Liao et al. Health Research Policy and Systems (2024) 22:96
https://doi.org/10.1186/s12961-024-01205-8

Health Research Policy and Systems

RESEARCH Open Access

Factors influencing the research impact in cancer research: a collaboration and knowledge network analysis

Shuang Liao[1*], Christopher Lavender[1] and Huiwen Zhai[1]

Abstract

Background Cancer is a major public health challenge globally. However, little is known about the evolution patterns of cancer research communities and the influencing factors of their research capacity and impact, which is affected not only by the social networks established through research collaboration but also by the knowledge networks in which the research projects are embedded.

Methods The focus of this study was narrowed to a specific topic – 'synthetic lethality' – in cancer research. This field has seen vibrant growth and multidisciplinary collaboration in the past decade. Multi-level collaboration and knowledge networks were established and analysed on the basis of bibliometric data from 'synthetic lethality'-related cancer research papers. Negative binomial regression analysis was further applied to explore how node attributes within these networks, along with other potential factors, affected paper citations, which are widely accepted as proxies for assessing research capacity and impact.

Results Our study revealed that the synthetic lethality-based cancer research field is characterized by a knowledge network with high integration, alongside a collaboration network exhibiting some clustering. We found significant correlations between certain factors and citation counts. Specifically, a leading status within the nation-level international collaboration network and industry involvement were both found to be significantly related to higher citations. In the individual-level collaboration networks, lead authors' degree centrality has an inverted U-shaped relationship with citations, while their structural holes exhibit a positive and significant effect. Within the knowledge network, however, only measures of structural holes have a positive and significant effect on the number of citations.

Conclusions To enhance cancer research capacity and impact, non-leading countries should take measures to enhance their international collaboration status. For early career researchers, increasing the number of collaborators seems to be more effective. University–industry cooperation should also be encouraged, enhancing the integration of human resources, technology, funding, research platforms and medical resources. Insights gained through this study also provide recommendations to researchers or administrators in designing future research directions from a knowledge network perspective. Focusing on unique issues especially interdisciplinary fields will improve output and influence their research work.

Keywords Research capacity and impact, Collaboration network, Knowledge network, Paper citations, Synthetic lethality, International collaboration, Cancer research

*Correspondence:
Shuang Liao
liaoshuang@sysucc.org.cn
Full list of author information is available at the end of the article

Background

Cancer is a major public health challenge globally [1]. According to WHO in 2019, cancer is estimated to be one of the top two leading causes of death in 112 out of 183 countries. To address this challenge, plenty of efforts have been made to accelerate progress and output in various field of cancer research, including cancer aetiology, vaccines and anti-cancer drug development, precision cancer medicine and treatment strategy innovations. However, little is known about the evolution patterns of cancer research communities and the influencing factors of their research capacity and impact, which is vital for establishing the reputation of cancer researchers and advancing their careers.

Although tumour mutational analysis has largely been exhausted for the identification of conventionally druggable targets, functional genomic screening based on synthetic lethality provides new avenues for discovering drug targets that were previously considered undruggable owing to their molecular structure or resulting functional loss [2]. In this study, we sought to reveal the factors that influence cancer research capacity, impact and the underlying logic through social network (research collaboration network) and knowledge network analysis based on synthetic lethality-related cancer research papers.

We focused our study in this area for several reasons. First, recent advances, such as CRISPR-based gene editing promoted the systematic screening for synthetic lethality-based cancer drug targets, resulting in unprecedented growth in this area. Second, an increasing number of researchers are contributing a wealth of knowledge elements in this area, along with the emergence of research achievements with high clinical value, such as the clinical use of PARP inhibitors in patients with BRCA mutant ovarian cancer. Third, there is a high level of collaboration in this field, including institutional, international and collaborations between academic institutions and industries, such as big pharmaceutical companies.

In the era of globalization, collaboration is the key to success in the fight against cancer, especially in the field of cancer research. Collaboration has been known to have an important influence in research capacity, output and impact [3]. Social networks, such as research collaboration networks, reflect the relationships and interactions among different levels of agents within the network – individuals, organizations or countries [4]. Their collaborative relationships serve as social capital [5]. According Li et al. [4], betweenness centrality, one of the structural social capital indicators, plays the most important role in leveraging resources in a co-authorship network. The positions of agents in the research collaboration network are vital for gaining and exchanging resources, ideas, knowledge and information from collaborating partners, thereby affecting their performance and innovative outputs [6]. Social network analysis is widely used to reveal the pattern of research collaboration and its impact on research citations.

Besides social capital, knowledge is also a vital resource for individuals and organizations in establishing competitiveness during innovation [7]. In addition to social networks, individuals and research organizations are also embedded in knowledge networks formed by the coupling of knowledge elements [8]. Formed through the combination of knowledge elements during the innovation process, knowledge networks serve as conduits for knowledge flow and influence future searches, the recombination of knowledge elements and, ultimately, the outcomes of innovation [9]. Therefore, the positions of individuals' or organizations' knowledge elements within the knowledge network will also impact their innovation outcomes.

In the fields of nano-energy and wind energy, studies have been conducted on the structural properties of both collaborative and knowledge networks, confirming the impact of node attributes in both networks on innovation performance, such as paper citation counts and patent counts [8, 10]. To the best of our knowledge, no studies have been conducted on the structural properties of both social and knowledge networks as well as their impact on research capacity and performance in the field of cancer research.

In our study, negative binomial regression analysis was applied to explore how collaboration and knowledge networks, together with other potential factors, affected the research capacity and impact in the field of 'synthetic lethality'-related cancer research. Regression model analysis was performed at the paper level, with citation number as the dependent variable. As greater research capacity and impact naturally leads to increased publication citations, citation-based measures are widely accepted as proxies for assessing the research capacity and impact of academic researchers [10, 11]. The positive influence of international collaboration on the research impact of publications has been documented in former studies; although, the effects may vary across disciplines and countries [12, 13]. We assume that the status of papers' corresponding countries within the international collaboration network may also influence their research impact in general, as cooperation between scholars from specific countries may exhibit some clustering in the field of cancer research. Therefore, instead of simply categorizing the papers into 'with' or 'without' international collaboration, we further included the 'country level international collaboration status' as a potential influencing factor, based on metrics from country level collaboration networks.

In the preliminary analysis, we also established institutional level collaboration networks too, but they failed to reveal any significant correlation between their node attributes and paper citations. Then, we analysed the impact of institution types on citations and found that the existence of a company within the affiliation list was significantly related to citations, with $P < 0.01$. Therefore, industry involvement was further included in our final model.

