Fig. 8** ROC curves for SpatialLeaderRank and other indices to identify top institutions with high academic impact

citation-based h-index, Altmetrics, and Altmetrics-based h-index) and compare its performance with that of other conventional indices. ROC curve analysis is a graphic method to illustrate the performance of a binary classification system under different recognition thresholds (Hassan et al., 2017; Zhang et al., 2019). The larger AUC (area under the ROC curve) is, the better performance the focal binary classification system has. Figure 8 presents the ROC curve of the SpatialLeaderRank and other indices for identifying the top 5% of institutions with high academic impact in terms of citation, citation-based h-index, Altmetrics, and Altmetrics-based h-index.

As shown in Fig. 8, all the indices have a reasonably high AUC in terms of all the four evaluation metrics, indicating the generally good performance of all the indices in identifying the top institutions with high academic impact. Note that the high AUC value is expected because of the majority (95%) of institutions are classified as not influential. The proposed SpatialLeaderRank consistently outperforms other indices with all measures of academic

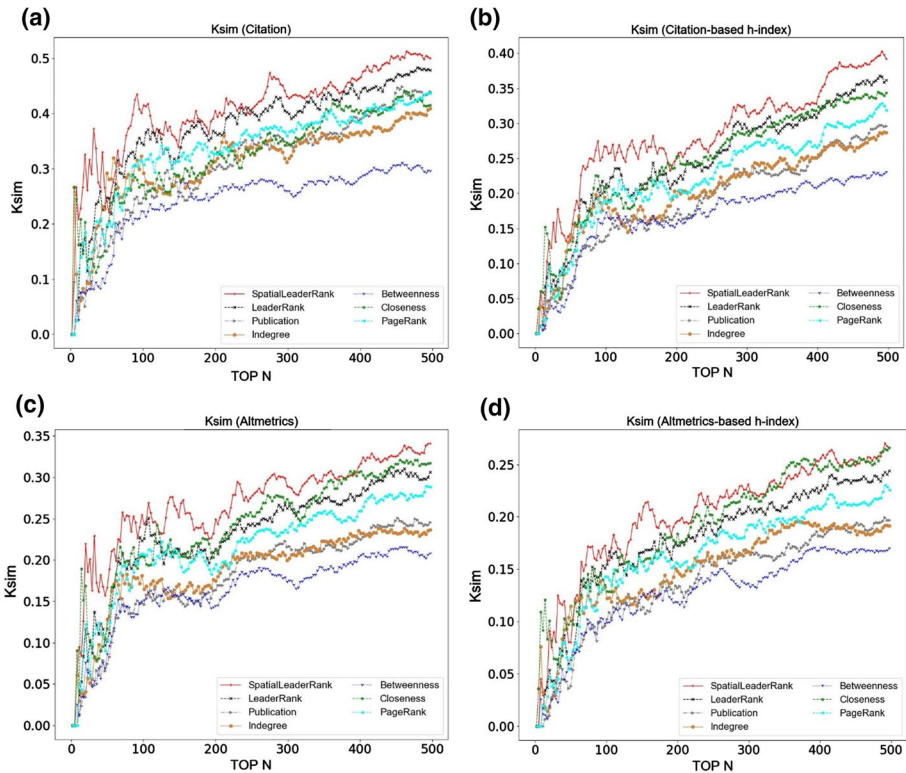
impact. This indicates that integrating the spatial features does help to improve the capability in identifying top institutions with high academic impact. The larger SpatialLeaderRank score an institution has, the greater the leadership status that it has, and the higher academic impact that it generates. It is noteworthy that the AUC gaps between SpatialLeaderRank and other indices are more significant in h-index (citation-based/Altmetrics-based) than that in citation/Altmetrics counts. This is because h-index (both the citation-based/Altmetrics-based) is more robust than citation/Altmetrics, and can more closely reflect the academic impact. SpatialLeaderRank has more advantages in identifying top institutions with high academic impact compared with other conventional indices.

The ROC curve analysis only examines the overlap of top-k ranked institutions according to different indices, as it considers these top-k ranked institutions as an unordered set. Thus, it is meaningful to adopt another metric to further examine the relative ordering of the top-k ranked institutions. In this study, we adopt the *KSim* metric (Haveliwala, 2003) which is based on Kendall’s  $\tau$  distance measure. Consider a ranking list by an index  $\tau_1$  and a ranking list by an academic impact evaluation metric  $\tau_2$ . Let  $U$  be the union of the institutions in  $\tau_1$  and  $\tau_2$ . Let  $\sigma_1$  be  $U - \tau_1$  and  $\tau'_1$  be the extension of  $\tau_1$ , where  $\tau'_1$  contains  $\sigma_1$  in addition to the existing ranked institutions in  $\tau_1$ . The rank of institutions in  $\sigma_1$  is set to have the same ordinal rank in the end of  $\tau_2$ . Similarly,  $\tau_2$  is extended to yield  $\tau'_2$ . The *Ksim* between  $\tau_1$  and  $\tau_2$  is defined as the Kendall’s distance between  $\tau'_1$  and  $\tau'_2$ , respectively:

$$KSim(\tau_1, \tau_2) = \frac{|(u, v) : \tau'_1, \tau'_2 \text{ agree on order of } (u, v), u \neq v|}{(|U|)(|U| - 1)}, \tag{10}$$

In other words, *KSim* measures the probability of  $\tau_1$  and  $\tau_2$ ’s agreement on the relative ordering of a randomly selected pair of institutions  $(u, v) \in U \times U$ .

Figure 9 visualizes the *KSim* curve between each index and the four academic impact evaluation metrics from the top 5 institutions to the top 500 institutions. In general, SpatialLeaderRank consistently outperforms all other indices. In particular, when  $N$  is small, the effect of extreme cases is more significant. In such a case, the advantage of SpatialLeaderRank over other indices is more obvious, indicating that SpatialLeaderRank is more robust to rare extreme cases. As  $N$  increases, all indices’ *Ksim* increase and become stable. Betweenness centrality has the worst performance, according to *KSim*, in all evaluation metrics. Betweenness centrality of an institution is proportional to the number of shortest paths traversing through the institution. The unsatisfactory performance of betweenness centrality is because most of the top-ranked institutions by betweenness centrality are the local hubs connecting to local institutions. Publication count and indegree have similar performance. Interestingly, we find that if the publications and leading behaviors are mostly associated with local institutions, the academic impact does not get much improved. It’s interesting that closeness centrality has a relatively high *Ksim* in Altmetrics and, particularly, Altmetrics-based h-index. In Altmetrics-based H-index, when  $N$  is larger than 300, closeness centrality’s performance is close to the SpatialLeaderRank, primarily because institutions with high closeness centrality have shorter distance from other institutions, so the spread of research output is faster through the collaboration network.



