Check for updates

OPEN

Mixing Patterns in Interdisciplinary Co-Authorship Networks at Multiple Scales

Shihui Feng[1,3] & Alec Kirkley[2,3] ✉

There are inherent challenges to interdisciplinary research collaboration, such as bridging cognitive gaps and balancing transaction costs with collaborative benefits. This raises the question: Does interdisciplinary research collaboration necessarily result in disciplinary diversity among collaborators? We aim to explore this question by assessing collaborative preferences in interdisciplinary research at multiple scales through the examinination of disciplinary mixing patterns at the individual, dyadic, and team level in a coauthor network from the field of artificial intelligence in education, an emerging interdisciplinary area. Our key finding is that disciplinary diversity is reflected by diverse research experiences of individual researchers rather than diversity within pairs or groups of researchers. We also examine intergroup mixing by applying a novel approach to classify the active and non-active researchers in the collaboration network based on participation in multiple teams. We find a significant difference in indicators of academic performance and experience between the clusters of active and non-active researchers, suggesting intergroup mixing as a key factor in academic success. Our results shed light on the nature of team formation in interdisciplinary research, as well as highlight the importance of interdisciplinary training.

There are many significant social and global problems that cross disciplinary boundaries. The scientific complexity of these problems calls for the synthesis of concepts, theories and methods from multiple disciplines, and new research areas beyond traditional disciplinary frameworks. With an exponentially growing amount of digital data, formulating data-informed decisions requires both subject domain expertise as well as fluency with computational techniques to process, analyze and interpret this large-scale data. As a result, a holistic approach to understanding these problems necessitates the integration of different branches of knowledge, which has resulted in an increasing trend towards interdisciplinary research in both the natural and social sciences since the mid-1980s[1].

However, there are some obstacles in developing interdisciplinary collaboration that cannot be neglected. Firstly, collaboration requires a common ground where a group of individual researchers have a certain level of shared understanding and mutual knowledge of the research problems[2,3]. In particular, social and natural scientists may have different perspectives and approaches to defining, solving and presenting problems, which introduces philosophical obstacles in interdisciplinary collaboration[4]. Committing to an interdisciplinary collaboration poses a risk for researchers from different disciplines, in terms of the balance of transaction costs and collaborative benefits. The motivation for researchers to participate in interdisciplinary collaboration could highly depend on the evaluation of the perceived risks and rewards. Secondly, obstacles relevant to psychosocial and practical perspectives hindering collaboration in general can also be applied to interdisciplinary collaboration. For instance, the "obstructive misconceptions or prejudices" between social and natural scientists[4] could result in a lack of appreciation of each other's value and contributions in collaboration, which could influence the effectiveness and continuity of interdisciplinary collaboration.

Given the challenges of interdisciplinary collaboration, an essential question is whether or not there is data-based evidence of homophily or diversity with respect to disciplinary background in interdisciplinary research collaboration. This study aims to assess disciplinary diversity in research collaboration at the individual, dyadic, and team level for research on artificial intelligence in education (AIED). AIED has been developing fast as an interdisciplinary research area in the last decade, focusing on applying computational techniques in analyzing large-scale educational data and developing intelligent systems for supporting teaching and learning

[1]Unit of Human Communication, Development, and Information Sciences, Faculty of Education, The University of Hong Kong, Hong Kong, China. [2]Department of Physics, University of Michigan, Ann Arbor, USA. [3]These authors contributed equally: Shihui Feng and Alec Kirkley. ✉e-mail: akirkley@umich.edu

activities. It is a demonstration of the newly emerging interdisciplinary research paradigm where statistical and computational knowledge is integrated into social and humanities contexts. Here, the mixing patterns of a coauthor collaboration network in AIED research are studied following four research questions: (1) Do individual researchers tend to have diverse experience in multiple disciplines? (2) Do researchers in an interdisciplinary area prefer to collaborate with others from a similar or different research background? (3) Do teams as a whole tend to be composed of researchers with similar or diverse research backgrounds? (4) Do researchers with different structural characteristics in a collaboration network have different research performance? Our findings provide data-informed evidence of the mechanisms underlying the formation of collaboration in interdisciplinary research, and these results can further yield insights for formulating strategies and training programs to facilitate effective collaboration.

Related Work

The paradox of "interdisciplinarity" in interdisciplinary collaboration.   A variety of group and organizational theories provide theoretical underpinnings for the formation, dynamics and complexity of academic collaboration. The formation of team members in research is vital to the success and effectiveness of collaboration. Lewin's group dynamics theory[5] suggests that the shared incentives among group members and task interdependence significantly affect the group process in a collaboration, and places higher priority on the shared incentives rather than the similarity or dissimilarity of individuals. However, a certain level of similarity in characteristics of individuals could positively affect the development of shared commitment towards a goal. Ruef, Aldrich, and Carter[6] provided supporting evidence that homophily, together with network ties, has the determining effects on group formation. Homophily in group composition refers to the tendency for people to collaborate with others who share a certain level of similarity on various attributes, for instance gender, age or ethnicity[7,8]. The homophily principle provides a theoretical underpinning to understanding the formation of various social ties[9], and has thus been studied extensively in different contexts.

In this study, we are interested in examining homophily with respect to academic backgrounds of coauthors in interdisciplinary research collaboration. In interdisciplinary research collaboration, diverse academic backgrounds within a research team extend research capacity but also may increase the complexity and disequilibrium of group dynamics in collaboration. Bringing individuals from different disciplines together introduces a heterogenous attribute to a group, which conflicts with the principle of homophily in group composition. An essential question that comes along with this line of thinking is whether or not homophily regarding academic backgrounds is still an applicable mechanism for group composition in interdisciplinary research collaboration.

Previous works studying interdisciplinary collaboration are largely focused on its effects on professional practices in the context of healthcare[10–12]. Regarding the factors associated with the success of interdisciplinary academic collaboration, a study conducted by Van Rijnsoever and Hessels[13] found that years of working experience, previous experience of working at other universities or firms, and being female are positively associated with interdisciplinary research collaboration. Cummings and Kiesler[14] also found that prior collaboration experience plays an important role in eliminating the barriers in interdisciplinary collaboration. However, there is still a lack of research studying the diversity of group composition in interdisciplinary collaboration with a focus on the homogeneity or heterogeneity of team members' research backgrounds based on historical publication records. This study aims to address the novel question of assessing the diversity of academic backgrounds in interdisciplinary collaboration from individual, dyadic, and team levels using network approaches.

