Information Processing and Management 62 (2025) 104170

Contents lists available at ScienceDirect

Information Processing and Management

journal homepage: www.elsevier.com/locate/infoproman

The attention inequality of scientists: A core-periphery structure perspective

Haoyang Wang [a], Win-bin Huang [b,c,d,*], Yi Bu [b,c,d,*]

[a] School of Information, University of Texas at Austin, Austin, TX 78701, USA
[b] Department of Information Management, Peking University, Beijing 100871, China
[c] Center for Informationalization and Information Management Research, Peking University, Beijing 100871, China
[d] Center for Digital Intelligence Science and Education Research, Peking University Chongqing Research Institute of Big Data, Chongqing 401332, China

ARTICLE INFO

Keywords:
Core-periphery structure
Attention inequality
Citation network
Rich club phenomenon
Scientometrics

ABSTRACT

This study investigates the dynamics of scientific attention within author citation networks, utilizing the Microsoft Academic Graph dataset. Three author citation networks were constructed within the domains of nanoscience, chemical physics, and human-computer interaction. We apply analytical measurements to reveal core-periphery structures, indicating a growing disparity in scientific interactions. Our analysis highlights a concerning trend: while connections among prominent authors are strengthening, interactions among "ordinary" scientists remain relatively weak. This trend is further corroborated by the application of the network percolation method. After removing the prominent authors in the citation networks, multilayered and complex relationships among authors are revealed. We observe a decreasing trend of connection strength among relatively "ordinary" authors. The observed inequality of attention raises significant concerns about neglecting diverse voices within the scientific community. In response to these phenomena, our research emphasizes the importance of cultivating an inclusive scientific environment for early-career and underrepresented scholars, aiming for long-term sustainability in the scientific community.

1. Introduction

In cognitive science, attention can be described as a mental activity that involves making selections and concentrating focus (Kahneman, 1975; Klamer & Dalen, 2002). During the process of scientific publications, cognitive attention seamlessly transforms into the referencing mechanism within the scientific ecosystem, leading to the fundamental phenomenon of citation. While traditional measurements such as author output, *h*-index, and social media altmetrics can measure the attention gained by an author and provide valuable insights, our research focuses on citation practices. The exploration of the motives and rationale behind citing a research paper has been a subject of extensive study within the disciplines of both sociology and information science in previous research (Garfield, 1965; Kunnath et al., 2021; Oppenheim & Renn, 1978). However, when scientists reference a publication, their decision might go beyond merely the quality of this publication, such as factors of author's reputation, personal relationships. This brings us to the self-reinforcing phenomenon in scientific impact known as the Matthew Effect, which describes the tendency for scientists' past

* Corresponding authors.
E-mail addresses: huangwb@pku.edu.cn (W.-b. Huang), buyi@pku.edu.cn (Y. Bu).

https://doi.org/10.1016/j.ipm.2025.104170
Received 11 April 2024; Received in revised form 15 February 2025; Accepted 24 March 2025
Available online 1 April 2025

accomplishments to contribute to their future success (Merton, 1968). The Matthew Effect has been extensively studied in the field of science of science, represented as cumulative advantage (Allison et al., 1982). Prior success can enhance future citations (Azoulay et al., 2014) and increase the likelihood of receiving science funding (Liao, 2021). In an academic landscape marked by a substantial surge in the volume of publications, this complex interplay of factors may contribute to uneven citation patterns. Consequently, content that may be genuinely relevant to the same specific topic can often go unnoticed or receive inadequate citation, thus attention, visibility, and recognition can sometimes overshadow the inherent merit of the research itself.

Understanding these dynamics is crucial when considering the broader context of networks. Scientific, technological, socio-economic, ecological, and geological systems can all be effectively depicted using network approaches (Network Science by Albert-László Barabási, 2016). Within the scientific community, the author citation network represents a specific type of knowledge exchange network. It becomes a vital tool for analyzing how attention is distributed and how certain patterns, such as the Matthew Effect, influence the flow of information within the scholarly communication network.