Node attributes in both social and knowledge networks, as well as other potential factors, including journal impact factors (IF), country level international collaboration status, publication duration and industry involvement, were taken as the independent variables. Some of these factors have been proposed in previous studies as relevant to the citation impact of papers [12, 14, 15].

In the collaboration network, degree centrality indicated the number of direct ties (partners) one researcher had. A structural hole was originally considered to be a lack of linkage between any pair of nodes in the network [16]. Specifically, an author occupies structural holes in the collaboration/social network if they connect with other collaborators who are not connected themselves [17]. Therefore, structural hole values in the co-authoring networks in this study illustrate the degree to which an author's partners are disconnected from each other, indicating non-redundant and efficient access to information for the focal author. Both metrics reflect an author's degree of prestige among their collaboration partners for efficient access of information, knowledge and resources [18]. We reasoned that the average node attributes of a paper's 'research guarantors' within the collaboration network would affect their research capacity and impact, making it an important factor for regression analysis.

Similarly, in the knowledge network, degree centrality indicates the number of direct linkages of a certain knowledge element, while structural holes represent the degree of disconnectedness among elements linked to a focal knowledge element, indicating information control advantages of the latter as researchers who study these uncombined elements are likely to use information in the focal element [19]. Besides, surprising combinations of research content from distant disciplines are also revealed as related to impactful research [20]. Differences in node attributes among knowledge elements of a certain paper reflect the author's choice of research topic. A project with a frontier, multi-disciplinary research topic might have a higher measure in structural holes, while an ordinary and popular research topic might increase the measure of degree centrality.

For both networks, only local metrics such as degree centrality and structural holes were included. Global-level centrality metrics, such as closeness and betweenness centrality, were not considered to avoid a possible suppressor effect, as revealed in previous studies [4, 21].

Methods

Data collection

The bibliographic data used in this study were extracted in March 2023 from Web of Science (WoS) and Journal Citation Report (JCR) databases. The following search terms were used in the WoS Core Collection to retrieve records of synthetic lethality-based cancer research papers:

1: (((((((TS = (cancer*)) OR TS = (tumor*)) OR TS = (tumour*)) OR TS = (carcinoma*)) OR TS = (neoplasm*)) OR TS = (Glioblastoma)) OR TS = (Melanoma)) OR TS = (Adenocarcinoma).
2: TS = ("Synthetic lethal*").
3: #2 AND #1
4: TS = ("correction:").
5: (#3) NOT #4

Only documents classified as 'articles' were included. A total of 2074 papers were collected from WoS, along with the title, journal, abstract, authors and their affiliations, keywords, publication year, citation count and other information. The 5-year impact factors of the journals were downloaded from the Journal Citation Report (JCR) database. For all the journals listed in the dataset, the 5-year impact factors for the corresponding year were added manually.

Measurements

Collaboration and knowledge networks were created on the basis of 5-year moving windows (2009–2013, 2010–2014, 2011–2015, 2012–2016, 2013–2017), following previous approaches [10, 22]. For example, if a paper was published in 2014, the metrics of its author and keywords were measured within the collaboration and knowledge networks for the years 2010–2014.

Dependent variables

The dependent variable was the number of citations for each paper in the sample, which was directly retrieved from the WoS dataset.

Independent variables

The Independent variables included degree centrality and structural holes in the collaboration and knowledge networks.

Construction of the collaboration and knowledge networks In this study, the collaborative and knowledge networks were constructed on the basis of paper-level data following previous approaches [10].

The collaboration networks were established on the basis of co-author data of the papers included in each 5-year time frame, with authors serving as nodes and co-authorship experiences as the ties. Authors who co-published a paper were considered collaboration partners. Sci2 software was used for data cleaning and constructing the collaboration (co-author) network [23]. Manual unification of author names was conducted using the merge table function of Sci2 before the final construction of the network.

In addition to the individual level, a national-level international collaboration network was further constructed on the basis of bibliometric data to classify different countries based on their node attributes within the network.

Based on the papers included in each 5-year time frame, co-keyword networks were established in the study as the knowledge networks, with article keywords suggested by Web of Science (ISI keywords) considered as the knowledge elements and their co-appearance in the same paper served as a tie between them. Manual unification of keywords, including synonyms, singular and plural forms, abbreviations and so on, was performed on the extracted file before the final network construction.

Measurement of degree centrality and structural holes For both the collaboration and knowledge networks, the measurement of degree centrality and structural holes followed previous approaches as adopted by Guan et al. [10]. The normalized degree centrality proposed by Freeman [24] employed in previous studies [10, 21, 25] was utilized to account for the size effect of different networks.

For the calculation of structural holes, we used Burt's constraint measure to determine the network constraint C_i. This measure indicates how strongly i can be constrained by its neighbours (Burt, 2009, 2004) [19, 26]. Subsequently, the network constraint measure (C_i) was subtracted from two to indicate the control advantage of each node in spanning structural holes, as based on previous studies [10, 25]. Both metrics, degree centrality and structural holes, were calculated using Pajek software [27].

$$\begin{aligned} Structural\ holes_i &= 2 - C_i \\ &= 2 - \sum\nolimits_j \Big(p_{ij} + \sum_{k,k\neq i,k\neq j} p_{ik}p_{kj}\Big)^2 \end{aligned}$$

In this formula, i is the focal element, p_{ij} indicates the proportion at which an element j accounts for element i's contacts. For instance, if i connects with j and five other elements, then p_{ij} is 1/6. Additionally, k is the third element which connects with both i and j. Therefore, the focal element i would have lower p_{ij} values if it connects with more elements, thereby being less constrained.

Aggregating the measures into paper level Since this study was conducted at the paper level, we needed to aggregate the degree centrality and structural holes values for each publication. For the knowledge network, the values of degree centrality and structural holes of all the keywords of a specific paper were averaged to obtain its paper-level measure. However, in the collaboration network, two different approaches were initially adopted for data aggregation. The first approach, termed the 'lead author' approach, entailed averaging only the measures of the first author and the corresponding/co-corresponding authors for each paper. The other approach, termed the 'all-author' approach, involved averaging the measures of all authors for the regression analysis. The 'lead author' approach is based on the academic consensus regarding the order of authorship in the field of biomedical research and the concept of a 'research guarantor'. We assume that the first and corresponding authors contribute the most to these papers, and therefore, their node attributes might be more relevant to the research impact of the publications than those of the other authors. A similar approach of allocating credits to corresponding authors, institutions or countries has also been adopted and proven to be useful in previous studies [28–30]. Regression results for the two models are compared in the results section of this article.