Assessing researcher interdisciplinarity.   Interdisciplinary research (IDR) can be understood as a variety of ways of bridging and integrating two or more disciplinary approaches and knowledge[15]. "Interdisciplinarity" emphasizes the integration of disciplinary knowledge, compared to the idea of "multidisciplinarity" in which components from different disciplines are assembled, or the creation of novel methods and concepts advocated by "transdisciplinarity"[16]. In interdisciplinary research collaboration, one normally expects disciplinary diversity of research team members with respect to their research backgrounds and experiences[17]. The most crucial aspect of assessing interdisciplinarity of research collaboration at various scales is to properly define the disciplinary profile of an individual researcher, which is an open and challenging problem in its own right[18].

Early studies define research disciplines of authors based on departmental affiliations[19]. However, in interdisciplinary research areas, departmental affiliations are poor representations of an individual's research experience, as by the nature of the area, authors may not be easily classified by a single field. Huutoniemi et al.[15] reviewed that the most accessible information used in previous studies to quantitatively define the disciplinary content of a researcher's profile include the ISI journal categories[16,20,21], research areas of funding organizations[22], and researchers' departmental affiliations[23]. Research papers are considered as the appropriate representative entity for gauging interdisciplinarity, as they serve as a proxy for individual output[16]. Consequently, ISI journal subject categories for a researcher's past publications are used for identifying the disciplinary profile of a researcher in this study. Here, we are interested in exploring the mechanisms underlying interdisciplinary research collaboration, particularly with respect to the diversity or homogeneity of research backgrounds of collaborators.

Porter et al.[16] propose measures for interdisciplinarity of a body of research that are derived from the subject categories of cited journals from reference lists, accounting for similarity in subject categories when computing their measures. It is found that individual researchers heavily utilize knowledge from different domains in their publications, which is consistent with our findings as we will discuss. Their method adopts the reference lists of a research article as a representation of the underlying cognitive space of the research work, and further considers the cognitive distance between subject categories for assessing the integration and specification of researchers' knowledge across different disciplines. This method provides a valuable blueprint for calculating the interdisciplinarity score at the paper and researcher levels. There are other proposed approaches for evaluating disciplinary content and assessing knowledge integration at the paper or field level using methods based on article keywords,

abstracts and reference lists including text-based clustering[24,25], word co-occurrence[26,27], and semantic structural analysis[28]. Here, we use subject categories to go beyond individual-level interdisciplinarity to assess the diversity of research collaborations at the dyadic and group level, which is not addressed in these studies. We do not attempt to redefine knowledge boundaries or develop a new classification of research topics, which is a well studied field of its own with a long history[29]. Instead, we use the existing classification of independent disciplines and see how the diversity of these disciplines manifests itself in interdisciplinary collaboration at multiple scales by proposing scalable, interpretable approaches.

Network approaches to studying research collaboration. A network is a mathematical object from graph theory consisting of nodes connected in pairs by edges. Networks are a useful tool for representing pairwise relationships in various social or physical systems in an abstract manner, and consequently, network approaches have been widely applied to study the structure of relationships and interconnection among components within and across systems[30]. Network structural properties can reveal the accessibility and diversity of resources embedded in social connections[31,32], as well as the effectiveness of information transfer and innovation diffusion[33,34]. Research collaboration can be well represented by networks consisting of researchers and the collaborative ties among them, and a large body of literature has studied research collaboration from a network science perspective[35–37]. Guimera[38] studied the temporal structures of research collaboration networks and found that prior collaboration experience and the recruitment of newcomers has a positive effect on the success of research collaboration in multiple fields. Moody[39] analyzed the cohesion of research collaboration in sociology by examining a sociology collaboration network from 1963 to 1999. Dahlander and McFarland[40] identified six attributes of collaborative ties that affect the formation and persistence of research collaboration across time. In general, previous studies have primarily focused on the following aspects of collaboration networks: 1) Descriptive structural characteristics; 2) Group formation; 3) Temporal group dynamics; and 4) Structural factors associated with the success of collaboration. In this study, we focus on providing new insights about interdisciplinarity in collaboration networks through aspects (2) and (4) using novel measures and approaches.

Methods

Data collection. The collaboration network data used in this study are collected from three representative journals on artificial intelligence in education (AIED), an emerging interdisciplinary research area. The three journals studied are *International Journal of Artificial Intelligence in Education*, *Proceedings of Educational Data Mining*, and *Proceedings of Learning at Scale*. The bibliometric information of all the available publications from these journals during the years 2010 to 2019 is obtained from the DBLP database. The collaboration network is constructed with the 2022 authors in the dataset as nodes, with an edge between two nodes if these authors coauthored a paper together.

The Scopus database classifies all journals and conference proceedings into 27 (ASJC) major categories and 334 minor categories. In this study, we employ the major categories for defining the researchers' disciplinary profile, as the subdivided minor categories with many cognate areas could inflate the diversity in the disciplinary profile of an researcher. These minor categories can not necessarily be treated orthogonally in a vector space representation of an author's research history, as they can be cognitively similar[16]. There have been other approaches proposed to correct for correlations among subject disciplines in computing interdisciplinarity measures[41,42], but there is no standard approach for how to quantify the cognitive overlap in these disciplines due to the problem's inherent complexity. Thus, here we choose to interpret each major subject classification as orthogonal to allow for an intuitive interpretation of an author's disciplinary history embedding (as we will see in the example below).

The Scopus database provides the number of papers per research field for indexed authors based on the groupings of the 27 major discipline categories, which are extracted for each author in the dataset to represent the disciplinary profile of their research background. The computational cost of analyzing the disciplinary profile of authors based on reference lists of all past publications poses scalability limitations for larger systems, and so the categories comprising an author's publication record, rather than all referenced journal categories from these papers, is used as a proxy for the disciplinary content of a researcher's output. Each author's publication counts were normalized to give the fraction of all of their work classified under a given category, which was represented with a vector with 27 entries, the number of major disciplines classified by the database. We also denote an author's primary discipline as the discipline in which they published the most. For example, if author i has 50 publications classified under 'Computer Science', 30 publications classified under 'Math', and 20 publications classified under 'Sociology', they would have a vector $\overrightarrow{x_i}$ with entries {0.5, 0.3, 0.2} for the entries corresponding to these disciplines respectively, and 0's elsewhere, with 'Computer Science' as their primary discipline.

Additionally, other author metadata is retrieved through the Scopus API, including their earliest and latest publication year, and h-index. We consider the research field with the highest number of publications of an author as their primary research discipline, but all the publication fields of an author are considered for assessing the interdisciplinarity of individual researchers. To explore the associations between structural properties of authors in the collaboration network and academic performance and experiences, the h-index is used as an indicator of academic success. Academic experience is measured based on the number of years between an author's first and latest publication.

Measures for assessing disciplinary diversity. Different measures were used to capture the diversity of research collaboration at the individual, dyadic and team level, which we discuss here. In addition, we detail a simple scheme to classify active and non-active collaborators in the network based on their tie patterns, which allows us to explore the associations between research collaboration and academic performance and experiences.

