



ARTICLE

<https://doi.org/10.1057/s41599-025-04996-5>

OPEN

 Check for updates

# The impact of collaboration networks constructed through common project experience on research output

Kaishun Li<sup>1</sup> & Kaiming Li<sup>2</sup>✉

Science funding is vital for driving national innovation and technology development. China's science funding supports a growing number of researchers, and they form collaboration networks through common project experience. However, existing studies lack focus on the impact of collaboration networks based on common project experience on research output due to data limitations. To address this gap, we analyzed a sample of 43,271 science funding projects approved by the National Natural Science Foundation of China from 2009 to 2014. Collaboration networks were constructed based on principal investigators' common project experience, and the centrality indicator was used to assess their ability to utilize these networks. Our findings revealed that centrality had a significant positive impact on project papers, citations, and efficiency, thus highlighting the role of such collaboration networks in promoting research output. However, the promotional effect varied; centrality had a stronger impact when investigators had project experience or worked on natural science research but weaker when projects were affiliated with top research institutions or involved foreign team members. These collaboration networks enhanced research output through improved relationships and larger teams. Moreover, they had a more significant impact on promoting English output than Chinese. This paper offers essential insights for optimizing science funding management from the perspectives of project experience, scientific field, institutional level, and international composition.

<sup>1</sup> School of Economics and Management, Shanghai Institute of Technology, 120 Caobao Road, Shanghai 200235, China. <sup>2</sup> Department of Architecture, Shanghai Academy of Fine Arts, Shanghai University, 99 Shangda road, Shanghai 200444, China. ✉email: [kaiming1239@shu.edu.cn](mailto:kaiming1239@shu.edu.cn)

## Introduction

Science funding is of paramount importance for enhancing a country's competitiveness, technological progress, and economic development (Hu, 2020). In 2022, science funding in China amounted to 32.699 billion Chinese yuan, supporting 51,600 projects<sup>1</sup>. Similarly, science funding in the US reached 8.601 billion US dollars in the same year, funding 18,900 projects<sup>2</sup>. These substantial funding amounts and project numbers indicate the significance both China and the United States attach to science funding, as well as the size of the funded research community. Science funding projects are typically carried out in the form of teams, forming common project experience (CPE) when multiple researchers collaborate on the same project. This experience can evolve further into collaboration networks, involving a large number of researchers and ultimately influencing research output. However, unlike co-author relationships, CPE involves greater uncertainty, looser collaboration, a relatively weak foundation, and a more indirect impact on paper output. Given the widespread existence and high uncertainty of CPE, discussing the collaboration network based on it can significantly enhance research output. CPE offers two advantages over co-authorship: (i) in actual research groups, many collaborations among researchers do not always result in co-authored publications. However, daily interactions and discussions significantly contribute to their research outputs, particularly within the same project. In such cases, CPE more accurately captures the essence of collaboration, and (ii) co-authorship networks of PIs may include collaborations that are unrelated to their NSFC projects. Using co-authorship to construct networks may, therefore, overestimate the effect of the PI's centrality on project outcomes.

Collaboration networks exert a significant impact on research output (Anderson & Richards-Shubik, 2022). As academic research specialization deepens, the demands on researchers' capabilities continue to grow, making independent research more challenging. Nevertheless, collaboration offers an effective approach to overcoming individual cognitive limitations and bounded rationality (Li et al., 2013). Within collaboration networks, researchers share workloads, expertise, specific skills, equipment, and resources to achieve common goals (Bozeman & Corley, 2004; He et al., 2009). Furthermore, the group norms generated by collaboration networks have supervisory and motivating effects, encouraging researchers to invest more energy in their work (Bozeman et al., 2013; Hoekman et al., 2010). However, collaboration networks can also have negative consequences, as researchers invest considerable time and resources in maintaining connections within the network (Sinnewe et al., 2016). While a consensus view on the impact of collaboration networks on research output has not yet been reached, the proportion of collaboration papers is increasing in reality (Kong et al., 2019). Further in-depth discussions are required to fully understand the impact of collaboration networks on research output.

Social capital theory and resource dependence theory suggest that differences at both the individual and organizational-environmental levels may lead to variations in the impact of collaboration networks on research output (Hillman et al., 2009; Inkpen & Tsang, 2005). First, researchers with specific experience may efficiently utilize networks to find resources and collaborators due to their accumulated expertise and clearer search direction (Dahlander & McFarland, 2013). Second, language and cultural differences between foreign and domestic members, along with spatial distance, may escalate communication costs and weaken the resource-sharing function of the network (Aldieri et al., 2018). Third, natural science research demands more resources compared to social science research, making the

resource acquisition function of collaboration networks more crucial in this scientific domain (Benz & Rossier, 2022). Fourth, top research institutions possess abundant resources, and their researchers are also highly capable, leading collaboration networks to provide them with relatively limited assistance (Zong & Zhang, 2019).

Additionally, collaboration networks influence research output through specific mechanisms. On the one hand, collaboration networks can strengthen the relationship bonds between researchers, leading to enhanced trust and reciprocity, reducing collaboration risks and transaction costs, and facilitating knowledge transfer (Liu et al., 2023). On the other hand, collaboration networks can expand the size of research teams, thereby optimizing individual workloads and enhancing team diversity, fostering new ideas (Yuan & Van Knippenberg, 2022). Finally, the impact of collaboration networks on the output of different language types varies. Chinese research institutions put more emphasis on English output in assessment, and English output itself has a broader readership than Chinese (Xu, 2020). However, the existing literature rarely addresses the relationship between collaboration networks and research output from the above perspective, potentially leading to a limited understanding of this relationship. Therefore, this paper addresses the research questions of whether the collaboration network constructed using CPE can enhance research output, the factors influencing the network's role, and the underlying mechanisms of the collaboration network. This paper analyzes research output at the project level, offering a novel perspective compared to the individual-level focus of previous studies (Anderson & Richards-Shubik, 2022; Li et al., 2013). This distinction expands the research output literature by broadening the analysis scope. Additionally, while some studies have examined the impact of collaboration networks on research output, the direction of this effect remains inconclusive (He et al., 2009; Sinnewe et al., 2016). This paper provides new empirical evidence to resolve this ambiguity, demonstrating that collaboration networks have a significant positive impact on research outputs.

To address the above questions, we conducted the following research. Firstly, we utilized crawler technology to collect data on 43,271 approved projects from the website of the National Natural Science Foundation of China (NSFC) from 2009 to 2014. We constructed CPE collaboration networks based on whether principal investigators (PIs) have participated in the same project. The PI refers to the individual responsible for managing the funded project. The PI oversees the project's planning, organization, and implementation, serving as its leader and primary decision-maker. We calculated the network centrality index of the PI based on the network to assess their ability to utilize it for resource and collaborator searches. Secondly, we collected project output data from the Web of Science (WOS) and China National Knowledge Infrastructure (CNKI). We developed three output indicators: the number of project papers, the citations of those papers, and the project data envelopment analysis (DEA) efficiency. This paper extends the applicability of variables from previous studies by incorporating novel data on PIs' CPE and large-scale bibliometric statistics, validating the robustness of the variables, and providing large-scale empirical evidence for both the collaboration network and research output literature (Hu, 2020; Li et al., 2013). Thirdly, we examined the impact of PI centrality on various project output indicators and further investigated the variation of this effect in relation to project experience, scientific field, institutional level, and international composition, as well as two potential mechanisms: tie strength and team size. This paper emphasizes key characteristics of the NSFC, enabling a detailed examination of how the impact of PI

collaboration networks on project research output varies across different conditions, an area that has received less attention in previous studies. Additionally, this analysis provides valuable insights for optimizing fund allocation strategies. Finally, we explored the influence of collaboration networks on research output across different language types.

This paper makes the following contributions: first, it utilizes researchers' CPE to uncover the impact of a more common and uncertain collaboration network on research output in reality. In the context of increasing science funding, CPE is becoming more widespread. However, to the best of our knowledge, due to data limitations, few studies have utilized CPE to construct collaboration networks. This contribution extends the research on collaboration networks (Bozeman & Corley, 2004; He et al., 2009; Hoekman et al., 2010; Kim & Kim, 2020; Sinnewe et al., 2016).

Second, our study discusses the factors that strengthen and weaken the influence of collaboration networks and analyzes the mechanisms of collaboration networks, supplementing the lack of discussions in the existing literature on these aspects (Anderson & Richards-Shubik, 2022; Li et al., 2013). Project experience, scientific field, institutional level, and international composition can either strengthen or weaken the utility of collaboration networks, thereby affecting the assessment of the relationship between collaboration networks and research output. Tie strength and team size are directly affected by the collaboration network, helping us to open the black box of the relationship between the collaboration network and research output. In addition, compared to previous studies, the mechanism analysis in this article reveals two noteworthy findings. On the one hand, in large teams, higher centrality combined with stronger tie strength has a negative effect on project research output, a phenomenon not observed in smaller teams. On the other hand, the positive effect of PI centrality on project research output exhibits a nonlinear relationship with team size. Specifically, in smaller teams, the promotion effect of centrality on project output weakens as team size increases. However, this effect strengthens in larger teams as the team size grows.

Third, previous studies have seldom discussed the differences in the impact of collaboration networks on the output of different language types. This paper finds that English-language articles exhibit relatively higher quality, innovation, and influence. Therefore, the role of PI centrality in promoting research output is stronger for English-language publications. Within the context of China's ongoing focus on research localization, discussions in this area can offer valuable insights for researchers in balancing the allocation of resources between international and domestic publications (Kwiek, 2020; Leydesdorff et al., 2019; Marginson, 2022; Xu, 2020; Zhang & Guo, 2017). Both international and domestic publications are of paramount importance, and finding a balance between the two is of great significance to the sustainable development of academia.

The rest of this article is structured as follows: Section 2 presents the literature review and research hypothesis. Section 3 describes the data and research methods. Section 4 discusses the impact of CPE collaboration networks on research output. Section 5 presents the heterogeneity analysis. Section 6 delves into mechanism analysis and further study. Section 7 provides the conclusion.

## Literature review and hypothesis development

**CPE collaboration networks and research output.** Collaboration networks formed through different relationships have varying impacts on research output. Collaboration in the form of CPE is widespread, connecting researchers to create large and complex collaboration networks that significantly influence research

output (Kardes et al., 2014). Based on CPE, these networks are more elusive and challenging to observe than co-authorship relationships, yet they are prevalent among researchers. The impact of the collaboration network built on the basis of CPE on research output is a more indirect process, operating from project to paper. In contrast, the collaboration network established through co-author relationships exerts a more direct effect on research output, operating from paper to paper (Anderson & Richards-Shubik, 2022). In the case of CPE-based networks, the foundation for collaboration between the two parties is weaker, introducing greater uncertainty regarding the production of papers. However, in networks established through co-author relationships, the two parties have already co-published papers, significantly reducing communication costs and increasing the likelihood of producing new papers. Evidently, the influence of the collaboration network is closely tied to the foundation of its formation. Nevertheless, due to data limitations, current studies predominantly rely on co-author relationships to construct networks (Ji et al., 2022; Krishen et al., 2021). Insufficient research on building collaboration networks based on CPE may lead to incomplete conclusions regarding collaboration networks.

With the increasing specialization of research, collaboration has become a prevalent practice in academia, contributing to forming collaboration networks that enhance research output (Bozeman & Corley, 2004; Katz & Martin, 1997). Collaborating researchers benefit from the overlap of expertise, facilitating skill enhancement. Furthermore, they engage in cross-validation and internal reviews of each other's work before paper submission (Smith et al., 2021). Collaboration networks foster an environment for enhanced knowledge creation by integrating expertise and technologies from diverse sources. Additionally, the learning experience of acquiring skills and technologies from partners sets the foundation for future research activities, while social networks with partners provide valuable information and research opportunities (He et al., 2009). These networks also offer face-to-face communication opportunities, intensifying and diversifying language or other forms of interaction. Such opportunities facilitate the generation of a shared language, common values, and the transfer of knowledge that may be difficult to express verbally (Filieri et al., 2014; Hoekman et al., 2010).