**Fig. 9** *KSim* between each index and four evaluation metrics from top 5 to top 500

### Ranking institutions by SpatialLeaderRank, other conventional indices, and academic impact indices

We rank the institutions according to SpatialLeaderRank, the other conventional indices (LeaderRank, publication, indegree, betweenness, closeness, and PageRank) and academic impact indices (citation, citation-based h-index, Altmetrics, and Altmetrics-based h-index). Table 2 presents the top 20 institutions. Table 3 presents the rank of the SpatialLeaderRank-based top 20 institutions in other rankings. The top institutions ranked by SpatialLeaderRank are also highly ranked according to all the four academic impact indices. This strong association indicates that leading long-distance and cross-linguistic-border collaborations generally lead to greater academic impact.

Four institutions (Harvard University, University of Oxford, University of Cambridge, and University of California San Diego) are among the top 20 according to all indices. Particularly, Harvard University is ranked as the top one institution according to four academic impact indices, and the top three according to the other indices. Chinese Academy of Sciences (CAS) and the French National Centre for Scientific Research (French: Centre national de la recherche scientifique, CNRS) are two state research organizations in China and France, respectively. These two mega organizations have multiple research institutions and have published many papers in the field (#1 for CAS and #7 for CNRS) (Table 2). It is not surprising to observe the high ranking of

**Table 2** Top 20 institutions' rank by SpatialLeaderRank and by other indices

Citation	Pagerank	Closeness	Betweenness	Indegree	Publication	LeaderRank	SpatialLeaderRank	Altmetrics-based h-index	Altmetrics	Citation-based h-index
Harvard Univ	Chinese Acad Sci	Chinese Acad Sci	Chinese Acad Sci	Chinese Acad Sci	Chinese Acad Sci	Chinese Acad Sci	Chinese Acad Sci	Harvard Univ	Harvard Univ	Harvard Univ
Chinese Acad Sci	Univ Sao Paulo	Harvard Univ	Harvard Univ	Harvard Univ	Harvard Univ	Harvard Univ	Harvard Univ	Stanford Univ	Stanford Univ	MIT
Univ Cambridge	Harvard Univ	UCSD	Univ Sao Paulo	Univ Michigan	Russian Acad Sci	CNRS	CNRS	MIT	MIT	UCSF
Univ Oxford	CSIC	Univ Michigan	Univ Tokyo	UCSD	Univ Sao Paulo	Univ Tokyo	Univ Oxford	UCSF	Univ Cambridge	Stanford Univ
UCSD	UCSD	Univ Oxford	CSIC	Univ Oxford	SJTU	Univ Toronto	Univ Tokyo	UCSD	UCSF	UCSD
UCSF	Univ Oxford	Univ Cambridge	Univ Oxford	Univ Cambridge	Univ Tokyo	Univ Copenhagen	Univ Cambridge	Univ Calif Berkeley	UCSD	Univ Cambridge
Stanford Univ	Univ Michigan	Univ Illinois	King Saud Univ	Univ Copenhagen	CNRS	CAMS	Univ Toronto	Univ Cambridge	Univ Oxford	Univ Calif Berkeley
Univ Toronto	Zhejiang Univ	Stanford Univ	Univ Cambridge	Univ Illinois	Univ Toronto	Kyoto Univ	UCSD	Univ Oxford	Univ Calif Berkeley	Univ Toronto
MIT	Univ Cambridge	Univ Copenhagen	Kyoto Univ	Univ Sao Paulo	Seoul Natl Univ	UCSD	Univ Copenhagen	Yale Univ	Chinese Acad Sci	Univ Washington
Univ Penn	Univ Tokyo	NCI	Univ Michigan	Zhejiang Univ	Zhejiang Univ	Univ Oxford	UCSF	HHMI	Univ Toronto	Univ Penn
CNRS	Univ Copenhagen	Univ Penn	Univ Copenhagen	CSIC	Fudan Univ	Univ Michigan	NCI	Univ Penn	Yale Univ	MGH
Univ Calif Berkeley	Peking Univ	Yale Univ	UCSD	Univ Tokyo	Univ Copenhagen	UCSF	Univ Michigan	Univ Toronto	Univ Penn	Yale Univ
Univ Copenhagen	Univ Illinois	Johns Hopkins Univ	Zhejiang Univ	Johns Hopkins Univ	UCSD	Johns Hopkins Univ	Univ Washington	MGH	Univ Washington	Univ Oxford
Univ Michigan	Kyoto Univ	Univ Toronto	Univ Illinois	Peking Univ	Univ Michigan	Univ Cambridge	Johns Hopkins Univ	Univ Washington	Univ Copenhagen	Chinese Acad Sci

**Table 2** (continued)

Citation	Pagerank	Closeness	Betweenness	Indegree	Publication	LeaderRank	SpatialLeaderRank	Altmetrics-based h-index	Altmetrics	Citation-based h-index
Univ Tokyo	SJTU	Ohio State Univ	Karolinska Inst	Univ N Carolina	Univ Oxford	Univ Penn	Stanford Univ	UCL	Univ Michigan	Johns Hopkins Univ
Univ Washington	Johns Hopkins Univ	Univ Wisconsin	CNRS	Stanford Univ	Peking Univ	Russian Acad Sci	Univ Penn	UCLA	UCL	HHMI
Duke Univ	UNAM	UCSF	Johns Hopkins Univ	Univ Maryland	Univ Illinois	Univ Sao Paulo	Duke Univ	Columbia Univ	UCLA	Univ Copenhagen
UCLA	Karolinska Inst	Fudan Univ	NCI	SJTU	Univ N Carolina	Univ Washington	Kyoto Univ	Johns Hopkins Univ	Johns Hopkins Univ	Univ Michigan
Johns Hopkins Univ	Univ N Carolina	McGill Univ	UCL	Univ Pittsburgh	Sun Yat-Sen Univ	Seoul Natl Univ	Univ Illinois	Washington Univ	CNRS	Columbia Univ
Yale Univ	Ohio State Univ	Univ Washington	Seoul Natl Univ	Univ Florida	Univ Cambridge	Osaka Univ	Univ N Carolina	Duke Univ	Duke Univ	CNRS

“CAMS”: Chinese Academy of Medical Sciences; “CNRS”: Centre national de la recherche scientifique (French National Centre for Scientific Research);

“CSIC”: Consejo Superior de Investigaciones Cientificas (Spanish National Research Council); “HHMI”: Howard Hughes Medical Institute;

“MGH”: Massachusetts General Hospital; “MIT”: Massachusetts Institute of Technology; “NCI”: National Cancer Institute, USA;