Individual disciplinary diversity. This refers to the extent to which an individual researcher's publication history spans its constituent set of research disciplines, allowing us to address our first question of whether or not individual researchers tend to have experience in multiple disciplines. As it is an intuitive measure for the diversity of categorical data with clear upper and lower bounds[43], entropy is used here to measure the variation of the fields comprising each individual researcher's publication history. Using the information from the publication count vector $\overrightarrow{x_i}$, the entropy for researcher *i*'s publication history is given by

$$H_i = -\frac{1}{\log(n_d^{(i)})}\sum_{d=1}^{N_d}\overrightarrow{x_{id}}\log(\overrightarrow{x_{id}}),$$

(1)

where $\overrightarrow{x_{id}}$ is the fraction of researcher *i*'s publications classified under field *d* (the *d*-th entry in the normalized publication count vector $\overrightarrow{x_i}$), and $n_d^{(i)}$ is the number of unique disciplines for author *i*. The prefactor $\log(n_d^{(i)})^{-1}$ is to ensure that we consider the entropy of an individual researcher's background relative to the maximum value it could have given a perfectly equal distribution of publications across the disciplines *i* participates in. This allows us to assess how high the entropy of a researcher's publication distribution is relative to its maximum possible value, conditioned on how many disciplines the author published in. Authors with only one publication field (1.2% of all authors) are excluded in the analysis. High values of this index (H_i close to 1) indicate researchers with a high level of individual interdisciplinarity in their publication record, and low values (H_i close to 0) indicate researchers with a low level of interdisciplinarity. We note that similar measures have been employed to assess individual interdisciplinarity[20], including measures based on the Stirling diversity index[44], the Herfindahl index[45], and the Shannon entropy[46], like our own. Our modification to the standard Shannon entropy allows us to assess the extent to which an individual's publication record is balanced among the disciplines they contribute to. This gives us a measure for assessing how an individual allocates their energy towards different fields.

Dyadic disciplinary diversity. The level of similarity of research background for a pair of researchers in the collaboration network is assessed with this measure, addressing the second research question of whether or not disciplinary homophily is an applicable mechanism for explaining the collaboration preferences in interdisciplinarity research. For a surface level assessment of pairwise interdisciplinarity, the fraction of all edges that are comprised of researchers with the same primary discipline (the discipline in which an author published the most) is computed. However, to account for imbalances in the global distribution of primary affiliations (i.e. how many ties we expect between authors of the same primary discipline by chance), we compare this fraction with the same fraction computed on all pairs of authors who *did not* collaborate. To see whether these fractions differ significantly, we use a two proportion *z*-test, the details of which we describe shortly.

However, given the nature of interdisciplinary research, it is essential to take the diversity within each individual's research experience into consideration while assessing collaboration patterns, as individuals are not well categorized into a single research domain. We thus employ cosine similarity to measure the dyadic interdisciplinarity in the network, by comparing the publication count vectors for each of the authors. Cosine similarity is a common measure for determining the similarity of two non-zero vectors depending on their orientations in some high dimensional space, and in our context is given by

$$S_{ij} = \frac{\overrightarrow{x_i}\cdot\overrightarrow{x_j}}{\left\|\overrightarrow{x_i}\right\|\left\|\overrightarrow{x_j}\right\|},$$

(2)

where $\left\|\overrightarrow{x_i}\right\|$ is the magnitude of $\overrightarrow{x_i}$. The value of S_{ij} is also restricted to [0, 1], and a high value of S_{ij} indicates a high similarity in the research backgrounds of authors *i* and *j*, while a low value indicates dissimilarity. We also compute Eq. (2) for both edges and non-edges to see whether researchers collaborate with others that are more or less similar than those they do not collaborate with.

Team disciplinary diversity. To address the third research question of whether research teams in an interdisciplinary area tend to be composed as a whole of researchers with similar or diverse backgrounds, we look at team disciplinary diversity. This is also assessed based on both primary discipline and publication vectors $\overrightarrow{x_i}$ to give results from multiple perspectives. Within-group entropy is employed to assess the team interdisciplinarity based on the primary publication fields for all authors in a research group. In a similar manner to Eq. (1), the within-group entropy $\tilde{H}_p$ for a paper *p* is given by

$$\tilde{H}_p = -\frac{1}{\log(\text{Min}\{|p|, N_d\})}\sum_{d=1}^{N_d}f_{pd}\log(f_{pd}),$$

(3)

where f_{pd} is the fraction of authors on the paper *p* with primary discipline *d*, and $|p|$ is the number of authors on paper *p*. The new normalization factor $\log(\text{Min}\{|p|, N_d\})^{-1}$ is introduced here because a tight upper bound on the entropy of collaboration *p* is restricted by either the size of the collaboration or the number of possible disciplines (whichever is smaller). Additionally, in a similar manner to the analysis on dyadic interdisciplinarity, the within-group average cosine similarity is used to assess the team interdisciplinarity beyond looking simply at primary discipline. The mean within-group cosine similarity $\tilde{S}_p$ for paper *p* is given by
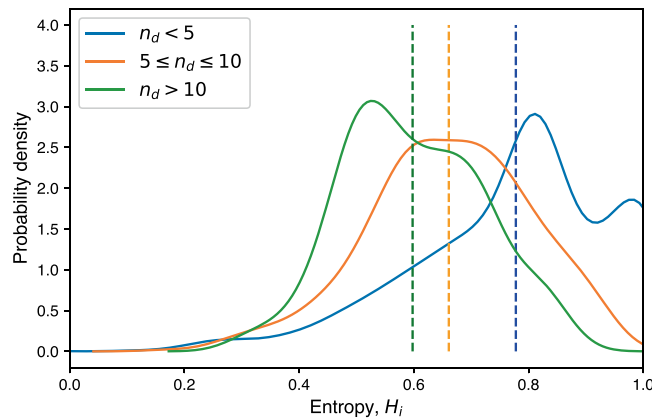
Figure 1. Probability densities of entropy (Eq. (1)) for researchers with varying numbers of publication disciplines, n_d. The mean of each distribution is indicated by a vertical line. Individual researchers had relatively high diversity in their research backgrounds, with those active in more fields showing a slightly less spread out research profile.

$$\tilde{S}_p = \frac{2}{|p|(|p| - 1)} \sum_{(i,j) \in p} S_{ij},$$

(4)

where the prefactor normalizes the measure to [0, 1], and the sum is over all pairs of nodes in p. The measures in Eqs. (3) and (4) can be interpreted in a similar manner as the measures in Eqs. (1) and (2) respectively, except they assess disciplinary diversity at the team-level rather than the individual or pairwise level.