Collaboration networks enable researchers to efficiently mobilize resources necessary for publications, serving as a basis for new funding and the initiation of further research. Thus, these networks help researchers surpass their cognitive and rational limitations, leading to more influential research outcomes (van Rijnsoever & Hessel, 2011). Within the network, valuable resources such as information and knowledge are exchanged through interactions, and the transfer process is optimized by minimizing redundancy (Li et al., 2013; Xu et al., 2014). Moreover, networks also serve to diminish opportunistic behavior among members by fostering a high level of trust, consequently reducing the monitoring costs associated with collaboration (Nahapiet & Ghoshal, 1998). Group norms within the network, such as exclusion and forgiveness, can influence individual behavior, promoting cooperation and improving research efficiency (Ali & Miller, 2016). Finally, several studies have shown that international collaboration positively influences research output. Researchers engaged in international collaborations tend to achieve significantly higher output levels compared to those without such partnerships (Kwiek, 2020). Furthermore, international collaborations can greatly enhance the impact of articles and help establish global leadership in various research fields (Leydesdorff et al., 2019; Zhang & Guo, 2017).

However, collaboration networks can also impede research output (Liang & Liu, 2018). Identifying and evaluating partners in the early stages, formulating contracts to manage collaboration

and distribute expected results, coordinating researchers in the mid-term, monitoring work progress, and evaluating the quality of research results in the later stages are all costs associated with collaboration networks (He et al., 2009; Sinnewe et al., 2016). While the network provides individuals with more connections, it can significantly increase the cost of maintaining these connections. Excessive connections may restrict an individual's vision and hinder communication with other external organizations (Inkpen & Tsang, 2005). Moreover, the strong group norms and mutual identity generated by collaboration networks may also restrict information openness and communication, potentially leading to collective blindness, which can have serious negative effects (Inkpen & Tsang, 2016). In research collaboration specific to the form of co-authors, costs such as nominal authorship and ghostwriting may be incurred. Furthermore, some researchers adopt these practices to increase their influence but refuse to take responsibility for the published articles (Bozeman et al., 2013). Another common form of research collaboration is inbreeding, which has been found to be linked to lower output and less openness to non-inbred researchers (Horta et al., 2010).

Based on the above analysis, we hypothesized that the CPE collaboration network can enhance research output. On the one hand, it facilitates the flow of resources, such as information and knowledge, among researchers; on the other hand, it fosters a high level of trust and inhibits opportunistic behavior. Although collaboration networks increase the cost of maintaining relationships for researchers, the benefits they offer are substantial. Considering the combined effect of the two, the collaboration network generally has a positive impact. Therefore, we propose hypothesis 1.

**H1: CPE collaboration networks will enhance research output.**

**Heterogeneity of collaboration networks.** Social capital theory posits that differences at the individual level can influence one's network position and accumulation of social capital (Inkpen & Tsang, 2005). Researchers with extensive project experience can effectively utilize the network. With sufficient project experience, the researcher can conduct targeted searches for resources within the collaboration network in a more targeted manner. Simultaneously, the researcher can more effectively assess the potential of partners and projects, selecting the most suitable collaborators and projects from the network to minimize transaction costs associated with collaboration. Furthermore, as project experience increases, resources start to accumulate for these senior and highly regarded researchers (Dahlander & McFarland, 2013), expanding their collaboration network and providing them with a wider range of options. Nonetheless, when the researcher lacks sufficient project experience, their ability to identify suitable collaborators and projects remains underdeveloped. Moreover, the multitude of opportunities presented by the collaboration network may consume significant time and energy, disrupting the researcher's normal research rhythm and hindering output.

The presence of foreign members can elevate the communication cost of the collaboration network. Firstly, language differences increase communication costs and can impact the enthusiasm of members both at home and abroad, hindering the timely exchange of ideas (Aldieri et al., 2018). Secondly, cultural differences make it more challenging to establish cohesive group norms, as members from different cultural backgrounds may have contrasting levels of acceptance toward the same working model. Lastly, spatial distance hinders the seamless integration of foreign members into the team and can even lead to a sense of alienation. Some studies suggest that researchers from countries with lower levels of scientific research may engage in free-riding

behavior by leveraging international collaborations to co-publish with researchers from countries with higher research capacities, thereby increasing their publication output (Kim & Kim, 2020). In contrast, when the team consists of domestic members only, communication costs are reduced due to the shared language. Moreover, members sharing the same cultural background tend to have a closer level of acceptance toward the work model, enhancing the likelihood of fostering strong group norms. Additionally, the increased opportunities for face-to-face communication lead to higher team integration among the members.

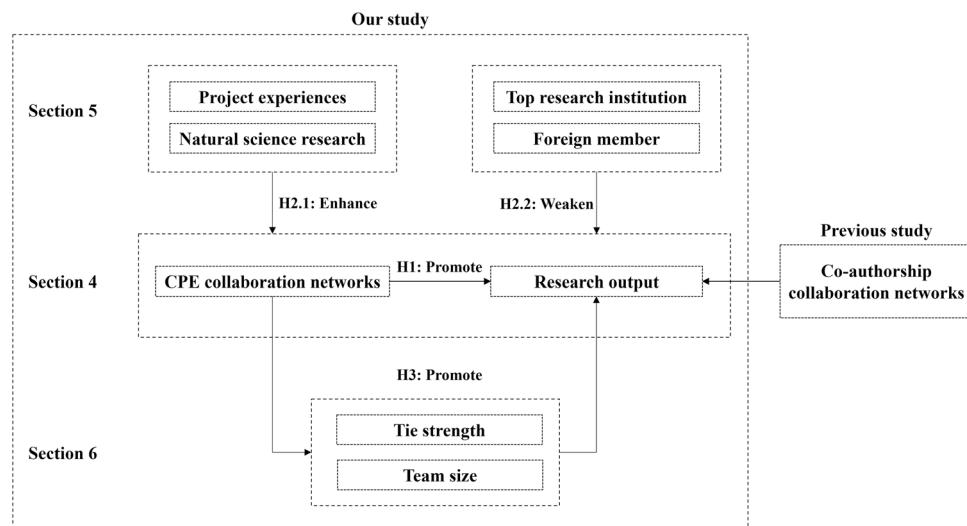
Resource dependence theory suggests that changes at the organizational and environmental level can impact a team's access to essential external resources (Hillman et al., 2009). Collaboration networks have varying effects on projects in different scientific fields. Natural sciences typically involve experimentation, necessitating significant resource consumption, including equipment, instruments, materials, and more (Benz & Rossier, 2022). In such cases, the researcher's collaboration network can assist in resource acquisition and integration. Simultaneously, the research methods and evaluation of results in natural sciences tend to be more objective, and the use of common standards can mitigate debates among researchers regarding correctness or quality, fostering academic exchange. In contrast, social sciences usually do not rely on experimentation, substantially reducing expenses and diminishing the impact of network resource availability. Moreover, social science tends to be more subjective, with diverse research methods and evaluation systems across different institutions, making comparisons of superiority challenging. The connections formed by collaboration networks fail to foster academic exchanges in such cases and can also increase communication costs among researchers.

Researchers' dependence on the collaboration network can decrease with abundant institutional resources. General research institutions often have limited resources and lack regular activities such as forums, seminars, and lectures. In such cases, the collaboration network offers researchers valuable opportunities to communicate and engage with other researchers in these institutions. The collaboration norms established within the network can enhance communication efficiency and encourage participation from all parties in knowledge exchange (Arregle et al., 2007). Moreover, the sense of identity fostered by the collaboration network can enhance communication opportunities and promote more frequent collaboration (Tsai et al., 2014). However, in top research institutions, abundant research resources and strong researchers make the collaboration network's assistance relatively limited. Furthermore, the presence of numerous collaborators attracted by the reputation of top research institutions increases the screening cost for the researcher. Selecting an unsuitable collaborator or project can impede their ongoing research progress. Based on the analysis above, we propose Hypothesis 2.1 and Hypothesis 2.2.

**H2.1: The promotion effect of the collaboration network on research output is stronger when the researcher has more project experience or when the project pertains to natural science research.**

**H2.2: The promotion effect of the collaboration network on research output is weaker when the host institution of a project is a top research institution or the project team includes foreign members.**

**Mechanism of collaboration networks.** The degree to which a researcher utilizes collaboration networks directly impacts the strength of their ties in the network. In social networks, the relationships between nodes are typically categorized into weak ties and strong ties based on their strength (Granovetter, 1973).



**Fig. 1 Analytical framework.** This figure presents the main research contents of Sections 4 to 6 and illustrates the logical relationship among these sections.

Weak ties are characterized by their short duration, infrequent interactions, and low emotional intimacy. They facilitate access to diverse social circles and enhance the potential for acquiring varied information (Lyu et al., 2019). Conversely, strong ties entail long-term, frequent interactions and high emotional intimacy. They offer opportunities for access and trust among like-minded individuals. Researchers with aligned goals establish connections within a research collaboration network to foster long-term partnerships. Strong ties are associated with dense network clusters, whereas weak ties serve as potential bridges between these clusters. Furthermore, strong ties foster increased trust and reciprocity, thereby mitigating the costs and risks involved in new collaborations. The combination of higher trust and reduced risk in these strong-tie relationships, in turn, facilitates knowledge transfer and creation (Guan et al., 2015; Wang, 2016). While weak ties encourage the collision and generation of novel ideas, scientific research ultimately requires a high level of knowledge intensity. Thus, stable and long-lasting strong ties are more favorable for fostering cooperative development.

Collaboration networks directly influence the size of research project teams. In general, the researcher's increased connections within the network enhance the possibility of contacting potential collaborators and selecting suitable individuals to join their research team (Yuan & Van Knippenberg, 2022). With an increasing team size, the workload assigned to individual researchers decreases, thereby expediting the output of research results. Moreover, enlarging the team size enhances the diversity of team members' backgrounds, fostering the generation of new ideas. However, expanding the team size may also present disadvantages, including heightened communication costs for the researcher and the potential for opportunistic behavior among team members. Despite these potential issues, we argue that the internal structure of the researcher-composed team remains relatively flat, which does not significantly elevate the management costs for the researcher. Additionally, opportunistic behavior yields limited benefits to researchers, as they strive to uphold their reputation and are unwilling to include individuals who have not contributed to their articles. Based on the above analysis, we propose Hypothesis 3.

**H3: The collaboration network enhances research output by improving the researcher's tie strength and expanding the project team size.**

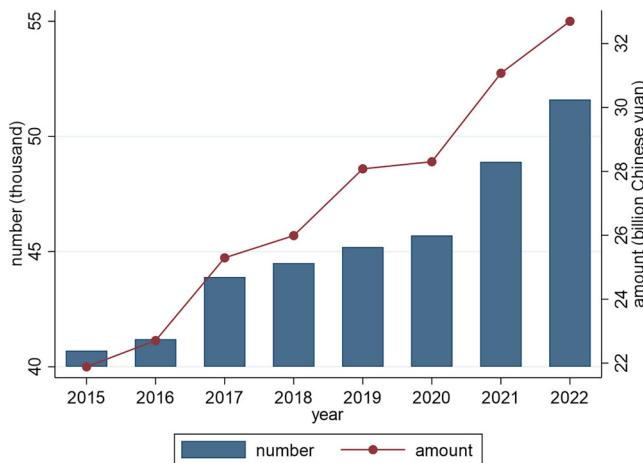
**Analytical framework.** This paper presents an analytical framework as depicted in Fig. 1. Firstly, it examines the impact of

collaboration networks formed through CPE on research output (hypothesis 1). While collaboration networks may entail transaction costs, they predominantly facilitate resource sharing, leading to an overall positive impact. Secondly, the study delves into the heterogeneity of this impact on research output (hypotheses 2.1 and 2.2). Project experience enhances researchers' efficiency in utilizing collaboration networks, and there is a heightened demand for resource-sharing functionalities in natural science research. However, the resource-rich top research institutions may diminish the reliance on collaboration networks for resource sharing, and the involvement of foreign members might increase communication costs within the networks. Thirdly, the paper explores how CPE collaboration networks influence research output (hypothesis 3). One key mechanism is the strength of researchers' ties, and another is the size of project teams.

## Data and methodology

**NSFC.** NSFC was established on February 14, 1986, as a vice-ministerial unit under the Ministry of Science and Technology of China. Its primary role is to manage natural science funding in compliance with the law. The NSFC is responsible for organizing and implementing funding plans, conducting project setting and review, granting project approval, and providing supervision. The NSFC funds scientific research in 9 domains: mathematics and physical science, chemical science, life science, earth science, engineering and material science, information science, management science, health science, and interdisciplinary science. There are 17 types of funding offered by the NSFC, with the general program and youth science fund encompassing the broadest range and receiving the highest amount of funding. As depicted in Fig. 2, the NSFC's funded projects rose from 40,700 in 2015 to 51,600 in 2022, with the annual funding amount increasing from 21.884 billion Chinese yuan in 2015 to 32.699 billion Chinese yuan in 2022<sup>3</sup>.