Control variables

The following variables were controlled for in the regression analysis: a journal's 5-year impact factor, year since publication, industry involvement and status in international collaboration (for countries). Dummy variables were created for the last two variables.

To classify different countries according to their status in international collaboration, country-level collaboration networks were further created on the basis of co-author data from the papers included in each 5-year network. In these networks, countries served as nodes, and co-authorship experiences formed the ties. Degree centrality and structural holes were calculated for each country, using the same approaches adopted in individual networks.

In the preliminary analysis, no significant relationship was found between country level degree centrality/structural holes (both as continuous variables) and paper citations. This is unsurprising, as significant variations exist in the number of citations among papers from the same country. To improve statistical power, countries were subsequently ranked and classified into two categories. Leading countries were defined as those with structural holes ranked in the top 20%, while the remaining countries were categorized as non-leading countries. The 20% cutoff was selected to ensure the stability of relevant metric values for leading countries over time, as well as good discrimination power between the leading countries and others. When calculating the international collaboration status for each paper, only the corresponding country was considered Table 1.

Results

Structural characteristics and evolution patterns of collaboration and knowledge networks

The structural characteristics of the collaboration network and knowledge network in the synthetic lethality-based cancer research field are presented in Table 2 by time period and visualized in Fig. 1 and Fig. 2, to illustrate the evolution pattern and parameter changes over time. Both networks are weighted, with edge attributes also presented in Table 2.

In the knowledge network, a rapid increase in network size and edge number was observed in both period 2 and period 3 (as presented in Table 2), indicating fast growth and high integration of knowledge in this area. A decreasing trend was observed in clustering coefficients in both period 2 and period 3, revealing an evolution pattern of the knowledge network towards lower integration and clustering (as shown in Fig. 1), which may be due to the expansion of knowledge fields involved and the growing importance of interdisciplinary research.

For the collaboration network, period 2 and period 3 also saw a dramatic increase in network size and edge number, revealing fast growth of the co-authoring network. Despite a moderate increase in period 2, the typical number of co-authoring relationships (average degree) and frequency (mean weight) remained relatively stable across the three periods compared to network growth, resulting in a drop in network density from 0.009 to 0.002. High network clustering coefficients were observed in all three periods, with the highest in period 2 (0.948), confirming vigorous and tight research cooperation in this field, as shown in Fig. 2.

When examining the country-level collaboration network (Fig. 3), the USA is undoubtedly at the centre of international cooperation, taking top priority. The most robust links are found between the USA and China, the USA and England, the USA and Germany and the USA and Canada, which align precisely with previous studies on global scientific collaboration networks [32]. European countries have also developed an important

Table 1 Definitions of distinct variables

Variables	Description
Dependent variable	
Citation	Number of citations of a specific paper
Independent variables	
Degree centrality in collaboration network	The mean value of normalized degree centrality of the first and corresponding authors of a specific paper in a collaboration network
Structural holes in collaboration network	The mean value of 2-C (network constraint measure) of the first and corresponding authors of a specific paper in a collaboration network
Degree centrality in knowledge network	The mean value of normalized degree centrality of all the ISI-keywords of a specific paper in a knowledge network
Structural holes in knowledge network	The mean value of 2-C (network constraint measure) of all the ISI-keywords of a specific paper in a knowledge network
Control variables	
Journal IF	Five-year impact factors downloaded from the JCR database for each journal in the publication year
Year since publication	A continuous variable was used to indicate the time duration between paper publication and analysis (2023)
Industry involvement	The dummy variable was set to one if a company was included in the affiliation list
Status in international collaboration	The dummy variable was set to one if the value of structural holes of a country were ranked in the top 20%

Table 2 Network structural characteristics by time periods

Network parameters	Network period		
	Period 1 2008–2012	Period 2 2013–2017	Period 3 2018–2022
Knowledge network			
Network size (no. of nodes)	930	2043	2636
No. of edges	8340	23124	30194
Edge attributes			
Minimum weight	1	1	1
Maximum weight	15	32	38
Mean weight	1.138	1.222	1.274
Average degree	17.936	22.637	22.909
Network structure			
Network density	0.019	0.011	0.009
Network clustering coefficient (transitivity)	0.229	0.171	0.142
No. of isolated nodes	1	1	5
Size of the largest component	929	2037	2631
Network diameter	4	5	4
Collaboration network			
Network size (no. of nodes)	1578	5725	10221
No. of edges	11409	65316	103710
Edge attributes			
Minimum weight	1	1	1
Maximum weight	11	18	10
Mean weight	1.046	1.025	1.045
Average degree	14.460	22.818	20.294
Network structure			
Network density	0.009	0.004	0.002
Network clustering coefficient (transitivity)	0.929	0.948	0.708
No. of isolated nodes	6	8	4
Size of the largest component	361	3851	8136
Network diameter	12	15	13

cluster within the network, with high density and integration, which is unsurprising considering their scientific advancements and the short geographic distances between them.

Regression analysis and results

Table 3 displays the descriptive statistics of the variables and their correlation data.

To address the issue of over-dispersion, a negative binomial regression model was used in this study for data analysis. VIF (variance inflation factor) values of all the variables were found to be below 2, indicating the absence of significant multicollinearity. Five regression models were constructed: model 1 includes only the control variables and models 2 and 3 incorporate the independent variables from the collaboration networks based on the lead author approach and the all-author approach, respectively. Models 4 and 5 encompass all the variables, with collaboration networks based on the lead author approach and the all-author approach, respectively. Table 4 presents the regression results for the number of citations for each paper.

We first compared the regression results between different aggregation approaches for collaboration network parameters, one based on lead author data (model 2 and model 4) and one on all author data (model 3 and model 5). The regression results based on 'lead author' and 'all-author' approaches are quite consistent regarding the influence of the author's degree centrality on citations. However, a notable difference was observed in the coefficient significance for structural holes. In the lead author models, the author's structural holes value