Core-shell decomposition. The last research question examines the associations of the structural characteristics and academic performance and experience in the collaboration network. We define active collaborators in the network as researchers who are active in collaborating with multiple research groups in multiple projects. These authors published more than one article with diverse groups and perform a significant role in contributing to the global connectivity of research collaboration in the field, but may have a low level of local transitivity. Local transitivity, which we denote C_i for an author i, refers to the fraction of all possible ties that exist among i's neighbors, and is given by

$$C_i = \frac{2}{|\partial_i|(|\partial_i| - 1)} \sum_{(j,k) \in \partial_i} A_{jk}$$

(5)

where ∂_i is the set of edges adjacent to i, and A_{jk} is the binary adjacency matrix such that $A_{jk} = 1$ if there is a connection between j and k, and $A_{jk} = 0$ if there is not. Collaboration networks constructed using co-authorship data tend to have a large number of fully connected cliques: co-authors of the same research paper are fully connected. Therefore, a high number of nodes have a maximum local transitivity ($C_i = 1$), as they only collaborate with members of their research group. Thus, simply by looking for nodes i with local transitivity $C_i < 1$, we can identify the nodes that act as bridges in the collaboration network by associating those with $C_i < 1$ as the "core" of the network and those with $C_i = 1$ as the "shell". In this way, we can see how the network separates into nodes with topologically diverse neighborhoods and nodes with homogeneous connectivity. There are other measures to assess the level of global connectivity a node facilitates (such as betweenness centrality), but here we are only concerned with a binary classification of whether a node is *active* in collaboration (has multiple distinct groups of collaborators) or *inactive* (has only one group of collaborators). As computation of local transitivity is fast on most networks, this method is a relatively cost-effective approach for performing a decomposition of a collaboration network into a core and a shell. Removing nodes i with $C_i = 1$ and iteratively identifying nodes of $C_i = 1$, we can decompose the network into nodes with different "coreness" values, which gives a more sophisticated means of identifying the importance of nodes for the global connectivity of the network, but we leave this and other extensions to future work.

Results
Individual disciplinary diversity.
The 2022 authors in the collaboration network are from 18 primary disciplines and have publications in journals spanning all 27 major disciplines. In Fig. 1, we plot the distribution of the entropies (Eq. (1)) for all researchers that contributed to a given number of subfields, n_d. For easier visualization, the histograms were smoothed using a kernel density estimate to obtain a probability density function. Based on the densities in the figure, we can see that authors in the interdisciplinarity area contribute relatively equally to all the fields they publish in (H_i is moderately high on average), but that the distributions vary depending on how many fields an author participated in. In particular, authors with more publication fields are not able to contribute equally to all of these fields, and so we see a systematic decrease in the position of the H_i values. The
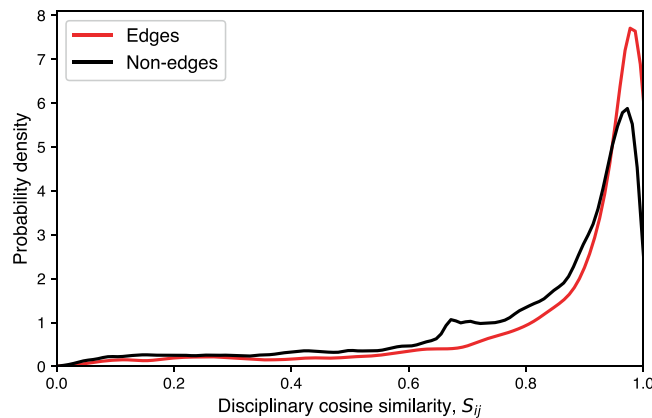
Figure 2. Probability densities of cosine similarity (Eq. (2)) for edges and non-edges. The S_{ij} values for edges tended to be higher than those for non-edges, an effect that is shown to be statistically significant through Mann-Whitney and Kolmogorov-Smirnov tests.

individual disciplinary diversity distributions for each individual journal all present similar results, and so the trends we see persist at the journal-level as well, although we do not present these results here.

Dyadic disciplinary diversity. Among 5002 edges between the 2022 authors in the network, 81% of pairs have the same primary discipline, while 75% of non-edge pairs have the same primary discipline (a majority of authors have computer science as a primary discipline). Using a two-proportion z-test, this difference is statistically significant at the 1% level ($z = 11.7$, $p < 0.01$). These results suggest that authors preferentially collaborate with others of the same primary discipline. However, as discussed in Methods, we need to go beyond primary disciplines to analyze interdisciplinarity in an interdisciplinary research field, so cosine similarity (Eq. (2)) is also examined across all edge and non-edge pairs. Figure 2 shows the probability densities of S_{ij} over these pairs, indicating a shift in the distribution for edges towards higher similarity values than for the non-edges. We test the null hypothesis that it is equally likely that a randomly selected value from the edge distribution is less than or greater than a randomly selected value from the non-edge distribution using a Mann-Whitney U test, finding that we can reject this null in favor of the alternative hypothesis that the cosine similarities on the edges are systematically higher than on the non-edges (Median 1 = 0.95, Median 2 = 0.89, $n_1 = 5002$, $n_2 = 2.04 \times 10^6$, $U \gg 10$, $p \ll 0.01$, one-tailed). We also report the results from a Kolmogorov-Smirnov test to determine whether the distributions are the same, which also indicates significant differences between edges and non-edges ($D = 0.22$, $p \ll 0.01$). These findings suggest that interdisciplinary researchers also prefer collaborators with similar interdisciplinary research backgrounds. We also plot the collaboration network, with edges colored according to their cosine similarity S_{ij}, for visual inspection, in Fig. 3.

Team disciplinary diversity. To assess whether or not homophily with respect to disciplinary profiles is an applicable mechanism to explain team formation in research collaboration, we examine disciplinary diversity at the team level. A group of co-authors of the same articles is considered a research team, represented as a fully connected clique in the collaboration network. To assess the disciplinary diversity of a team solely considering primary disciplines, we compute $\tilde{H}_p$ from Eq. (3) on all the research teams p in the network. We also compute Eq. (3) on 1000 randomized teams (drawn uniformly at random without replacement from all researchers in the network) for each unique team size present in the network. Then, for every team p in the network, we take the difference of the observed value of $\tilde{H}_p$ and the average value $\mu_{|p|}^{(H)}$ from simulations of random teams of the same size, and divide by the standard deviation $\sigma_{|p|}^{(H)}$ of the results for the randomized teams. This gives us the z-score $z_p^{(H)}$ of the observed result $\tilde{H}_p$ in the null ensemble where researchers have no collaboration preferences, thus

$$z_p^{(H)} = \frac{\tilde{H}_p - \mu_{|p|}^{(H)}}{\sigma_{|p|}^{(H)}}.$$
(6)