The selection of the NSFC as the data source is primarily based on three key considerations. First, natural science funding has extensive coverage and involves numerous researchers. Second, research institutions and researchers highly value natural science funding, and securing funding projects has become a critical criterion for evaluating the research capabilities of institutions and individuals. Third, the funding projects typically have long durations, usually 3–5 years, fostering long-term CPE, which in



**Fig. 2 The number and amount of projects funded by the NSFC from 2015 to 2022.**

The blue bars indicate the annual number of projects funded by the NSFC, with the left scale representing units in thousands. The red line depicts the annual funding amount of projects by the NSFC, with the right scale representing units in billion Chinese yuan.

turn increases the likelihood of real collaborations among researchers.

**Sample.** We collected three types of data for this study: science fund project data, records of projects in which the project leader has participated, and project paper output data. Initially, we obtained data on 43,271 science funding projects approved by NSFC from 2009 to 2014. This dataset includes information such as funding amount, funding period, number of project members, disciplines, project types, host institution, and other pertinent details. We selected 2009 as the starting year since WOS began labeling papers funded by NSFC in July 2008. The cut-off year was set as 2014 due to the delayed disclosure of project information by NSFC. We found that most of the project information approved in and after 2015 had not been disclosed during the collection process. Subsequently, we obtained records of the projects in which the PI has participated from NSFC, forming the basis for constructing the PI's CPE collaboration network. The 43,271 projects examined in this study involved 36,706 researchers, collectively participating in 194,039 projects during or before 2014. Lastly, we acquired the paper output data for the projects in both English and Chinese languages from the WOS and CNKI, respectively. We collected 436,772 English articles and 407,285 Chinese articles funded by the 43,271 projects, along with their corresponding citations until 2021.

The final sample consists of 43,271 funding project observations. We winsorized all the continuous variables at the 1% and 99% levels to make sure the sample followed a normal distribution.

## Variables

**Independent variables.** The construction of the CPE collaboration network involves a yearly process. It starts by selecting all PIs from the projects within a specific year and filtering their project participation up to that year. This process results in a researcher-project matrix, with rows representing researcher IDs and columns representing project IDs. A value of 1 is assigned at the intersection of rows and columns if the PI participated in the project; otherwise, it is marked as 0. The researcher-project matrix is then transposed to create a project-researcher matrix. Multiplying this matrix with the researcher-project matrix generates a researcher-researcher matrix, where rows and columns

represent researcher IDs, and values at the intersections indicate collaboration counts between researchers. To calculate the independent variable index, all diagonal elements in the researcher-researcher matrix are set to 0. The resulting researcher-researcher matrix is imported into Pajek software for visualization. Figure 3 displays the CPE collaboration network of PIs from 2009 to 2014. Each dot represents a PI, and connections between dots signify their participation in the same project. The dot size reflects the frequency of collaboration with other PIs. The network exhibits increasing node density from 2009 to 2011, indicating a growing number of PIs joining the network. From 2012 to 2014, the node density stabilizes, but the number of larger dots increases, suggesting the emergence of several highly productive researchers during this period. It is important to note that the NSFC project regulations do not restrict PIs from participating in multiple projects simultaneously<sup>4</sup>. In the original data for this paper, there are numerous instances of PIs participating in multiple programs within the same year.

This paper aims to explore the influence of collaboration networks formed through researchers' CPE on the research output of the projects they lead. To achieve this objective, it is necessary to use indicators that can assess individuals' resource acquisition capacity within the network. The centrality index, derived from social network theory, is well-suited for this purpose as it effectively captures the connections among individuals within the network (Li et al., 2013). Centrality is closely linked to an individual's potential ability, influence, and reputation. Existing literature commonly employs three types of centrality indicators (Borgatti, 2005; Freeman, 1978):

(i) Degree centrality. It represents the number of connections between a focal node and other nodes in the network. In our context, it directly measures the number of collaborations with other researchers.

$$\text{Degree}_p = \frac{d_p}{n-1} \quad (1)$$

Where  $d_p$  represents the number of direct connections PI  $p$  has,  $n$  represents the total number of nodes in the network.

(ii) Betweenness centrality. It is defined as the ratio of the number of shortest paths between pairs of nodes that include node  $p$  to the total number of shortest paths between pairs of nodes. This measure assigns higher values to researchers who bridge different clusters within the network. It indicates that individuals have more influence if other researchers rely on them to connect with one another.

$$\text{Betweenness}_p = \frac{\sum_{q,r} x(p)_{qr}}{(n-1)(n-2)} \quad (2)$$

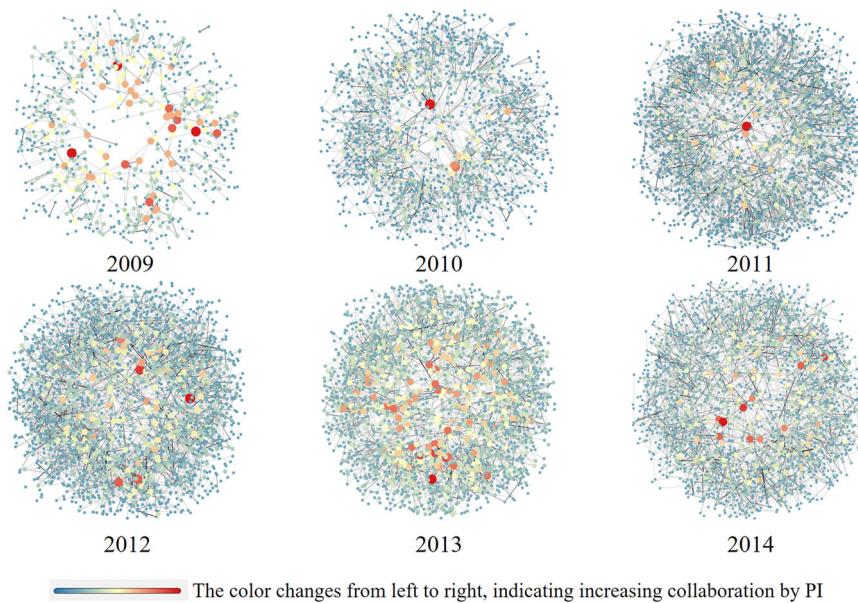
Where  $x(p)_{qr}$  is the shortest path between PI  $q$  and PI  $r$  on which PI  $p$  lies,  $x_{qr}$  is the shortest path between PI  $q$  and PI  $r$  with or without PI  $p$ .

(iii) Closeness centrality. It is defined as the inverse of the sum of all shortest distances between a given node and other nodes within the network. In our study, it indicates the proximity of a researcher to all other researchers in the network, whether directly or indirectly. Researchers with a high closeness centrality demonstrate a greater ability to access or acquire resources possessed by others in the network.

$$\text{Closeness}_p = \frac{n-1}{\sum_{q \neq p} g_{pq}} \quad (3)$$

Where  $g_{pq}$  represents the length of the shortest path between PI  $p$  and PI  $q$ .

All three measures of centrality hold significance in various contexts, albeit with varying interpretations. Generally,



**Fig. 3 The CPE collaboration network map of PIs from 2009 to 2014.** Each dot represents a PI. Lines between the dots indicate collaborative projects between two PIs. The larger the dot, the more collaborations the PI has. The dot colors use a gradient from blue to red. The closer a dot is to red, the more collaborations the PI has.

individuals with high centrality possess broader, deeper, more focused, and extensive approaches to acquiring and disseminating information, as well as influencing others. To facilitate comparisons among individuals, we rank the centrality of each individual in each year and assign a percentile value ranging from 1 to 100. Subsequently, we calculate the average of the three percentile values for each individual in each year to determine their annual comprehensive centrality index (Fogel et al., 2021).

$$Cen_{p,t} = (Degree_{p,t} + Betweenness_{p,t} + Closeness_{p,t})/3 \quad (4)$$

**Dependent variables.** Since papers constitute the primary output of NSFC projects, we employ paper data to construct indicators for evaluating project output (Hu, 2020; Yu et al., 2022). The first indicator, TNum, quantifies the number of papers produced by the project. It reflects the total number of papers funded by the project, sourced from WOS and CNKI. We select these databases for data collection due to their authoritative status in cataloging English and Chinese papers, respectively. Additionally, they include project information for funded papers, facilitating the matching of papers to specific projects.

Given the varying quality of each article, we developed a second output indicator based on the number of citations an article has received. This second indicator, termed TQuo, quantifies the citations accrued by project papers. It denotes the total number of citations garnered by articles funded through the project. Considering that older articles typically accumulate more citations, we adjusted for this by dividing the number of citations of each article by the number of years since its publication (calculated as 2021 minus the publication year, with 2021 being the year in which we collected the citation data). We then averaged these adjusted citation counts and summed them up.

The third indicator, denoted as Vrs, quantifies the efficiency and is computed using DEA (Amara et al., 2020). The calculation of Vrs involves treating a single project as the decision-making unit, with its funding amount, total number of project members, and project duration serving as input indicators, as well as the number of project papers and the total citations of project papers as output indicators. The project's output-input efficiency value is obtained by calculating the ratio of the weighted output to the

weighted input and comparing it with the maximum efficiency value among all projects in the current year. An optimization problem is formulated and solved to obtain a weight matrix used to calculate the project's DEA efficiency. The DEA efficiency of a project is a relative measure indicating how the project's output-input ratio relates to the maximum output-input ratio achieved in the given year, thus ranging from 0 to 1. As funding projects typically strive to maximize research outcomes based on a range of inputs (e.g., funds, personnel, and time), this paper adopts an output-oriented DEA model. This article employs the BCC model to calculate DEA efficiency, as the BCC model assumes variable returns to scale. In practice, returns to scale often fluctuate with changes in input levels, making the BCC model a more appropriate choice for evaluating the efficiency of fund projects (Banker et al., 1984).

**Analytical approach.** To test hypothesis 1, we employ the following fixed effect model:

$$Y_{i,t} = \alpha + \beta_1 Cen_{p,t} + \beta_2 Tit_{p,t} + \beta_3 Fund_{i,t} + \beta_4 Time_{i,t} + \beta_5 WHum_{i,t} + \text{fixed effects} + \varepsilon_{i,t} \quad (5)$$

Subscript  $i$  represents the project,  $p$  stands for the researcher, and  $t$  stands for the year the project started. The dependent variable  $Y$  includes the number of project papers, citations of project papers, and project DEA efficiency (as defined in 3.3.2).  $Cen$  is an independent variable, indicating the centrality of the PI of project  $i$  in the current year's collaboration network (as defined in 3.3.1).  $Tit$  represents the professional title of researcher  $p$  in year  $t$ , with 1 for assistant professor, 2 for associate professor, and 3 for professor.  $Fund$  represents the funding amount of project  $i$ .  $Time$  represents the funding period of project  $i$ .  $WHum$  represents the team size of project  $i$ . Additionally, we controlled for four fixed effects: project start year, discipline type, funding type, and host institution. Controlling for fixed effects related to discipline type can help mitigate the influence of discipline on paper citations. The regressions were performed with cluster-robust standard errors at the host institution level.

**Table 1** Summary statistics for variables.

Variable	N	Mean	SD	Min	Max	Median
TNum	43,271	17.170	15.383	1.000	88.000	13.000
TQuo	43,271	328.117	510.500	3.000	3,194.000	151.000
Vrs	43,271	22.257	19.742	1.220	100.000	16.471
Cen	43,271	29.654	32.325	1.000	100.000	30.333
Tit	43,271	2.645	0.547	1.000	3.000	3.000
Fund	43,271	48.928	41.546	2.000	990.000	35.000
Time	43,271	3.414	0.607	0.250	6.000	3.000
WHum	43,271	13.664	3.689	2.000	29.500	13.500

To test hypotheses 2.1 and 2.2, we introduce additional adjustment variables:

$$Y_{i,t} = \alpha + \beta_1 Cen_{p,t} + \beta_2 Adj_{p,t} + \beta_3 Cen_{p,t} * Adj_{p,t} + \beta_4 Tit_{p,t} + \beta_5 Fund_{i,t} + \beta_6 Time_{i,t} + \beta_7 WHum_{i,t} + \text{fixed effects} + \varepsilon_{i,t} \quad (6)$$

Adj comprises a set of adjustment variables, including Exper, Natu, Top, and Fore. Exper is a binary variable representing project experience. It takes the value 1 if PI p has previously led an NSFC project before the start year of the current project and 0 otherwise. Natu is a binary variable indicating whether project i belongs to natural science. It takes the value 1 if the project falls under natural science and 0 otherwise. Top is a binary variable representing whether the host institution of project i affiliates with a top institution. It takes the value 1 if the host institution belongs to the 985 Project or the Chinese Academy of Sciences and 0 otherwise. Fore is a binary variable indicating the presence of foreign members in the project team. It takes the value 1 if there are foreign members and 0 otherwise. The definitions of the other variables remain the same as in model (5).

