



The effect of collaborations on scientific research output: the case of nanoscience in Chinese regions

Stefano Scarazzati¹ · Lili Wang¹ 

Received: 31 January 2019 / Published online: 14 September 2019
© The Author(s) 2019

Abstract

The pattern and impact of scientific collaborations depends on a country's or region's development stage. Prior to promoting scientific collaborations, it is crucial to understand what types of collaboration are more beneficial. By taking the stage of regional scientific development into consideration, this paper investigates the effect of various collaborations on regional scientific output. Our results highlight the differences between collaboration effects across regions at various capability levels. Scientifically weak regions benefit from more concentrated collaboration, in particular with advanced domestic regions. However, scientifically strong regions can benefit from a centrally located position in a broader collaboration network. Findings from this analysis indicate that international collaborations might not be beneficial a priori. Whether or not one region benefits from international collaborations depends on this region's local capacity. Being embedded in a broad network structure can help increase one region's publication output. However, such benefit is limited only to regions that have relatively higher research capacity.

Keywords Scientific output · Collaborations · Regional capacity · Network structure · Chinese regions · Nanoscience

Introduction

Collaborations have been recognised to play a pivotal role, alongside competition, in the innovation process of a country (de Solla Price and Beaver 1966). The primary benefit from collaborations is the share of knowledge, skills and techniques among partners, and such inflow of knowledge contributes to the “accumulation of knowledge” process that leads to economic development and growth (Katz and Martin 1997; Luukkonen et al. 1992). Not surprisingly, governments tend to encourage research collaborations among authors, including it in the funding conditions, in an attempt to increase the country's scientific output (Lee and Bozeman 2005). Literature has shown that collaborations, and

✉ Lili Wang
wang@merit.unu.edu

Stefano Scarazzati
stefano.scarazzati@gmail.com

¹ UNU-MERIT, Maastricht University, Maastricht, The Netherlands

especially international collaborations, have a positive correlation with the number of publications (Lee and Bozeman 2005; McFadyen and Cannella 2004) and the publications' impact (He et al. 2009; Wuchty et al. 2007).

However, having more collaborations not necessarily translates into higher scientific output, as not every collaboration is ideal (Lee and Bozeman 2005). While collaborations bring benefits to researchers, building a collaboration network also entails significant costs (Jeong et al. 2011; Katz and Martin 1997), hence authors must make trade-offs to maximise their utility. Authors in scientifically developed regions do not seek the same collaborations as those in scientifically weak regions. According to the centre-periphery hypothesis, researchers in less developed areas are willing to collaborate with those in more advanced areas in order to gain access to resources, knowledge, expertise etc., while authors in advanced regions seek for complementarities (Acosta et al. 2011). Empirical studies testing this hypothesis found positive results (Schubert and Sooryamoorthy 2010), although some scepticism remains (Acosta et al. 2011; Wagner and Leydesdorff 2005). Given that countries tend to follow different types of collaborations (Meyer and Persson 1998; Ozcan and Islam 2014), it is crucial to understand which types of collaborations are more beneficial for the studies countries/regions.

To further explain the mechanism of collaborations, few studies emphasize the need of differentiating the sub-unit *level of advancement* (Moed 2016; Duque et al. 2005; Moed and Halevi 2014; UNESCO 2014). The pattern and impact of collaborations is associated with a nation's scientific development stage. However, it is important to point out that a generalised estimation assuming that a nation presents a homogeneous level of development can lead to inaccurate suggestions. This is in particular the case of China where large internal disparities persist (Fan et al. 2011; Wang et al. 2013b; Wang et al. 2019) but are often overlooked in scientific collaboration studies. A region's development stage associated with its absorptive capacity has been shown to be closely connected with its innovation output (Wang et al. 2019). However, the relationship between regional development stage and scientific collaborative performance has hardly been investigated.

Since the opening-up reform, China has experienced an expansion in collaborations, especially in the international context. In the past decades, China established collaborations with more than 150 countries and the number of co-published articles in China increased faster than the average (Niu and Qiu 2014; Zhou and Glänzel 2010). While more and more collaboration cooperation agreements with different countries have been established, the Chinese government has been actively promoting international collaborations between Chinese and foreign researchers, aiming to enhance the internationalization of China's scientific research activities (Andreosso-O'Callaghan 1999; Bound et al. 2013; Wang et al. 2017b). Yet, little is known about what types of collaboration are more beneficial for Chinese regions at various development stages.

Considering that the pattern of international collaborations also varies greatly across different disciplines (Luukkonen et al. 1992; Coccia and Wang 2016), our study focuses on collaborations in one field, i.e. nanotechnology. As a promising, rapidly developing high-tech sector (Zheng et al. 2014), nanotechnology has a great potential for the future development of a wide-range of areas. This is also the field where extensive research funding has been fuelled by Chinese government (Bai 2005) and remarkable development has been observed in different regions (Wang et al. 2019).

This study will contribute to the existing literature, by further exploring the correlation between collaborations and scientific output, in the context of nanotechnology-specific

knowledge production, among the Chinese regions. This paper aims to answer the following questions:

1. Does domestic collaboration positively affect the scientific research output in Chinese regions?
2. Does international collaboration positively affect the scientific research output in Chinese regions?
3. Does a centrally located position positively affect a region's scientific research output?
4. Does regional development stage matter in benefiting from domestic/international collaborations? If so, what types of collaborations are more suitable for low-capability regions and what types of collaborations are more beneficial for high-capability regions?

The structure of the paper is as follows. The next section will present a theoretical and empirical background to the study. Section three will present the data used and will discuss the methodology, including the variables and model specification adopted. Section four will discuss the results from the econometric analysis. Finally, the last section will present conclusions, limitations and some policy advice.

Background

Domestic/international collaborations and research output

Most literature holds a positive view on the contribution of collaborations to scientific output. Half a century ago, de Solla Price and Beaver (1966) found a positive correlation between the amount of collaborations an author has, and his or her total number of publications, suggesting that the most productive researchers are also the ones collaborating the most. This theory has been supported by recent empirical studies, which found a positive association of collaborations with scientific output (i.e. the number of publications) and with their social impact (i.e. the number of citations) (Adams et al. 2005; Glänzel 2001; McFadyen and Cannella 2004; McFadyen et al. 2009).

The strength of the collaborations in particular has been considered as an important element. Having strong partnerships entails higher trust and reciprocity, which reduces the costs and risks from new collaborations, and in turns positively affects the scientific output (Gonzalez-Brambila et al. 2013; Guan et al. 2015a).

Collaborations are differentiated in the literature between domestic and international collaborations. Existing literature argues that international collaborations are preferable over domestic ones as they seem to produce higher impact (Jeong et al. 2014; Leydesdorff et al. 2014; Tang and Shapira 2012; Wagner et al. 2018). Ebersberger et al. (2014) found that, while a high level of regional technological specialization and technological variety were negatively correlated with domestic extra-regional collaborations, technological variety was positively associated with the propensity to collaborate with foreign partners. Such findings suggest a strong preference of researchers towards international collaborations, rather than domestic networks. Chinese policies in recent years have been in line with this view, actively promoting transnational collaborations through the participation to different international research projects, such as Framework Programme 7, Horizon 2020 and World-Class 2.0 (European Commission 2007, 2018; Zhao 2018).

Given the unbalanced development across China, whether different regions have the capability to manipulate and benefit from such collaborations remains a question. As aforementioned in the introduction section, an important gap in the existing literature consists in the focus on most studies on country level analyses, while only few studies have differentiated for the sub-unit *level of advancement* when exploring ‘who’ benefits from those collaborations (Duque et al. 2005). With the aim to fill this gap, this study investigates the impact of different types of collaborations on the scientific output of Chinese regions at different development stages.

Diversified collaborations and research output

Scholars have also studied how diversified collaborations impact research performance. Some studies suggest that differences between the collaborators negatively impact on their performance, as they could hamper the knowledge flow and exchange (Hoskisson et al. 2002). Nevertheless, a large body of literature has shown that collaborative diversity potentially contributes to the innovation process, allowing partners to access a diverse range of skills, expertise and knowledge (Gonzalez-Brambila et al. 2013; Guan et al. 2015a).

To examine the diversity of collaborations, a large body of the literature has applied social network analysis to study a partner’s position embedded in the collaboration networks (Abbasi et al. 2012). Partners involved in a social network share workload, information, skills and expertise, equipment, etc. (Lee and Bozeman 2005; Li et al. 2013). The configuration of linkages among partners, i.e. who you reach and how (Nahapiet and Ghoshal 1998), is a main determinant in the process of knowledge creation (McFadyen and Cannella 2004). Researchers attempt to answer the question of how partners should change their interactions to acquire a more desirable and advantageous position within a network (Li et al. 2013). Given that the most influential partner(s) are often the ones located in central places, researchers have developed several centrality measures to determine the importance of a partner within a network (Abbasi et al. 2011). The main measures used include degree centrality, closeness centrality, betweenness centrality and eigenvector centrality, which measure the following respectively: the number of collaborators each node has; the distance of a node to other nodes in the network; the number of times a node lies ‘between’ other pairs of nodes; the value of centrality of the nodes connected to a node (Abbasi et al. 2011).

Actors embedded in a large and cohesive network (with a high degree of centrality) are believed to hold an advantageous position, which allows them to access new, richer and diverse information (Gonzalez-Brambila et al. 2013; Phelps et al. 2012). Having more ties in the network (i.e. being in a centralized position) can promote the generation of novel and useful ideas (McFadyen and Cannella 2004; McFadyen et al. 2009).

A large stream of the literature has supported the importance of holding a bridging position within the network (high betweenness centrality), as it allows the actors to benefit from the access to new, non-redundant knowledge and resources (Burt 1992; Guan et al. 2015b; Li et al. 2013; Gonzalez-Brambila et al. 2013). In their studies, Gonzalez-Brambila et al. (2013) measured a striking prevalence for the role of brokerage over cohesion on both citations and publications output, while Li et al. (2013) found betweenness centrality to have the largest and most significant impact on citations, out of the six indicators analysed. Guan et al. (2015b) built a multilevel collaboration network to measure the effect of network position in the process of innovation production; using betweenness centrality as a measure of structural centrality, they found a positive significant effect of the latter on

innovation output. In line with the literature, we chose to use betweenness centrality as a measure of verified collaboration with an important position.

The analysis contributes to the existing literature in different ways. First, most studies in the literature focused on the impact of network characteristics on knowledge transfer and flow, or on publications impact (McFadyen et al. 2009), while only a few studies examined the impact on knowledge creation. This work contributes to the literature by estimating the relationship between the network position and knowledge creation—measured by publications—in the field of nanotechnology. In addition, this work provides evidence of the impact that a broad collaboration structure has on the regional knowledge creation in China, over a 17-year dataset.