In the late 1990s, Borgatti and Everett (2000) first quantitatively defined the core-periphery network structure. Their basic concept is that the "core" is a densely connected network entity that cannot be subdivided into cohesive factions, while the "periphery" represents a sparsely connected set of nodes, typically located on the periphery of the network. The core-periphery structure is widely identified in many fields, including social networks (Alba & Moore, 1978; Corradini, 2024; Granovetter, 1978; Kojaku & Masuda, 2017), neural networks (Bassett et al., 2013; Bassett & Bullmore, 2009; Tunç & Verma, 2015), protein-protein interaction networks (Bruckner et al., 2015; Da Silva et al., 2008), transportation networks (Lee et al., 2014; Rossa et al., 2013), economic networks (Kojaku & Masuda, 2017; Krugman, 1996), and bibliographical networks (Bianchi et al., 2021; Freeman, 1984; Mullins et al., 1977; Safadi et al., 2021; Wedell et al., 2022; Willis & McNamee, 1990). The study of core-periphery structure provides information about the network at an intermediate level, allowing researchers to better understand the complex structure of networks. This segregation enables a more accurate categorization of the functional and dynamic roles of nodes based on their structural positions (Gallagher et al., 2021).

In science of science studies, core-periphery structure has been commonly observed in many types of networks including citation networks (Chen & Guan, 2016; Weng & Daim, 2012; Zhang & Zhang, 2015), co-citation networks (Sedita et al., 2020), bibliographic coupling networks (Painter et al., 2021), and co-authorship networks (Zelnio, 2012). The core-periphery structure can be examined in networks across diverse contexts, and common topics related to this structure include the rich club phenomenon (Colizza et al., 2006), knowledge diffusion (Balland et al., 2010), and innovation (Eder, 2019; Safadi et al., 2021). The differentiation between core and periphery in a network reflects the distinct roles of nodes, while observing the evolution of core and periphery over time helps to discern trends and evolutionary patterns among specific groups within the network. As highlighted by Hojman and Szeidl (2008), there is a strong connection between inequality and the core-periphery structure. The core-periphery model is effective in illustrating attention inequality by showing how core nodes receive significantly more attention and are significantly more influential compared to peripheral nodes. This model provides a clear representation of the uneven distribution of attention and how it is spread across the network.

There are various techniques and algorithms available for identifying and extracting core-periphery structures from networks. Gallagher et al. (2021) summarize two main types of measurements. The first method, known as the "two-block model", is based on a definition originally suggested by Borgatti and Everett (2000). Another way to measure core-periphery structure is k-cores decomposition (Alvarez-Hamelin et al., 2005), a commonly used method in network analysis. This method proposes a hierarchical core-periphery structure consisting of layers that are nested within each other and converge toward a central core. The typical methods are rich clubs (Zhou & Mondragon, 2004), nestedness (Almeida-Neto et al., 2008; Atmar & Patterson, 1993), and onion layers (Schneider et al., 2011; Wu & Holme, 2011). Assortativity, a common indicator capturing the tendency of similar nodes to connect, plays a crucial role in recognizing the core-periphery structure from subtly different perspectives (Catanzaro et al., 2004). The core-periphery structure concept posits that the periphery nodes are supposed to connect to the dissimilar core nodes, a notion that might initially appear to be a slight contradiction in terms of assortativity. However, assortative networks and core-periphery structures often coexist and reinforce each other (Ahuja et al., 2012; Khanna & Guler, 2022). Therefore, we are driven to incorporate the assortativity method to gain a more diverse and multifaceted understanding of the structure of the author citation network.

2. Research objectives

As described earlier, prior core-periphery structure studies in the science of science have analyzed a variety of networks, focusing on journal articles, patents, institutions, and countries. However, studies at the author level are less common, and even existing author-level studies are predominantly focused mostly on the author co-citation network (Khelfaoui & Gingras, 2019). To this end, within the broader context of attention mechanisms, this study focuses on the core-periphery structure of author citation networks. Employing multiple metrics to assess the core-periphery structure in the author citation network, our approach enables a more detailed exploration of the dynamics of knowledge dissemination, attention inequality, and their valuable implications for the scientific ecosystem. This study aims to provide insights into the science policies and practices that promote the long-term sustainability and vibrancy of scientific inquiry and collaboration.