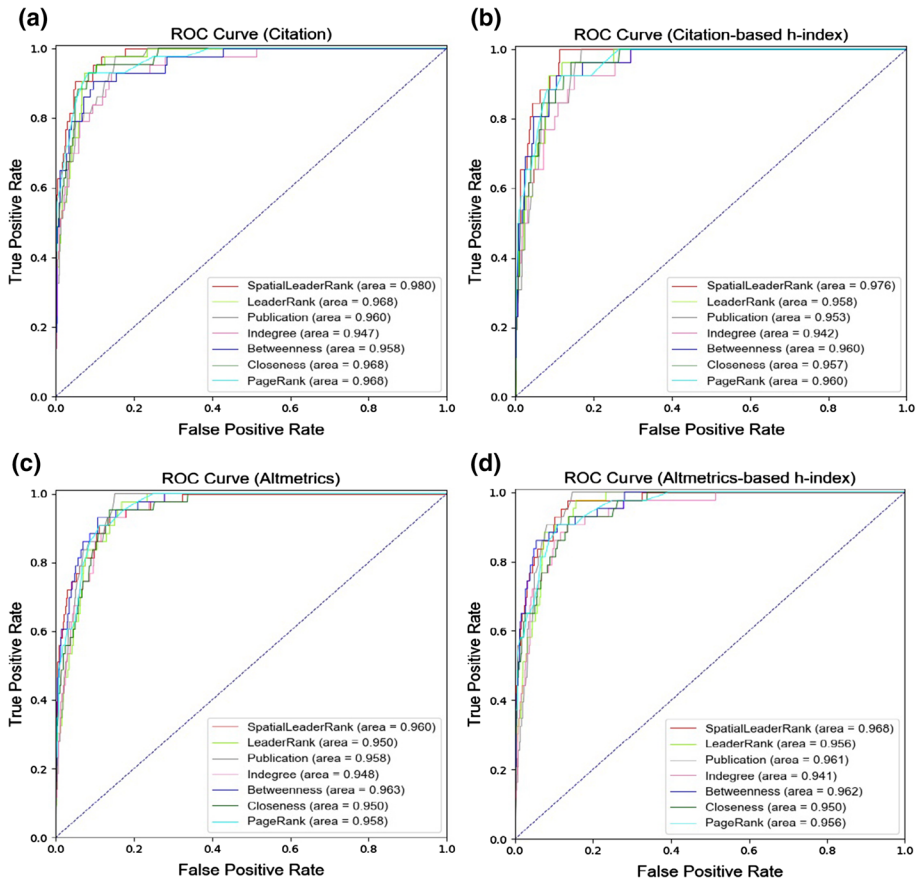
“SJTU”: Shanghai Jiao Tong University; “UCLA”: University of California, Los Angeles; “UCSD”: University of California San Diego;

“UCSF”: The University of California, San Francisco; “UCL”: University College London; “UCLA”: University of California, Los Angeles;

“UNAM”: National Autonomous University of Mexico

**Table 3** Comparison of institution rank by SpatialLeaderRank and other indices

Institution	Spatial- Leader- Rank	Citation	Citation- based h-index	Altmetrics	Altmetrics- based h-index	LeaderRank	Publication	Indegree	Betweenness	Closeness	PageRank
Chinese Acad Sci	1	2	14	9	27	1	1	1	1	1	1
Harvard Univ	2	1	1	1	1	2	2	2	2	2	3
CNRS	3	11	20	19	26	3	7	81	16	64	73
Univ Oxford	4	4	13	7	8	10	15	5	6	5	6
Univ Tokyo	5	15	23	31	35	4	6	12	4	49	10
Univ Cambridge	6	3	6	4	7	14	20	6	8	6	9
Univ Toronto	7	8	8	10	12	5	8	32	23	14	31
UCSD	8	5	5	6	5	9	13	4	12	3	5
Univ Copenhagen	9	13	17	14	21	6	12	7	11	9	11
UCSF	10	6	3	5	4	12	26	29	30	17	32
NCI	11	27	35	30	34	21	49	21	18	10	26
Univ Michigan	12	14	18	15	25	11	14	3	10	4	7
Univ Washington	13	16	9	13	14	18	22	24	44	21	36
Johns Hopkins Univ	14	19	15	18	18	13	21	13	17	13	16
Stanford Univ	15	7	4	2	2	29	33	16	24	8	27
Univ Penn	16	10	10	12	11	15	24	25	32	11	28
Duke Univ	17	17	25	20	20	31	40	40	53	40	44
Kyoto Univ	18	35	57	77	93	8	23	35	9	69	14
Univ Illinois	19	24	44	37	56	22	17	8	14	7	13
Univ N Carolina	20	21	21	22	30	24	18	15	35	26	19



**Fig. 10** ROC curve for SpatialLeaderRank and other indices to identify top 5% institutions in ISLS

CAS and CNRS according to the collaboration indices. However, their performance in research impact is less significant than that of collaboration, indicating that CAS and CNRS are widely leading the research through collaboration with many other institutions, the academic impact still has room for further improvement. Similarly, we found that many other Chinese institutions are highly ranked (top 20) in terms of the conventional indices such as Shanghai Jiaotong University, Zhejiang University, Fudan University, Peking University, and Sun Yat-sen University in terms of publication, indegree, betweenness centrality, closeness centrality, and PageRank. However, neither their SpatialLeaderRank nor academic impact is highly ranked (top 20). Their publications do not match their academic impact status, suggesting that Chinese institutions should focus on improving the quality of research instead of quantity only. Notably, building on the result of Table 2 and Fig. 7, from a SpatialLeaderRank point of view, Chinese institutions are recommended to expand their research leadership radius and integrate international research resources to embed themselves into the global research collaboration network.