The case of nanoscience development among Chinese regions

Nanotechnology is a promising, rapidly developing high-tech sector (Zheng et al. 2014). Nanoscience results from the cooperation of multidisciplinary fields, i.e. chemistry, physics, biotechnology, engineering and material science, towards the study of atoms and molecules (Schummer 2004). This field has a great potential for the future development of a wide-range of areas including energy, healthcare, pharmaceutical industry, food industry and climate change (Sozer and Kokini 2009; Zheng et al. 2014). Not surprisingly, nanotechnology is progressing rapidly, and international collaborations have been found to play a significant role in its development (Zheng et al. 2014). Over the past decades, China significantly invested in nanotechnology, and it rapidly became one of the leading nations for what concerns the share of the world's publication on nanotechnology (Tang and Shapira 2011b; Zhou and Leydesdorff 2006). In 2004, China's world share of nano-related publications – 8.34%—was higher than the country's average world share of publications – 6.52%—(Zhou and Leydesdorff 2006). The development of nanotechnology in China reflects the desire from the government to move a developing country towards the global technology-economic frontier (Wang et al. 2019) and gain a leading role within the international context. Therefore, studying the extent to which collaborations impact on China's regional development in a highly relevant-to-the-government sector can provide guidance for more effective investments.

Along with the rapid growth in scientific and technological output related to nanotechnology, an unbalanced regional development is also observed in China. Wang et al. (2012) studied nanotechnology collaborations between China and the US and they measured that such collaborations were asymmetrical, with a small group of Chinese scientists working with a large number of US scientists. Tang and Shapira (2011a) found that a small number of “elite” universities in China were collaborating with a wide range of universities in the US, while the majority of the Chinese universities had few (or no) partnerships. Using patent network data, Ozcan and Islam (2014) find that the highly centralized networks in China were dominated by a few large players.

A possible explanation can be found in the role played by technological proximity, which refers to the shared knowledge base of different collaborators (Cunningham and Werker 2012). According to Cunningham and Werker (2012), actors must be different enough to benefit from each other's knowledge, but sufficiently similar to understand each other. Therefore, not all the regions might benefit from the same type of collaborations, simply because they might be too much similar to/different from each other. Less developed regions in China are too different from foreign leading countries, hence they could not maximise the benefits from such collaborations. While developed Chinese regions

might have the necessary capacity and resources to optimise the benefits coming from foreign partners, other less-developed regions do not have the capability necessary to benefit from those collaborations, due to barriers and costs, such as financial resources, administration, etc. (Katz and Martin 1997). Moreover, international collaborations involve significantly higher costs than shorter-distance collaborations (Wagner 2006). Thus, regions face trade-offs between costs and benefits to decide the best strategy to implement (Jeong et al. 2014). In this study we will follow the approach of Duque et al. (2005) and Lee and Bozeman (2005), and we investigate “who” benefits from “what”.

Data collection and methodology

Data collection

For the bibliometric analysis, this paper uses a panel dataset, from 1999 to 2015, containing records of the collaborations for co-publications on nanotechnology, among 30 Chinese mainland regions¹ and between the mainland regions and 27 non-mainland/foreign regions.² The dataset contains 419,910 total publications in the nanotechnology field, obtained from Clarivate Analytics Web of Science (WoS), Science Citation Index Expanded. Moreover, the dataset was constructed using a lexical query searching and defining strategy developed by the Georgia Institute of Technology (see more in Porter et al. 2008; Wang et al. 2013a). The query used to search for publications on nanotechnology is reported in the “Appendix”. The analysis is based on based on 280,543 meso- and macro-level collaboration connections, 190,126 of which are among the Chinese mainland provinces, while 90,417 collaborations are between the mainland provinces and 27 non-mainland/foreign regions. The data for the control variables implemented in the analysis was obtained and re-elaborated from: China National Intellectual Property Administration (CNIPA),³ China Statistical Yearbook (various issues) and China Statistical Yearbook on Science and Technology (various issues). See Fig. 1 for the steps of data collection and processing.

As aforementioned, the unbalance in regional development is as one of the main issues characterising Chinese regions; hence, this paper will need to separate the estimations, accounting for different regional levels, in order to obtain more significant and unbiased results. As there is no standard classification for Chinese regions, various attempts have been made, including the twofold (i.e. coastal and inland) and threefold (coastal, middle and western) geographical groups (Hao and Wei 2010; Wang and Szirmai 2013). To reflect the scientific research ability, we contend that it is more appropriate to make the regional division based on the volume of scientific knowledge. Therefore, following the methodology used by Wang et al. (2019), this work uses the total number of scientific publications

¹ Tibet is not included due to the lack of data. See Wang et al. (2019) for more details.

² We include only the major partners in this non-mainland/foreign group, and they include Australia, Austria, Belgium, Brazil, Canada, England, France, Germany, Hong Kong, India, Ireland, Japan, Macao, Malaysia, Netherlands, Norway, Pakistan, Portugal, Russia, Scotland, Singapore, South Korea, Spain, Sweden, Switzerland, Taiwan and USA.

³ Which was formerly known as China’s State Intellectual Property Office (SIPO).

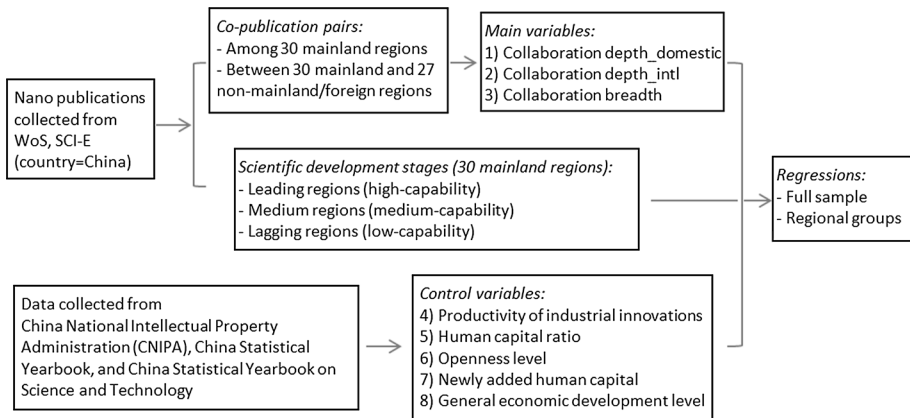


Fig. 1 Steps of data collection and processing

as the criterion to divide the 30 Chinese regions under study into three groups,⁴ respectively “advanced”, “medium” and “lagging” regions, based on their knowledge capability level, as measured by the number of total annual regional publications. These three regional groups are also referred to high-, medium- and low-capability regions.

Variables

Dependent variable

The dependent variable in this analysis is the total number of regional nanotechnology-related publications in each year, which is used to represent the level of nanotech-knowledge generated in the specific region and specific year.

Collaboration variables

Depth of domestic collaborations Based on the matrix of scientific collaborations between Chinese regions, we calculate the intensity of such collaborations using the Jaccard index, to measure the strength of bilateral relationships between different regions (Luukkonen et al. 1993; Wang et al. 2017b). The index was firstly introduced in 1973 by Henry Small; given two different sets of papers for regions X and Y, the Jaccard index is defined as the intersection (number of co-authored papers) divided by the union of the two sample sets (Leydesdorff 2008; Small 1973). Numerically, the domestic collaboration intensity can be written as:

$$CI_{xy} = \frac{Coll_{xy}}{Pub_x + Pub_y - Coll_{xy}} \quad (1)$$

⁴ “Advanced” regions include Beijing, Liaoning, Jilin, Shanghai, Jiangsu, Zhejiang, Anhui, Shandong, Hubei, Guangdong; “medium” regions include Tianjin, Hebei, Heilongjiang, Fujian, Henan, Hunan, Chongqing, Sichuan, Shaanxi, Gansu; “Lagging” regions include Shanxi, Inner Mongolia, Jiangxi, Guangxi, Hainan, Guizhou, Yunnan, Qinghai, Ningxia, Xinjiang.

where CI is the intensity of collaboration (Jaccard) index, $Coll_{xy}$ is the total number of collaborations between regions X and Y; Pub_x is the total number of scientific publications in region X and Pub_y is the total number of scientific publications in region Y. The index can take values between 0 and 1, where 0 indicates that there are no collaborations between the two regions, and value 1 means that the total number of publications equals the number of collaborations; i.e. 100% of the publications in X and Y are the result of the collaborations between the two regions (Boschma et al. 2014). To obtain more meaningful results, the CI variable for collaboration intensity with other regions is not calculated out of the total sample, implying that there is not one single variable for “domestic collaboration intensity”. Instead, three variables are created, each measuring the intensity of collaborations with advanced regions, with medium regions, and with lagging regions respectively.

Depth of international collaborations Similar to the index to capture the depth of domestic collaborations (see Eq. 1), one can also construct the variable to measure the international collaborations. However, the correlation between the two types of collaboration intensities will influence the regression quality. Hence we build the external–internal ($E-I$) index to examine the depth of international collaborations. This measures the dominance of external ties in a network, over the internal ones (Krackhardt and Stern 1988). The $E-I$ index, proposed by Krackhardt and Stern (1988), focuses on the relative stronger role of “friendships” between organisations’ subunits, to those within units, and is obtained by the following formula:

$$E-I \text{ index} = \frac{E_i - I_i}{E_i + I_i} \quad (2)$$

where E_i is the number of ties external to the network, while I_i is the number of internal ties. The index ranges from -1 to $+1$, and it indicates the extent to which the network ties cut across the group boundaries: a value of “ -1 ” indicates that all the network ties are within the group (with strong domestic collaborations), while “ $+1$ ” means that all the network ties happen across group (with strong international collaborations). The value “ 0 ” is a special case in which the number of ties within and across groups are identical (Gonzalez-Brambila et al. 2013; Krackhardt and Stern 1988; Tortoriello and Krackhardt 2010).

We calculate the $E-I$ index, measuring the dominance of external over internal network ties. In every region, external ties are all the connections that involve collaborations between this region and foreign partner(s), while internal ties are the collaborations between the region and other domestic partners. Regions might be constrained in the population size and amount of resources they can dedicate to building network collaborations (Krackhardt and Stern 1988), resulting in a trade-off between the amount of internal and external relationships they can maintain. Hence the $E-I$ index adopted in this analysis allows us to consider the impact of external collaborations *relative* to the internal ones.

Breadth of collaborations (with importantly embedded positions) Besides the strength of domestic and international collaborations mentioned above, it is also important to understand how broadly and importantly a player is embedded in the collaboration network. To this end, as mentioned in “[Diversified collaborations and research output](#)” section, several centrality measures (including degree centrality, closeness centrality, betweenness centrality and eigenvector centrality) were used. These centrality measures are highly correlated with each other and share a great deal of similarity (Guan et al. 2015b; Valente et al. 2008). Moreover, existing literature has shown that, in capturing the advantageous position in a

network, betweenness centrality is more relevant than other centrality measures (see for example Gonzalez-Brambila et al. 2013; Guan et al. 2015b; Li et al. 2013). Hence, only betweenness centrality will be used for the estimations in the regressions.