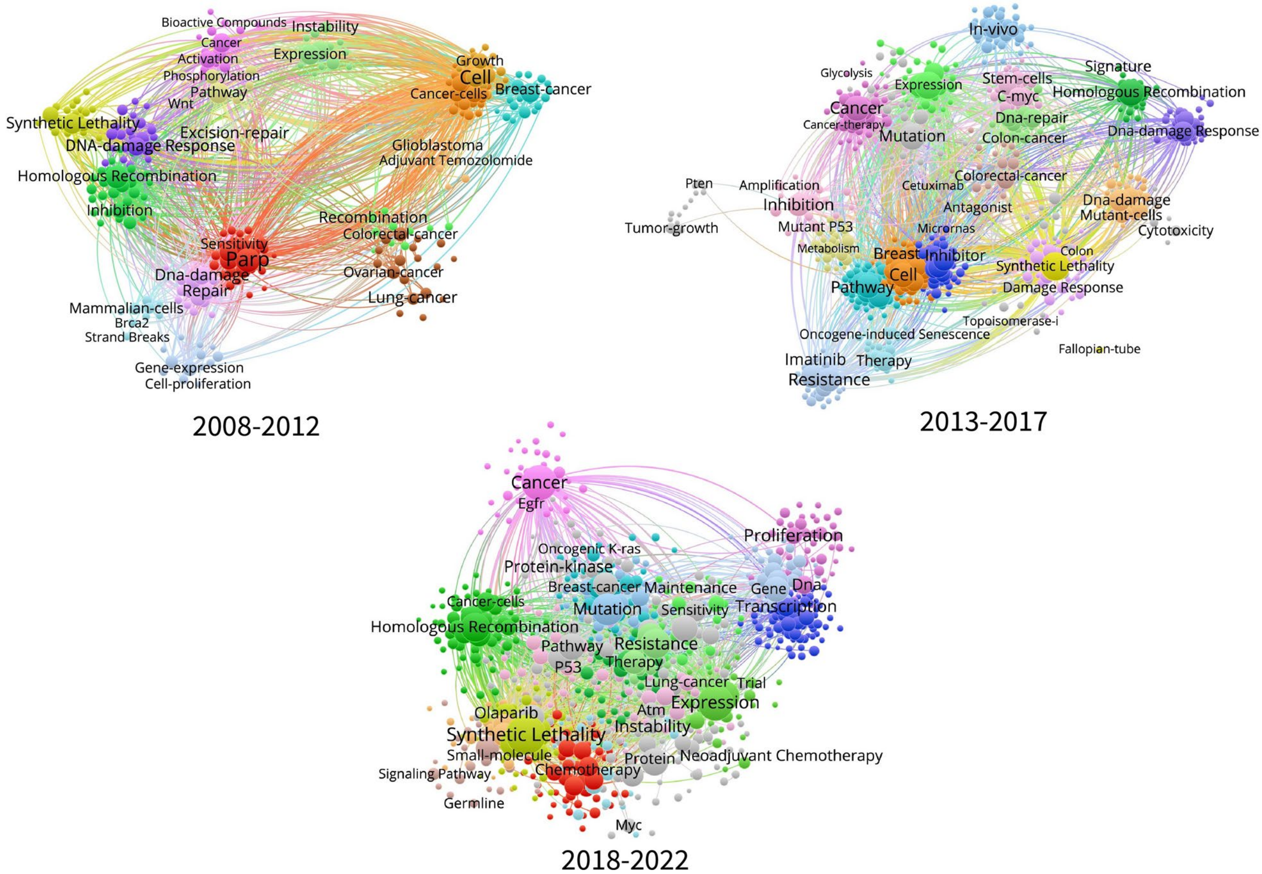

Fig. 1 Evolution and pattern of the knowledge network in synthetic lethality-based cancer research papers (2008–2022). The main component of the knowledge network in synthetic lethality-based cancer research papers (2008–2022) are presented for each 5-year duration. The nodes represent keywords, which are further clustered according to relatedness. Nodes from the same cluster are presented in the same colour. The size of each node indicates the frequency of occurrence. Only items with a total link strength above 10 are included. VOSviewer was used for visualization [31]

is significantly related to the citation number, regardless of whether knowledge network parameters are included. However, no significant relationship was found between the author's structural holes value and the citation number in either of the all-author models. We believe that the all-author aggregation method might have smoothed out the significant correlation between lead author structure holes and citation number. Therefore, we chose the lead author aggregation method over the all-author one for the final model (model 4).

As shown in model 2 and model 4, the coefficient for degree centrality in the collaboration network is positive and significant ($P < 0.05$ in model 2, $P < 0.1$ in model 4). Conversely, the coefficient for (degree centrality in the collaboration network)^2 is negative at the significance level of $P < 0.05$ in both models. According to an analysis on this phenomenon in a previous study in the field of wind energy [10], this indicates that the average degree centrality of the lead authors in the collaboration network has an inverted U-shaped relationship with the number of citations for the paper. To assess the validity of this relationship, we set all other variables to their average values and presented the inverted U-shaped relationship between author degree centrality and citation number in Fig 4, following a previous approach. Utilizing the delta method, we calculated the turning point to be 0.0418 (95% confidence interval [CI] 0.0211–0.0626), which falls within our normalized degree centrality data range of 0.0000–0.0743. Consequently, a moderate number of ties among the lead authors in a collaboration network could increase the paper's citation count (the academic influence) to some extent, while degree centrality beyond a certain threshold can have a negative impact on citations.

Furthermore, the coefficient for structural holes in the collaboration network is positive and significant (adjusted odds ratio 1.32, 95% CI 0.96–1.81; $P = 0.09$ in model 2 and adjusted odds ratio 1.32, 95% CI 0.97–1.81; $P = 0.08$ in model 4). This suggests that the average