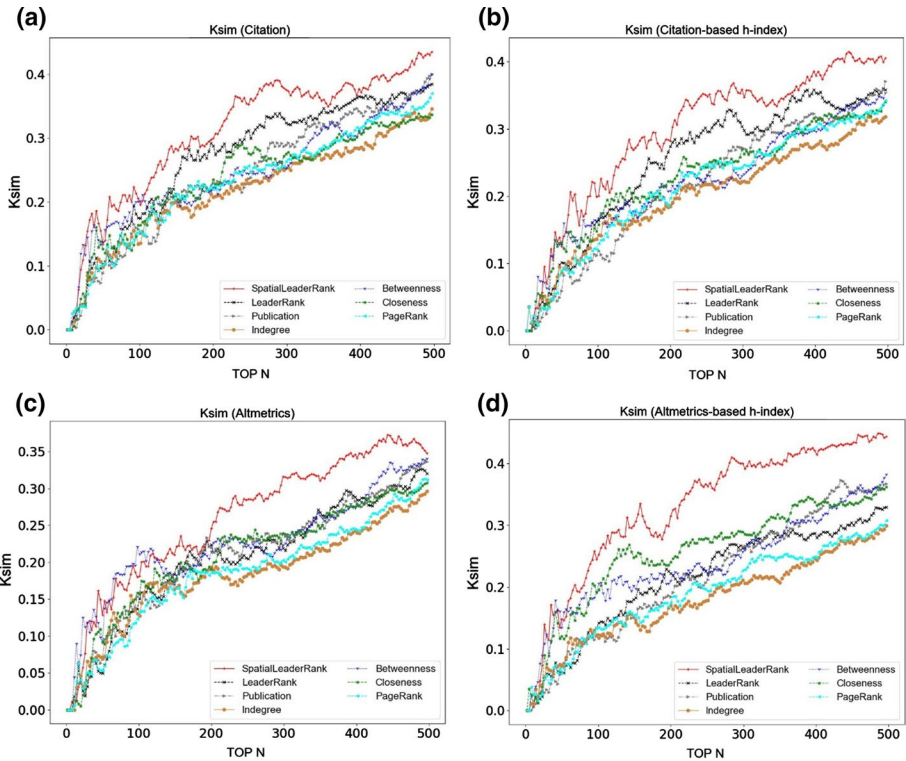


Fig. 11 *KSim* between each index and four evaluation metrics from the top 5 to top 500 in ISLS

### Results in Information Science & Library Science

To check the robustness of the study, we implement the same ROC curve analysis and *Ksim* measure analysis in ISLS field, as shown in Figs. 10 and 11. The results are consistent with those for pharmaceutical sciences. The proposed SpatialLeaderRank consistently outperformed other indices in predicting the academic impact of institutions. Similar to analysis in pharmaceutical sciences, we present the top 20 institutions in ISLS in Table 4. Interestingly, we found a few reasonable changes in the ranks of certain key players in the field. Specifically, we found that Wuhan University is top-ranked by traditional metrics such as publication, indegree, closeness, and LeaderRank, but only 20 as ranked by the proposed SpatialLeaderRank. The University of Illinois at Urbana-Champaign, the long-standing leader in the field, is ranked #9 by LeaderRank, but #1 by the proposed SpatialLeaderRank. Such results make much sense because although Wuhan University has one of the best ISLS programs in China, its leading efforts are mostly among domestic universities so far. The University of Illinois at Urbana-Champaign, on the other hand, has the top-ranked ISLS program for decades. Its leading efforts are reaching more non-US institutions. This finding echoes the previous discussion for the pharmaceutical field: although the research performance of Chinese institutions has been growing very fast in ISLS field, it is suggested that they expand their leadership to more international collaborations.

**Table 4** Top 20 institutions' rank by SpatialLeaderRank and by other indices in ISLS

Pagerank	Closeness	Betweenness	Indegree	Publication	LeaderRank	SpatialLeaderRank	Altmetrics-based h-index	Altmetrics	Citation-based h-index	Citation
Indiana Univ	Wuhan Univ	Univ Wisconsin	Wuhan Univ	Wuhan Univ	Wuhan Univ	Univ Illinois	Vanderbilt Univ	Univ Maryland	Univ Maryland	Univ Maryland
Univ Wisconsin	Indiana Univ	Univ Illinois	Indiana Univ	Univ Illinois	Univ Granada	City Univ Hong Kong	Harvard Univ	Wolverhampton Univ	Leiden Univ	Indiana Univ
Wuhan Univ	Univ Wisconsin	Indiana Univ	Univ Wisconsin	Indiana Univ	City Univ Hong Kong	Univ Washington	Indiana Univ	Leiden Univ	Indiana Univ	Leiden Univ
City Univ Hong Kong	Nanyang Technol Univ	Univ Sheffield	Univ Granada	Univ Granada	Indiana Univ	Univ Wisconsin	Univ Maryland	Delft Univ Technol	Univ Amsterdam	Univ Arizona
Univ Illinois	City Univ Hong Kong	Wuhan Univ	City Univ Hong Kong	Nanyang Technol Univ	Univ Wisconsin	Univ Michigan	Univ Michigan	Univ Sheffield	City Univ Hong Kong	Univ Arkansas
Univ Maryland	Univ Illinois	Univ Amsterdam	Univ Illinois	Univ Wisconsin	Chinese Acad Sci	Indiana Univ	Wolverhampton Univ	Univ Michigan	Wolverhampton Univ	Wolverhampton Univ
Nanyang Technol Univ	Mcgill Univ	Univ Maryland	Beihang Univ	Univ Malaysia	Univ Amsterdam	Univ Texas Austin	City Univ Hong Kong	Indiana Univ	Vanderbilt Univ	Univ Amsterdam
Arizona State Univ	Chinese Acad Sci	Univ N Carolina	Univ Carlos III Madrid	Russian Acad Sci	Univ Illinois	Georgia State Univ	Univ Amsterdam	Vanderbilt Univ	Harvard Univ	City Univ Hong Kong
Univ Arizona	Univ Arizona	Univ Carlos III Madrid	Nanyang Technol Univ	Univ N Carolina	Nanyang Technol Univ	Harvard Univ	Leiden Univ	Brunel Univ	Univ Arkansas	Vanderbilt Univ
Univ Amsterdam	Univ Amsterdam	Mcgill Univ	Chinese Acad Sci	Univ Maryland	Univ N Carolina	Univ N Carolina	Columbia Univ	Univ Amsterdam	Univ Wisconsin	Harvard Univ
Univ Texas Austin	Univ Sheffield	City Univ Hong Kong	Univ Maryland	Univ Carlos III Madrid	Mcgill Univ	Univ British Columbia	Univ Wisconsin	Univ N Carolina	Temple Univ	Univ Wisconsin