3. Data

This paper adopts the Microsoft Academic Graph (MAG) as an empirical bibliometric dataset (Wang et al., 2020). The MAG dataset has been widely used in bibliometric studies (e.g., Huang et al., 2022; Zhao et al., 2021) and is known for its broad coverage, openness, and well-designed domain classification system (Paszcza, 2016). MAG offers precise citation records of publications and the authorship information which are vital in our study. MAG has completed the author disambiguation process which adopts the state-of-the-art machine learning methods and incorporates supplementary public information such as personal websites and scholars' *curricula vitae* (Wang et al., 2020). Therefore, we are confident in our ability to distinguish research outputs by different authors. The MAG copy we adopted ranges from 1800 to May 2022.

MAG's hierarchical disciplinary classification system consists of six levels, with each publication assigned to one or more levels based on its research topic and subject. The assignment of fields of study labels is automated using a data-driven approach, with Sinha et al. (2015) reporting an impressive accuracy rate of 98 %. The field of study labels are then organized into the six-level hierarchical structure. These six levels are labeled as L0 (highest level), L1, L2, L3, L4, and L5 (lowest level), with the L0 and L1 levels being manually curated to ensure accuracy. In practice, it is common for a single research paper to be associated with multiple disciplines, which enhances the breadth of disciplinary coverage. In this study, due to the limitations of computational resources, we focus on three typical L1-level disciplines (see details in the later paragraphs) to study the changing trend of potential core-periphery structure. The L1-level (containing 292 disciplines) represents the second-highest level in the MAG classification system, and is typically used to group related fields of study within a broad subject area.

The disciplines we focus on are nanoscience, chemical physics, and human-computer interaction. **Nanoscience (Nano)** refers to the science, engineering, and technology conducted at the nanoscale (10^{-9} meter). Nanoscience emerges from the interdisciplinary formation of different disciplines such as surface science, organic chemistry, molecular biology, semiconductor physics, and molecular engineering.

Chemical physics (Chem. Phys.) is a complex combination of chemistry and physics that uses physics research methods to study chemical processes. It originated early in the nineteenth century and accumulated a vast amount of provocative and far-reaching research that impacted many surrounding disciplines in the later decades. Until the late nineteenth century, this discipline experienced significant change, marked by a rapid increase in the number of research papers (Hiebert, 1996). The interdisciplinary nature of chemical physics inspires us to select it as a target domain in the following analyses.

The subject of **Human Computer Interaction (HCI)** refers to a multidisciplinary field that focuses on designing and evaluating computer systems for human use, especially their interactions (Sinha et al., 2010). While the first conception of HCI originated in the 1940s, combining knowledge from computer science and psychology, etc., HCI research boomed in the late twentieth century, making it an ideal discipline for our analysis.

4. Methods

4.1. Author citation network

This study focuses on the core-periphery structure in the author citation network. The author citation network constructed in this study is an undirected weighted graph where nodes represent disambiguated authors and edge weights represent the number of mutual citation links between two authors. Self-loops are included in our study to retain the information from the author's self-citations. Author-level citations are derived from paper-paper citation relationships with a full counting-based method (Waltman, 2016), which assigns each co-author the same credit. In an author citation network, for example, if author A cites author B five times and author B cites author A seven times, the edge weight between nodes A and B in this author citation network would be $5 + 7 = 12$.

In this study, author citation networks are constructed separately for the three disciplines. As of May 2022, the latest time covered by our dataset, the basic descriptive statistics of the author citation networks for the three disciplines were as presented in Table 1.