For hypothesis 3, we replaced the dependent variables in the model (5) with the PI's tie strength (TieStrength) and the team size (WHum). TieStrength is calculated by dividing the number of collaborations the PI engages in during the year by the number of collaborators. A higher ratio indicates a stronger relationship between the PI and others (Liu et al., 2023). WHum represents the total number of members of the project team.

### The impact of CPE collaboration networks on research output

**Descriptive statistics.** The results presented in Table 1 deviate from previous studies. The average number of project papers (TNum) is 17.170, slightly surpassing the count of 12.830 reported by Hu (2020). This disparity could be attributed to the different periods considered; Hu (2020) covered project approvals in 2010–2011, while this article examines the period of 2009–2014, suggesting an increase in individual project paper outputs over time. The average citations of project papers (TQuo) amount to 328.117, implying that, on average, a single paper funded by the project has received 19.110 citations by 2021 (328.117/17.170). The average value of project DEA efficiency (Vrs) is 22.257, which is lower than the average of 54.265 reported by Amara et al. (2020) for the efficiency of 202 management scholars. While the Vrs values originally ranged from 0 to 1, directly using them in this form would result in coefficients that were too small, making analysis inconvenient. Therefore, we multiplied the Vrs values by 100, making the coefficients easier to interpret without affecting the regression results. The wide range in efficiency may stem from two factors. First, the varying abilities of PIs likely contribute to significant differences in the input-output ratio across projects. Second, inherent differences in input-output efficiency exist between disciplines. For instance, the average output in engineering and materials science is typically

**Table 2** Impact of CPE collaboration networks of PIs on research output.

Dependent variable	TNum	TQuo	Vrs
	(1)	(2)	(3)
Cen	0.018*** (7.16)	1.056*** (9.75)	0.027*** (7.30)
Tit	4.076*** (32.07)	138.849*** (19.45)	6.321*** (28.56)
Fund	0.084*** (10.96)	1.864*** (7.45)	0.089*** (10.28)
Time	2.528*** (8.93)	22.244 (1.64)	-0.356 (-0.60)
WHum	0.573*** (16.79)	4.342*** (3.38)	0.284*** (6.39)
Constant	-14.684*** (-15.11)	-296.126*** (-6.55)	-2.229 (-1.05)
Year_FE	Yes	Yes	Yes
Discipline_FE	Yes	Yes	Yes
Funding_FE	Yes	Yes	Yes
Institution_FE	Yes	Yes	Yes
Adj. R <sup>2</sup>	0.350	0.214	0.254
N	42996	42996	42996

All regressions include project start year, discipline type, funding type, and host institution fixed effects. Standard errors are clustered by institution. T statistic is presented in parenthesis.

\*\*\*p < 0.01, \*\*p < 0.05, \*p < 0.1.

much higher than in management-related disciplines. The mean value of Cen is 29.654, signifying a low degree of connection among the PIs in the collaboration network and suggesting a relatively sparse network. Furthermore, the average value of the PI's professional title is 2.645, indicating that a majority of the PIs hold the title of professor.

**Basic estimation.** The results in Table 2 demonstrate a significant positive impact of the PI's centrality on research output, thus verifying Hypothesis 1. In column 1, the coefficient of Cen is statistically significant at the 1% level, suggesting that a 1% increase in the network centrality of the PI leads to a 0.018 increase in the number of project papers. Similarly, in column 2, the coefficient of Cen is statistically significant at the 1% level, indicating that a 1% increase in the centrality of the PI results in a 1.056 times increase in the citations of project papers. Likewise, in column 3, the coefficient of Cen remains statistically significant at the 1% level, suggesting that a 1% increase in the centrality of the PI leads to a 0.027 increase in the DEA efficiency of the project. These findings align with the findings of Li et al. (2013), providing further evidence that collaboration networks play a role in fostering research output. Moreover, they highlight the positive significance of CPE.

The regression results of the control variables indicate a positive relationship between the researcher's professional title, project funding, funding period, team size, and project output. Specifically, the coefficients of Tit are significantly positive in all three columns, suggesting a positive association between the researcher's professional title and the project output. These results align with the findings of Amara et al. (2020). Similarly, the coefficients of Fund are significant in columns 1–3, indicating a positive relationship between the amount of project funding and the quantity and quality of produced papers. These findings are consistent with the findings of Hu (2020). Moreover, the coefficient of Time is significantly positive in column 1. It exhibits a near 10% statistical significance in column 2, suggesting a positive correlation between the funding period and the

**Table 3 Robust test: propensity score matching.**

Dependent variable	TNum	TQuo	Vrs
	(1)	(2)	(3)
DumyCen	1.153*** (6.22)	64.605*** (7.81)	1.728*** (6.29)
Tit	4.243*** (24.91)	155.027*** (16.25)	6.718*** (23.63)
Fund	0.083*** (9.69)	2.040*** (7.09)	0.091*** (9.14)
Time	2.669*** (6.72)	16.145 (0.86)	-0.792 (-1.19)
WHum	0.705*** (14.96)	7.244*** (3.72)	0.459*** (7.05)
Constant	-17.094*** (-12.15)	-338.876*** (-5.74)	-3.738* (-1.66)
Year_FE	Yes	Yes	Yes
Discipline_FE	Yes	Yes	Yes
Funding_FE	Yes	Yes	Yes
Institution_FE	Yes	Yes	Yes
Adj. R <sup>2</sup>	0.405	0.238	0.299
N	23484	23484	23484

All regressions include project start year, discipline type, funding type, and host institution fixed effects. Standard errors are clustered by institution. T statistic is presented in parenthesis.  
\*\*\*p < 0.01, \*\*p < 0.05, \*p < 0.1.

quantity and quality of project output. Likewise, the coefficients of WHum are significantly positive in columns 1–3, indicating a positive relationship between the project team size and the research output.

**Robust test.** When discussing the impact of collaboration networks on research output, it is crucial to consider the potential presence of uncontrolled variables that may introduce a non-random distribution of the PI's centrality in the sample. These uncontrolled variables can also influence research output, leading to a self-selection problem in the model. To address this issue, we employ the propensity score matching method (Fan et al., 2022). Firstly, the annual median of the centrality variable is calculated, and a dummy variable called DumyCen is created. The dummy variable takes a value of 1 if the centrality of a sample is greater than the median and 0 otherwise. Secondly, samples with a DumyCen value of 1 are matched with those having a DumyCen value of 0 by year, considering covariates such as the PI's title, funding amount, funding period, discipline type, project type, and host institution. The matching process utilizes a 1:1 proximity match method. Thirdly, the newly matched samples undergo retesting.

The results in Table 3 demonstrate the robustness of the findings after employing the matching method to process the samples. In column 1, the coefficient of DumyCen is statistically significant at the 1% level, suggesting that researchers with higher centrality produce an average of 1.153 more papers compared to those with lower centrality. Similarly, in column 2, the coefficient of DumyCen is significantly positive at the 1% level, indicating that researchers with higher centrality receive an average of 64.605 more citations than those with lower centrality. Additionally, in column 3, the coefficient of DumyCen remains significantly positive, implying that researchers with higher centrality exhibit an average efficiency that is 1.728 percentage points higher than those with lower centrality.

A comparison of the coefficient of DumyCen in columns 1–3 with the means of TNum, TQuo, and Vrs reveals that the number of project papers, the citations of project papers, and the DEA

**Table 4 Robust test: instrumental variable.**

Dependent variable	TNum	TQuo	Vrs
	(1)	(2)	(2)
Cen	0.020*** (3.06)	0.940*** (4.64)	0.029*** (3.18)
Tit	4.074*** (31.94)	138.958*** (19.41)	6.318*** (28.44)
Fund	0.084*** (10.92)	1.876*** (7.47)	0.089*** (10.24)
Time	2.530*** (8.95)	22.124 (1.63)	-0.353 (-0.59)
WHum	0.573*** (16.81)	4.331*** (3.38)	0.285*** (6.39)
KleibergenPaap rk	145.505		
Wald F			
Year_FE	Yes	Yes	Yes
Discipline_FE	Yes	Yes	Yes
Funding_FE	Yes	Yes	Yes
Institution_FE	Yes	Yes	Yes
Adj. R <sup>2</sup>	0.084	0.037	0.048
N	42996	42996	42996

All regressions include project start year, discipline type, funding type, and host institution fixed effects. Standard errors are clustered by institution. T statistic is presented in parenthesis.

\*\*\*p < 0.01, \*\*p < 0.05, \*p < 0.1.

efficiency for PIs with higher network centrality are 6.72% (1.153/17.170), 19.69% (64.605/328.117), and 7.76% (1.728/22.257) higher, respectively, than those for PIs with lower network centrality. The regression results after propensity score matching are consistent with the basic findings and partially mitigate the endogeneity issue that may arise from self-selection, indicating that the impact of the PI's network centrality on project research output is robust.

The greater the research output of a project, the more likely it is to attract collaborators to work with the PI, thereby increasing the PI's centrality and potentially leading to a reverse causality issue. While the centrality measure of the PI in this study is based on the collaboration network data from the year prior to the publication year of the project's paper, which helps mitigate the reverse causality concern to some extent, we further apply the instrumental variable method to address this issue more robustly (Yu et al., 2022). The instrumental variables must meet the requirements of correlation and exclusivity. Accordingly, this study selects the average host institution centrality (IvCen) as the instrumental variable. The mean centrality value exhibits a strong correlation with the centrality of the PI. Moreover, it primarily impacts research output through the pathway influencing PI centrality, thus fulfilling the exclusivity condition.

The findings in Table 4 demonstrate the robustness of the results after retesting with the instrumental variable method. The Kleibergen-Paap F value of 145.505 exceeds the 10% critical value (16.38) suggested by Stock-Yogo, indicating a minimal concern regarding weak instrumental variables. The coefficients of Cen in columns 1–3 are all significantly positive at the 1% level, suggesting that using the average centrality of the host institution as an instrumental variable for the PI's network centrality confirms the robustness of its role in promoting research output. Moreover, the coefficients of Cen are similar to those in Table 2, suggesting that the basic regression results are less influenced by the reverse causality issue. The R<sup>2</sup> in Table 4 is significantly lower than in Table 2, which can be attributed to the fact that the instrumental variable is at the host institution level, where variability is considerably lower than at the PI level, leading to a significant reduction in their explanatory power for the research output indicators.

## Heterogeneity discussion of CPE collaboration networks

Social capital theory posits that individual factors, such as the PI's experience and team members' backgrounds, influence their position within the CPE network and the accumulation of social capital (Inkpen & Tsang, 2005). Resource dependence theory suggests that organizational and environmental factors, such as the project's scientific field and the level of the host institution, impact the team's access to necessary external resources (Hillman et al., 2009). Therefore, analyzing the effect of the PI's centrality in the CPE network on research output from both individual and organizational perspectives allows for a better understanding of how centrality yields heterogeneous effects in different contexts.

**Individual factors.** The first individual level factor influencing heterogeneity is the PI's project experience. Researchers with more experience in a specific field are also more efficient at resource search and allocation. In a CPE collaboration network, experienced researchers benefit in two ways: their familiarity with project management reduces time spent on non-research tasks; their prior leadership experience enables quicker resource identification and collaborator selection. Furthermore, the experience can lead to a Matthew effect, concentrating resources on these researchers and expanding their collaboration network (Hu, 2020). We introduce the Exper variable to examine how project experience influences the relationship between collaboration network centrality and research output. If PI  $p$  of project  $i$  has previously led an NSFC project before year  $t$ , the Exper variable is set to 1; otherwise, it is set to 0.

The results in Table 5 indicate that the impact of CPE collaboration networks on research output is stronger for researchers with prior project experience. In column 1, the coefficients for the cross-product term (Cen\*Expert) and Cen are significantly positive, suggesting that an increase in the centrality

of experienced researchers leads to a greater rise in the number of project papers compared to those without prior project experience. In column 2, a similar positive effect is observed, with experienced researchers showing a stronger increase in project paper citations as their network centrality rises. Column 3 confirms that project experience enhances the impact of researcher network centrality on project efficiency.