For example, if a team is of size $|p| = 4$, we run 1,000 simulations drawing teams of 4 at random from all authors in the network to get a vector $\vec{H}_{|p|}$ of simulation results, the mean and standard deviation of which we use in Eq. (6) as $\mu_{|p|}^{(H)}$ and $\sigma_{|p|}^{(H)}$ respectively. In the same manner, we compute a z-score $z_p^{(S)}$ using the same simulations, but take the measure of interest to be $\tilde{S}_p$ rather than $\tilde{H}_p$
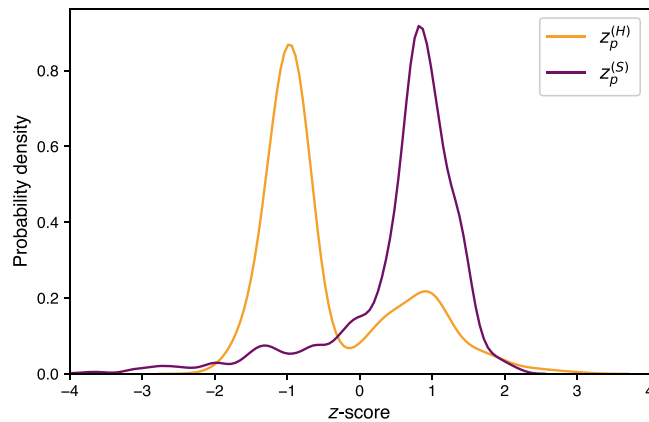
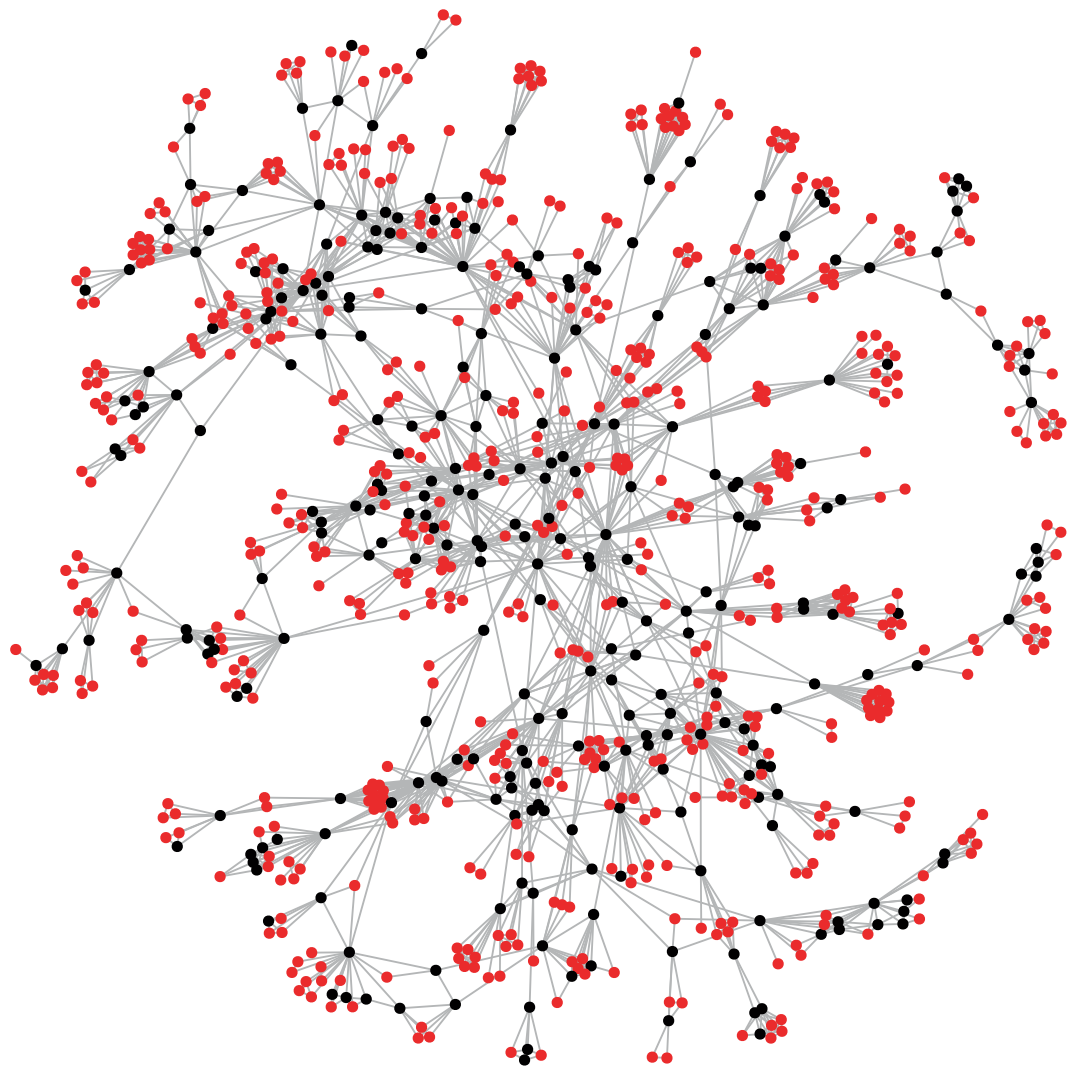$$z_p^{(S)} = \frac{\tilde{S}_p - \mu_{|p|}^{(S)}}{\sigma_{|p|}^{(S)}}.$$
(7)
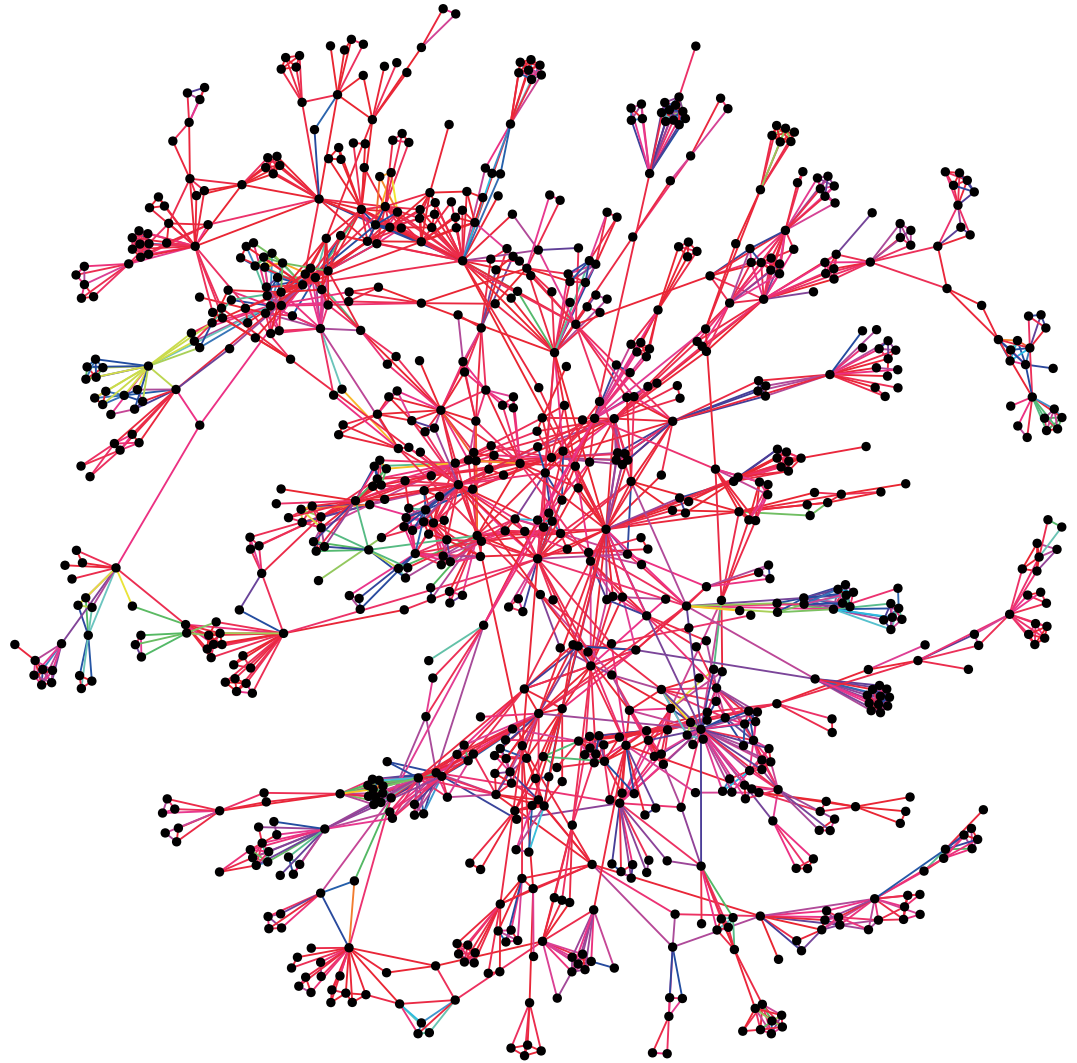
Figure 3. The giant component of the collaboration network used in this study, with edges colored according to interdisciplinary cosine similarity S_{ij} (Eq. (2)). The values on the edges range from $S_{ij} = 0$ (violet) to $S_{ij} = 1$ (red). Most edges have a high level of interdisciplinary similarity, as verified in the analysis in Results.