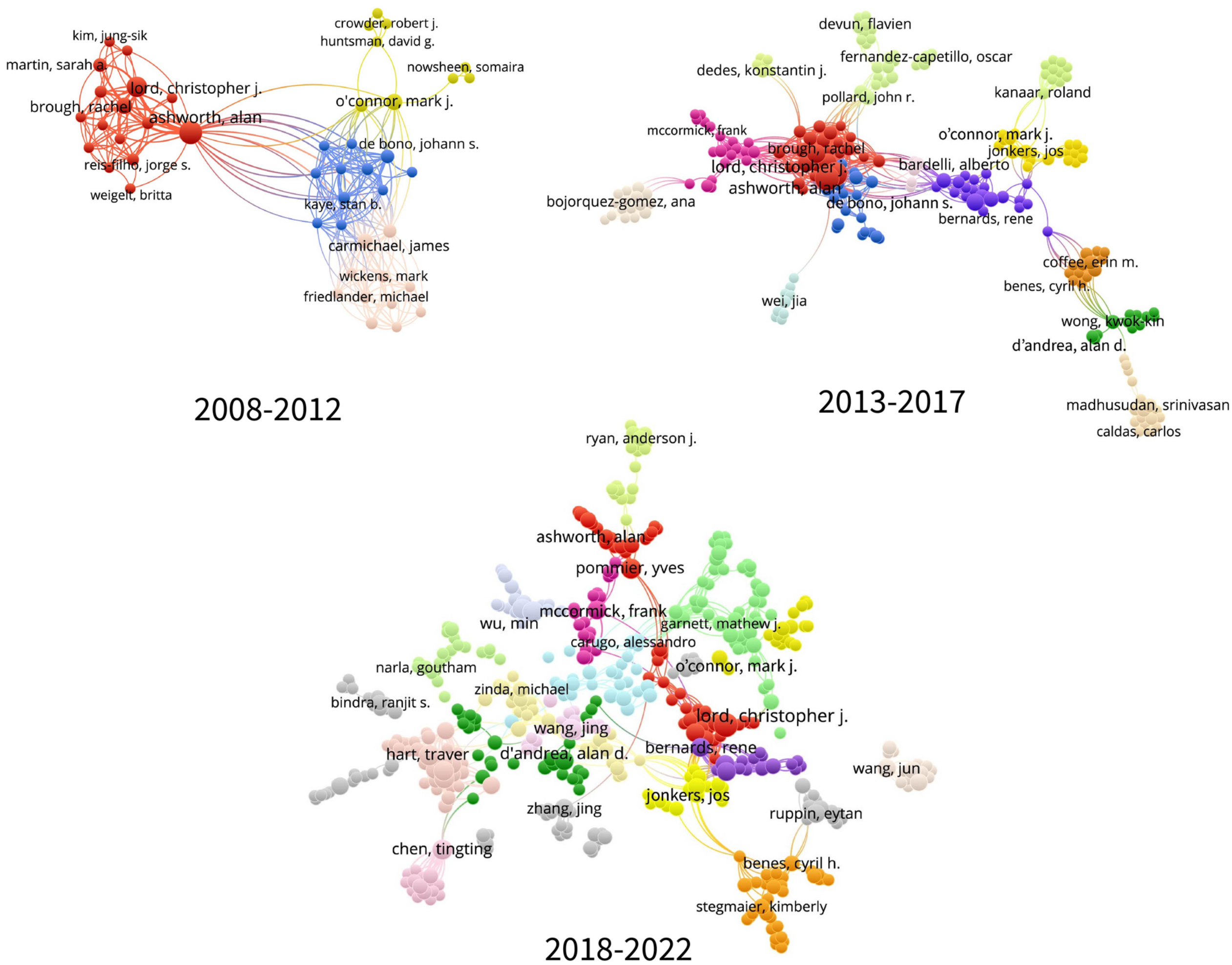

Fig. 2 Evolution and pattern of the collaboration network in synthetic lethality-based cancer research papers (2008–2022). The main component of the collaboration network of authors in synthetic lethality-based cancer research papers (2008–2022) are presented for each 5-year duration. The nodes represent authors, who are further clustered according to relatedness. Nodes from the same cluster are presented in the same colour. The size of each node indicates the number of publications by the author, and the strongest collaboration links are represented by lines. Only the largest set of connected items (authors) is included. VOSviewer was used for visualization

structural holes of the lead authors in the collaboration network are positively associated with the number of citations for the paper.

Unlike in the collaboration network, the coefficients for degree centrality in knowledge networks are negative and non-significant in model 4, suggesting that there is no significant relationship between a paper's knowledge element's average degree centrality in the knowledge network and its number of citations. However, positive and significant coefficients are found for structural holes in knowledge network in model 4 (adjusted odds ratio 4.37, 95% CI 1.28–14.96; $P = 0.02$ in model 4). This indicates that a paper's knowledge element's average structural holes in the knowledge network are positively related to its number of citations.

It is worth noting that the Akaike information criterion (AIC) for model 4 is the lowest among all models. Therefore, the addition of the independent variables based on the paper's degree centrality and structural holes in both the collaboration and knowledge networks successfully improved the model.

All the control variables included in this study are significantly associated with the number of paper citations, indicating the important influence of journal reputation, citation duration, the international collaboration status of the corresponding country and collaboration with pharmaceutical companies.

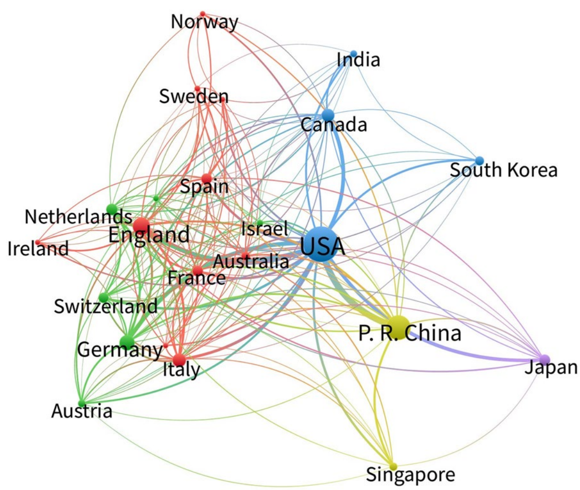

Fig. 3 Main component of the collaboration network of countries in synthetic lethality-based cancer research papers (2018–2022). The nodes represent countries, and the size of each node indicates the number of publications from that country. Nodes from the same cluster are presented in the same colour. Only countries with a total link strength above 10 are included. VOSviewer was used for visualization

Discussion

It has been well-documented that academic collaboration in the form of co-authorship may influence research capacity and impact [33], but the extent varies across countries, research fields and even research topics [34]. To eliminate the potential influence of 'hot' or 'cold' research topics, this study narrowed its focus to a specific topic within the field of cancer research – synthetic lethality-based cancer research.

Regression model establishment

Based on the regression results, the inclusion of node attributes of lead authors in the collaboration network and knowledge elements in the knowledge network effectively improved the regression model. Therefore, we have demonstrated that centrality and structural holes in these two networks are highly influential factors for the research impact of synthetic lethality-based cancer research papers. Additionally, all of the control variables in the regression model are statistically significant. Aside from journal IF and publication year, which have long been recognized as vital factors affecting a paper's citation count, this study also reveals a strong relationship between international collaboration status and industry involvement with paper citations.

Initially, no significant variance was observed between papers with or without international collaboration in the regression models (data not shown). We attribute this to the intricate nature of the relationship between international collaboration, specifically international co-authorship and cancer research impact, in terms of citation count, for different entities. At the national level, the proportion of internationally co-authored cancer research papers by a country may be influenced by factors such as its stage of scientific development, country size, economic development level, funding policies and medical resource availability [11].

More importantly, it is crucial to consider the different roles of the collaborating institutions/countries when analysing the impact of international collaboration, especially the role of the lead collaborators (research guarantors) [35]. We found that papers authored by leading countries (according to the corresponding country), which ranked high in structural holes ranking, not degree centrality, in the international collaboration network, exhibited significantly higher citation counts compared with others. Therefore, bridging structural holes can increase a country's paper citation count. It is the status of a country within the cancer research network that matters, not the number of collaborators.