Betweenness centrality measure, first proposed by Freeman (1977), is defined as the extent to which a node lies on the paths between others. In other words, it is a measure of the fraction of shortest paths (or ‘geodesic’ distance) between pairs of nodes in the network that pass through the given node (Newman 2005). It is calculated based on the formula below:

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}} \quad (3)$$

where the betweenness centrality (C_B) of a node v is equal to the sum of the number of shortest paths (σ) from any node s to a node t , which the node v lies on, divided by the number of paths from s to t , with v , s and t belonging to the set of vertices/nodes V (Brandes 2001). Moreover, the betweenness centrality measure is normalised to obtain a value between 0 and 1, where 0 is the lowest betweenness, and 1 is the highest (Freeman 1977).

Control variables

Productivity of industrial innovations—granted patents to R&D ratio Researchers have extensively made use of patenting counts as indicators for the current state of a technology level (Archibugi and Pianta 1996; Youtie et al. 2008). Therefore, the model includes the regional yearly number of industry patents, divided by the total regional level of R&D expenditure, as a proxy for the region’s productivity of industrial innovations.

Human capital ratio—share of university personnel Evidence suggests that the level of research personnel has a significant impact on technological output (Zhang 2017). Therefore, the model controls for the regional level of university personnel over the regional population, lagged by 1 year. There are two main reasons for observing the personnel level at time $t-1$: first, the lagged value of the explanatory variable reflects the fact that the knowledge production in a specific year is the result of a research in the past (most likely in the previous year) (Fu 2008). Second, using 1 year lag for the independent variable, allows us to eliminate possible endogeneity bias for the variables of interest (Zhang 2017).

Openness level—FDI to GDP ratio Literature has widely confirmed the significant role of foreign investment in the development of scientific knowledge and output, and there is large evidence also for the specific case of China (Fu 2008; Kuo and Yang 2008; Zhang 2017). The annual regional FDI is normalized to make it comparable, by dividing it for the regional level of yearly GDP. Same as the previous variable, in order to capture the lagged contribution of this factor and to eliminate possible endogeneity bias from the model, this indicator is lagged by 1 year.

Newly added human capital—graduates The total value of regional graduates, i.e. the number of individuals with a Masters or PhD in a region in a given year, captures the effect

of newly added human capital in the process of knowledge production.⁵ In the analysis, the total number of regional graduates is normalized by taking the natural logarithm. Moreover, similarly to the university personnel and FDI variables, a 1 year lag is considered for possible delayed impacts of newly-graduated students on scientific productivity (time for finding a job, time to work on a publication etc.), as well as to eliminate endogeneity biases.

General economic development level—regional GDP Given the unbalance in development across the Chinese provinces, it is important to control for regional GDP, which considers the different stages of economic advancement of the regions (Zhang 2017). Taking the natural logarithm of GDP is necessary to make the results comparable, given the highly skewed distribution of gross domestic product among the Chinese provinces. Similar to human capital and openness variables, 1 year lag is also taken for this indicator.

Model

As discussed in previous section, the dependent variable in this analysis is the number of nanotechnology publications, which is a non-negative integer. Therefore, count data models, such as Poisson or Negative Binomial (NB) models are preferred to traditionally used models such as Ordinary Least Square (OLS) models (Gonzalez-Brambila et al. 2013; Guan and Liu 2016; Guan et al. 2015b; Wang et al. 2019).

The Poisson model is a maximum likelihood estimation model, used in the literature for non-negative count data, which follows a Poisson distribution (Wooldridge 2002). The Poisson model is a restricted version of the Negative Binomial, and it imposes a strong assumption of constant variance; i.e. it requires the mean of the dependent variable to be equal to its variance (Gardner et al. 1995; Guan et al. 2015b). However, the dataset in our analysis suffers from over-dispersion, as shown by the standard deviation of nanotechnology publications ($\sigma = 1332.21$) significantly exceeding its mean ($\mu = 823.4$). It presents a strongly right-skewed distribution, with many regions producing a small number of publications, and few regions producing a significantly high amount of publications. Since this could lead to low standard errors and p values, the Negative Binomial model results to be the most appropriate model, as it allows to correct for over-dispersion in the dependent variable (Guan and Liu 2016).

The Negative Binomial model was proposed by Hausman et al. (1984), to estimate the relationship between patents and R&D expenditure, between 1968 and 1974, using a non-negative integer as the dependent variable, and accounting for the yearly random and fixed effects. Such model addresses the limitations from the Poisson model's assumptions, by adding an alpha (α) parameter to the regression, which accounts for the unobserved heterogeneity among the observations (Cameron and Trivedi 2005). The negative binomial regression model in our analysis is as follows:

$$P(Y = y_{it} | x_{it}) = \frac{\Gamma(y_{it} + (1/\alpha))}{\Gamma(1/\alpha)\Gamma(1 + y_{it})} \left(\frac{1}{1 + \alpha\mu_{it}} \right)^{1/\alpha} \left(\frac{\alpha\mu_{it}}{1 + \alpha\mu_{it}} \right)^{y_{it}} \quad (4)$$

⁵ Given the Hukou system in China, it is common that graduates tend to find a job in the same region after their studies.

where y_{it} is the number of nano-publications in region i at year t , Γ is the gamma function, μ_{it} is the conditional mean, α is the degree of dispersion. The conditional mean μ_{it} can be expressed as:

$$\begin{aligned} \mu_{it} = & \exp(\beta_0 + \beta_1 * CI_{it}^{high} + \beta_2 * CI_{it}^{medium} + \beta_3 * CI_{it}^{low} + \beta_4 * (E - I)_{it} \\ & + \beta_5 * BETWEENNESS_{it} + \beta_6 * (TOTPAT/R\&D)_{it} + \beta_7 * UNIEMPL_{it-1}^{share} \quad (5) \\ & + \beta_{8*} * (FDI/GDP)_{it-1} + \beta_9 * lnGRAD_{it-1} + \beta_{10} * lnGDP_{it} + \delta_{it}) \end{aligned}$$

where CI_{it}^{high} , CI_{it}^{medium} , CI_{it}^{low} are the collaboration intensities with advanced, medium and lagging regions respectively, for region i in year t . $(E - I)_{it}$ is the internationalization index, while $BETWEENNESS_{it}$ is the normalized betweenness centrality. $(TOTPAT/R\&D)_{it}$, $UNIEMPL_{it-1}^{share}$, $(FDI/GDP)_{it-1}$, $lnGRAD_{it-1}$ and $lnGDP_{it}$ are the control variables: total granted patents ratio to R&D expenditure, the share of university personnel, the FDI over GDP ratio, the share of graduates and the log of regional GDP. Apart from the total granted patents ratio to R&D expenditure $(TOTPAT/R\&D)_{it}$, and regional GDP $(lnGDP_{it})$, which are measured for year t , the other control variables are taken at time year $t - 1$, for the reasons explained above.

Although it is widely used in the literature, we must acknowledge two main limitations of the Hausman et al. (1984) model. First, Allison and Waterman (2002) argued how the fixed-effect negative binomial model by Hausman et al. (1984) is not a “true” fixed-effect model. They demonstrated that the model does not control for all the covariates that remain stable over time. Given such an issue with the fixed-effect specification for the NB model, we adopted the population averaged (PA) Negative Binomial model (Cameron and Trivedi 2005; Wang et al. 2019). The PA specification model considers the average of the random effects, instead of using the observation-specific random effects, and it provides estimates for the average of the observations (Cameron and Trivedi 2005).⁶ To prove for robustness of the NB model, the analysis compares the estimates with those from a fixed-effect (FE) Poisson regression. The FE Poisson is a more advanced version of the normal Poisson, and it allows estimating correlations in longitudinal datasets. Although Poisson model’s fit for the sample in analysis is not ideal and produces biased estimators, the FE Poisson has strong robustness properties for the estimation of parameters (Wooldridge 2002), thus allowing us to compare estimates between models.

The second limitation is that the results of the Negative Binomial model might suffer from relative lower precision, thus reducing the accuracy of the conclusions, due to the larger standard errors estimated by the model (Gonzalez-Brambila et al. 2013). However, in testing the strength of the estimations, through a cross-models comparison, it is possible to control also for the consistency in the significance of the estimations. To provide an additional robustness check for the estimated model, the results from NB and FE Poisson models will also be compared to estimations using an OLS model, where the dependent variable is the natural logarithm of the nano-technology publications. Although the magnitude of the in the OLS estimations are not comparable to those from the other two models, given the different dependent variables adopted, nevertheless it is used in the literature to

⁶ Although it does not constitute a problem for this analysis, since it estimates the average effect on the dependent variable across provinces, nevertheless a problem might raise due to the choice of *random effect* over *fixed-effect* model. It could be the case that individual-specific effects are non-random, but correlated with the regressors in the equation (Cameron and Trivedi 2005).

Table 1 Descriptive statistics for the full sample *Source:* Author's own calculations

Variables	Mean	SD	Min	Max
Nano publications	823.3529	1332.2136	0.00	11,305.00
Coll. Intensity with leading regions	0.0066	0.0049	0.00	0.03
Coll. Intensity with medium regions	0.0045	0.0030	0.00	0.02
Coll. Intensity with lagging regions	0.0018	0.0017	0.00	0.02
<i>E-I</i> index	-0.4862	0.2523	-1.00	1.00
Betweenness centrality normalized	0.0189	0.0305	0.00	0.29
Granted patents/R&D	0.9376	0.6175	0.20	5.78
University personnel/regional population ($t-1$)	16.2832	13.0457	3.54	86.09
FDI/GDP ($t-1$)	0.0619	0.0776	0.01	0.77
Log regional graduates ($t-1$)	8.2535	1.5962	2.77	11.33
Log GDP	8.7653	1.1355	5.47	11.20

provide information about the general direction and magnitude of the regressors' estimated coefficients (Guan et al. 2015b).

Results

Descriptive statistics

Table 1 summarises the general mean, standard deviation, minimum and maximum values for the variables used in the analysis, for the whole dataset. It is possible to observe that while the mean of nano-publications was equal to 823.35 for the period 1999–2015, the standard deviation ($\sigma = 1332.21$) was higher than the mean, suggesting a highly skewed distribution across the regions, as already mentioned in the methodology discussion (“[Data collection and methodology](#)” section). Table 1 shows that the standard deviation is significantly high for most of the variables used in the analysis, and in few cases, it exceeds the mean, suggesting for over-dispersion, as in the case of ‘nano-publications’, ‘betweenness centrality’ and ‘FDI/GDP’ ratio.

To better describe the situation across regions, Table 2 presents the descriptive statistics, divided by the three groups. The table shows that the mean (μ) of nano-publications was 110.34 for lagging regions, while it was 6–16 times higher for medium and advanced regions ($\mu = 617.45$ and 1742.27, respectively).