**Table 4** (continued)

Pagerank	Closeness	Betweenness	Indegree	Publication	LeaderRank	SpatialLead-erRank	Almetrics-based h-index	Almetrics	Citation-based h-index	Citation
Chinese Acad Sci	Univ N Carolina	Univ Western Ontario	Univ Amster-dam	Univ Fed Minas Gerais	Arizona State Univ	Univ Maryland	Univ Wash-ington	Univ Granada	Georgia State Univ	Temple Univ
Mcgill Univ	Arizona State Univ	Univ Granada	Natl Taiwan Univ	Univ Complutense Madrid	Univ Maryland	Natl Univ Singapore	Delft Univ Technol	Univ Wisconsin	Univ Arizona	Georgia State Univ
Univ N Carolina	Univ British Columbia	Univ Texas Austin	Natl Univ Singapore	Univ Amster-dam	Natl Univ Singapore	Univ Pittsburgh	Univ Granada	City Univ Hong Kong	Univ Wash-ington	Univ Illinois
Univ Min-nesota	Univ Mary-land	Univ Wash-ington	Univ N Carolina	Univ Shef-field	Zhejiang Univ	Nanyang Technol Univ	Stanford Univ	Univ Illinois	Michigan State Univ	Wuhan Univ
Univ Pitts-burgh	Univ Hong Kong	Georgia State Univ	Yonsei Univ	Chinese Acad Sci	Nanjing Univ	Univ Amster-dam	Univ Shef-field	Univ Sydney	Univ Illinois	Univ Granada
Univ Michi-gan	Natl Univ Singapore	Natl Univ Singapore	Arizona State Univ	Mcgill Univ	Univ Arizona	Univ Man-chester	Univ Sydney	Nanyang Technol Univ	Univ Grana-da	Univ British Columbia
Georgia State Univ	Univ Western Ontario	Univ Michi-gan	Univ Shef-field	Penn State Univ	Univ Min-nesota	Georgia Inst Technol	Univ N Carolina	Univ Twente	Northwestern Univ	Univ N Carolina
Univ Carlos III Madrid	Univ Michi-gan	Chinese Acad Sci	Univ Min-nesota	Natl Taiwan Univ	Univ Politecn Valencia	Mcgill Univ	Georgia State Univ	Univ Malaysia	Univ Texas Austin	Univ Michigan
Univ Hong Kong	Univ Min-nesota	Nanyang Technol Univ	Drexel Univ	Univ Fed Santa Catarina	Max Planck Gesell	Wuhan Univ	Univ Man-chester	Harvard Univ	Univ N Carolina	Univ Wash-ington

## Discussion and conclusions

In this paper, to address the spatial bias in research leadership flows in research collaboration, we examine the spatial distribution and the dynamic trend of research leadership flows in pharmaceutical sciences. We observe that developing countries have been playing an increasingly important role in leading the research in pharmaceutical sciences. However, the cross-linguistic-border research is still dominated by developed countries.

Then, we construct a geographically weighted network, based on which we propose the SpatialLeaderRank. SpatialLeaderRank ranks the institutions integrating both topological features and spatial features. Comprehensive experiments with the data in both pharmaceutical sciences and ISLS demonstrate the superior performance of the proposed SpatialLeaderRank in predicting the academic impact of institutions. Leading institutions are identified and presented.

This study sheds light on the important association between long-distance and cross-linguistic-border collaborations and academic impact. With the growing trend of cross-linguistic-border collaborations, the distance between collaborators, particularly between the research leader and participators, should be an integral part to consider while examining the research leadership.

From a policy perspective, we found a clear rebalancing process between the research leadership mass in developed and developing countries. Many Eastern Asian countries, particularly China, is quickly emerging as a new global leader in pharmaceutical sciences. There are two main reasons for the change. First, the expenditure on research has been increasing rapidly in China (Basu et al., 2018). As of 2018, the research and development (R&D) expenditure in China was 1967.79 billion CNY,<sup>3</sup> a 179% increase from 2010. China recently passed the European Union in R&D investment (Basu et al., 2018). Meanwhile, the research funding in the United States only increased by 57%.<sup>4</sup> Apparently, with abundant funding, Chinese institutions are playing the role as the leader more often. Second, the evaluation of research output in China is primarily based on quantitative measures such as publication number, the impact factor of journals, and citation count. The emphasis on these quantitative measures drove the whole academic community to publish as many papers as possible. Despite the success in publication number, the academic impact of Chinese institutions is still laid behind the Western institutions. It is suggested that policymakers in China shift the focus of the research evaluation towards the actual academic impact from quantitative measures. On 18/02/2020, the Ministry of Science and Technology and the Ministry of Education of China jointly published an announcement to urge the Chinese institutions to adopt a more scientific and influence-driven research evaluation approach.<sup>5</sup> This indicates the start of the transformation from quantity to quality in China.

For the Western countries, including Europe and the United States, they are still playing the role of the major leaders in the field. More research expenditure is needed to maintain a good status. In general, cross-linguistic-border collaboration is playing an increasingly important role in pharmaceutical research. Given the higher impact of long-distance collaborations, cross-linguistic-border collaboration should be encouraged utilizing joint-funding schemes and academic exchanges.

---

<sup>3</sup> <http://www.stats.gov.cn>.

<sup>4</sup> <https://www.statista.com>.

<sup>5</sup> [http://www.moe.gov.cn/srcsite/A16/moe\\_784/202002/t20200223\\_423334.html](http://www.moe.gov.cn/srcsite/A16/moe_784/202002/t20200223_423334.html).

Long-distance and cross-linguistic-border collaborations shall be encouraged to increase the chance to generate high impact research. However, the impact of research should be purely based on its scientific merit.

The proposed SpatialLeaderRank is a general method that can be used to evaluate the leadership at the author and country levels. It is also applicable to examine other types of relationships, such as to evaluate the academic influence of scientific journals, to identify the key innovator in the industry by analyzing the patent-citation network, and to identify the key moderator in the financial system by analyzing the guarantee-relationship network.

## Limitations and future research

This study still has some limitations, which provide some potential directions for future research: first, we only study two disciplines: the pharmaceutical field and ISLS. It is recommended that future research should cover a broad set of fields and even all disciplines to enhance the generalizability of our proposed metric. Second, in data processing, for institutions with multiple addresses, we only adopt the first search result of Google Map. But this may lead to a bias for spatial-feature-based studies, particularly for those mega organizations such as CAS with many branches distributed nationwide. Publications divided into mega organizations' branches would result in different rankings. Future research should identify and locate each author's school/department address information instead of the institution address information so that we can depict the spatial research leadership more accurately. Since we are studying at the institution level, we ignore individuals' information. For example, for an author with multiple institutions, we treat them equally. What is more, we do not take researchers' mobility into consideration. Particularly, there is a time lag between when the research being executed and when the paper in print. Leading authors may visit other countries/institutions, and completed the research, and returned his home institution. In this paper, the leading credit is assigned to his home institution, which will also lead to a bias. In the future, we should measure the research leadership from an individual point of view. Third, we leverage WOS to perform data collection, but only WOS data sources may not have enough coverage of publications in pharmaceutical sciences or ISLS (Tang & Shapira, 2011). We may underestimate the absolute counts of publications in pharmaceutical sciences and ISLS. Future research should merge publication data from multiple data sources, such as Scopus<sup>6</sup> and PubMed.<sup>7</sup>

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11192-021-03943-w>.

**Acknowledgements** An early version of this paper has been posted in arxiv: (Chaocheng et al., 2020). We would like to thank the editor and reviewers for their valuable comments. This work was supported by grants from Key Projects of Philosophy and Social Sciences Research, Ministry of Education (Grant No. 20JZD024), the General Project of National Natural Science Foundation of China (Grant No. 71874131), and the National Natural Science Foundation of China (Grant No. 71972164).