As shown in Fig 3, in the field of synthetic lethality-based cancer research, the collaboration network at the

Table 3 Descriptive statistics and correlation matrix

Variables	Mean	SD	VIF	1	2	3	4	5	6	7	8	9
1. Number of citations	48.60	64.89	–	1								
2. Degree centrality in collaboration network	0.004	0.005	1.52	0.27**	1							
3. Structural holes in collaboration network	1.55	0.29	1.52	0.24**	0.53**	1						
4. Degree centrality in knowledge network	0.06	0.02	1.46	0.04	0.10*	0.11**	1					
5. Structural holes in knowledge network	1.91	0.08	1.46	0.03	−0.002	0.07	0.55**	1				
6. Journal IF	7.58	7.27	1.12	0.69**	0.22**	0.28**	−0.01	−0.001	1			
7. Year since publication	7.72	1.37	1.10	0.11**	0.19**	−0.08*	0.03	−0.05	−0.03	1		
8. Status in international collaboration	0.49	0.50	1.04	0.12**	0.04	0.004	−0.10**	−0.10**	0.12**	0.09*	1	
9. Industry involvement	0.08	0.28	1.02	0.17**	0.08*	0.13**	0.06	0.06	0.08*	−0.026	−0.021	1

* $P < 0.05$, ** $P < 0.01$

Table 4 Negative binominal regression results: a comparison between aggregation of lead author & all-author data

Models	1	2	3	4	5
Variables	**Dependent variable: number of citations**				
		Lead author	**All author**	**Lead author**	**All author**
Control variables					
Journal IF	0.077***	0.070***	0.070***	0.069***	0.069***
Years since publication	0.143***	0.127***	0.120***	0.135***	0.127***
Status in international collaboration	0.161**	0.181***	0.176***	0.192***	0.187***
Industry involvement	0.423***	0.346***	0.309***	0.340***	0.302***
Independent variables					
Degree centrality in collaboration network		31.136**	51.287**	29.790*	51.498**
(Degree centrality in collaboration network)^2		−447.438**	−824.960***	−414.719**	−808.241**
Structural holes in collaboration network		0.276*	0.205	0.280*	0.200
Degree centrality in knowledge network				−3.800	−4.630
(Degree centrality in knowledge network)^2				12.391	17.395
Structural holes in knowledge network				1.475**	1.550**

*** $P<0.01$, ** $P<0.05$, * $P<0.1$

country level exhibits a pattern with noticeable clustering. We speculate that leading countries prioritize obtaining non-redundant cancer research information, ideas and resources across these clusters, enhancing their overall research capacity and influence more effectively. Despite having some collaborators, non-leading countries may have limited access to resources and information beyond their own cluster due to barriers such as geographical distance, language difficulties or a lack of cooperation history. To improve overall research capacity and influence by enhancing positions in the cancer research network, governments should consider offering more international research fund supporting leadership in cancer research or taking a more active role in international health/cancer research organizations, such as the World Health Organization and the Union of International Cancer Control (UICC).

Moreover, the significant relationship between industry involvement and citations underscores the importance of integrating industry and academic researchers in cancer research. Since most companies involved in synthetic lethality-based cancer research are pharmaceutical or biotechnology companies, we conducted literature searches on university-industry collaborations in related fields.

Pharmaceutical involvement not only signifies the profound integration of human resources, knowledge, technology, funding, research platforms and medical resources but also serves as a robust predictor of the clinical experimentation of basic research findings [36]. In a study on university–industry collaborations within the Irish pharmaceutical industry, the significance of government funding was emphasized in motivating collaborations during the initiation phase, while the establishment of intellectual property (IP) agreements was identified as the impetus for knowledge sharing in the engagement phase [37]. This is supported by a survey of 105 university–industry collaborations within the US biotechnology industry, where researchers found that transparent IP policies enabled the formation of trust, which is vital for effective knowledge transfer and achieving success in university–industry collaborations [38]. To encourage university–industry cooperation in synthetic lethality-based cancer research, attention should be paid to the establishment of reasonable research policies and efficient processes, such as those for the establishment of intellectual property rights protection. Government investment is also crucial, with more national funds for university–industry led cooperation needed.

The relationship between lead authors' average degree centrality in the collaboration network and paper citations

The observed inverted U-shaped relationship between the average degree centrality of lead authors in the collaboration network and the number of paper citations is consistent with previous research [10]. Central authors typically have numerous connections with others, enabling them priority in acquiring ideas, the latest information and resources, which can effectively promote their own research [39]. As indicated from the rising part of the inverted U-shaped curve before reaching the vertex, engaging in cancer research collaboration, whether in a leading or supporting role, enhances an author's

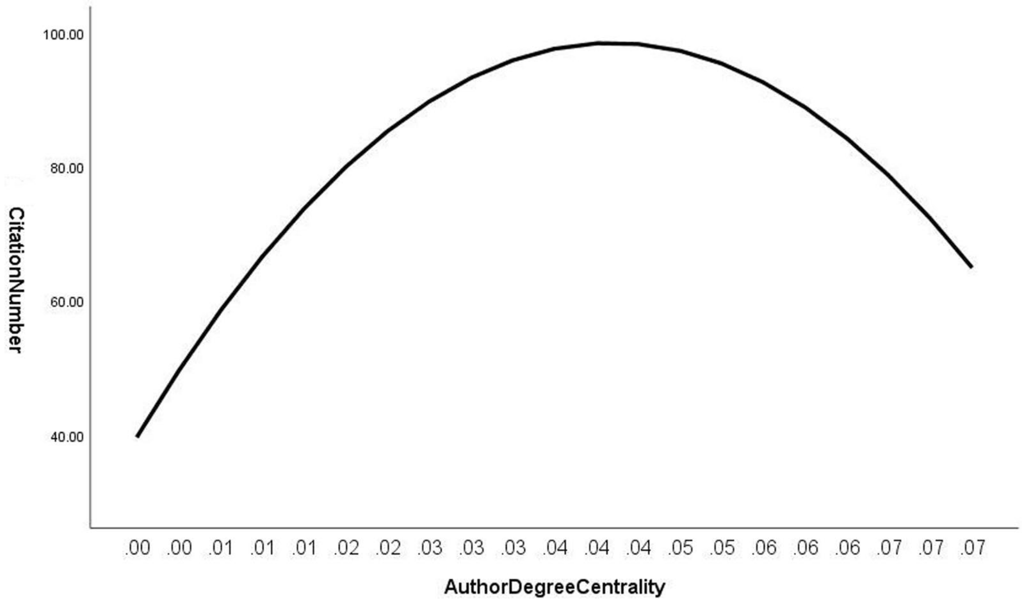

Fig. 4 The relationship between author degree centrality and citation number

academic research capacity, influence and reputation, ultimately contributing to increased paper citations.