We plot kernel density estimated probability densities of Eq. (6) and Eq. (7) for the full collaboration network in Fig. 4. We can see from these results that research teams tend to be composed of people with more homogeneous backgrounds than expected by chance, both with respect to primary discipline and full research profile. In particular, the distribution of $z_p^{(H)}$ has its mass centered at $z = -1$, indicating that most research teams have $\tilde{H}_p$ about one standard deviation lower (more concentrated) than expected on average for an uncorrelated random network. Additionally, the distribution of $z_p^{(S)}$ has its mass centered at $z = +1$, suggesting that many research teams have an $\tilde{S}_p$ about one standard deviation above (more similar discipline vectors $\vec{x}_i$ than) what is expected for a random team configuration. These results suggest that, in the interdisciplinary area, research teams as a whole tend to be composed of researchers with similar research backgrounds.

Academic performance and collaboration diversity.   We apply the core-shell decomposition discussed in the Methods section to separate the active inter-group collaborators from the inactive ones, which is visualized in Fig. 5. The decomposition reveals a shell of 1,602 nodes and a core of 420 nodes, indicating that most nodes in the network only participate in a single collaboration, and a smaller portion actively work with multiple groups. To examine the associations between structural diversity of authors in the collaboration network and academic experience and performance, we plot the distributions of h-index and years of publication experience (the difference between the earliest and latest publication on record for the author) for the core and shell nodes in Fig. 6. The results indicate that researchers in the core tend to have a systematically higher h-index and more publication experience than those in the shell. To statistically validate this claim, we apply both Mann-Whitney and Kolmogorov-Smirnov tests (as done for the cosine similarity densities), finding that in all cases the results are statistically significant (h-index: Median 1 = 10, Median 2 = 4, $n_1 = 420$, $n_2 = 1602$, $U \gg 10$, $p \ll 0.01$, for one-tailed Mann-Whitney U test; $D = 0.31$, $p \ll 0.01$ for KS test); (publication years: Median 1 = 13, Median 2 = 9, $n_1 = 420$,

**Figure 4.** Probability densities of team disciplinary diversity *z*-scores in Eqs. (6) and (7). Both distributions suggest that research teams are more homogeneous than expected by chance.

Figure 5. The giant component of the collaboration network used in this study, with nodes colored according to core-shell classification. Core nodes are colored black, while shell nodes are colored red. This decomposition helps to separate the network into active and inactive collaborators.
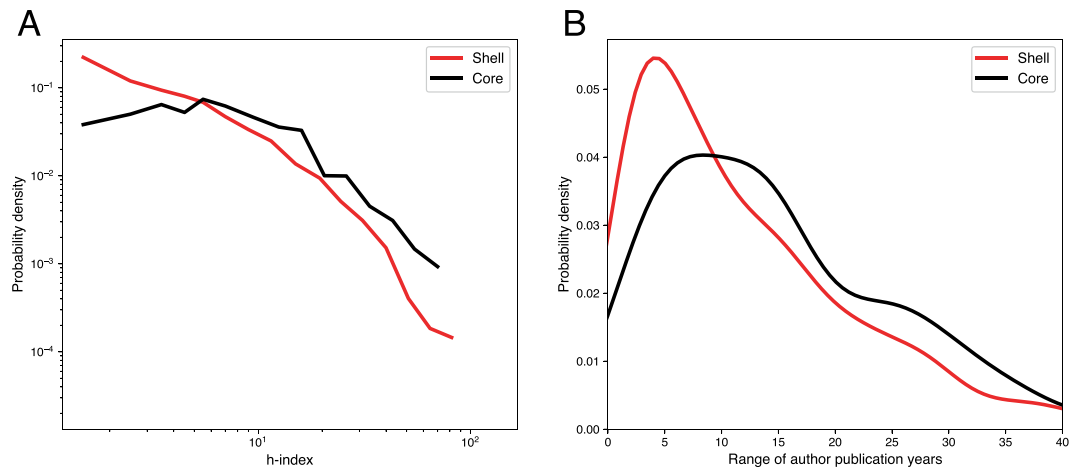
Figure 6. Distributions of (**A**) h-index and (**B**) years of publication experience for the core and shell in the decomposition shown in Fig. 5. The distributions for the core were significantly shifted above the distributions for the shell in both cases, indicating significantly more academic experience and success among the actively collaborating core researchers.