<sup>6</sup> <https://www.scopus.com/search/form.uri?display=basic>.

<sup>7</sup> <https://pubmed.ncbi.nlm.nih.gov/>.

## References

- Adams, J. D., Black, G. C., Clemmons, J. R., & Stephan, P. E. (2005). Scientific teams and institutional collaborations: Evidence from US universities, 1981–1999. *Research Policy*, *34*(3), 259–285.
- Aduku, K. J., Thelwall, M., & Kousha, K. (2017). Do Mendeley reader counts reflect the scholarly impact of conference papers? An investigation of computer science and engineering. *Scientometrics*, *112*(1), 573–581.
- Agrawal, A., Cockburn, I., & McHale, J. (2006). Gone but not forgotten: Knowledge flows, labor mobility, and enduring social relationships. *Journal of Economic Geography*, *6*(5), 571–591.
- Askeridis, J. M. (2018). An h index for Mendeley: Comparison of citation-based h indices and a readership-based h men index for 29 authors. *Scientometrics*, *117*(1), 615–624.
- Babbar, S., Koufteros, X., Behara, R. S., & Wong, C. W. Y. (2019). SCM research leadership: The ranked agents and their networks. *Supply Chain Management: An International Journal*, *24*(6), 821–854.
- Barjak, F., & Robinson, S. (2008). International collaboration, mobility and team diversity in the life sciences: Impact on research performance. *Social Geography*, *3*(1), 23–36.
- Baruffaldi, S. H., & Landoni, P. (2012). Return mobility and scientific productivity of researchers working abroad: The role of home country linkages. *Research Policy*, *41*(9), 1655–1665.
- Basu, A., Foland, P., Holdridge, G., & Shelton, R. D. (2018). China's rising leadership in science and technology: Quantitative and qualitative indicators. *Scientometrics*, *117*(1), 249–269.
- Bercovitz, J., & Feldman, M. (2011). The mechanisms of collaboration in inventive teams: Composition, social networks, and geography. *Research Policy*, *40*(1), 81–93.
- Boschma, R. (2005). Proximity and innovation: A critical assessment. *Regional Studies*, *39*(1), 61–74.
- Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, *30*(1), 107–117.
- Calero-Medina, C., López-Illescas, C., Visser, M. S., & Moed, H. F. (2008). Important factors when interpreting bibliometric rankings of world universities: An example from oncology. *Research Evaluation*, *17*(1), 71–81.
- Callaway, E. (2020). Will the pandemic permanently alter scientific publishing? *Nature*, *582*, 167–168.
- Cantner, U., & Rake, B. (2014). International research networks in pharmaceuticals: Structure and dynamics. *Research Policy*, *43*(2), 333–348.
- Campbell, E. G., Weissman, J. S., Moy, E., & Blumenthal, D. (2001). Status of clinical research in academic health centers: Views from the research leadership. *JAMA*, *286*(7), 800–806.
- Chaocheng, H., Jiang, W., & Qingpeng, Z. (2019). Research leadership flow determinants and the role of proximity in research collaborations networks. *Journal of the Association for Information Science and Technology*. <https://doi.org/10.1002/asi.24331>
- Chaocheng, H., W. Jiang and Z. Qingpeng (2020). Characterizing Research Leadership on Geographical Weighted Collaboration Network. arXiv preprint [arXiv:2004.07102](https://arxiv.org/abs/2004.07102)
- Chen, K. H., Zhang, Y., & Fu, X. L. (2019). International research collaboration: An emerging domain of innovation studies? *Research Policy*, *48*(1), 149–168.
- Chinchilla-Rodríguez, Z., Sugimoto, C. R., & Larivière, V. (2019). Follow the leader: On the relationship between leadership and scholarly impact in international collaborations. *PLoS ONE*, *14*(6), e0218309.
- Collins, H. M. (2001). Tacit knowledge, trust and the Q of sapphire. *Social Studies of Science*, *31*(1), 71–85.
- Cooke, P. (2006). Global bioregional networks: A new economic geography of bioscientific knowledge. *European Planning Studies*, *14*(9), 1265–1285.
- Downs, J. A., & Horner, M. W. (2012). Analysing infrequently sampled animal tracking data by incorporating generalized movement trajectories with kernel density estimation. *Computers, Environment and Urban Systems*, *36*(4), 302–310.
- International Committee of Medical Journal Editors. (2018). Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals. Retrieved October 5, 2019 from <http://www.icmje.org/icmje-recommendations.pdf>.
- Ellis, P. D., & Zhan, G. (2011). How international are the international business journals? *International Business Review*, *20*(1), 100–112.
- Fang, W., Dai, S., & Tang, L. (2020). The impact of international research collaboration network evolution on Chinese business school research quality. *Complexity*. <https://doi.org/10.1155/2020/7528387>
- Fernandez, A., Ferrandiz, E., & Leon, M. D. (2016). Proximity dimensions and scientific collaboration among academic institutions in Europe: The closer, the better? *Scientometrics*, *106*(3), 1073–1092.
- Fiala, D., Rousselot, F., & Jezek, K. (2008). PageRank for bibliographic networks. *Scientometrics*, *76*(1), 135–158.
- Forman, C., & van Zeebroeck, N. (2019). Digital technology adoption and knowledge flows within firms: Can the internet overcome geographic and technological distance? *Research Policy*, *48*(8), 103697.