For what concerns the explanatory variables of the model, the summary statistics table highlights a significant difference in the intensity of collaborations, both nationally and internationally, among groups. For example, the average of collaboration intensity with advanced, medium and lagging regions, for medium regions is generally two to three times larger than for the lagging ones; for advanced regions, it is two to five times larger than the lagging ones. A similar trend applies to betweenness centrality, with lagging regions having a mean significantly lower than medium and advanced regions.

Regarding the control variables, Table 2 shows that the share of research personnel is remarkably unbalanced across regions, with lagging regions having on average less than half of the personnel in advanced regions. Moreover, the mean of the ratio of FDI to regional GDP in lagging and medium regions is significantly lower than that in advanced ones.

Table 2 Descriptive statistics by regional groups *Source:* Authors' own calculations

Variables	Lagging regions			Medium regions			Advanced regions					
	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max
	Nano publications	110.3412	169.65	0.00	980.00	617.4471	612.17	11.00	2752.00	1742.2706	1881.06	71.00
Coll. Intensity with leading regions	0.0020	0.00	0.00	0.01	0.0064	0.00	0.00	0.01	0.0114	0.00	0.00	0.03
Coll. Intensity with medium regions	0.0023	0.00	0.00	0.02	0.0047	0.00	0.00	0.02	0.0064	0.00	0.00	0.02
Coll. Intensity with lagging regions	0.0012	0.00	0.00	0.01	0.0023	0.00	0.00	0.02	0.0020	0.00	0.00	0.01
<i>E-I</i> index	-0.6670	0.28	-1.00	1.00	-0.4920	0.19	-1.00	-0.02	-0.3081	0.13	-0.63	0.04
Betweenness centrality normalized	0.0021	0.00	0.00	0.03	0.0130	0.01	0.00	0.08	0.0416	0.04	0.00	0.29
Granted patents/R&D	1.0360	0.74	0.23	5.78	0.8362	0.41	0.20	2.97	0.9408	0.64	0.26	5.65
University personnel/regional population (<i>t-1</i>)	10.5930	3.63	3.54	17.01	15.2022	7.93	4.62	39.28	23.0546	18.89	4.53	86.09
FDI/GDP (<i>t-1</i>)	0.0513	0.11	0.01	0.77	0.0449	0.04	0.01	0.20	0.0894	0.06	0.02	0.25
Log regional graduates (<i>t-1</i>)	6.8529	1.57	2.77	9.10	8.6019	1.00	6.34	10.22	9.3058	0.99	7.03	11.33
Log GDP	7.9609	1.09	5.47	9.79	8.8898	0.88	6.84	10.52	9.4453	0.89	7.42	11.20
Observations	170				170				170			

Table 3 Correlation matrix *Source:* Authors' own calculations

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1. Coll. Intensity with leading regions	1									
2. Coll. Intensity with medium regions	0.703***	1								
3. Coll. Intensity with lagging regions	0.261***	0.250***	1							
4. <i>E-I</i> index	0.503***	0.296***	0.122**	1						
5. Betweenness centrality normalized	0.536***	0.427***	0.0888	0.439***	1					
6. Granted patents/R&D	-0.112*	-0.191***	0.00273	0.0663	-0.154***	1				
7. University personnel/regional population ($t-1$)	0.634***	0.518***	0.190***	0.351***	0.673***	-0.311***	1			
8. FDI/GDP ($t-1$)	0.200***	0.0820	-0.0161	0.262***	0.276***	0.128**	0.305***	1		
9. Log regional graduates ($t-1$)	0.734***	0.598***	0.421***	0.504***	0.376***	-0.211***	0.515***	0.0086	1	
10. Log GDP	0.606***	0.472***	0.392***	0.375***	0.164***	0.0382	0.192***	-0.0235	0.850***	1

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

It is important to note that Table 2 presents a particularly high standard deviation for many of the variables, hinting that even after dividing the regions in sub-groups, these are still not perfectly homogeneous. Although the groups division and the estimates that will be presented are considered robust, it is necessary to be careful in the interpretation and generalisation of results.

Table 3 shows the correlation matrix for the explanatory variables of the model. While most of the variables have a correlation lower than 0.6, the table shows few pairs of variables significantly correlated. Those particularly concern the first variable (collaboration intensity with leading regions), which is correlated with CI_{it}^{medium} ($\rho = 0.703$), university personnel ($\rho = 0.634$), regional graduates ($\rho = 0.734$), and GDP ($\rho = 0.606$). Also, betweenness centrality is correlated with university personnel ($\rho = 0.673$), while regional graduates are also correlated with GDP ($\rho = 0.850$).

Since the negative binomial model does not provide statistics to control for possible multicollinearity that might raise from such high correlation (McFadyen et al. 2009), it is necessary to make robustness checks for the correlated variables. First, since the highest correlation concerned two of the control parameters, Table 7 in “Appendix” reports the results from checks for multicollinearity between the control variables by regressing them together (Model 1), as well as taking out only log GDP (Model 2), or regional graduates (Model 3). No significant changes in the coefficient signs or significance levels are revealed for the variables. Moreover, the Wald test at the bottom of the table indicates that models (2) and (3) are less significant than (1), suggesting that all the control variables should be included. Models (4), (5) and (6) in Table 7 control for the correlation between CI_{it}^{high} and CI_{it}^{medium} , showing no significant change in the coefficients. The same applies for the regressions in columns (7), (8) and (9), which control for the correlation of CI_{it}^{high} with university personnel, graduates and GDP. Finally, columns (10) and (11) show no significant effects for the correlation between betweenness and university personnel. It follows that there is no important multicollinearity in our model, hence conclusions can be drawn.

Econometric analysis

The impact of explanatory variables

After controlling for biases due to multicollinearity, it is possible to analyse the regressions output. Table 4 presents the results from the negative binomial model for the full sample, with different combinations of independent variables. Table 5, instead, shows the results for the main regressions in Table 4, divided for the three regional groups—respectively advanced, medium and lagging regions. This was calculated using Stata’s split coefficient command, to gain a better understanding of the impact of the explanatory variables on knowledge creation, for different regions. The models include all the control variables, as well as dummy variables for the year fixed-effects, which have been omitted from the regression tables. In Tables 4 and 5, column (1) provides estimates just for the control variables. Columns (2) to (5) present the estimates of the collaboration intensities with advanced, medium and lagging regions, first individually, then combined. Column (6) shows the estimate for $E-I$ index, and column (7) estimates betweenness centrality impact alone. Column (8) shows the estimates for the full model.

Starting from the collaboration indicators, in Table 4 it is possible to see how collaborating with advanced regions has a strong positive effect on publications, which is consistent

Table 4 Results of negative binomial regressions for the full dataset *Source:* Authors' own calculations

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Coll. Intensity with advanced regions		63.27*** (10.48)			62.08*** (10.39)			70.94*** (9.062)
Coll. Intensity with medium regions			21.06** (9.236)		21.13* (10.87)			28.04*** (10.87)
Coll. Intensity with lagging regions				-11.78 (11.52)	-7.754 (12.00)			-8.272 (10.82)
<i>E-I</i> index						0.230** (0.106)		0.318** (0.128)
Betweenness centrality normalized							2.308** (1.038)	4.169*** (0.713)
Granted patents/R&D	-0.175 (0.151)	-0.221* (0.128)	-0.164 (0.158)	-0.178 (0.150)	-0.209 (0.133)	-0.148 (0.149)	-0.156 (0.138)	-0.154 (0.110)
University personnel/regional population ($t-1$)	0.0118 (0.00957)	0.0157*** (0.00601)	0.0123 (0.00935)	0.0117 (0.00954)	0.0154*** (0.00584)	0.0115 (0.0102)	0.00420 (0.00891)	0.00839 (0.00673)
FDI/GDP ($t-1$)	0.279* (0.163)	0.225 (0.148)	0.315* (0.186)	0.260 (0.161)	0.236 (0.186)	0.263 (0.180)	0.182 (0.119)	0.105 (0.173)
Log regional graduates ($t-1$)	0.654*** (0.129)	0.723*** (0.114)	0.679*** (0.130)	0.657*** (0.127)	0.739*** (0.113)	0.601*** (0.128)	0.603*** (0.137)	0.658*** (0.121)
Log GDP	0.426** (0.167)	0.327** (0.136)	0.391** (0.172)	0.427** (0.166)	0.291** (0.143)	0.453*** (0.162)	0.480*** (0.169)	0.335** (0.138)
Constant	-3.812*** (0.910)	-3.889*** (0.701)	-3.840*** (0.885)	-3.831*** (0.900)	-3.811*** (0.701)	-3.587*** (0.959)	-3.852*** (0.927)	-3.619*** (0.675)
Observations	480	480	480	480	480	475	480	475
Wald chi2	11,494.5	7139.0	20,338.0	11,886.2	8365.6	12,633.0	18,469.1	6673.5
Prob > chi2	0	0	0	0	0	0	0	0

(1) Dependent variable is nano-publications. (2) Robust standard errors in parentheses

(3) * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. (4) Year dummy variables are omitted from the table

Table 5 Results of negative binomial regressions by regional groups. *Source:* Authors' own calculations

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Coll. Intensity with leading regions (lagging)		142.0*** (19.68)			132.6*** (20.15)			142.7*** (20.66)
Coll. Intensity with leading regions (medium)		68.25*** (15.82)			87.31*** (15.89)			88.89*** (14.55)
Coll. Intensity with leading regions (advanced)		36.66* (19.83)			49.43*** (16.72)			64.00*** (16.81)
Coll. Intensity with medium regions (lagging)			39.84 (24.72)		8.389 (16.37)			7.164 (12.13)
Coll. Intensity with medium regions (medium)			28.29*** (10.26)		41.65*** (9.962)			43.72*** (8.481)
Coll. Intensity with medium regions (advanced)			6.699 (18.65)		1.059 (19.15)			17.16 (17.78)
Coll. Intensity with lagging regions (lagging)				26.38*** (8.365)	6.051 (7.195)			4.141 (7.704)
Coll. Intensity with lagging regions (medium)				-17.93 (12.49)	-10.13 (12.09)			-15.17 (11.41)
Coll. Intensity with lagging regions (advanced)				-106.6*** (28.18)	-128.2*** (31.62)			-92.16*** (25.33)
<i>E-I</i> index (lagging)						-0.00271 (0.122)		0.161 (0.132)
<i>E-I</i> index (medium)						-0.230 (0.252)		0.0879 (0.206)
<i>E-I</i> index (advanced)						0.570** (0.252)		0.741*** (0.207)
Betweenness centrality normalized (lagging)							4.537 (7.400)	0.940 (2.894)

Table 5 (continued)