$n_2 = 1602$, $U \gg 10$, $p \ll 0.01$, for one-tailed Mann-Whitney U test; $D = 0.15$, $p \ll 0.01$ for KS test). These results suggest that, in the interdisciplinary area, the researchers who have longer working experience and better academic performance tend to be more active in collaborating with diverse groups on more projects.

Discussion

Disciplinary diversity is mainly reflected by individual researchers in an interdisciplinary research area.
Our results suggest that disciplinary diversity is better demonstrated at the individual level than the dyadic or group level in an interdisciplinary research area. This implies that perhaps interdisciplinary research topics attract researchers who have experience in multiple fields, but this does not necessarily lead to diverse collaborations. Research experience in multiple fields strengthens the flexibility and adaptability of a researcher for engaging in projects that cross disciplinary boundaries. The capability of connecting knowledge across different disciplines also enables researchers to develop novel questions and analysis methods, which are central to interdisciplinary research. One potential challenge faced by interdisciplinary researchers is the competing demands of time and effort for each field they participate in. The findings of this study indicate that interdisciplinary researchers involved in less than five disciplines can contribute relatively equally to all the fields they are engaged in. However, the capacity to contribute to all fields equally is diminished as the number of research fields they participate in increases. Our findings on the prevalence of high disciplinary diversity of individual researchers in this interdisciplinary research area highlight the importance of interdisciplinary training, which not only prepares individuals with a comprehensive knowledge base, but also supports them to collaborate in interdisciplinary fields.

Disciplinary homophily is stronger than diversity for collaboration in interdisciplinary research.
Despite the presumed benefits for collaboration with people from diverse academic backgrounds in interdisciplinary research areas, our study finds that researchers still prefer to collaborate with others who are alike in terms of their research background. Given that individual researchers tend to have interdisciplinary research backgrounds, we consider the multiple fields that individuals participate in while assessing pairwise similarity in the collaboration network, and we find that researchers prefer to collaborate with others who work in a similar set of fields. These findings indicate that homogeneity in pairwise collaborations is not constrained to the primary disciplines of individuals in interdisciplinary research, and that the diverse research experiences of individuals should be taken into consideration. Dyadic homogeneity and individual-level interdisciplinarity reduce transaction costs, ensure the diversity of the body of knowledge within a research group, and facilitate the development of a shared collaborative grounding. Our results may thus provide a theoretical contribution to understanding the development of collaboration in interdisciplinary research, as well as insights for characterizing interdisciplinary research. A previous study[47] considers the diversity of disciplines of researchers in a project as a dimension for defining interdisciplinary research. Based on the findings of this study, it is not necessary to have researchers from diverse disciplines in an interdisciplinary research, rather, disciplinary diversity can be reflected at the individual level instead of the group level.

Diversity in collaborating with multiple groups is beneficial.
Based on the core-shell analysis of the collaboration network, we find that researchers with a diverse neighborhood structure tend to have a better academic performance and longer working years. This makes sense, as researchers with reputable track records and more experience have a greater pool of resources that facilitate the development of research collaborations with diverse groups on multiple projects. In a complementary way, collaborating with many teams on more projects

can also enhance a researcher's academic performance, which unsurprisingly is positively associated with number of years publishing. However, confirming a causal relationship from this finding requires further research.

Conclusions

This study proposes novel and cost-effective measures for assessing disciplinary diversity at three scales within research collaborations in an interdisciplinary area. Our findings contribute to the conceptual, theoretical, and methodological aspects of understanding research collaboration in interdisciplinary areas.

Firstly, we introduce new measures for assessing disciplinary diversity at the individual, dyadic, and team levels based on the categories of researchers' past publications which could be further employed in future studies on other datasets. These measures could theoretically be applied to a wide variety of networks with categorical node metadata, but they are used in this study particularly for addressing disciplinary diversity. Secondly, a new cost-effective approach for identifying a core of nodes with diverse neighborhood structure in a network is proposed, which is especially effective on networks that are tree-like at the clique level, such as collaboration networks. In terms of theoretical contributions, this study strengthens our understanding of the underlying principles involved in developing collaborations in interdisciplinary research. Our results indicate that homophily with respect to researchers' academic backgrounds is an applicable principle for explaining collaborative relationships, and that additionally, individual interdisciplinarity and dyadic homogeneity together form the theoretical underpinnings of developing collaborations in interdisciplinary research. Thirdly, the findings of the study shed light on the nature of team formation in practice, as well as highlight the importance of interdisciplinary programs.

It is important to support the development of interdisciplinary programs at both the institutional and national levels, as researchers with interdisciplinary backgrounds can better contribute to interdisciplinary collaboration. Regarding team formation in interdisciplinary research, it is important to consider the diverse research experiences of individuals as well as the overlapping of individual disciplines among group members. Future studies are suggested to assess and compare the interdisciplinarity of researchers by considering the publication records as well as their citing and cited publication records, which can provide further evidence about the integration of knowledge from multiple areas and interdisciplinary contributions[20]. Further research is also needed to explore the factors affecting the success of research collaborations in interdisciplinary research areas.

Data availability

The datasets generated during and analysed during the current study are available from the corresponding author upon request.