The team's international composition is the second individual level factor contributing to heterogeneity. Foreign team members bring differences in language, culture, and cognitive approaches, which may affect collaboration network effectiveness. While NSFC projects primarily comprise domestic members, including foreign members can increase communication costs and hinder information flow due to language barriers (Aldieri et al., 2018). Cultural differences may also impede integration and the development of positive group norms. In contrast, teams with more domestic members benefit from enhanced communication efficiency, better transfer of tacit knowledge, and smoother integration and alignment in group norms. To examine how team international composition influences the relationship between collaboration networks and research output, we introduce a binary variable, Fore, which equals 1 if the team includes foreign members and 0 otherwise.

The results in Table 6 show that the presence of foreign members in the team reduces the positive impact of CPE collaboration networks on research output. In column 1, the positive coefficient of Cen and the negative coefficient for the interaction term between Cen and Fore indicate that an increase in centrality leads to fewer project papers when foreign members are included in the team. In column 3, the significantly positive coefficient of Cen and the negative interaction term imply that the positive impact of centrality on research output is nearly neutralized when the team includes foreign members.

**Table 5 Project experience, collaboration networks, and research output.**

Dependent variable	TNum	TQuo	Vrs
	(1)	(2)	(3)
Cen	0.006** (2.36)	0.457*** (4.26)	0.009** (2.35)
Exper	2.970*** (11.49)	49.283*** (5.47)	3.217*** (9.17)
Cen*Exper	0.017*** (3.94)	1.073*** (6.42)	0.026*** (4.28)
Tit	3.763*** (30.31)	131.399*** (18.27)	5.961*** (26.70)
Fund	0.079*** (10.85)	1.740*** (6.88)	0.084*** (9.84)
Time	2.386** (8.76)	18.187 (1.36)	-0.526 (-0.88)
WHum	0.577*** (17.26)	4.475*** (3.52)	0.290*** (6.61)
Constant	-14.227*** (-15.09)	-274.647*** (-6.19)	-1.605 (-0.76)
Year_FE	Yes	Yes	Yes
Discipline_FE	Yes	Yes	Yes
Funding_FE	Yes	Yes	Yes
Institution_FE	Yes	Yes	Yes
Adj. R <sup>2</sup>	0.356	0.218	0.259
N	42996	42996	42996

Note: All regressions include project start year, discipline type, funding type, and host institution fixed effects. Standard errors are clustered by institution. T statistic is presented in parenthesis.  
\*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$ .

**Table 6 International composition, collaboration networks, and research output.**

Dependent variable	TNum	TQuo	Vrs
	(1)	(2)	(3)
Cen	0.019*** (7.35)	1.070*** (9.91)	0.027*** (7.50)
Fore	-1.469*** (-2.75)	-14.392 (-0.78)	-1.867*** (-2.69)
Cen*Fore	-0.019** (-2.19)	-0.396 (-1.17)	-0.024** (-2.11)
Tit	4.072*** (32.00)	138.777*** (19.45)	6.316*** (28.57)
Fund	0.083*** (10.91)	1.855*** (7.41)	0.088*** (10.20)
Time	2.554*** (9.07)	22.648* (1.67)	-0.323 (-0.54)
WHum	0.584*** (17.25)	4.492*** (3.51)	0.298*** (6.75)
Constant	-14.837*** (-15.44)	-298.475*** (-6.63)	-2.419 (-1.14)
Year_FE	Yes	Yes	Yes
Discipline_FE	Yes	Yes	Yes
Funding_FE	Yes	Yes	Yes
Institution_FE	Yes	Yes	Yes
Adj. R <sup>2</sup>	0.351	0.214	0.255
N	42996	42996	42996

Note: All regressions include project start year, discipline type, funding type, and host institution fixed effects. Standard errors are clustered by institution. T statistic is presented in parenthesis.  
\*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$ .

**Table 7 Science field, collaboration networks, and research output.**

Dependent variable	TNum	TQuo	Vrs
	(1)	(2)	(3)
Cen	0.017*** (3.51)	0.764*** (3.26)	0.025*** (3.27)
Natu	3.866*** (7.72)	221.906*** (11.33)	4.946*** (6.75)
Cen*Natu	0.015* (1.94)	1.149*** (3.25)	0.024** (2.01)
Tit	3.930*** (17.97)	167.619*** (15.78)	6.899*** (18.71)
Fund	0.070*** (7.53)	2.338*** (5.94)	0.075*** (6.61)
Time	3.082** (8.86)	38.456** (2.12)	-1.125* (-1.86)
WHum	0.787*** (14.67)	18.197*** (8.26)	0.586*** (7.78)
Constant	-20.213*** (-14.80)	-668.422*** (-10.05)	-5.658** (-2.48)
Year_FE	Yes	Yes	Yes
Funding_FE	Yes	Yes	Yes
Institution_FE	Yes	Yes	Yes
Adj. R <sup>2</sup>	0.363	0.232	0.245
N	15962	15962	15962

Note: All regressions include project start year, funding type, and host institution fixed effects. Standard errors are clustered by institution. T statistic is presented in parenthesis. \*\*\*p < 0.01, \*\*p < 0.05, \*p < 0.1. Discipline type fixed effect and variable Natu have complete collinearity, so it is not controlled here.

**Organizational and environmental factors.** The project's scientific field is the first organizational and environmental factor contributing to heterogeneity. Natural and social sciences research differs significantly regarding resource requirements and evaluation criteria. Natural science research often involves high-cost resources, such as expensive instruments and large quantities of consumables (Benz & Rossier, 2022), making the resource-sharing effect of collaboration networks particularly important. These networks enable researchers to share costly equipment, improving resource efficiency. Additionally, natural science achievements are evaluated more objectively, facilitating smoother knowledge exchange. In contrast, social science research requires fewer resources, and the resource-sharing effect is less pronounced. Moreover, the more diverse evaluation criteria in social sciences can lead to debates, hindering communication efficiency and research progress. We construct a dummy variable, Natu, to examine how the scientific field influences the relationship between collaboration networks and research output. When project *i* is a chemistry project, it takes the value of 1, representing natural sciences. When project *i* is an economics and management project, it takes the value of 0, representing social sciences.

The results in Table 7 show that CPE collaboration networks have a stronger impact on research output in natural science projects. In column 1, the positive coefficients of Cen and the interaction term Cen\*Expert indicate that the effect of PI centrality on the number of project papers is more pronounced in natural science projects than in social science projects. In column 2, a similar pattern is observed for project citations, with centrality having a greater impact on natural science projects. Column 3 confirms that PI centrality more significantly promotes project efficiency in natural science projects.

The second organizational and environmental factor contributing to heterogeneity analyzed in this paper is the level of the

**Table 8 Institution level, collaboration networks, and research output.**

Dependent variable	TNum	TQuo	Vrs
	(1)	(2)	(3)
Cen	0.033*** (6.30)	1.251*** (7.40)	0.039*** (5.96)
Top	-0.556*** (-2.64)	41.389*** (7.33)	-0.436* (-1.71)
Cen*Top	-0.024*** (-3.21)	-0.169 (-0.61)	-0.018* (-1.88)
Tit	4.305*** (32.34)	139.570*** (20.01)	6.482*** (29.38)
Fund	0.081*** (10.20)	1.930*** (7.79)	0.086*** (9.61)
Time	2.454*** (8.33)	22.673* (1.74)	-0.410 (-0.71)
WHum	0.575*** (16.64)	1.406 (1.09)	0.247*** (5.49)
Constant	-14.369*** (-14.70)	-313.280*** (-7.64)	-1.469 (-0.74)
Year_FE	Yes	Yes	Yes
Discipline_FE	Yes	Yes	Yes
Funding_FE	Yes	Yes	Yes
Adj. R <sup>2</sup>	0.307	0.185	0.218
N	43264	43264	43264

All regressions include project start year, discipline type, funding type fixed effects. Standard errors are clustered by institution. T statistic is presented in parenthesis. \*\*\*p < 0.01, \*\*p < 0.05, \*p < 0.1. Host institution fixed effect and Top have complete collinearity, so it is not controlled here.

host institution. Institutions differ significantly in research resources and researchers' capabilities. In China, top-tier institutions like 985 project<sup>5</sup> universities and the Chinese Academy of Sciences<sup>6</sup> receive substantial government funding and have more stringent recruitment standards, resulting in researchers with higher qualifications and stronger research capabilities (Zong & Zhang, 2019). As a result, researchers in these institutions have less need to search for resources or build collaboration networks, as their institutions already provide ample resources. In contrast, researchers in general institutions, facing resource constraints and lower capabilities, rely more on collaboration networks to access resources and connect with proficient collaborators. These networks' collaboration norms and sense of identity also motivate researchers to invest more in their scientific work. To examine the impact of institutional level on the relationship between collaboration networks and research output, we introduce the Top variable, assigning a value of 1 when the project's host institution belongs to 985 project universities or the Chinese Academy of Sciences and 0 otherwise.

The results in Table 8 show that CPE collaboration networks have a weaker impact on research output in top institutions. In column 1, the positive coefficient of Cen and the negative coefficient of the interaction term between Cen and Top suggest that PI centrality has a weaker effect on project papers in top institutions compared to general institutions. In column 3, the positive effect of PI centrality on project efficiency is weaker in top institutions, as indicated by the negative interaction term between Cen and Top.

### Mechanism analysis and further studies

**Tie strength.** Within the network, nodes form both strong and weak ties. Nodes with strong ties exhibit higher interaction frequency and engage in more profound collaborations. Additionally, strong ties foster trust and reciprocity, which mitigate the

risks and costs associated with new collaborations, thereby facilitating the exchange and interaction of knowledge (Liu et al., 2023). On the other hand, weak ties offer greater collaboration diversification, serving as bridges that connect diverse groups within the network and foster the integration of different fields (Lyu et al., 2019). While weak ties may facilitate the cross-pollination of ideas across fields, they encounter challenges during actual implementation. Scientific research demands a high knowledge density and requires frequent discussions and communication during collaborations. Therefore, strong ties play a crucial role and significantly contribute to the development of subsequent projects. To investigate the impact of tie strength on research output within collaboration networks, we computed the tie strength of PI  $p$  in year  $t$ . Specifically, we determine the tie strength by dividing the number of collaborations the PI engages in with others in the year by the total number of individuals the PI collaborates with. A higher ratio indicates a stronger relationship between the PI and others. We introduced the following model to test the mediating effect of tie strength as a mediating variable.

$$\text{TieStrength}_{p,t} = \alpha + \beta_1 \text{Cen}_{p,t} + \beta_2 \text{Tit}_{p,t} + \beta_3 \text{Fund}_{i,t} + \beta_4 \text{Time}_{i,t} + \beta_5 \text{WHum}_{i,t} + \text{fixed effects} + \varepsilon_{i,t} \quad (7)$$

$$Y_{i,t} = \alpha + \beta_1 \text{Cen}_{p,t} + \beta_2 \text{TieStrength}_{p,t} + \beta_3 \text{Tit}_{p,t} + \beta_4 \text{Fund}_{i,t} + \beta_5 \text{Time}_{i,t} + \beta_6 \text{WHum}_{i,t} + \text{fixed effects} + \varepsilon_{i,t} \quad (8)$$

The results in Table 9 indicate that PI centrality influences the number of project papers and project output efficiency through tie strength. In column 1, the coefficient of Cen is significantly positive, suggesting that PI centrality has a significant positive effect on tie strength. In columns 2–4, TieStrength is added as an independent variable. The coefficients of PI centrality remain significantly positive but are reduced compared to the baseline regression results in Table 2. However, the coefficient of TieStrength is significantly positive only in columns 2 and 4, and only these columns pass the Sobel test, indicating that tie strength mediates the relationship between PI centrality and the number of project papers and project output efficiency. However, it does not mediate the relationship between PI centrality and the citations of project papers, which could be because citations rely more on factors such as the paper's innovation, cutting-edge research, and the attention drawn to the research topic, which may not be directly related to tie strength.