- Fu, H.-Z., Chuang, K.-Y., Wang, M.-H., & Ho, Y.-S. (2011). Characteristics of research in China assessed with Essential Science Indicators. *Scientometrics*, 88(3), 841–862.
- Gauffriau, M., Larsen, P., Maye, I., Roulin-Perriard, A., & von Ins, M. (2007). Publication, cooperation and productivity measures in scientific research. *Scientometrics*, 73(2), 175–214.
- 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(2), 323–335.
- Glänzel, W., & Schubert, A. (2001). Double effort = double impact? A critical view at international co-authorship in chemistry. *Scientometrics*, 50(2), 199–214.
- Glänzel, W., & Thijs, B. (2004). Does co-authorship inflate the share of self-citations? *Scientometrics*, 61(3), 395–404.
- Gleich, D. F. (2015). PageRank beyond the web. *Siam Review*, 57(3), 321–363.
- González-Alcaide, G., Park, J., Huamaní, C., & Ramos, J. M. (2017). Dominance and leadership in research activities: Collaboration between countries of differing human development is reflected through authorship order and designation as corresponding authors in scientific publications. *PLoS ONE*, 12(8), e0182513.
- Guan, J., Zuo, K., Chen, K., & Yam, R. C. (2016). Does country-level R&D efficiency benefit from the collaboration network structure? *Research Policy*, 45(4), 770–784.
- Gui, Q. C., Liu, C. L., & Du, D. B. (2018). International knowledge flows and the role of proximity. *Growth and Change*, 49(3), 532–547.
- Hansson, F., & Mønsted, M. (2008). Research leadership as entrepreneurial organizing for research. *Higher Education*, 55(6), 651–670.
- Hassan, S.-U., Imran, M., Gillani, U., Aljohani, N. R., Bowman, T. D., & Didegah, F. (2017). Measuring social media activity of scientific literature: An exhaustive comparison of scopus and novel altmetrics big data. *Scientometrics*, 113(2), 1037–1057.
- Haveliwala, T. H. (2003). Topic-sensitive pagerank: A context-sensitive ranking algorithm for web search. *IEEE Transactions on Knowledge and Data Engineering*, 15(4), 784–796.
- Hemlin, S., Allwood, C. M., Martin, B. R., & Mumford, M. D. (2013). Why is leadership important for creativity in science, technology, and innovation? Introduction. In S. Hemlin, C. M. Allwood, B. Martin, & M. D. Mumford (Eds.), *Creativity and leadership in science, technology, and innovation* (pp. 1–26). Routledge.
- Herrling, P. L. (1998). Maximizing pharmaceutical research by collaboration. *Nature*, 392(6679), 32–35.
- Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, 102(46), 16569–16572.
- 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.
- Holste, J. S., & Fields, D. (2010). Trust and tacit knowledge sharing and use. *Journal of Knowledge Management*, 14(1), 128–140.
- Jia, H. (2018). China's plan to recruit talented researchers. *Nature*, 553(7688), S8.
- Jiang, L., Zhu, N. B., Yang, Z. L., Xu, S., & Jun, M. (2018). The relationships between distance factors and international collaborative research outcomes: A bibliometric examination. *Journal of Informetrics*, 12(3), 618–630.
- Jeong, S., Choi, J. Y., & Kim, J. (2011). The determinants of research collaboration modes: Exploring the effects of research and researcher characteristics on co-authorship. *Scientometrics*, 89(3), 967–983.
- Jeong, S., Choi, J. Y., & Kim, J. Y. (2014). On the drivers of international collaboration: The impact of informal communication, motivation, and research resources. *Science and Public Policy*, 41(4), 520–531.
- Jokić, M. (2020). Productivity, visibility, authorship, and collaboration in library and information science journals: Central and Eastern European authors. *Scientometrics*, 122(2), 1189–1219.
- Jones, B. F., Wuchty, S., & Uzzi, B. (2008). Multi-university research teams: Shifting impact, geography, and stratification in science. *Science*, 322(5905), 1259–1262.
- Jonsen, K., Butler, C. L., Mäkelä, K., Piekkari, R., Drogendijk, R., Lauring, J., Lervik, J. E., Pahlberg, C., Vodosek, M., & Zander, L. (2013). Processes of international collaboration in management research: A reflexive, autoethnographic approach. *Journal of Management Inquiry*, 22(4), 394–413.
- Katz, J. S., & Martin, B. R. (1997). What is research collaboration? *Research Policy*, 26(1), 1–18.
- Kim, J., & Diesner, J. (2015). Coauthorship networks: A directed network approach considering the order and number of coauthors. *Journal of the Association for Information Science and Technology*, 66(12), 2685–2696.
- Klavans, R., & Boyack, K. (2008). Thought leadership: A new indicator for national and institutional comparison. *Scientometrics*, 75(2), 239–250.

- Klavans, R., & Boyack, K. (2010). Toward an objective, reliable and accurate method for measuring research leadership. *Scientometrics*, 82(3), 539–553.
- Kwiek, M. (2015). The internationalization of research in Europe: A quantitative study of 11 national systems from a micro-level perspective. *Journal of Studies in International Education*, 19(4), 341–359.
- Laband, D. N., & Tollison, R. D. (2000). Intellectual collaboration. *Journal of Political Economy*, 108(3), 632–662.
- Lander, B. (2015). Proximity at a distance: The role of institutional and geographical proximities in Vancouver's infection and immunity research collaborations. *Industry and Innovation*, 22(7), 575–596.
- Larivière, V., Gingras, Y., Sugimoto, C. R., & Tsou, A. (2015). Team size matters: Collaboration and scientific impact since 1900. *Journal of the Association for Information Science and Technology*, 66(7), 1323–1332.
- Li, J., Zhang, Y., Wang, X., & Ho, Y.-S. (2009). Bibliometric analysis of atmospheric simulation trends in meteorology and atmospheric science journals. *Croatica Chemica Acta*, 82(3), 695–705.
- Li, Q., Zhou, T., Lu, L. Y., & Chen, D. B. (2014). Identifying influential practitioners by weighted Leader-Rank. *Physica A-Statistical Mechanics and Its Applications*, 404, 47–55.
- Li, Y. P., Wang, Z. T., Zhong, X. Y., & Zou, F. (2019). Identification of influential function modules within complex products and systems based on weighted and directed complex networks. *Journal of Intelligent Manufacturing*, 30(6), 2375–2390.
- Liang, L., Zhang, L., Kretschmer, H., & Scharnhorst, A. (2006). Geographical and lingual preferences in scientific collaboration of the European Union (1994–2003) In: *Proceedings international workshop on webometrics, informetrics and scientometrics & seventh COLLNET meeting*, Nancy, France.
- Liu, W., Hu, G., & Tang, L. (2018). Missing author address information in Web of Science—An explorative study. *Journal of Informetrics*, 12(3), 985–997.
- Liu, X., Bollen, J., Nelson, M. L., & Van de Sompel, H. (2005). Co-authorship networks in the digital library research community. *Information Processing & Management*, 41(6), 1462–1480.
- Liu, X. Z., & Fang, H. (2014). Scientific group leaders' authorship preferences: An empirical investigation. *Scientometrics*, 98(2), 909–925.
- Lu, L. Y., Chen, D. B., Ren, X. L., Zhang, Q. M., Zhang, Y. C., & Zhou, T. (2016). Vital nodes identification in complex networks. *Physics Reports-Review Section of Physics Letters*, 650, 1–63.
- Lu, L. Y., & Zhou, T. (2011). Leaders in social networks, the delicious case. *PLoS ONE*, 6(6), 9.
- Lund, B. (2019). Examination of correlates of H-index as a measure of research productivity for library and information science faculty in the United States and Canada. *Scientometrics*, 120(2), 897–915.
- Luo, Q., Xia, J. C., Haddow, G., Willson, M., & Yang, J. (2018). Does distance hinder the collaboration between Australian universities in the humanities, arts and social sciences? *Scientometrics*, 115(2), 695–715.
- Kwiek, M. (2020a). Internationalists and locals: international research collaboration in a resource-poor system. *Scientometrics*, 124(1), 57–105.
- Kwiek, M. (2020b). What large-scale publication and citation data tell us about international research collaboration in Europe: Changing national patterns in global contexts. *Studies in Higher Education*, 45(4), 1–21. <https://doi.org/10.1080/03075079.2020.1749254>.
- McKelvey, M., Alm, H., & Riccaboni, M. (2003). Does co-location matter for formal knowledge collaboration in the Swedish biotechnology–pharmaceutical sector? *Research Policy*, 32(3), 483–501.
- Melitz, J., & Toubal, F. (2014). Native language, spoken language, translation and trade. *Journal of International Economics*, 93(2), 351–363.
- Moya-Anegón, F., Guerrero-Bote, V. P., Bornmann, L., & Moed, H. F. (2013). The research guarantors of scientific papers and the output counting: A promising new approach. *Scientometrics*, 97(2), 421–434.
- Pajić, D. (2015). Globalization of the social sciences in Eastern Europe: Genuine breakthrough or a slippery slope of the research evaluation practice? *Scientometrics*, 102(3), 2131–2150.
- Parreira, M. R., Machado, K. B., Logares, R., Diniz, J. A. F., & Nabout, J. C. (2017). The roles of geographic distance and socioeconomic factors on international collaboration among ecologists. *Scientometrics*, 113(3), 1539–1550.
- Persson, O., Glänzel, W., & Danell, R. (2004). Inflationary bibliometric values: The role of scientific collaboration and the need for relative indicators in evaluative studies. *Scientometrics*, 60(3), 421–432.
- Plotnikova, T., & Rake, B. (2014). Collaboration in pharmaceutical research: Exploration of country-level determinants. *Scientometrics*, 98(2), 1173–1202.
- Quan, W., Mongeon, P., Sainte-Marie, M., Zhao, R., & Larivière, V. (2019). On the development of China's leadership in international collaborations. *Scientometrics*, 120(2), 707–721.
- Rennie, D., Yank, V., & Emanuel, L. (1997). When authorship fails: a proposal to make contributors accountable. *JAMA*, 278(7), 579–585.