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Betweenness centrality normalized (medium)							3.278**	4.131**
							(1.342)	(1.759)
Betweenness centrality normalized (advanced)							1.861**	2.884***
							(0.904)	(0.612)
Granted patents/R&D (lagging)	-0.0219	-0.0000814	-0.0391	-0.0285	-0.000341	0.0293	-0.0211	0.0775*
	(0.0868)	(0.0583)	(0.0775)	(0.0872)	(0.0578)	(0.0762)	(0.0842)	(0.0422)
Granted patents/R&D (medium)	-0.174	-0.198	-0.0901	-0.204	-0.147	-0.154	-0.182	-0.108
	(0.175)	(0.171)	(0.116)	(0.172)	(0.0894)	(0.168)	(0.178)	(0.0786)
Granted patents/R&D (advanced)	-0.173	-0.198	-0.171	-0.145	-0.189	-0.214	-0.148	-0.210
	(0.246)	(0.236)	(0.252)	(0.204)	(0.196)	(0.235)	(0.226)	(0.174)
University personnel/regional population ($t-1$) (lagging)	0.0684*	0.0692***	0.0641*	0.0621	0.0641***	0.0766*	0.0646	0.0698***
	(0.0400)	(0.0193)	(0.0337)	(0.0422)	(0.0182)	(0.0417)	(0.0404)	(0.0235)
University personnel/regional population ($t-1$) (medium)	0.0484***	0.0521***	0.0405**	0.0477***	0.0354***	0.0533***	0.0501***	0.0409***
	(0.0159)	(0.0147)	(0.0162)	(0.0147)	(0.0119)	(0.0161)	(0.0156)	(0.0114)
University personnel/regional population ($t-1$) (advanced)	0.00613	0.00677	0.00623	0.00394	0.00439	0.00288	0.0000172	-0.00869*
	(0.00947)	(0.00823)	(0.00961)	(0.00905)	(0.00694)	(0.00893)	(0.00868)	(0.00483)
FDI/GDP ($t-1$) (lagging)	-0.0251	-0.0169	-0.0211	0.0823	0.0110	0.102	-0.0511	0.0480
	(0.100)	(0.135)	(0.114)	(0.0938)	(0.127)	(0.0950)	(0.0967)	(0.111)
FDI/GDP ($t-1$) (medium)	1.160	1.410	0.676	1.821	1.619	1.611	1.525	1.787
	(2.464)	(2.129)	(2.696)	(2.156)	(1.961)	(2.391)	(2.390)	(1.757)
FDI/GDP ($t-1$) (advanced)	2.157	2.400*	2.412	1.888*	2.578**	2.287	1.426	2.385**
	(1.404)	(1.384)	(1.568)	(1.085)	(1.112)	(1.480)	(1.099)	(1.114)
Log regional graduates ($t-1$) (lagging)	0.742***	0.735***	0.728***	0.662***	0.681***	0.666***	0.702***	0.592***
	(0.0958)	(0.103)	(0.0938)	(0.0787)	(0.0832)	(0.0942)	(0.0948)	(0.0859)

Table 5 (continued)

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log regional graduates ($t-1$) (medium)	0.519*** (0.175)	0.455** (0.180)	0.640*** (0.150)	0.472*** (0.171)	0.591*** (0.117)	0.529*** (0.181)	0.463*** (0.179)	0.544*** (0.116)
Log regional graduates ($t-1$) (advanced)	0.493 (0.313)	0.480* (0.283)	0.477 (0.318)	0.517*** (0.249)	0.505** (0.211)	0.514* (0.294)	0.479 (0.303)	0.535*** (0.180)
Log GDP (lagging)	-0.0411 (0.156)	-0.0718 (0.158)	-0.0237 (0.143)	0.00747 (0.151)	-0.00948 (0.137)	0.0655 (0.154)	-0.000207 (0.149)	0.173 (0.127)
Log GDP (medium)	0.285* (0.166)	0.298* (0.155)	0.173 (0.158)	0.311* (0.161)	0.165 (0.122)	0.306* (0.170)	0.331** (0.164)	0.294*** (0.116)
Log GDP (advanced)	0.403 (0.298)	0.374 (0.263)	0.421 (0.300)	0.392 (0.247)	0.380* (0.202)	0.460 (0.297)	0.431 (0.288)	0.454** (0.185)
Constant	-2.140*** (1.003)	-2.155** (0.917)	-2.230*** (0.913)	-2.045*** (0.966)	-2.360*** (0.731)	-2.550*** (1.020)	-2.177** (0.987)	-3.154*** (0.666)
Observations	480	480	480	480	480	475	480	475

(1) Dependent variable is nano-publications. (2) Robust standard errors in parentheses

(3) * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. (4) Year dummy variables are omitted from the table

across specifications. As predicted, Chinese regions benefit significantly from collaborating with regions that have higher knowledge stock, resources and experience. This result is stable across groups, as shown in Table 5, and interestingly lagging regions have a higher relative impact from such collaborations as compared to the other groups.

With regard to collaboration intensity with medium regions, Table 4 provides a positive strong effect on publications; however, in the tripartite analysis, a positive significant effect is found only for medium regions, namely within the same regional group. The coefficients are not significant for other regional groups (i.e. lagging and advanced groups). This could reflect the fact that many medium regions tend to cooperate among themselves, given the similar background and level of knowledge capabilities. In fact, sharing similar level of knowledge, expertise and understanding between the parties is a crucial element to build a fruitful collaboration (Li et al. 2013). This within-group collaboration effect could also explain the significant coefficient in Table 5, Col. 4, presenting the collaboration impact of lagging regions with other lagging regions ($b_{(lagging)} = 26.38$). However, the coefficient of collaboration with lagging regions becomes not significant in the whole model (Table 4, Col. 8), suggesting that the negative effect received by more developed regions completely offsets the positive effect received by less developed regions. Advanced regions present a significantly negative coefficient from the collaboration with lagging regions ($b_{(lagging)} = -92.16$). Since more advanced regions have higher knowledge stock and resources, they do not benefit from collaborations with less developed regions, although they still bear the costs from such collaborations, and that is where the negative impact might derive from. From another perspective, the negative effect might be associated with collaboration patterns (e.g. who initiated the research, who leads the research, what is required by the funding sponsor, etc). Due to the lack of matching data between authors and their affiliations in our earlier data, this study does not examine roles of collaboration. This, however, is of interest to be tested in future research.

Moving on to the internationalisation index (or $E-I$ index), Table 4 shows a positive and statistically significant coefficient in the full model specification. However, from the tripartite analysis (Table 5) it appears that $E-I$ index has a positive impact only for advanced regions. Medium and lagging regions do not show any statistically significant coefficient for the internationalisation index, suggesting that there might be significant costs associated with the opening to international collaborations, such as financial, administrative and cultural costs (Wagner 2006). In line with previous arguments from the literature, the regression results suggest that such costs might offset the benefits of less developed regions. Scientifically advanced areas are characterised by higher human capital, financial resources and technological capabilities, which ensure a high absorptive capacity (Fu 2008), thus allowing them to benefit more from a broad network structure. In contrast, lagging regions in China do not seem to have reached that stage yet. Considering the discrepancy of regional development in China, this finding calls attention for Chinese government in implementing collaboration policies. Not all Chinese regions, at least at the current stage, can leverage the knowledge flows generated via collaboration.

Finally, the index of collaboration breadth (proxied by betweenness centrality) shows a strong positive effect in the pooled estimation in Table 4, both in model (7) only including betweenness, and in the full model (8). This finding is in line with the literature, which suggests that actors in a more central position have access to diverse and non-redundant information, which is ultimately reflected by a higher number of nanotechnology-related publications. However, Table 5 highlights that the positive effect from betweenness is not homogeneously distributed across regions; the positive coefficient found in the pooled estimation is hiding a strongly significant impact for advanced and medium regions only. The

coefficient for lagging regions is not statistically significant, indicating that those regions do not benefit from filling the structural holes in the network. Perhaps, the low level of absorptive capacity in lagging regions causes the costs from maintaining central positions in the network to exceed the benefits, thus having a reversal effect on their performance.

Together with the regression results regarding collaboration depth variables explained above, this calls attention from Chinese government to take regional capability into consideration while implementing collaboration policies. High capability regions can benefit from broad collaboration networks and leveraging knowledge from diversified partners, while low capability regions seem to benefit more from concentrated collaborations with advanced domestic partners.

The impact of control variables

In general, the ratio of total granted patents to R&D expenditures has a negative, statistically non-significant coefficient in almost all the regression models. This holds for both the pooled analysis and the estimations by regional groups. A possible explanation for the lack in significance is that the patents data are industry-related, thus a change in the parameter does not directly affect academic publications.

With regard to university personnel, the coefficient is not statistically significant in the pooled estimation; however, from Table 5 we can see that such insignificance is due to the negative sign for the advanced regions' group, which offsets the effect of lagging and medium regions. This may suggest that leading regions already have a high level of university personnel, and more personnel input cannot be used effectively, which indicates diminishing returns in the knowledge production process in such regions. However, in the lagging and medium regions, there might be still increasing returns to scale, hence one unit of personnel input will produce more than one unit of scientific output. Therefore, an increase in the personnel inputs is more valuable in lagging and medium regions than in advanced regions.

The openness variable (FDI/GDP) is also not significant in the pooled analysis; however, in Table 5 we can see that while the coefficient is not significant for lagging and medium regions, it is positively significant for advanced regions. This result corresponds with the effect of aforementioned international collaboration depth. Considering that advanced regions have higher scientific capability and can transform the interaction with foreign countries into positive gain in knowledge production. This finding is in line with that of Wang et al. (2017a), i.e. whether or not one region benefits from FDI depends on this region's local capacity. Finally, the variable of newly added human capital (regional graduates) and general economic/development level (region GDP) are both significant in the pooled analysis, and such significance is further explored in Table 5. The variable of regional graduates is positively significant in lagging and medium regions. This, similar to the personnel variable explained earlier, shows that human capital input is more needed in lagging and medium regions than in advanced regions.