We hypothesized that the effect of tie strength on the relationship between PI centrality and project output varies significantly with team size. In large teams, PIs with high tie strength and centrality often take on more communication and coordination tasks, leading other members to become overly dependent on the leader. This dependence can stifle the independence and innovation of team members, ultimately reducing the team's overall innovation capacity. Additionally, PIs with high tie strength and centrality typically serve as information hubs in larger teams. They are more susceptible to information overload, which can heighten the risk of conflicts during decision-making, exacerbate differences in opinion, and create communication barriers. These factors collectively lead to research output falling short of expectations. To test this inference, we divided the sample into larger and smaller groups based on team size and introduced an empirical model with an interaction term between centrality and tie strength to conduct group testing. When the size of the project team exceeds the annual median, it is classified into the larger team group.

**Table 9 Mechanism analysis: tie strength.**

Dependent variable	TieStrength	Tnum	Tquo	Vrs
	(1)	(2)	(3)	(4)
Cen	0.003*** (47.76)	0.016*** (5.99)	1.027*** (8.79)	0.023*** (6.19)
TieStrength		0.774*** (2.90)	8.725 (0.87)	0.997*** (2.73)
Tit	-0.015*** (-3.93)	4.087*** (32.11)	138.978*** (19.46)	6.336*** (28.59)
Fund	0.000** (2.05)	0.084*** (10.93)	1.863*** (7.45)	0.089*** (10.25)
Time	-0.005 (-0.96)	2.533*** (8.95)	22.290 (1.64)	-0.350 (-0.59)
WHum	0.003*** (5.64)	0.570*** (16.66)	4.312*** (3.36)	0.281*** (6.30)
Constant	1.033*** (49.82)	-15.484*** (-15.61)	-305.141*** (-6.81)	-3.259 (-1.47)
Year_FE	Yes	Yes	Yes	Yes
Discipline_FE	Yes	Yes	Yes	Yes
Funding_FE	Yes	Yes	Yes	Yes
Institution_FE	Yes	Yes	Yes	Yes
Sobel test P value		0.004	0.391	0.007
Adj. R <sup>2</sup>	0.134	0.351	0.214	0.254
N	42996	42996	42996	42996

All regressions include project start year, discipline type, funding type, and host institution fixed effects. Standard errors are clustered by institution. T statistic is presented in parenthesis.

\*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$ .

Otherwise, it is categorized into the smaller team group.

$$Y_{i,t} = \alpha + \beta_1 \text{Cen}_{p,t} + \beta_2 \text{TieStrength}_{p,t} + \beta_3 \text{Cen}_{p,t} * \text{TieStrength}_{p,t} + \beta_4 \text{Tit}_{p,t} + \beta_5 \text{Fund}_{i,t} + \beta_6 \text{Time}_{i,t} + \beta_7 \text{WHum}_{i,t} + \text{fixed effects} + \varepsilon_{i,t} \quad (9)$$

The results in Table 10 show that when team size is large, the interaction between PI tie strength and PI centrality negatively impacts research output, consistent with our inference. In column 1, the interaction term ( $\text{Cen} * \text{TieStrength}$ ) coefficient is significantly negative, while it is not significant in column 2. The  $p$  value of the coefficient difference test between groups is 0.010, indicating that the coefficient difference is significant. Therefore, columns 1 and 2 suggest that when team size is large, higher tie strength interacting with centrality may negatively affect the number of project papers. The coefficient of the interaction term in column 3 is significantly negative at the 10% level, whereas it is not significant in column 4. The  $p$  value of the coefficient difference test between groups is 0.080, indicating that the coefficient difference is significant. Thus, columns 3 and 4 indicate that when team size is large, higher tie strength interacting with centrality may have a negative impact on the citations of project papers. In column 5, the coefficient of the interaction term is significantly negative, while it is not significant in column 6. The  $p$  value of the coefficient difference test between groups is 0.000, confirming that the coefficient difference is significant. Therefore, columns 5 and 6 demonstrate that higher tie strength and centrality may negatively affect project output efficiency when team size is large.

**Team size.** The collaboration network is closely tied to the size of the project team. In general, researchers with higher centrality

**Table 10 The impact of the interaction between tie strength and centrality on research output under different team sizes.**

Group	Larger team	Smaller team	Larger team	Smaller team	Larger team	Smaller team
Dependent variable	TNum	TNum	TQuo	TQuo	Vrs	Vrs
	(1)	(2)	(3)	(4)	(5)	(6)
Cen	0.055*** (3.85)	-0.004 (-0.36)	2.153** (4.27)	0.983** (2.26)	0.088*** (4.98)	0.005 (0.25)
TieStrength	2.138*** (2.61)	0.288 (0.47)	47.203* (1.77)	21.514 (0.98)	3.425*** (3.35)	0.570 (0.61)
Cen*TieStrength	-0.028** (-2.19)	0.013 (1.22)	-0.747* (-1.72)	-0.228 (-0.60)	-0.049*** (-3.15)	0.011 (0.70)
Tit	4.789*** (22.30)	3.933*** (25.77)	117.372*** (15.70)	155.978*** (17.26)	6.342*** (22.92)	6.682** (22.56)
Fund	0.081*** (10.39)	0.063*** (4.17)	1.985*** (7.40)	0.897 (1.48)	0.090*** (9.09)	0.103*** (3.45)
Time	3.078*** (7.54)	2.440*** (5.17)	27.217* (1.66)	42.433** (2.25)	-0.182 (-0.30)	-1.499 (-1.08)
WHum	0.600*** (9.55)	0.546*** (10.54)	3.437* (1.68)	7.155*** (3.00)	0.605*** (8.01)	-0.401*** (-4.91)
Constant	-21.392*** (-10.97)	-13.058*** (-8.31)	-329.272*** (-4.56)	-388.222*** (-5.94)	-11.813*** (-4.41)	6.718 (1.61)
Year_FE	Yes	Yes	Yes	Yes	Yes	Yes
Discipline_FE	Yes	Yes	Yes	Yes	Yes	Yes
Funding_FE	Yes	Yes	Yes	Yes	Yes	Yes
Institution_FE	Yes	Yes	Yes	Yes	Yes	Yes
Empirical P Value	0.010		0.080		0.000	
Adj. R <sup>2</sup>	0.373	0.203	0.280	0.140	0.324	0.135
N	20440	22348	20440	22348	20440	22348

All regressions include project start year, discipline type, funding type, and host institution fixed effects. Standard errors are clustered by institution. T statistic is presented in parenthesis. \*\*\*p < 0.01, \*\*p < 0.05, \*p < 0.1.

have greater access to collaborations, increasing the likelihood of selecting suitable partners (Yuan & Van Knippenberg, 2022). Increasing team size can further alleviate individual workloads and expedite research output. Additionally, the continuous inclusion of new members can enhance the diversity of team members' backgrounds, thereby facilitating the exploration of new research directions. However, the increase in team size also brings disadvantages, including increased management complexity and the potential for exacerbated free-riding behavior among some members. Nevertheless, scientific research teams typically have a relatively flat organizational structure, and the associated increase in management complexity resulting from an expanded team size is relatively minimal. Furthermore, once an article is published, it becomes a lifelong accomplishment for the researcher. Due to reputational considerations, most researchers handle instances of free-riding behavior by others with care. To investigate the impact of collaboration networks on research output through team size, we consider the number of project members as the dependent variable. It should be noted here that the issue of reverse causality between PI centrality and team size is relatively minor. On the one hand, when PIs can engage in multiple projects within the same year, their network centrality is less influenced by team size. On the other hand, when constructing the collaboration network for each year, we incorporate not only each PI's project experience from the current year but also their experience from previous years, which further mitigates the impact of team size on the collaboration network. We introduced the following model to test the mediating effect of team size as a mediating variable.

$$WHum_{i,t} = \alpha + \beta_1 Cen_{p,t} + \beta_2 Tit_{p,t} + \beta_3 Fund_{i,t} + \beta_4 Time_{i,t} + \text{fixed effects} + \varepsilon_{i,t} \quad (10)$$

$$Y_{i,t} = \alpha + \beta_1 Cen_{p,t} + \beta_2 WHum_{i,t} + \beta_3 Tit_{p,t} + \beta_4 Fund_{i,t} + \beta_5 Time_{i,t} + \text{fixed effects} + \varepsilon_{i,t} \quad (11)$$

The results in Table 11 demonstrate that PI centrality influences research output through team size. The coefficient of Cen in column 1 is significantly positive at the 1% level, indicating that as PI centrality increases, team size also increases. When team size is added to the model as an independent variable, the coefficients of Cen in columns 2–4 remain significantly positive. However, compared to the centrality coefficients in Table 2, they are reduced. Additionally, the coefficients of WHum in columns 2–4 are all significantly positive, and each passes the Sobel test. These results indicate that team size mediates the relationship between PI centrality and the number of project papers, citations, and project output efficiency.

We propose the hypothesis that the impact of PI centrality on research output changes nonlinearly as team size expands. When the team size is small, increasing the team size weakens the impact of PI centrality on research output. This is because, with a smaller team, the PI can easily manage the team's overall dynamics, coordinate tasks, and foster innovation. However, as the team expands, the complexity of communication and management increases, making it difficult for the PI to effectively oversee all members, which diminishes the utility of PI centrality. Once the team reaches a certain critical size, structural and managerial changes may occur. For instance, the team might be reorganized into smaller sub-teams with clearer task allocation and role division. This allows the PI to focus on key decisions while delegating the management of routine tasks. In this scenario, the hierarchical structure enhances management and innovation, leading to a recovery, or even an increase, in the utility of PI centrality as the team continues to grow. To verify

**Table 11 Mechanism analysis: team size.**

Dependent variable	WHum	Tnum	Tquo	Vrs
	(1)	(2)	(3)	(4)
Cen	0.001*** (2.93)	0.017*** (6.56)	1.032*** (9.54)	0.026*** (7.08)
WHum		1.363*** (21.47)	22.390*** (8.38)	0.788*** (8.67)
Tit	0.210*** (12.45)	4.201*** (33.34)	137.263*** (19.44)	6.360*** (28.92)
Fund	0.009*** (4.57)	0.086*** (11.34)	1.777*** (7.09)	0.089*** (10.39)
Time	0.158** (1.99)	2.421*** (8.63)	19.525 (1.42)	-0.427 (-0.71)
Constant	6.026*** (33.61)	-17.220*** (-17.88)	-388.043*** (-8.18)	-4.160* (-1.88)
Year_FE	Yes	Yes	Yes	Yes
Discipline_FE	Yes	Yes	Yes	Yes
Funding_FE	Yes	Yes	Yes	Yes
Institution_FE	Yes	Yes	Yes	Yes
Sobel test P value		0.000	0.000	0.000
Adj. R <sup>2</sup>	0.364	0.354	0.217	0.255
N	42996	42996	42996	42996

All regressions include project start year, discipline type, funding type, and host institution fixed effects. Standard errors are clustered by institution. T statistic is presented in parenthesis.

\*\*\*p < 0.01, \*\*p < 0.05, \*p < 0.1.

this hypothesis, we construct the following model.

$$\begin{aligned}
 Y_{i,t} = \alpha + \beta_1 Cen_{p,t} + \beta_2 WHum_{i,t} + \beta_3 Cen_{p,t} * WHum_{i,t} \\
 + \beta_4 WHumSqu_{i,t} + \beta_5 Cen_{p,t} * WHumSqu_{i,t} + \beta_6 Tit_{p,t} \\
 + \beta_7 Fund_{i,t} + \beta_8 Time_{i,t} + \text{fixed effects} + \varepsilon_{i,t}
 \end{aligned} \quad (12)$$

WHumSqu represents the quadratic term of WHum, and the definitions of other variables remain consistent with the baseline model. If the coefficient of  $\beta_3$  is negative and significant, and the coefficient of  $\beta_5$  is positive and significant, then the hypothesis is supported.

The results of Table 12 demonstrate that PI centrality's effect on the number of project papers and project output efficiency follows a nonlinear pattern as team size expands. In column 1, the coefficient of the interaction term between Cen and WHum is significantly negative, while the coefficient of the interaction term between Cen and WHumSqu is significantly positive. These results support the hypothesis, indicating that when the team size is small, increasing the number of team members weakens the positive effect of PI centrality on the number of project papers. Conversely, as the team size becomes larger, the increase in team members strengthens the positive effect of PI centrality on the number of project papers.

When the team size is small, the negative effect of expanding the team may be due to increased management complexity or member laziness. In terms of management complexity, a larger team requires more communication and coordination, which may slow down information exchange and complicate decision-making. These challenges could ultimately hinder the execution efficiency and responsiveness of the project. Disagreements and conflicts between team members may also occur more frequently and become harder to manage, requiring more time and resources for resolution. On the member laziness side, individual contributions may become less visible as the team grows, weakening members' sense of responsibility and reducing work enthusiasm and productivity. Additionally, team members may rely more on others, diminishing personal initiative and work efficiency.