However, when the inverted U-shaped curve reaches its peak, citations will start to decline as degree centrality increases. As in our study, if an author's centrality exceeds a certain threshold, the surplus of information can overwhelm them, resulting in a decrease in the quality of knowledge, which, in turn, negatively impacts research output and influence. The above result again emphasizes the importance of encouraging cancer research collaboration, especially for young cancer researchers at an early stage of their careers.

The relationship between lead authors' average structural holes in the collaboration network and paper citations

Based on the regression results, a positive relationship between the average structural holes of lead authors in the collaboration network and paper citation counts was found in both model 2 and model 4, although at a lower significant level compared with degree centrality. This suggests that by connecting with unique and diverse collaborators, central authors bridging structural holes can enhance the efficiency of their information acquisition, when conducting research on synthetic lethality-related cancer projects. Authors who bridge more structural holes have access to more non-redundant information and gain more control benefits [40], resulting in increased research capacity and impact and higher citation rates for their papers, as found in our study.

A significant difference was also observed between country and individual level collaboration networks. While degree centrality is the most important factor for individuals, the value of structural holes seems to be more influential at the country level. This reflects differences in the impact of social networks at different dimensions. At the macro level, the overall international collaboration status of a country is more important than the number of cooperating partners, while at the individual level, especially for young researchers, the number of collaborators directly determines the quality of resources and information obtained, thereby affecting the building of their research capabilities and the influence of research achievements.

The relationship between structural holes in the knowledge network and paper citations

This study revealed a positive relationship between the average structural holes of the knowledge elements in the knowledge network and paper citations, at the significance level of $P < 0.05$ in model 4. However, degree centrality showed no significant influence on citation counts. We posit that knowledge elements bridging rich structural holes in the knowledge network offer non-redundant information to cancer researchers who explore them. Additionally, they also create combinatorial opportunities between unconnected knowledge elements during the searching process, ultimately increasing citations from researchers across various disciplines. Therefore, elements bridging structural holes enjoy advantages in controlling the flow of knowledge during searches, thereby enhancing the citation opportunities for the papers involved. The above results provide suggestions for the selection of cancer research directions and keywords. Focusing on unique issues that connect with diverse information in cancer, especially interdisciplinary fields, will improve citations and the influence of research work.

These finding aligns with a previous study that analysed the influence of collaboration and knowledge network structures on organizational exploratory innovations in the field of nano-energy [8]. Although the two studies differ in terms of research field, level of analysis, variables and regression model design, they both reveal that in the context of exploratory research practices, the centrality of researchers/organizations in collaboration networks and the structural holes in their knowledge elements have a significant positive impact on their academic achievements and influence.

Conclusions

There are several innovations of this study.

A new method in aggregating collaboration network measures at the paper level was adopted in this paper. In previous studies, the 'all-author' approach was typically used for data aggregation at the paper level in social network analysis [10, 41, 42]. As a result, an inverted U-shaped relationship was revealed between authors' centrality and paper citations, while the effects from the structural holes value remained non-significant [10].

After comparing the regression results of the 'all-author' and 'lead author' approaches, this study focused solely on the degree centrality and structural holes values of the lead authors, specifically the first and corresponding authors. Our regression results confirm the significant relationship between the authors' structural holes value and paper citations, which might have been smoothed out by the 'all-author' approach. This choice reflects a distinctive characteristic of the biomedical research field, where the first and corresponding authors typically make core contributions to the article and better represent the research strengths compared with the other authors. This method has broad generalizability to other bibliometric studies based on biomedical data.

Multi-level collaboration networks were established, including individual-level collaboration networks based on authors and national-level collaboration networks. To account for the influence of a corresponding country's international collaboration status, countries were ranked and categorized on the basis of their structural holes values.

We revealed the vital importance of collaboration and knowledge networks, international collaboration status and university–industry cooperation in influencing cancer research capacity and impact, as well as the differences in their contributing factors. For non-leading countries, measures should be taken to enhance the international collaboration status – to become connected with unique collaborators across various clusters. For individuals, especially early career researchers, increasing the number of their collaborators seems to be sufficiently effective. University–industry cooperation should also be encouraged, enhancing the integration of human resources, technology, funding, research platforms and medical resources in the fight against cancer.

Insights gained through this study provide recommendations to research workers or administrators for the design of research directions through a knowledge network perspective. Focusing on unique issues especially interdisciplinary fields, will ultimately improve output and the influence of their research work.

Limitations and follow-up research

This study is based on a relatively limited number of articles published on one specific topic within the field of cancer research: synthetic lethality, which is characterized by rapid growth and multi-disciplinary interaction from around the world. Future research may explore other cancer research fields or topics and the various patterns of collaboration and knowledge networks, as well as other potential factors affecting the impact of cancer research, such as unexpected combinations of keywords[20].

Analyses may also be conducted to explore the availability of funding and its impact on citations. Considering the rise of gold open access publishing, it may also be necessary to consider the impact of access modality while carrying-out regression analysis based on citation data.

Finally, this study averaged only the measures of the first and corresponding authors while calculating the collaboration network metrics for each publication, emphasizing their significant contributions and impact on the related article. Future studies may further analyse the impact of other authors in different positions on the author list.

Abbreviation

IF Impact factor

Acknowledgements

Not applicable.

Author contributions
S.L. conceived of and designed the study and contributed to the study design, data analysis, data interpretation and writing of the manuscript. C.L. contributed to writing of the manuscript. H.W.Z. contributed to data acquisition.

Funding
This work was supported by the Department of Science and Technology of Guangdong Province of China under the grant 2020A1414010027.

Availability of data and materials
The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate
Ethics approval was not required as the material is publicly available.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]State Key Laboratory of Oncology in South China, Guangdong Provincial Clinical Research Center for Cancer, Sun Yat-sen University Cancer Center, No. 651 Dongfeng East Road, Guangzhou 510060, Guangdong, People's Republic of China.