Robustness checks

The robustness of the specified model was discussed, and the model was proved to be sound and to provide consistent estimates after running separate regressions testing correlated variables one at the time (see Table 7 in the "Appendix"). Some additional analyses have been conducted to confirm the reliability of the estimations obtained from the negative binomial model. Firstly, the model was re-estimated using Poisson fixed-effect specification, with robust

Table 6 Estimations for robustness checks—cross models comparison *Source:* Authors' own calculations

Variables	(1) Negative binomial	(2) Poisson FE	(3) OLS
Coll. Intensity with leading regions	70.94*** (9.062)	41.88*** (11.51)	64.85*** (10.19)
Coll. Intensity with medium regions	28.04*** (10.87)	8.097 (11.18)	18.76* (10.46)
Coll. Intensity with lagging regions	-8.272 (10.82)	-23.11** (10.36)	-2.450 (12.98)
<i>E-I</i> index	0.318** (0.128)	0.651*** (0.169)	0.218 (0.146)
Betweenness centrality normalized	4.169*** (0.713)	3.321*** (0.451)	3.871*** (0.718)
Granted patents/R&D	-0.154 (0.110)	-0.0290 (0.0772)	-0.101 (0.0679)
University personnel/regional population ($t-1$)	0.00839 (0.00673)	-0.00111 (0.00376)	0.00925 (0.00711)
FDI/GDP ($t-1$)	0.105 (0.173)	-0.291 (0.544)	0.483*** (0.155)
Log regional graduates ($t-1$)	0.658*** (0.121)	0.411*** (0.138)	0.653*** (0.0985)
Log GDP	0.335** (0.138)	0.210 (0.187)	0.361*** (0.126)
Constant	-3.619*** (0.675)		-3.973*** (0.702)
Observations	475	475	475
Wald chi2	6673.5	3,340,523.8	19,967.8
Prob > chi2	0	0	0

(1) Dependent variable is nano-publications in columns 1 and 2. Log of nano-publications in column 3. (2) Robust standard errors in parentheses. (3) * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. (4) Year dummy variables are omitted from the table

standard errors. As explained above, although our data does not follow a Poisson distribution, and it is characterised by over-dispersion, the Poisson FE model has been indicated by the literature to have some robustness properties (Wooldridge 2002), and it should provide some comparable estimates. Secondly, the regression was re-estimated using ordinary least square (OLS) methodology, after taking the natural logarithm of the dependent variable: the log of nanotechnology publications. Although the model uses a different dependent variable, making it hard to compare the estimation coefficients, it is used in the literature to test the robustness and consistency of the NB results (see for example Guan et al. 2015b). Table 6 reports the results from the two control regressions, as well as the results from the original NB model.

Collaboration intensity with advanced regions is highly significant and positive in all the three models. Collaboration intensity with medium regions, instead, is only significant in the NB and OLS models, it is not significant in the Poisson model. Probably the over-dispersion problem caused estimations to be inaccurate; however, the coefficient of C_{it}^{medium} has a positive sign in all the three models, suggesting a similar direction of the impact. Collaboration

intensity with lagging regions maintains a negative direction in all the three estimation models, although it does not show a significant coefficient for the NB and OLS models. Similarly, the internationalization index maintains a positive estimation coefficient in all the three models, although it is only significant for models (1) and (2). Finally, betweenness centrality is positive and significant in all the three models. For what concerns the control variables, Table 6 also shows consistency across different estimation models.

Although the coefficients estimated are not identical for the three models, we can conclude that overall the NB model used is robust. In fact, despite the enormous issues and limitations coming from the application of the OLS and Poisson FE models for the dataset in analysis, the robustness checks proved that the estimates from the NB model are consistent, and we found a good matching rate with the control models.

Discussion and conclusions

Collaboration patterns are believed to be connected with a country's scientific development stages (Moed 2016; Moed and Halevi 2014; UNESCO 2014). In this context, it is crucial to understand what types of collaboration are more beneficial for what types of country/region. Employing the scientific articles published by Chinese researchers in the nanotechnology field, this paper examined the impact of collaboration depth (with domestic and foreign partners) and collaboration breadth on the regional scientific output.

Results from our analysis suggest that benefits from collaborations depend on the development stages of collaborative partners. The strong positive effect of collaborations with more advanced domestic regions has been confirmed for all the subgroups; lagging regions have been found to benefit the most from collaborations with scientifically advanced regions. However, the contribution of collaborations with medium regions and lagging regions is very limited, existing only for the within-group collaborations. These results prove that the primary benefit from collaborations is the sharing of skills and knowledge between the parties (Katz and Martin 1997). If there is not much knowledge for advanced regions to learn from the lagging regions, such collaborations are not beneficial to the former.

Regarding the strength of international collaborations, our study finds a significant positive effect in advanced regions; this is in line with the literature supporting international collaborations as a mean to achieve diverse knowledge and information, which can foster the national innovation capacity (see for example Glänzel 2001; Guan et al. 2015b; Leydesdorff et al. 2014). However, such positive and significant effect does not hold for less developed regions. This seems to be because less developed Chinese regions and foreign countries are characterised by very different levels of knowledge and capability. Therefore, the costs of collaborating internationally exceed the benefits for less developed regions, and they hamper the effectiveness of such collaborations, resulting in little/no knowledge creation (Katz and Martin 1997; Wagner 2006).

This confirms that, in order to benefit from scientific collaboration, it is important that the two actors involved share a similar level of knowledge and understanding (Gonzalez-Brambila et al. 2013). Through collaboration, partners can fill the knowledge gap between them, and create new knowledge. However, if the gap between collaborators is too wide in terms of capability, skills and competences (as in the case of collaborations between lagging regions and foreign countries), such cooperation does not lead to effective knowledge production (McFadyen et al. 2009).

Being centrally located in a verified collaboration network is found to bring significant and positive effect to regional scientific output. Again, however, such effect is conditional on a region's scientific capability. This study suggests that only medium and advanced regions can benefit from being embedded in a broad collaboration network. Such added value is not observable for lagging regions, suggesting that their low absorptive capacity level prevents them from exploiting the knowledge inflow deriving from a diversified network structure.

The above findings have important theoretical implications. First, our results complement the scientific development model discussed by Moed (2016), Duque et al. (2005), Moed and Halevi (2014), and UNESCO (2014). In accordance with their findings that collaboration patterns reflect development stages, our study emphasises the differences of collaboration effect received by different regions at different development stages. This result also provides empirical support to the issue raised by Lee and Bozeman (2005), who argue that not all collaborations are ideal. Although several studies have found a positive impact deriving from an increase in collaborations, very few differentiated the effect of such collaborations, by considering the level of scientific capacity of the actors involved in the collaboration process (Duque et al. 2005). This study suggests that generalising impact evaluations could result in limited or inaccurate interpretations, especially for countries with high internal development unbalances. In the case of Chinese regional collaborations, different partnerships (between different regional groups) lead to different effects. Second, findings from this analysis provided evidence that international collaborations might not be beneficial a priori. Costs and barriers increase when a region or country moves to the international context, hence not all the actors might benefit from such collaborations, and trade-offs based on the capabilities of the actors involved should be considered (Jeong et al. 2014).

Our results also present important policy implications. In recent decades, the Chinese government incurred in significant investments in nanotechnology to incentivise research and development, as well as collaborations within and outside the national borders. The interest for engaging in international partnerships, indicated by the large investments in collaboration programmes such as the Framework Programme 7, the Horizon 2020 and the World-Class 2.0 (European Commission 2007, 2018; Zhao 2018) aimed to improve the Chinese regions' knowledge capability, and promote future development. While providing some support to these policies' objectives, as shown by the positive significant correlation between collaborations variables and nano-publications, our results also call attention to the regional disparities that have been on the rise in the past decades. When such differences are taken into analysis, the policy scenario changes substantially.

In scientifically weaker regions, in order to increase knowledge production and catch up with more advanced regions, local authorities should be aware that domestic collaborations, in particular with more advanced regions, are more efficient than international collaborations, at least at current stage.

Considering the positive impact of collaboration breadth index in advanced and medium regions, our study verifies that regions with relatively high research capacity can manipulate (and benefit from) knowledge generated by more diversified partners. However, low-capability regions seem to benefit more from concentrated collaborations with less broad partners. If one wants to promote more collaborations between different organizations in China, as suggested by Ozcan and Islam (2014), our study shows that matching development stages or capability levels should be taken in consideration in order to really benefit from such collaborations.

Finally, we must acknowledge the limitations to the above approach in measuring knowledge creation. First, co-authored papers are only partial indicators of collaborations, as remarked by Katz and Martin (1997). There are many different types of collaborations

that take place without ultimately resulting in publications. Likewise, there are cases of co-authored papers resulted from indirect and marginal forms of collaborations (Katz and Martin 1997), thus it is important to be cautious in interpreting the results. Nevertheless, literature conveys that collaborations on publications are one of the best documented and most quantifiable measurements of scientific collaborations (Abramo et al. 2011; Glänzel and Schubert 2005). Second, the breadth of collaborations is only proxied by betweenness centrality to be included in the analysis. Previous research found that including different centrality measures in the same model might cause high multicollinearity (see Guan et al. 2015b), hence the decision to include only one measure. However, not considering all the relational measures in a network might lead to inaccurate estimations, as more interactions happen between actors, which are not accounted for (Gonzalez-Brambila et al. 2013). Therefore, attention must be paid in the interpretation of results. Additional research exploring the impact of other network structures' measures on knowledge creation might overcome the limitations in interpretations from this study, due to possible omitted variable bias. Third, this study focuses on collaborations at regional and national levels, without distinguishing subtopics of nanoscience. Considering the fact that nanoscience has been widely applied into various fields (Wang et al. 2013a), further research exploring collaborations in different scientific nano domains is encouraged to bring support to our results, and to provide further evidence of the correlation between collaborations and knowledge creation.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix

Query used to search publications on nanotechnology

KEY(monolayer* OR (mono-layer*) OR film* OR quantum* OR multilayer* OR (multi-layer*) OR array* OR molecu* OR polymer* OR (co-polymer*) OR copolymer* OR mater* OR biolog* OR supramolecul*) AND KEY(nano*) AND KEY(((quantum dot* OR quantum well* OR quantum wire*) AND NOT nano*) AND KEY((((SELF ASSEMBL*) OR (SELF ORGANIZ*) OR (DIRECTED ASSEMBL*)) AND MolEnv-I) AND NOT nano*) AND KEY(((molecu* motor*) OR (molecu* ruler*) OR (molecu* wir*) OR (molecu* devic*) OR (molecular engineering) OR (molecular electronic*) OR (single molecu*) OR (fullerene*) OR (coulomb blockad*) OR (bionano*) OR (langmuir-blodgett) OR (Coulombstaircase*) OR (PDMS stamp*)) AND NOT nano*) AND KEY(((TEM OR STM OR EDX OR AFM OR HRTEM OR SEM OR EELS) OR (atom* force microscop*) OR (tunnel* microscop*) OR (scanning probe microscop*) OR (transmission electron microscop*) OR (scanning electron microscop*) OR (energy dispersive X-ray) OR (Xray photoelectron*) OR (electron energy loss spectroscop*)) AND MolEnv-I) AND NOT nano*) AND KEY(((pebbles OR NEMS OR Quasicrystal* OR (quasi-crystal*)) AND MolEnv-I) AND NOT nano*) AND KEY(((biosensOR* OR (sol gel* OR solgel*) OR dendrimer* OR soft lithograph* OR molecular simul* OR quantum effect* OR molecular sieve* OR mesoporous material*) AND (MolEnv-R)) AND NOT nano*) AND (MolEnv-R)) AND NOT nano*)

See Table 7.