In column 2, the coefficient of the interaction term between Cen and WHum is negative but not significant. In contrast, the coefficient of the interaction term between Cen and WHumSqu is significantly positive. These results do not provide strong evidence that team expansion significantly impacts the relationship between PI centrality and project paper citations. In column 3, the coefficient of the interaction term between Cen and WHum is significantly negative, while the coefficient of the interaction term between Cen and WHumSqu is significantly positive. These results suggest that as team size grows, the effect of PI centrality on project output efficiency follows a nonlinear trajectory.

The following strategies can be considered to address the issues of increased management complexity and member laziness that may arise as team size expands: (i) improving communication efficiency is crucial. Clear communication mechanisms and tools should be implemented to ensure the timely transmission and feedback of information. For instance, project management software and regular team meetings, when used effectively, can assist in tracking progress and resolving issues. Team members should be encouraged to share knowledge and information actively. Internal training sessions or workshops can also be organized to enhance communication skills and overall team efficiency. (ii) Optimizing the decision-making process is essential. Decision-making responsibilities should be streamlined by reducing unnecessary approval steps and enhancing decision-making efficiency. For example, creating a decision-making group or assigning a representative responsible for decisions can

**Table 12 Centrality and research output: nonlinear changes with team size.**

Dependent variable	TNum	TQuo	Vrs
	(1)	(2)	(3)
Cen	0.039 (1.60)	-0.142 (-0.14)	0.204*** (4.09)
WHum	0.608*** (3.91)	16.671*** (3.50)	-0.957*** (-4.17)
Cen*WHum	-0.020*** (-2.97)	-0.247 (-0.86)	-0.067*** (-5.07)
WHumSqu	-0.006 (-1.17)	-0.652** (-3.93)	0.038*** (4.98)
Cen*WHumSqu	0.002*** (4.65)	0.050** (2.57)	0.005*** (6.37)
Tit	4.105*** (32.62)	138.983*** (19.41)	6.477*** (29.47)
Fund	0.083*** (11.16)	1.809*** (7.31)	0.090*** (10.09)
Time	2.456*** (8.89)	20.706 (1.52)	-0.598 (-1.03)
Constant	-13.706*** (-10.09)	-326.913*** (-6.02)	7.391*** (2.83)
Year_FE	Yes	Yes	Yes
Discipline_FE	Yes	Yes	Yes
Funding_FE	Yes	Yes	Yes
Institution_FE	Yes	Yes	Yes
Adj. R <sup>2</sup>	0.353	0.217	0.259
N	42996	42996	42996

All regressions include project start year, discipline type, funding type, and host institution fixed effects. Standard errors are clustered by institution. T statistic is presented in parenthesis.

\*\*\*p < 0.01, \*\*p < 0.05, \*p < 0.1.

contribute to reducing hierarchical complexity. A transparent decision-making process, in which team members understand the rationale and basis for decisions, will likely to foster trust and support within the team.

(iii) Minimizing team conflicts is crucial. A clear conflict resolution process should be established, and training should be provided to enhance team members' conflict management skills. Conflicts should be addressed promptly to prevent negative impacts on teamwork. When conducted consistently, regular team-building activities and periodic feedback sessions can strengthen collaboration, increase trust among members, and reduce potential friction. (iv) Enhancing team responsibility is essential. Each team member's roles and responsibilities should be clearly defined to ensure that the importance of their contributions is understood. Such clarity fosters a sense of personal accountability and encourages active participation. Additionally, effective incentives, including performance evaluations and reward systems, should be implemented to motivate team members to maintain enthusiasm and commitment to their tasks.

**Different language output.** The impact of CPE collaboration networks on research output may vary depending on the language type of the output. The articles funded by the NSFC project are predominantly in English and Chinese. To foster academic exchanges with foreign countries, research institutions often prioritize and incentivize the publication of English articles through bonuses and promotion assessments (Xu, 2020). This phenomenon may motivate researchers to increase their publication of English articles after acquiring resources through collaboration networks. Additionally, English articles have a broader readership, potentially resulting in higher citation rates. Moreover, Chinese researchers face challenges in publishing Chinese articles due to academic inbreeding, as it can be difficult to publish articles without connections to journal editors or other influential individuals. However, the situation is different for Chinese researchers when

publishing English articles. On the one hand, they have access to a wider range of English journals. On the other hand, the competition for publication is more equitable since individuals have fewer connections with foreign journals. In conclusion, the influence of collaboration networks on English research output is expected to be more pronounced. We have comprehensively analyzed why PI centrality's effect on project output is stronger for English papers. This analysis is structured around two key dimensions: research quality and innovation and research influence.

English journals often have stricter editorial and review processes, higher language standards, and more rigorous evaluation criteria, which can result in superior quality and innovation in English papers. The publication of such high-quality or innovative English articles enables researchers to establish broader international collaboration networks, enrich their research perspectives, and gain easier access to global research resources, ultimately enhancing their research output. We analyzed the number of SCI and SSCI papers published within each project to test this inference. These two databases include prominent journals worldwide, predominantly English-language publications. Based on whether the number of SCI and SSCI papers published by a project exceeds the median for all projects in a given year, we categorized the sample into two groups: one with higher quality and innovation and the other with lower quality and innovation.

The results of Table 13 demonstrate that PI centrality has a stronger positive impact on project research output in the group with higher quality and innovation, suggesting that English publications enhance the positive effect of PI centrality on research output through higher quality and stronger innovation. Columns 1 and 2 present the effects of PI centrality on the number of project papers grouped by quality and innovation. The coefficients of Cen are significantly positive across both groups, consistent with the primary findings of this study. However, the coefficient of Cen in Column 1 is larger than in Column 2, with a p value of 0.030 for the inter-group coefficient difference test, indicating that PI centrality has a stronger positive effect on the

**Table 13 Centrality and research output based on different quality and innovation.**

Group	Higher quality and innovation	Lower quality and innovation	Higher quality and innovation	Lower quality and innovation	Higher quality and innovation	Lower quality and innovation
Dependent variable	TNum	TNum	TQuo	TQuo	Vrs	Vrs
	(1)	(2)	(3)	(4)	(5)	(6)
Cen	0.017*** (4.91)	0.010*** (3.25)	1.157*** (8.70)	0.740*** (5.73)	0.024*** (5.66)	0.017*** (3.56)
Tit	3.699*** (16.08)	2.958*** (20.94)	174.644*** (13.54)	82.972*** (13.69)	6.432*** (16.15)	4.394*** (19.24)
Fund	0.079*** (9.32)	0.075*** (6.11)	2.197*** (6.79)	1.283*** (3.19)	0.086*** (8.46)	0.075*** (5.19)
Time	2.760*** (6.21)	2.108*** (5.16)	0.298 (0.01)	33.703** (2.15)	-1.457** (-1.98)	0.353 (0.43)
WHum	0.794*** (17.53)	0.489*** (13.48)	5.929*** (2.86)	5.747*** (5.20)	0.501*** (8.21)	0.232*** (4.75)
Constant	-14.236*** (-9.57)	-11.368*** (-8.75)	-283.372*** (-4.28)	-250.098*** (-4.94)	2.238 (0.87)	-1.566 (-0.57)
Year_FE	Yes	Yes	Yes	Yes	Yes	Yes
Discipline_FE	Yes	Yes	Yes	Yes	Yes	Yes
Funding_FE	Yes	Yes	Yes	Yes	Yes	Yes
Institution_FE	Yes	Yes	Yes	Yes	Yes	Yes
Empirical P Value	0.030		0.000		0.080	
Adj. R <sup>2</sup>	0.320	0.381	0.185	0.244	0.221	0.285
N	20832	22007	20832	22007	20832	22007

All regressions include project start year, discipline type, funding type, and host institution fixed effects. Standard errors are clustered by institution. T statistic is presented in parenthesis. \*\*\*p < 0.01, \*\*p < 0.05, \*p < 0.1.

**Table 14** Centrality and research output based on different output language structures.

Group	Higher English output	Lower English output	Higher English output	Lower English output	Higher English output	Lower English output
Dependent variable	TNum	TNum	TQuo	TQuo	Vrs	Vrs
	(1)	(2)	(3)	(4)	(5)	(6)
Cen	0.018*** (4.93)	0.013*** (4.10)	1.184*** (7.51)	0.654*** (6.85)	0.027*** (4.99)	0.018*** (4.48)
Tit	4.117*** (20.83)	3.594*** (20.98)	183.397*** (16.95)	69.822*** (17.41)	7.123*** (18.97)	4.835*** (22.81)
Fund	0.077*** (9.74)	0.082*** (5.86)	2.147*** (6.36)	0.122 (0.36)	0.088*** (8.70)	0.081*** (5.34)
Time	3.181*** (7.79)	2.102*** (4.58)	40.765* (1.93)	42.108*** (3.28)	-0.863 (-1.10)	0.725 (1.11)
WHum	0.707*** (14.27)	0.601*** (15.66)	10.134*** (4.61)	9.397*** (9.99)	0.349*** (5.09)	0.419*** (8.86)
Constant	-16.752*** (-12.46)	-13.669*** (-9.60)	-449.999*** (-6.79)	-282.046*** (-7.47)	-0.953 (-0.35)	-5.819*** (-2.67)
Year_FE	Yes	Yes	Yes	Yes	Yes	Yes
Discipline_FE	Yes	Yes	Yes	Yes	Yes	Yes
Funding_FE	Yes	Yes	Yes	Yes	Yes	Yes
Institution_FE	Yes	Yes	Yes	Yes	Yes	Yes
Empirical P Value	0.080		0.000		0.030	
Adj. R <sup>2</sup>	0.369	0.355	0.210	0.230	0.249	0.286
N	21431	21393	21431	21393	21431	21393

All regressions include project start year, discipline type, funding type, and host institution fixed effects. Standard errors are clustered by institution. T statistic is presented in parenthesis. \*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$ .

number of project papers in projects with more high-quality and innovative English articles. Columns 3 and 4 report the grouped regression results on the effect of PI centrality on the citations of project papers. The coefficient of Cen in Column 3 is greater than in Column 4, and the  $p$  value of the inter-group coefficient difference test is 0.000. These results indicate that PI centrality has a stronger positive effect on the citations of project papers in projects containing more high-quality and innovative English articles. Finally, Columns 5 and 6 show that the positive impact of PI centrality on project output efficiency is also stronger in projects with more high-quality and innovative English publications.

As the dominant language in international academic exchanges, English offers several advantages over other languages, such as greater international visibility, stronger academic influence, enhanced resource integration, and a more pronounced reputational effect for its publications. When PI centrality is high, the collaboration network of the PI can better capitalize on these advantages, thereby amplifying the influence of English papers. These enhanced opportunities for dissemination and higher citation rates strengthen the role of centrality in driving project output. To verify this inference, we calculated the average number of English and Chinese papers per project, which were 11.381 and 8.677, respectively, indicating that NSFC projects clearly prefer publishing in English. Then, the average citations per project were examined. English papers received an average of 290.711 citations compared to 101.806 for Chinese papers, demonstrating that English publications garner more citations and thus contribute more to personal academic influence.

Finally, we utilized the output language structure of a project to measure the influence of English output. The output language structure is measured by the proportion of English output of a project, calculated as the ratio of the average citations of the project's English papers to the average citations of its Chinese papers. If a project's proportion of English output exceeds the annual median, it is classified as having a high proportion of

English output; otherwise, it is classified as having a low proportion of English output.

The results in Table 14 show that the positive impact of PI centrality on project output is stronger when the proportion of English output in the project is higher. Columns 1 and 2 present the group test results for the impact of PI centrality on the number of project papers. The coefficient of centrality (Cen) is significantly positive in both groups, consistent with the basic regression results of this study. From an economic perspective, the coefficient for the group with a higher proportion of English output is greater than that for the group with a lower proportion. The inter-group coefficient difference test shows that the  $p$  value for the difference in Cen between columns 1 and 2 is 0.080, which is significant at the 10% level. Thus, the impact of centrality on the number of project papers is stronger in the group with a higher proportion of English output. English-language outputs have greater dissemination potential and more citation opportunities. By leveraging this advantage, PIs can further enhance their influence within the collaboration network, consolidate more resources, and increase the number of project papers.