To improve validity, we opt to commence our investigation during a period when the three disciplines experienced more intensive publication and their development gradually matured. This decision led us to establish a focal period starting from 1980, as this timeframe aligns with the heightened publication activity and developmental milestones. It is important to clarify that our dataset spans a continuous range of years, encompassing time cohorts both before and after 1980. However, our primary focus lies in the period after 1980, to capture the most significant developments. It is worth noting that our author citation networks are constructed up to a specific year using citation records up to that year, thus capturing a meaningful cumulative snapshot of the evolutionary dynamics within the academic landscape.

4.2. Null model

In the following section, we calculate the rich club coefficient of the network and the connection strength of the author citation network. To establish a comparative benchmark, we construct a null model of the author citation network by swapping citation relationships between pairs of publications. By comparing the relevant metrics between the real and random author citation networks,

Table 1
The descriptive statistics of the author citation network. The three disciplines under focus exhibit similar levels of citation intensity and network scale.

Descriptive statistics	Nano	Chem. Phys.	HCI
# of nodes (authors)	1547,918	607,127	617,894
# of edges	309,471,562	38,417,632	20,367,818

we can identify significant differences relative to a random network with similar properties. The null model assists in assessing whether the real network has any significant structural properties and provides fundamental yet latent information on degree orders with respect to an ideally random but analogous network. Therefore, it is necessary to use the null model to normalize as the original values in a real network may not be comparable.

When constructing the null model, we randomized citation relationships at the publication level rather than directly swapping edges in the author citation network. This decision is based on the fundamental nature of how author citation networks are formed in the real world. Since the author citation network emerges as a cumulative structure from paper-level citations over time, directly altering author-level connections may introduce artificial distortions that do not reflect how citation networks naturally evolve. This ensures that the chronological progression of scientific knowledge remains intact while still disrupting specific citation relationships. Additionally, this approach helps to preserve the in-degree and out-degree distributions of authors.

As mentioned earlier, we construct author citation networks based on publication citation relationships. In the randomization process, we select two pairs of publication citation records in MAG where the two citing papers were published in the same year (Fontana et al., 2020) and randomly exchange the citation relation between the two pairs of publications, repeating the total number of edges in the network multiplied by 100 times to reach the state of sufficient randomness. For example, suppose paper A cites B, paper C cites D, and A and C were published in the same year in a real paper citation network. In the null model, we then switch the citation relationship—A is connected to D, and B is connected to C. This swapping technique is a well-established method in network science (Clough et al., 2015; Foster et al., 2010; Hao & Kovács, 2024) and has been widely applied to evaluate the significance of structural properties in complex networks.

4.3. Core-periphery structure

Corresponding to the definition of the core-periphery structure mentioned in the Introduction section, we construct an ideal core-periphery structure (Fig. 1) for better intuitive understanding. In this toy example that illustrates an author citation network, following the concept that a set of peripheral nodes with sparse connections is mainly linked to a densely interconnected core, nodes 1–4 are considered core authors while nodes 5–10 are considered periphery authors.

To better identify the core-periphery structure in the citation network, this research carefully selects three distinct yet inter-connected perspectives, namely connection strength, rich club, and assortativity perspectives. This selection is inspired by Gallagher et al. (2021), who summarized two types of strategies quantifying the core-periphery structure in a network, namely “two-block model” and “hierarchical and layered structure.” The first hypothesizes that nodes are arranged into two groups, i.e. the core and the periphery, while the latter emphasizes that core nodes have a hierarchical nature. These two strategies correspond to connection strength and the rich club perspectives we adopt, respectively. The connection strength perspective measures the strength of connections between nodes in the network, while the rich club perspective focuses on highly interconnected nodes that form a “rich club” in the network. As mentioned in the Introduction section, combining both the characteristics of hub-and-fringe and hierarchical structure, we further expand this theoretical framework by using assortativity metrics in our research. Assortativity metrics is applied to capture the core-periphery feature of the whole network which examines the tendency of nodes with similar attributes to connect with each other. Through these three perspectives, we conduct a relatively multifaceted core-periphery structure analysis and obtain similar but cohesive results.