Received: 13 March 2024 Accepted: 30 July 2024
Published online: 06 August 2024

References

1. Deo SVS, Sharma J, Kumar S. GLOBOCAN 2020 report on global cancer burden: challenges and opportunities for surgical oncologists. Ann Surg Oncol. 2022;29:6497–500.
2. Huang A, Garraway LA, Ashworth A, Weber B. Synthetic lethality as an engine for cancer drug target discovery. Nat Rev Drug Discovery. 2020;19:23–38.
3. Bailie J, Matous P, Bailie R, Passey ME. Patterns of collaboration and knowledge generated by an Australian rural research centre over 20 years: a co-authorship network analysis. Health Res Policy Syst. 2023;21:87.
4. Li EY, Liao CH, Yen HR. Co-authorship networks and research impact: a social capital perspective. Res Policy. 2013;42:1515–30.
5. Nahapiet J, Ghoshal S. Social capital, intellectual capital, and the organizational advantage. Academy Manag Rev. 1998;23:242–66.
6. Phelps C, Heidl R, Wadhwa A. Knowledge, networks, and knowledge networks: a review and research agenda. J Manag. 2012;38:1115–66.
7. Moorthy S, Polley DE. Technological knowledge breadth and depth: performance impacts. J Knowl Manag. 2010;14:359–77.
8. Guan JC, Liu N. Exploitative and exploratory innovations in knowledge network and collaboration network: a patent analysis in the technological field of nano-energy. Res Policy. 2016;45:97–112.
9. Yayavaram S, Ahuja G. Decomposability in knowledge structures and its impact on the usefulness of inventions and knowledge-base malleability. Adm Sci Q. 2008;53:333–62.
10. Guan JC, Yan Y, Zhang JJ. The impact of collaboration and knowledge networks on citations. J Informet. 2017;11:407–22.
11. de Moya-Anegon F, Guerrero-Bote VP, Lopez-Illescas C, Moed HF. Statistical relationships between corresponding authorship, international co-authorship and citation impact of national research systems. J Informet. 2018;12:1251–62.
12. Khor KA, Yu LG. Influence of international co-authorship on the research citation impact of young universities. Scientometrics. 2016;107:1095–110.
13. Moed HF. Citation analysis in research evaluation. Berlin: Springer; 2005.
14. Tahamtan I, Afshar AS, Ahamdzadeh K. Factors affecting number of citations: a comprehensive review of the literature. Scientometrics. 2016;107:1195–225.
15. Wang L, Thijs B, Glänzel W. Characteristics of international collaboration in sport sciences publications and its influence on citation impact. Scientometrics. 2015;105:843–62.
16. Abbasi A, Wigand RT, Hossain L. Measuring social capital through network analysis and its influence on individual performance. Libr Inf Sci Res. 2014;36:66–73.
17. Burt RS. The contingent value of social capital. Adm Sci Q. 1997;42:339–65.
18. Tsai W. Knowledge transfer in intraorganizational networks: effects of network position and absorptive capacity on business unit innovation and performance. Acad Manag J. 2001;44:996–1004.
19. Burt RS. Structural holes: the social structure of competition. Cambridge: Harvard University Press; 2009.
20. Shi F, Evans J. Surprising combinations of research contents and contexts are related to impact and emerge with scientific outsiders from distant disciplines. Nat Commun. 2023. https://doi.org/10.1038/s41467-023-36741-4.
21. Abbasi A, Altmann J, Hossain L. Identifying the effects of co-authorship networks on the performance of scholars: a correlation and regression analysis of performance measures and social network analysis measures. J Informet. 2011;5:594–607.
22. Wang J. Knowledge creation in collaboration networks: effects of tie configuration. Res Policy. 2016;45:68–80.
23. **Science of Science (Sci2) Tool** [http://sci.slis.indiana.edu]
24. Freeman LC. Centrality in social networks conceptual clarification. Soc Networks. 1978;1:215–39.
25. Wang CL, Rodan S, Fruin M, Xu XY. Knowledge networks, collaboration networks, and exploratory innovation. Acad Manag J. 2014;57:484–514.
26. Burt RS. Structural holes and good ideas. Am J Sociol. 2004;110:349–99.
27. **Exploratory social network analysis with Pajek (2nd ed.)** [http://mrvar.fdv.uni-lj.si/pajek/]
28. Gazni A, Ghaseminik Z. Author practices in citing other authors, institutions, and journals. J Asso Info Sci Technol. 2016;67:2536–49.
29. Cova T, Jarmelo S, Formosinho SJ, de Melo JSS, Pais A. Unsupervised characterization of research institutions with task-force estimation. J Informet. 2015;9:59–68.
30. Guerrero-Bote VP, Olmeda-Gómez C, Moya-Anegón F. Atlas of scientific institutions in food science (Scopus, 2003–2013). Lwt-Food Science And Technology. 2016;67:133–42.
31. van Eck NJ, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics. 2010;84:523–38.
32. Isfandyari-Moghaddam A, Saberi MK, Tahmasebi-Limoni S, Mohammadian S, Naderbeigi F. Global scientific collaboration: a social network analysis and data mining of the co-authorship networks. J Inf Sci. 2023;49:1126–41.
33. Polyakov M, Polyakov S, Iftekhar MS. Does academic collaboration equally benefit impact of research across topics? the case of agricultural, resource, environmental and ecological economics. Scientometrics. 2017;113:1385–405.
34. Sud P, Thelwall M. Not all international collaboration is beneficial: the mendeley readership and citation impact of biochemical research collaboration. J Asso Infor Sci Technol. 2016;67:1849–57.
35. Moya-Anegón F, Guerrero-Bote VP, Bornmann L, Moed HF. The research guarantors of scientific papers and the output counting: a promising new approach. Scientometrics. 2013;97:421–34.
36. Contopoulos-Ioannidis DG, Ntzani EE, Ioannidis JPA. Translation of highly promising basic science research into clinical applications. Am J Med. 2003;114:477–84.
37. O'Dwyer M, Filieri R, O'Malley L. Establishing successful university-industry collaborations: barriers and enablers deconstructed. J Technol Transfer. 2023;48:900–31.
38. Bstieler L, Hemmert M, Barczak G. Trust formation in university-industry collaborations in the us biotechnology industry: ip policies, shared governance, and champions. J Prod Innov Manag. 2015;32:111–21.
39. Abbasi A, Hossain L, Leydesdorff L. Betweenness centrality as a driver of preferential attachment in the evolution of research collaboration networks. J Informet. 2012;6:403–12.
40. Rodan S, Galunic C. More than network structure: how knowledge heterogeneity influences managerial performance and innovativeness. Strateg Manag J. 2004;25:541–62.

41. Zhang Y, Ding JL, Yan H, He M, Wang W. A study of the influence of collaboration networks and knowledge networks on the citations of papers in sports industry in china. Complexity. 2022. https://doi.org/10.1155/2022/9236743.
42. Colladon AF, D'Angelo CA, Gloor PA. Predicting the future success of scientific publications through social network and semantic analysis. Scientometrics. 2020;124:357–77.

Publisher's Note