Table 7 Results of negative binomial regressions—robustness checks for highly correlated variables *Source*: Authors' own calculations

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Coll. Intensity with leading regions	62.078*** (10.388)	68.615*** (10.859)	43.417*** (13.757)	63.269*** (10.482)		62.013*** (10.376)	62.558*** (12.485)	50.866*** (13.541)	74.078*** (10.855)		
Coll. Intensity with medium regions	21.126* (10.873)	21.233** (9.947)	10.829 (10.866)		21.062** (9.236)	21.072* (10.782)					
Coll. Intensity with lagging regions	-7.754 (11.998)	-8.657 (12.079)	-21.173*** (8.137)								
<i>E-I</i> index							0.330** (0.131)	0.457*** (0.113)	0.248** (0.117)	2.523*** (0.896)	2.308** (1.038)
Betweenness centrality normalized											
Granted patents/R&D	-0.209 (0.133)	-0.175 (0.129)	-0.244** (0.110)	-0.221* (0.128)	-0.164 (0.158)	-0.208 (0.133)	-0.181 (0.112)	-0.239** (0.103)	-0.158 (0.123)	-0.152 (0.135)	-0.156 (0.138)
University personnel/regional population (<i>t</i> -1)	0.015*** (0.006)	0.016*** (0.006)	0.029*** (0.009)	0.016*** (0.006)	0.012 (0.009)	0.015*** (0.006)		0.027*** (0.009)	0.016*** (0.006)		0.004 (0.009)
FDI/GDP (<i>t</i> -1)	0.236 (0.186)	0.245 (0.219)	1.271*** (0.223)	0.225 (0.148)	0.315* (0.186)	0.247 (0.181)	0.220 (0.221)	0.878*** (0.234)	0.160 (0.173)	0.185 (0.119)	0.182 (0.119)
Log regional graduates (<i>t</i> -1)	0.739*** (0.113)	0.807*** (0.084)		0.723*** (0.114)	0.679*** (0.130)	0.737*** (0.114)	0.785*** (0.116)		0.762*** (0.086)	0.621*** (0.134)	0.603*** (0.137)
Log GDP	0.291** (0.143)		0.544*** (0.152)	0.327** (0.136)	0.391** (0.172)	0.292** (0.143)	0.334** (0.138)	0.576*** (0.137)		0.490*** (0.169)	0.480*** (0.169)
Constant	-3.811*** (0.701)	-2.087*** (0.593)	-0.599 (1.197)	-3.889*** (0.701)	-3.840*** (0.885)	-3.815*** (0.698)	-4.104*** (0.727)	-0.651 (1.106)	-1.639** (0.644)	-4.008*** (0.952)	-3.852*** (0.927)
Observations	480	480	480	480	480	480	475	475	475	480	480
Wald chi2	8365.59	8146.31	7356.48	7139.01	20,337.98	6197.52	10,105.51	12,229.91	9873.59	20,349.28	18,469.14
Prob > chi2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

(1) Dependent variable is nano-publications. (2) Robust standard errors in parentheses

(3) **p* < 0.1, ***p* < 0.05, ****p* < 0.01. (4) Year dummy variables are omitted from the table

References

- Abbasi, A., Altmann, J., & Hossain, L. (2011). Identifying the effects of co-authorship networks on the performance of scholars: A correlation and regression analysis of performance measures and social network analysis measures. *Journal of Informetrics*, 5(4), 594–607. <https://doi.org/10.1016/j.joi.2011.05.007>.
- Abbasi, A., Hossain, L., & Leydesdorff, L. (2012). Betweenness centrality as a driver of preferential attachment in the evolution of research collaboration networks. *Journal of Informetrics*, 6, 403–412. <https://doi.org/10.1016/j.joi.2012.01.002>.
- Abramo, G., D'Angelo, C. A., & Solazzi, M. (2011). The relationship between scientists' research performance and the degree of internationalization of their research. *Scientometrics*, 86(3), 629–643. <https://doi.org/10.1007/s11192-010-0284-7>.
- Acosta, M., Coronado, D., Ferrándiz, E., & León, M. D. (2011). Factors affecting inter-regional academic scientific collaboration within Europe: The role of economic distance. *Scientometrics*, 87, 63–74. <https://doi.org/10.1007/s11192-010-0305-6>.
- Adams, J. D., Black, G. C., Clemmons, J. R., & Stephan, P. E. (2005). Scientific teams and institutional collaborations: Evidence from U.S. universities, 1981–1999. *Research Policy*, 34(3), 259–285. <https://doi.org/10.1016/j.respol.2005.01.014>.
- Allison, P. D., & Waterman, R. (2002). Fixed-effects negative binomial regression models. *Sociology Methodology*, 32, 247–265.
- Andreosso-O'Callaghan, B. (1999). Technology transfer: A mode of collaboration between the European Union and China. *Europe-Asia Studies*, 51(1), 123–142. <https://doi.org/10.1080/09668139999155>.
- Archibugi, D., & Pianta, M. (1996). Measuring technological change through patents and innovation surveys. *Technovation*, 16(9), 451–468. [https://doi.org/10.1016/0166-4972\(96\)00031-4](https://doi.org/10.1016/0166-4972(96)00031-4).
- Bai, C. (2005). Ascent of nanoscience in China. *Science*, 309(5731), 61–63.
- Boschma, R., Heimeriks, G., & Balland, P.-A. (2014). Scientific knowledge dynamics and relatedness in biotech cities. *Research Policy*, 43(1), 107–114. <https://doi.org/10.1016/j.respol.2013.07.009>.
- Bound, K., Saunders, T., Wilsdon, J., & Adams, J. (2013). China's Absorptive State: Research, innovation and the prospects for China-UK collaboration. Nesta Report.
- Brandes, U. (2001). A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology*, 25(2), 163–177. <https://doi.org/10.1080/0022250X.2001.9990249>.
- Burt, R. S. (1992). *Structural holes: The social structure of competition*. Cambridge, MA: Harvard University Press.
- Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: Methods and applications*. New York: Cambridge University Press. [https://doi.org/10.1016/S0304-4076\(00\)00050-6](https://doi.org/10.1016/S0304-4076(00)00050-6).
- Coccia, M., & Wang, L. (2016). Evolution and convergence of the patterns of international scientific collaboration. *Proceedings of the National Academy of Sciences*, 113(8), 2057–2061. <https://doi.org/10.1073/pnas.1510820113>.
- Cunningham, S. W., & Werker, C. (2012). Proximity and collaboration in European nanotechnology. *Papers in Regional Science*, 91(4), 723–742. <https://doi.org/10.1111/j.1435-5957.2012.00416>.
- de Solla Price, D. J., & Beaver, D. (1966). Collaboration in an invisible college. *American Psychologist*, 21(11), 1011–1018. <https://doi.org/10.1037/h0024051>.
- Duque, R. B., Ynalvez, M., Sooryamoorthy, R., Mbatia, P., Dzorgbo, D. B. S., & Shrum, W. (2005). Collaboration paradox: Scientific productivity, the Internet, and problems of research in developing areas. *Social Studies of Science*, 35(5), 755–785. <https://doi.org/10.1177/0306312705053048>.
- Ebersberger, B., Herstad, S. J., & Koller, C. (2014). Does the composition of regional knowledge bases influence extra-regional collaboration for innovation? *Applied Economics Letters*, 21(3), 201–204. <https://doi.org/10.1080/13504851.2013.848019>.
- European Commission. (2007). *The seventh framework programme (FP7)*. Brussels: European Commission.
- European Commission. (2018). Horizon 2020—What's in it for China? <https://ec.europa.eu/programmes/horizon2020/en/horizon-2020-whats-it-china>. Accessed Sept 18, 2018.
- Fan, S., Kanbur, R., & Zhang, X. (2011). China's regional disparities: Experience and policy. *Review of Development Finance*. <https://doi.org/10.1016/j.rdf.2010.10.001>.
- Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, 40(1), 35–41.
- Fu, X. (2008). Foreign direct investment, absorptive capacity and regional innovation capabilities: Evidence from China. *Oxford Development Studies*, 36(1), 89–110. <https://doi.org/10.1080/13600810701848193>.
- Gardner, W., Mulvey, E. P., & Shaw, E. C. (1995). Regression analyses of counts and rates: Poisson, overdispersed poisson, and negative binomial models. *Psychological Bulletin*, 118(3), 392–404. <https://doi.org/10.1037/0033-2909.118.3.392>.