4.3.1. Connection strength

To identify authors who are considered "rich" within the author citation network, we examine the degree of each author node. In the undirected graph we constructed for this study, the degree of a node represents the total number of connections it has with other author nodes. This degree of an author node reflects both the number of times the author has been cited by colleagues and the number of times the author has cited others. Extant studies have pointed out that citing-cited flows reflect shifts of attention in the scientific ecosystem (Zhai et al., 2018) and reveal the knowledge import-export patterns among science workers (Hessey & Willett, 2013). That being said, we suppose that the total degree of an author node (the sum of all connections) may represent patterns of attention received (being cited) and offered (citing) by this author within his/her scientific community. Authors with a high degree are more well-connected within the scientific community, placing them in a prioritized class and justifying their characterization as “rich” authors; conversely, authors with a relatively low degree have less impact on attention shifts within the community, aligning them with what we term “ordinary” authors.

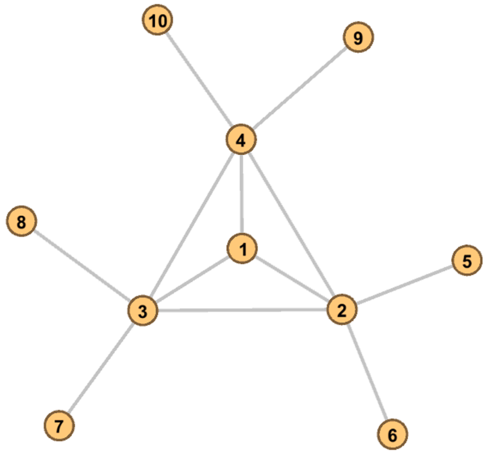

Fig. 1. An example of an ideal core-periphery structure in author citation network. Nodes 1–4 are considered perfectly classified as core nodes, as a fully connected structure is observed among them. In contrast, the other nodes (nodes 5–10) exhibit minimal connections with these four core nodes.

Given the variability in the degree distribution among authors, it is quite challenging to establish a universally applicable threshold on nodes' degrees to define a "rich" author. Simply choosing the authors with the greatest fixed number or percentage of degrees in these networks as "rich" people may oversimplify the division method and might not accurately portray the social standing of authors.[1] Thus, inspired by extant studies (Guo et al., 2015), we here rank all authors based upon their total degree in descending order and calculate the cumulative degree, rather than the original degrees, in the network. We then carefully select nodes whose cumulative degree is greater than a certain threshold (10 % in the later analyses) as "rich" authors and classify the rest as "ordinary" authors.[2] The choice of the 90th percentile, or the top 10 %, as a specific threshold is supported by its widespread use in the literature to distinguish between the most influential subjects and the rest (Bornmann, 2014; Waltman et al., 2012). This threshold effectively captures a significant yet manageable proportion of the population, allowing for a focused analysis of the elite group while maintaining statistical robustness. The pseudocode of this process can be found in the appendix.

The connection strength between rich and rich, rich and ordinary, and ordinary and ordinary authors is calculated in the author citation network based on the sum of edge weights. To compare the connection strength among citation networks from different years, we employ the randomized null model discussed earlier for normalization. We then calculate the sum of edge weights between rich and rich, rich and ordinary, and ordinary and ordinary authors in the randomized network, and divide the connection strength in the real author citation network by the connection strength in the randomized network to obtain the final normalized connection strength.