Columns 3 and 4 report the group test results for the impact of PI centrality on the citations of project papers. The Cen coefficient in column 3 is greater than in column 4, and the  $p$  value for the inter-group coefficient difference test is 0.000, indicating that the positive impact of centrality on project paper citations is also stronger when the proportion of English output is higher. Columns 5 and 6 present the group test results for the impact of PI centrality on project output efficiency. As expected, the impact of PI centrality on output efficiency is stronger when the proportion of English output in the project is higher. The greater impact of English-language output can provide PIs with more collaboration opportunities, enhance their influence within the collaboration network, and subsequently increase both citations and the efficiency of project output.

## Conclusions

This paper investigates the relationship between CPE collaboration networks and research output. Firstly, the study reveals that a higher network centrality of the PI is associated with increased output of the science funding projects they lead, as evidenced by the number of project papers, citations of project papers, and project DEA efficiency. Secondly, the positive correlation between the PI's centrality and the project's research output is more pronounced when the PI has project experience or the project belongs to natural science research. However, it is weaker when the project relies on top research institutions or includes foreign members in the project team. Thirdly, the study finds that a higher centrality of the PI is associated with stronger average tie strength within the collaboration network and a larger project team size. Finally, the centrality of the PI has a more substantial impact on English-language output compared to Chinese-language output.

In contrast to previous literature, this paper's findings highlight the significance of CPE collaboration networks in enhancing research output. Moreover, it identifies ways to improve the promotion role of CPE collaboration networks, offering valuable insights for optimizing science funding management. This article has the following policy implications: first, enhance the transparency of information disclosure for science funding projects and provide convenience for researchers in searching for collaboration partners. The premise of establishing a CPE collaboration network relies on researchers knowing where to find suitable collaborators. Currently, the common approach is to rely on introductions through acquaintances, which somewhat restricts individual choices. Improving the transparency of project information disclosure enables researchers to broaden their channels for finding partners, thus promoting the formation of collaboration networks. Second, optimize the allocation of science funds among different institutions and project teams to minimize the negative impact of unreasonable arrangements on the promotion of CPE collaboration networks. While top institutions have abundant resources, and their researchers possess stronger abilities, collaboration networks' utility is weaker. Additionally, teams with foreign members may experience weaker network utility due to language and cultural differences. Therefore, the NSFC must balance funding allocation between top and general institutions. Furthermore, when funding teams that include international scholars, the NSFC should establish specific foreign language proficiency requirements for local members. Third, the proportion of output in different language types during the assessment should be adjusted to incentivize localization research. Since most institutions prioritize English-related publications in researcher assessments, and English-language readership is more extensive, the promotion effect of collaboration networks on English-language output is more pronounced. By guiding institutions to proportionately increase the weight of Chinese output when formulating assessment standards, the government can contribute to the localization of research to a certain extent.

The limitations of this study are acknowledged. Firstly, despite utilizing NSFC project data, this study remains exploratory. The confidence in the results would increase with access to additional details, such as the demographics of PIs. Secondly, this study constructs collaboration networks based on CPE. However, it is important to exercise caution when interpreting the findings, considering the existence of other forms of collaboration among researchers. In the future, if data on alternative forms of collaboration can be acquired, it would allow for a more comprehensive exploration of the collaboration network. Finally, the conclusions drawn in this article are specific to Chinese science funding projects, which are greatly influenced by China's

institutional background. Future research will involve comparative studies utilizing multi-country data to enhance the understanding of collaboration network impacts.

## Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Received: 1 June 2024; Accepted: 30 April 2025;

Published online: 01 December 2025

## Notes

- 1 . China science funding official website: <https://www.nsfc.gov.cn/publish/portal0/tabc535/>.
- 2 . US science funding official website: <https://dellweb.bfa.nsf.gov/AwdLst2/default.asp>.
- 3 . The data comes from NSFC official website: <https://www.nsfc.gov.cn/>.
- 4 . The NSFC's regulations on application project limits: <https://www.nsfc.gov.cn/publish/portal0/wd/01-06/>.
- 5 . In May 1998, the Chinese government announced its plan to establish several world-class universities in the 21st century. Subsequently, a total of 39 colleges and universities were selected for funding under this project, and the government allocates substantial funds to these institutions annually to support their development.
- 6 . The Chinese Academy of Sciences was established in November 1949 and stands as the premier academic institution for natural sciences in China. It comprises 11 branches and over 100 scientific research institutes, with an annual budget exceeding 100 billion Chinese yuan.

## References

- Aldieri L, Kotsemir M, Vinci CP (2018) The impact of research collaboration on academic performance: An empirical analysis for some european countries. *Socio-Econ Plan Sci* 62:13–30
- Ali SN, Miller DA (2016) Ostracism and forgiveness. *Am Econ Rev* 106(8):2329–2348
- Amara N, Rhaiem M, Halilem N (2020) Assessing the research efficiency of canadian scholars in the management field: Evidence from the dea and fsqca. *J Bus Res* 115:296–306
- Anderson KA, Richards-Shubik S (2022) Collaborative production in science: An empirical analysis of coauthorships in economics. *Rev Econ Stat* 104(6):1241–1255
- Arregle JL, Hitt MA, Sirmon DG, Very P (2007) The development of organizational social capital: Attributes of family firms. *J Manag Stud* 44(1):73–95
- Banker RD, Charnes A, Cooper WW (1984) Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Manag Sci* 30(9):1078–1092
- Benz P, Rossier T (2022) Is interdisciplinarity distinctive? Scientific collaborations through research projects in natural sciences. *Soc Sci Inf* 61(1):179–214
- Borgatti SP (2005) Centrality and network flow. *Soc Netw* 27(1):55–71
- Bozeman B, Corley E (2004) Scientists' collaboration strategies: Implications for scientific and technical human capital. *Res Policy* 33(4):599–616
- Bozeman B, Fay D, Slade CP (2013) Research collaboration in universities and academic entrepreneurship: The-state-of-the-art. *J Technol Transf* 38(1):1–67
- Dahlander L, McFarland DA (2013) Ties that last: Tie formation and persistence in research collaborations over time. *Adm Sci Q* 58(1):69–110
- Fan L, Guo L, Wang X, Xu L, Liu F (2022) Does the author's collaboration mode lead to papers' different citation impacts? An empirical analysis based on propensity score matching. *J Informetr* 16(4):101350
- Filieri R, McNally RC, O'Dwyer M, O'Malley L (2014) Structural social capital evolution and knowledge transfer: Evidence from an irish pharmaceutical network. *Ind Mark Manag* 43(3):429–440
- Fogel K, Ma L, Morck R (2021) Powerful independent directors. *Financial Manag* 50(4):935–983
- Freeman LC (1978) Centrality in social networks conceptual clarification. *Soc Netw* 1(3):215–239
- Granovetter MS (1973) The strength of weak ties. *Am J Sociol* 78(6):1360–1380
- Guan J, Zhang J, Yan Y (2015) The impact of multilevel networks on innovation. *Res Policy* 44(3):545–559
- He ZL, Geng XS, Campbell-Hunt C (2009) Research collaboration and research output: A longitudinal study of 65 biomedical scientists in a new zealand university. *Res Policy* 38(2):306–317
- Hillman AJ, Withers MC, Collins BJ (2009) Resource dependence theory: A review. *J Manag* 35(6):1404–1427

- Hoekman J, Frenken K, Tijssen RJW (2010) Research collaboration at a distance: Changing spatial patterns of scientific collaboration within europe. *Res Policy* 39(5):662–673
- Horta H, Veloso FM, Grediaga R (2010) Navel gazing: Academic inbreeding and scientific productivity. *Manag Sci* 56(3):414–429
- Hu AG (2020) Public funding and the ascent of chinese science: Evidence from the national natural science foundation of china. *Res Policy* 49(5):103983
- Inkpen AC, Tsang EW (2005) Social capital, networks, and knowledge transfer. *Acad Manag Rev* 30(1):146–165
- Inkpen AC, Tsang EW (2016) Reflections on the 2015 decade award—social capital, networks, and knowledge transfer: An emergent stream of research. *Acad Manag Rev* 41(4):573–588
- Ji P, Jin J, Ke ZT, Li W (2022) Co-citation and co-authorship networks of statisticians. *J Bus Economic Stat* 40(2):469–485
- Kardes, H, Sevincer, A, Gunes, MH, & Yuksel, M (2014). Complex network analysis of research funding: A case study of nsf grants. *State Applications Social Network Analysis*, 163–187
- Katz JS, Martin BR (1997) What is research collaboration? *Res Policy* 26(1):1–18
- Kim E, Kim ES (2020) A critical examination of international research conducted by north korean authors: Increasing trends of collaborative research between china and north korea. *Scientometrics* 124(1):429–450
- Kong X, Mao M, Jiang H, Yu S, Wan L (2019) How does collaboration affect researchers' positions in co-authorship networks? *J Informatr* 13(3):887–900
- Krishen AS, Dwivedi YK, Bindu N, Kumar KS (2021) A broad overview of interactive digital marketing: A bibliometric network analysis. *J Bus Res* 131:183–195
- Kwiek M (2020) Internationalists and locals: International research collaboration in a resource-poor system. *Scientometrics* 124(1):57–105
- Leydesdorff L, Bornmann L, Wagner CS (2019) The relative influences of government funding and international collaboration on citation impact. *J Assoc Inf Sci Technol* 70(2):198–201
- Li EY, Liao CH, Yen HR (2013) Co-authorship networks and research impact: A social capital perspective. *Res Policy* 42(9):1515–1530
- Liang X, Liu AM (2018) The evolution of government sponsored collaboration network and its impact on innovation: A bibliometric analysis in the chinese solar pv sector. *Res Policy* 47(7):1295–1308
- Liu J, Guo X, Xu S, Zhang Y (2023) Quantifying the impact of strong ties in international scientific research collaboration. *PLoS one* 18(1):e0280521
- Lyu L, Wu W, Hu H, Huang R (2019) An evolving regional innovation network: Collaboration among industry, university, and research institution in china's first technology hub. *J Technol Transf* 44:659–680
- Marginson S (2022) All things are in flux: China in global science. *High Educ* 83(4):881–910
- Nahapiet J, Ghoshal S (1998) Social capital, intellectual capital, and the organizational advantage. *Acad Manag Rev* 23(2):242–266
- Sinnewe E, Charles MB, Keast R (2016) Australia's cooperative research centre program: A transaction cost theory perspective. *Res Policy* 45(1):195–204
- Smith TB, Vacca R, Krenz T, McCarty C (2021) Great minds think alike, or do they often differ? Research topic overlap and the formation of scientific teams. *J Informatr* 15(1):101104
- Tsai Y-H, Ma H-C, Lin C-P, Chiu C-K, Chen S-C (2014) Group social capital in virtual teaming contexts: A moderating role of positive affective tone in knowledge sharing. *Technol Forecast Soc Change* 86:13–20
- van Rijnsoever FJ, Hessel LK (2011) Factors associated with disciplinary and interdisciplinary research collaboration. *Res Policy* 40(3):463–472
- Wang J (2016) Knowledge creation in collaboration networks: Effects of tie configuration. *Res Policy* 45(1):68–80
- Xu J, Chau M, Tan BC (2014) The development of social capital in the collaboration network of information systems scholars. *J Assoc Inf Syst* 15(12):3
- Xu X (2020) China 'goes out' in a centre-periphery world: Incentivizing international publications in the humanities and social sciences. *High Educ* 80(1):157–172
- Yu N, Dong Y, De Jong M (2022) A helping hand from the government? How public research funding affects academic output in less-prestigious universities in china. *Res Policy* 51(10):104591
- Yuan, Y, & Van Knippenberg, D (2022). Leader network centrality and team performance: Team size as moderator and collaboration as mediator. *J Business Psychol*, 1–14
- Zhang C, Guo J (2017) China's international research collaboration: Evidence from a panel gravity model. *Scientometrics* 113:1129–1139
- Zong X, Zhang W (2019) Establishing world-class universities in china: Deploying a quasi-experimental design to evaluate the net effects of project 985. *Stud High Educ* 44(3):417–431

## Acknowledgements

This work was supported by the National Social Science Fund of China (Grant No. 24BGL203).

## Author contributions

First Author: Conceptualization, data curation, writing-original draft, and editing. Second Author: Conceptualization, supervision, writing-review, and editing.

## Competing interests

The authors declare no competing interests.

## Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

## Informed consent

This article does not contain any studies with human participants performed by any of the authors.

## Additional information

**Correspondence** and requests for materials should be addressed to Kaiming Li.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025