- Glänzel, W. (2001). National characteristics in international scientific co-authorship relations. *Scientometrics*, 51(1), 69–115. <https://doi.org/10.1023/A:1010512628145>.
- Glänzel, W., & Schubert, A. (2005). Analyzing scientific networks through co-authorship. In H. F. Moed, W. Glänzel & U. Schmoch (Eds.), *Handbook of quantitative science and technology research* (pp. 257–276). Dordrecht: Springer. <https://doi.org/10.1007/1-4020-2755-9>.
- Gonzalez-Brambila, C. N., Veloso, F. M., & Krackhardt, D. (2013). The impact of network embeddedness on research output. *Research Policy*, 42(9), 1555–1567. <https://doi.org/10.1016/j.respol.2013.07.008>.
- Guan, J., & Liu, N. (2016). Exploitative and exploratory innovations in knowledge network and collaboration network: A patent analysis in the technological field of nano-energy. *Research Policy*, 45(1), 97–112. <https://doi.org/10.1016/j.respol.2015.08.002>.
- Guan, J., Yan, Y., & Zhang, J. (2015a). How do collaborative features affect scientific output? Evidences from wind power field. *Scientometrics*, 102, 333–355. <https://doi.org/10.1007/s11192-014-1311-x>.
- Guan, J., Zhang, J., & Yan, Y. (2015b). The impact of multilevel networks on innovation. *Research Policy*, 44(3), 545–559. <https://doi.org/10.1016/j.respol.2014.12.007>.
- Hao, R., & Wei, Z. (2010). Fundamental causes of inland-coastal income inequality in post-reform China. *Annals of Regional Science*, 45(1), 181–206. <https://doi.org/10.1007/s00168-008-0281-4>.
- Hausman, J. A., Hall, B., & Griliches, Z. (1984). Econometric models for count data with an application to the patents-R&D relationship. *Econometrica*, 52, 909–938. <https://doi.org/10.2307/1911191>.
- He, Z. L., Geng, X. S., & Campbell-Hunt, C. (2009). Research collaboration and research output: A longitudinal study of 65 biomedical scientists in a New Zealand university. *Research Policy*, 38(2), 306–317. <https://doi.org/10.1016/j.respol.2008.11.011>.
- Hoskisson, R. E., Hitt, M. A., Johnson, R. A., & Grossman, W. (2002). Conflicting voices: The effects of institutional ownership heterogeneity and internal governance on corporate innovation strategies. *Academy of Management Journal*, 45, 697–716. <https://doi.org/10.2307/3069305>.
- 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, 967–983. <https://doi.org/10.1007/s11192-011-0474-y>.
- 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. <https://doi.org/10.1093/scipol/sct079>.
- Katz, J. S., & Martin, B. R. (1997). What is research collaboration? *Research Policy*, 26, 1–18. [https://doi.org/10.1016/S0048-7333\(96\)00917-1](https://doi.org/10.1016/S0048-7333(96)00917-1).
- Krackhardt, D., & Stern, R. N. (1988). Informal networks and organizational crises: An experimental simulation. *Social Psychology Quarterly*, 51(2), 123–140. <https://doi.org/10.2307/2786835>.
- Kuo, C. C., & Yang, C. H. (2008). Knowledge capital and spillover on regional economic growth: Evidence from China. *China Economic Review*, 19(4), 594–604. <https://doi.org/10.1016/j.chieco.2008.06.004>.
- Lee, S., & Bozeman, B. (2005). The impact of research collaboration on scientific productivity. *Social Studies of Science*, 35(5), 673–702. <https://doi.org/10.1177/0306312705052359>.
- Leydesdorff, L. (2008). On the normalization and visualization of author co-citation data: Salton's cosine versus the Jaccard index. *Journal of the American Society for Information Science and Technology*, 59(1), 77–85. <https://doi.org/10.1002/asi.20732>.
- Leydesdorff, L., Park, H. W., & Wagner, C. (2014). International coauthorship relations in the Social Sciences Citation Index: Is internationalization leading the Network? *Journal of the Association for Information Science and Technology*, 65(10), 2111–2126. <https://doi.org/10.1002/asi.23102>.
- Li, E. Y., Liao, C. H., & Yen, H. R. (2013). Co-authorship networks and research impact: A social capital perspective. *Research Policy*, 42(9), 1515–1530. <https://doi.org/10.1016/j.respol.2013.06.012>.
- Luukkonen, T., Persson, O., & Sivertsen, G. (1992). Understanding patterns of international scientific collaboration. *Science, Technology and Human Values*, 17(1), 101–126.
- Luukkonen, T., Tijssen, R. J. W., Persson, O., & Sivertsen, G. (1993). The measurement of international scientific collaboration. *Scientometrics*, 28(1), 15–36. <https://doi.org/10.1007/BF02016282>.
- McFadyen, M. A., & Cannella, A. A. J. (2004). Social capital and knowledge creation: Diminishing returns of the number and strength of exchange. *The Academy of Management Journal*, 47(5), 735–746.
- McFadyen, M. A., Semadeni, M., & Cannella, A. A., Jr. (2009). Value of strong ties to disconnected others: Examining knowledge creation in biomedicine. *Organization Science*, 20(3), 552–564. <https://doi.org/10.1287/orsc.1080.0388>.
- Meyer, M., & Persson, O. (1998). Nanotechnology—Interdisciplinarity, patterns of collaboration and differences in application. *Scientometrics*, 42(2), 195–205. <https://doi.org/10.1007/BF02458355>.

- Moed, H. F. (2016). Iran's scientific dominance and the emergence of South-East Asian countries as scientific collaborators in the Persian Gulf Region. *Scientometrics*, 108, 305–314. <https://doi.org/10.1007/s11192-016-1946-x>.
- Moed, H. F., & Halevi, G. (2014). Tracking scientific development and collaborations: The case of 25 Asian countries. *Research Trends* (38). <https://www.researchtrends.com/issue-38-september-2014/tracking-scientific-development-and-collaborations-the-case-of-25-asian-countries/>.
- Nahapiet, J., & Ghoshal, S. (1998). Social capital, intellectual capital, and the organizational advantage. *Academy of Management Review*, 23(2), 242–266. <https://doi.org/10.5465/AMR.1998.533225>.
- Newman, M. E. J. (2005). A measure of betweenness centrality based on random walks. *Social Networks*, 27(1), 39–54. <https://doi.org/10.1016/j.socnet.2004.11.009>.
- Niu, F., & Qiu, J. (2014). Network structure, distribution and the growth of Chinese international research collaboration. *Scientometrics*, 98(2), 1221–1233. <https://doi.org/10.1007/s11192-013-1170-x>.
- Ozcan, S., & Islam, N. (2014). Collaborative networks and technology clusters - the case of nanowire. *Technological Forecasting and Social Change*, 82(1), 115–131. <https://doi.org/10.1016/j.techfore.2013.08.008>.
- Phelps, C., Heidl, R., & Wadhwa, A. (2012). Knowledge, networks, and knowledge networks: A review and research agenda. *Journal of Management*, 38(4), 1115–1166. <https://doi.org/10.1177/0149206311432640>.
- Porter, A. L., Youtie, J., Shapira, P., & Schoeneck, D. J. (2008). Refining search terms for nanotechnology. *Journal of Nanoparticle Research*, 10(5), 715–728. <https://doi.org/10.1007/s11051-007-9266-y>.
- Schubert, T., & Sooryamoorthy, R. (2010). Can the centre-periphery model explain patterns of international scientific collaboration among threshold and industrialised countries? The case of South Africa and Germany. *Scientometrics*, 83, 181–203. <https://doi.org/10.1007/s11192-009-0074-2>.
- Schummer, J. (2004). Multidisciplinarity, interdisciplinarity, and patterns of research collaboration in nanoscience and nanotechnology. *Scientometrics*, 59(3), 425–465. <https://doi.org/10.1023/B:SCIE.0000018542.71314.38>.
- Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science*, 24(4), 265–269. <https://doi.org/10.1002/asi.4630240406>.
- Sozer, N., & Kokini, J. L. (2009). Nanotechnology and its applications in the food sector. *Trends in Biotechnology*, 27(2), 82–89. <https://doi.org/10.1016/J.TIBTECH.2008.10.010>.
- Tang, L., & Shapira, P. (2011a). China–US scientific collaboration in nanotechnology: Patterns and dynamics. *Scientometrics*. <https://doi.org/10.1007/s11192-011-0376-z>.
- Tang, L., & Shapira, P. (2011b). Regional development and interregional collaboration in the growth of nanotechnology research in China. *Scientometrics*, 86(2), 299–315. <https://doi.org/10.1007/s11192-010-0274-9>.
- Tang, L., & Shapira, P. (2012). Effects of international collaboration and knowledge moderation on China's nanotechnology research impacts. *Journal of Technology Management in China*, 7(1), 94–110. <https://doi.org/10.1108/17468771211207376>.
- Tortoriello, M., & Krackhardt, D. (2010). Activating cross-boundary knowledge: The role of Simmelian ties in the generation of innovations. *Academy of Management Journal*, 53(1), 167–181.
- UNESCO. (2014). *Higher education in Asia: Expanding out, expanding up*. ISBN 978-92-9189-147-4 licensed under CC-BY-SA 3.0 IGO. Montreal.
- Valente, T. W., Coronges, K., Lakon, C., & Costenbader, E. (2008). How correlated are network centrality measures? *Connections*, 28(1), 16–26. <https://doi.org/10.1016/j.bbi.2008.05.010>.
- Wagner, C. (2006). International collaboration in science and technology: Promises and pitfalls. In B. Louk & R. Engelhard (Eds.), *Science and technology policy for development, dialogues at the interface* (pp. 165–176). London: Anthem Press.
- Wagner, C. S., & Leydesdorff, L. (2005). Mapping the network of global science: Comparing international co-authorships from 1990 to 2000. *International Journal of Technology and Globalisation*, 1, 185–208. <https://doi.org/10.1504/IJTG.2005.007050>.
- Wagner, C. S., Whetsell, T., Baas, J., & Jonkers, K. (2018). Openness and impact of leading scientific countries. *Frontiers in Research Metrics and Analytics*, 3(10), 1–10. <https://doi.org/10.3389/frma.2018.00010>.
- Wang, L., Jacob, J., & Li, Z. (2019). Exploring the spatial dimensions of nanotechnology development in China: The effects of funding and spillovers. *Regional Studies*, 53(2), 245–260. <https://doi.org/10.1080/00343404.2018.1457216>.
- Wang, L., Meijers, H., & Szirmai, A. (2017a). Technological spillovers and industrial growth in Chinese regions. *Industrial and Corporate Change*, 26(2), 233–257. <https://doi.org/10.1093/icc/dtw022>.

- Wang, L., Notten, A., & Surpatean, A. (2013a). Interdisciplinarity of nano research fields: A keyword mining approach. *Scientometrics*, *94*(3), 877–892. <https://doi.org/10.1007/s11192-012-0856-9>.
- Wang, L., & Szirmai, A. (2013). The unexpected convergence of regional productivity in Chinese industry, 1978–2005. *Oxford Development Studies*, *41*(1), 29–53. <https://doi.org/10.1080/13600818.2012.756464>.
- Wang, L., Wang, X., & Philipsen, N. J. (2017b). Network structure of scientific collaborations between China and the EU member states. *Scientometrics*, *113*(2), 765–781. <https://doi.org/10.1007/s11192-017-2488-6>.
- Wang, X., Xu, S., Wang, Z., Peng, L., & Wang, C. (2013b). International scientific collaboration of China: Collaborating countries, institutions and individuals. *Scientometrics*, *95*(3), 885–894. <https://doi.org/10.1007/s11192-012-0877-4>.
- Wang, X., Xu, S., Liu, D., & Liang, Y. (2012). The role of Chinese–American scientists in China–US scientific collaboration: A study in nanotechnology. *Scientometrics*, *91*(3), 737–749. <https://doi.org/10.1007/s11192-012-0693-x>.
- Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data*. Cambridge: The MIT Press. <https://doi.org/10.1515/humr.2003.021>.
- Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, *316*(5827), 1036–1039. <https://doi.org/10.1126/science.1136099>.
- Youtie, J., Iacopetta, M., & Graham, S. (2008). Assessing the nature of nanotechnology: Can we uncover an emerging general purpose technology? *Journal of Technology Transfer*, *33*, 315–329. <https://doi.org/10.1007/s10961-007-9030-6>.
- Zhang, L. (2017). The knowledge spillover effects of FDI on the productivity and efficiency of research activities in China. *China Economic Review*, *42*, 1–14. <https://doi.org/10.1016/j.chieco.2016.11.001>.
- Zhao, L. (2018). World Class 2.0: China’s major initiative in Higher Education. *East Asian Policy*, *10*(02), 79–90. <https://doi.org/10.1142/S1793930518000193>.
- Zheng, J., Zhao, Z. Y., Zhang, X., Chen, D. Z., & Huang, M. H. (2014). International collaboration development in nanotechnology: A perspective of patent network analysis. *Scientometrics*, *98*(1), 683–702. <https://doi.org/10.1007/s11192-013-1081-x>.
- Zhou, P., & Glänzel, W. (2010). In-depth analysis on China’s international cooperation in science. *Scientometrics*, *82*(3), 597–612. <https://doi.org/10.1007/s11192-010-0174-z>.
- Zhou, P., & Leydesdorff, L. (2006). The emergence of China as a leading nation in science. *Research Policy*, *35*, 83–104. <https://doi.org/10.1016/j.respol.2005.08.006>.