4.3.2. Rich club

This study uses the weighted rich club coefficient to quantify the relative connection density among rich authors within the author citation network (Alstott et al., 2014). The weighted rich club coefficient is defined as follows:

$$\phi^w(k) = \frac{W_{>k}}{\sum_{l=1}^{E_{>k}} w_l^{rank}} \tag{1}$$

where $\phi^w(k)$ denotes the weighted rich club coefficient of the author citation network at degree k. $E_{>k}$ represents the number of edges between nodes which have a degree greater than k, while the $W_{>k}$ is the sum of the weights of the edges between these nodes. The denominator is the sum of the weights of the top $E_{>k}$ maximum weight edges in the network, where the notation w_l^{rank} follows that $w_l^{rank} \geq w_{l-1}^{rank}$ ($l = 1,\ldots,E_{>k}$), denoting the descending weight of the edges. In other words, the $\phi^w(k)$ measures the ratio of the total weights of edges between nodes with a degree larger than k to the sum of the weights of the top $E_{>k}$ edges, ranked in descending order by weight. As a result, the ratio $\phi^w(k)$ helps to calculate the potential "density" of connections among specific authors.

[1] Consider an extreme case where the author citation network consists of 1,000 authors. Suppose that one rich author has a degree of 1,000 (including self-loop) while the remaining 999 authors who only link to the rich author have a degree of one. If we use a fixed value of their original node degree values/percentiles in the network to define "richness", many authors with a degree of one will easily be mistakenly considered as "rich". However, with the utilization of the cumulative degree-based strategy, we can identify and address this abnormal scenario.

[2] Consider an author citation network with a descending degree sequence of {10, 8, 5, 4, 3, 3, 2, 2, 2, 1, 1, 1, …, 1}, where the total degree sum is 200. The rich author threshold is set at 10% of the total degree sum, equating to 20. To determine the rich authors, we rank authors in descending order of degree and progressively sum their values until exceeding the threshold. The first author, with a degree of 10, does not reach the threshold and is classified as rich. Adding the second author (10 + 8 = 18) still falls below the threshold, so they are also classified as rich. Including the third author (10 + 8 + 5 = 23) surpasses the threshold, at which point the process terminates. Consequently, the top three authors (with degrees 10, 8, and 5) are designated as rich authors, while all remaining authors are categorized as ordinary authors.

The weighted rich club coefficient is computed for both the real and the random networks (using the null model method as aforementioned). To assess the extent to which the intensity of the rich club phenomenon in the real network exceeds that of the random model, the normalized rich club coefficient (divided weighted rich club coefficient) is obtained by dividing the weighted rich club coefficient in the real network $\phi^w_{real}(k)$ by the weighted rich club coefficient in the random network $\phi^w_{random}(k)$. The resulting value indicates whether the rich club phenomenon exists in the real network, and, if so, to what extent. A value greater than one signifies the presence of the rich club phenomenon in the real network. A higher value of the normalized rich club coefficient connotes a stronger rich club phenomenon in the real network compared to the random network.

To quantitatively compare different $\phi^w(k)$ curves, we need a specific metric to characterize the curve. The $\phi^w(k) \sim k$ coordinate inspires us to search for a best-fit curve to understand the underlying relationships. Considering various sections of the curve, we put forward two measurements as follows:

Measurement 1 (α): In practice, using all data points for fitting does not typically yield a simple, easy-to-understand, and interpretable function (see Fig. 6). However, focusing on subsets of the data, particularly those corresponding to ordinary authors with smoothly increasing data samples, a more straightforward relationship emerges. Specifically, the outcome curves tend to appear mostly linear when utilizing the bottom 90 % of the data points. This approach is consistent with our classification method, which designates rich authors as those within the top 10 % based on cumulative degree, ensuring uniformity throughout our analysis. The linear trend indicates a simpler relationship between the variables under consideration. By concentrating on subsets that exhibit traits reflecting ordinary authors, we are able to identify clearer and more interpretable results. The fitted slope from OLS, $\boldsymbol{\alpha}$, is derived as the rate at which the normalized rich club coefficient increases as the author degree increases by one. Specifically, $\boldsymbol{\alpha}$ measures the change in the normalized rich club coefficient per unit increase in author degree. Practically, the OLS is implemented using the *LinearRegression* method from the Python library *sklearn* (Pedregosa et al., 2011).