References

1. Van Noorden, R. Interdisciplinary research by the numbers. *Nature* **525**, 306–307 (2015).
2. Clark, H. H. & Brennan, S. E. *Grounding in Communication*. (1991).
3. Hertzum, M. Collaborative information seeking: The combined activity of information seeking and collaborative grounding. *Information Processing & Management* **44**, 957–962 (2008).
4. Campbell, L. M. Overcoming obstacles to interdisciplinary research. *Conservation Biology* **19**, 574–577 (2005).
5. Lewin, K. *Resolving social conflicts; Selected papers on group dynamics*. (1948).
6. Ruef, M., Aldrich, H. E. & Carter, N. M. The structure of founding teams: Homophily, strong ties, and isolation among us entrepreneurs. *American Sociological Review* 195–222 (2003).
7. Ibarra, H. Homophily and differential returns: Sex differences in network structure and access in an advertising firm. *Administrative Science Quarterly* 422–447 (1992).
8. Araújo, E. B., Araújo, N. A., Moreira, A. A., Herrmann, H. J. & Andrade, J. S. Jr Gender differences in scientific collaborations: Women are more egalitarian than men. *PLOS One* 12 (2017).
9. McPherson, M., Smith-Lovin, L. & Cook, J. M. Birds of a feather: Homophily in social networks. *Annual Review of Sociology* **27**, 415–444 (2001).
10. Baggs, J. G., Ryan, S. A., Phelps, C., Richeson, J. & Johnson, J. The association between interdisciplinary collaboration and patient outcomes in a medical intensive care unit. *Heart & Lung: The Journal of Acute and Critical Care* **21**, 18–24 (1992).
11. Fewster-Thuente, L. & Velsor-Friedrich, B. Interdisciplinary collaboration for healthcare professionals. *Nursing Administration Quarterly* **32**, 40–48 (2008).
12. Petri, L. Concept analysis of interdisciplinary collaboration. *Nursing Forum*, vol. **45**, 73–82 (Wiley Online Library, 2010).
13. Van Rijnsoever, F. J. & Hessels, L. K. Factors associated with disciplinary and interdisciplinary research collaboration. *Research Policy* **40**, 463–472 (2011).
14. Cummings, J. N. & Kiesler, S. Who collaborates successfully? prior experience reduces collaboration barriers in distributed interdisciplinary research. In *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work*, 437–446 (2008).
15. Huutoniemi, K., Klein, J. T., Bruun, H. & Hukkinen, J. Analyzing interdisciplinarity: Typology and indicators. *Research Policy* **39**, 79–88 (2010).
16. Porter, A., Cohen, A., David Roessner, J. & Perreault, M. Measuring researcher interdisciplinarity. *Scientometrics* 72 (2007).
17. Yong, K., Sauer, S. J. & Mannix, E. A. Conflict and creativity in interdisciplinary teams. *Small Group Research* **45**, 266–289 (2014).
18. Adams, J., Loach, T. & Szomszor, M. Interdisciplinary research: Methodologies for identification and assessment. *Digital Research Reports* (2016).
19. Qin, J., Lancaster, F. W. & Allen, B. Types and levels of collaboration in interdisciplinary research in the sciences. *Journal of the Association for Information Science and Technology* **48**, 893–916 (1997).
20. Porter, A. & Rafols, I. Is science becoming more interdisciplinary? Measuring and mapping six research fields over time. *Scientometrics* **81**, 719–745 (2009).
21. Rinia, E. J., Van Leeuwen, T. N., Bruins, E. E., Van Vuren, H. G. & Van Raan, A. F. Measuring knowledge transfer between fields of science. *Scientometrics* **54**, 347–362 (2002).
22. Song, C.-H. Interdisciplinarity and knowledge inflow/outflow structure among science and engineering research in korea. *Scientometrics* **58**, 129–141 (2003).

23. Morrison, P. S., Dobbie, G. & McDonald, F. J. Research collaboration among university scientists. *Higher Education Research &. Development* **22**, 275–296 (2003).
24. Gowanlock, M. & Gazan, R. Assessing researcher interdisciplinary: A case study of the University of Hawaii NASA Astrobiology Institute. *Scientometrics* **94**, 133–161 (2013).
25. Guan, R., Yang, C., Marchese, M., Liang, Y. & Shi, X. Full text clustering and relationship network analysis of biomedical publications. *PLOS One* 9 (2014).
26. Ravikumar, S., Agrahari, A. & Singh, S. Mapping the intellectual structure of scientometrics: A co-word analysis of the journal scientometrics (2005–2010). *Scientometrics* **102**, 929–955 (2015).
27. Sedighi, M. Application of word co-occurrence analysis method in mapping of the scientific fields (case study: the field of infometrics). *Library Review* (2016).
28. Parinov, S. & Kogalovsky, M. Semantic linkages in research information systems as a new data source for scientometric studies. *Scientometrics* **98**, 927–943 (2014).
29. Carpenter, M. P. & Narin, F. Clustering of scientific journals. *Journal of the Association for Information Science and Technology* **24**, 425–436 (1973).
30. Newman, M. *Networks* (Oxford University Press, 2018).
31. Burt, R. S. *Structural holes: The social structure of competition* (Harvard University Press, 2009).
32. Lin, N. *Social capital: A theory of social structure and action*, vol. 19 (Cambridge University Press, 2002).
33. Reagans, R. & McEvily, B. Network structure and knowledge transfer: The effects of cohesion and range. *Administrative Science Quarterly* **48**, 240–267 (2003).
34. Rodan, S. & Galunic, C. More than network structure: How knowledge heterogeneity influences managerial performance and innovativeness. *Strategic Management Journal* **25**, 541–562 (2004).
35. Newman, M. Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences* **101**, 5200–5205 (2004).
36. Wagner, C. S. & Leydesdorff, L. Network structure, self-organization, and the growth of international collaboration in science. *Research Policy* **34**, 1608–1618 (2005).
37. Acedo, F. J., Barroso, C., Casanueva, C. & Galán, J. L. Co-authorship in management and organizational studies: An empirical and network analysis. *Journal of Management Studies* **43**, 957–983 (2006).
38. Guimera, R., Uzzi, B., Spiro, J. & Amaral, L. A. N. Team assembly mechanisms determine collaboration network structure and team performance. *Science* **308**, 697–702 (2005).
39. Moody, J. The structure of a social science collaboration network: Disciplinary cohesion from 1963 to 1999. *American Sociological Review* **69**, 213–238 (2004).
40. Dahlander, L. & McFarland, D. A. Ties that last: Tie formation and persistence in research collaborations over time. *Administrative Science Quarterly* **58**, 69–110 (2013).
41. Zhang, L., Liu, X., Janssens, F., Liang, L. & Glänzel, W. Subject clustering analysis based on isi category classification. *Journal of Infometrics* **4**, 185–193 (2010).
42. Zhang, L., Rousseau, R. & Glänzel, W. Diversity of references as an indicator of the interdisciplinarity of journals: Taking similarity between subject fields into account. *Journal of the Association for Information Science and Technology* **67**, 1257–1265 (2016).
43. Kumar Nayak, I. On diversity measures based on entropy functions. *Communications in Statistics - Theory and Methods* **14**, 203–215 (1985).
44. Stirling, A. A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society Interface* **4**, 707–719 (2007).
45. Rhoades, S. A. The Herfindahl-Hirschman index. *Federal Reserve Bulletin* **79**, 188 (1993).
46. Gray, R. M. Entropy and Information Theory (Springer Science & Business Media, 2011).
47. Aboelela, S. W. *et al.* Defining interdisciplinary research: Conclusions from a critical review of the literature. *Health Services Research* **42**, 329–346 (2007).

Author contributions

Both authors contributed significantly to this work. S.F. developed the study design, A.K. performed the experiments, S.F. and A.K. analysed the results and wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information