- Sánchez-Jiménez, R., Guerrero-Bote, V. P., & Moya-Anegón, F. (2017). The role of guarantor in scientific collaboration: The neighbourhood matters. *Journal of Informetrics*, *11*(1), 103–116.
- Sekara, V., Deville, P., Ahnert, S. E., Barabasi, A. L., Sinatra, R., & Lehmann, S. (2018). The chaperone effect in scientific publishing. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(50), 12603–12607.
- Sidone, O. J. G., Haddad, E. A., & Mena-Chalco, J. P. (2017). Scholarly publication and collaboration in Brazil: The role of geography. *Journal of the Association for Information Science and Technology*, *68*(1), 243–258.
- Sinnott, R. W. (1984). Virtues of the Haversine. *Sky and Telescope*, *68*(2), 158.
- Tang, L. (2013). Does “birds of a feather flock together” matter—Evidence from a longitudinal study on US–China scientific collaboration. *Journal of Informetrics*, *7*(2), 330–344.
- Tang, L., & Shapira, P. (2011). Regional development and interregional collaboration in the growth of nano-technology research in China. *Scientometrics*, *86*(2), 299–315.
- Ukrainski, K., Masso, J., & Kanep, H. (2014). Cooperation patterns in science within Europe: The standpoint of small countries. *Scientometrics*, *99*(3), 845–863.
- Wagner, C. S., Brahmakulam, I., Jackson, B., Wong, A., & Yoda, T. (2001). *Science and technology collaboration: Building capability in developing countries*. RAND.
- Wagner, C. S., Whetsell, T. A., & Leydesdorff, L. (2017). Growth of international collaboration in science: Revisiting six specialties. *Scientometrics*, *110*(3), 1633–1652.
- Wagner, C. S., Whetsell, T. A., & Mukherjee, S. (2019). International research collaboration: Novelty, conventionality, and atypicality in knowledge recombination. *Research Policy*, *48*(5), 1260–1270.
- Wang, L., & Wang, X. (2017). Who sets up the bridge? Tracking scientific collaborations between China and the European Union. *Research Evaluation*, *26*(2), 124–131.
- Wang, W., Wu, Y., & Pan, Y. (2014). An investigation of collaborations between top Chinese universities: A new quantitative approach. *Scientometrics*, *98*(2), 1535–1545.
- Wang, X., Xu, S., Wang, Z., Peng, L., & Wang, C. (2013). International scientific collaboration of China: Collaborating countries, institutions and individuals. *Scientometrics*, *95*(3), 885–894.
- Wu, J. (2013). Geographical knowledge diffusion and spatial diversity citation rank. *Scientometrics*, *94*(1), 181–201.
- Wu, Y. N., Peng, G. Z., Wang, H. W., & Zhang, H. M. (2019). A two-stage fault tolerance method for large-scale manufacturing network. *IEEE Access*, *7*, 81574–81592.
- Xu, S., Mariani, M. S., Lü, L., & Medo, M. (2020). Unbiased evaluation of ranking metrics reveals consistent performance in science and technology citation data. *Journal of Informetrics*, *14*(1), 101005.
- Yan, E., & Ding, Y. (2011). Discovering author impact: A PageRank perspective. *Information Processing & Management*, *47*(1), 125–134.
- Yu, D. E., Razon, L. F., & Tan, R. R. (2020). Can global pharmaceutical supply chains scale up sustainably for the COVID-19 crisis? *Resources, Conservation, and Recycling*. <https://doi.org/10.1016/j.resourcon.2020.104868>
- Zhang, C., & Guo, J. (2017). China’s international research collaboration: Evidence from a panel gravity model. *Scientometrics*, *113*(2), 1129–1139.
- Zhang, X., Wang, X., Zhao, H., de Pablos, P. O., Sun, Y., & Xiong, H. (2019). An effectiveness analysis of altmetrics indices for different levels of artificial intelligence publications. *Scientometrics*, *119*(3), 1311–1344.
- Zhou, B., Lei, Y. F., Li, C. B., Fang, B. L., Wu, Q. W., Li, L. J., & Li, Z. Q. (2019). Electrical LeaderRank method for node importance evaluation of power grids considering uncertainties of renewable energy. *International Journal of Electrical Power & Energy Systems*, *106*, 45–55.
- Zhou, J. L., Zeng, A., Fan, Y., & Di, Z. R. (2018). Identifying important scholars via directed scientific collaboration networks. *Scientometrics*, *114*(3), 1327–1343.