Measurement 2 (s): For the normalized rich club coefficient curves, the curved section is excluded from calculation of $\boldsymbol{\alpha}$. Nonetheless, we still need a metric to describe this section, facilitating a more effective comparison of connection density among authors with larger degrees. The metric **s** quantifies the relative area under the curve, computed for the segment of the curve with a cumulative degree greater than 90 % (corresponding to the rich authors) and <99.9 %.[3] Specifically, **s** is calculated by dividing the integral under the curve by the area encompassing the curve:

$$s = \frac{\int_{k_{90\%}}^{k_{99.9\%}} \phi^w(k)dk}{\phi^w(k)_{max}(k_{99.9\%} - k_{90\%})} \tag{2}$$

where $\phi^w(k)$ denotes the normalized rich club coefficient and the $\phi^w(k)_{max}$ corresponds to the maximum value of $\phi^w(k)$ for the range of degrees, k, limited to cumulative values between 90 % ($k_{90\%}$) and 99.9 % ($k_{99.9\%}$).

We use the *simps* function from the Python library *SciPy* to calculate the integral (Virtanen et al., 2020). The metric **s** ranges between [0,1], with higher values indicating a more pronounced normalized rich club phenomenon among rich authors, reflecting a more intense rich club effect within affluent communities. The range of normalized rich club coefficients used for calculating **s** is selected to focus on rich authors, thereby providing a better description of the rich club phenomenon across affluent communities.

4.3.3. Assortativity

First introduced in 2002, assortativity measures the tendency of nodes to connect with other nodes that have similar properties within a network (Newman, 2002; Noldus & Van Mieghem, 2015). The quantification of assortativity has two distinct, though akin, perspectives, namely topologies and node properties. Since this paper focuses on network topology, we adopt the simplest approach of measuring network topology, specifically node degree. In this context, the degree assortativity index assesses whether nodes with similar degrees are more likely to connect with each other. A network is considered assortative if nodes with higher degrees tend to connect more frequently with other high-degree nodes.

To calculate degree assortativity in the author citation network, denoted as $N = (\mathrm{V}, \mathrm{E})$, where V is the set of nodes representing authors and E is the set of citing edges in the network, we define P(v) as the sum of weights of the edges between a node v and its adjacent neighbors. Then we define the mixing matrix, M_{ij}, as representing the number of edges between nodes with property $\mathrm{P}(i)$ and $\mathrm{P}(j)$. To normalize the mixing matrix, we divide M_{ij} by total number of edges $|\mathrm{E}|$ to derive e_{ij}, where $e_{ij} = \frac{M_{ij}}{|\mathrm{E}|}$. Now we denote proportion of edges (u, v) such that $\mathrm{P}(u) = P(i)$ as $a_i = \sum_j e_{ij}$, while denoting proportion of edges (u, v) such that $\mathrm{P}(v) = P(i)$ as $b_i = \sum_j e_{ji}$. Meanwhile, σ_a and σ_b represent the standard deviation of $\mathrm{P}(i)a_i$ and $\mathrm{P}(i)b_i$ respectively. Based on the Pearson correlation coefficient, we define the assortativity coefficient r for this property P as follow.[4]

$$r = \frac{\sum_{i,j} P_i P_j (e_{ij} - a_i b_j)}{\sigma_a \sigma_b} \tag{3}$$

Empirically, we utilize the *degree_assortativity_coefficient* method from the Python package *networkx* to calculate the degree

[3] This is attributed to the significantly skewed distribution of authors' citation numbers in our exploratory data analysis. To enhance accuracy and reliability, we exclude outlier authors within the top 0.1 % cumulative degree.

[4] The calculation process for degree assortativity closely follows the tutorial provided by the Python package network: https://networkx.org/nx-guides/content/algorithms/assortativity/correlation.html

assortativity of the author citation network (Hagberg et al., 2008).

Our methodology incorporates a multifaceted analysis of core-periphery structures, examining connection strength, rich club phenomenon, and degree assortativity. While some current approaches may focus on one of these aspects, our approach provides a more nuanced understanding of the dynamics within scientific networks.

5. Results

5.1. Overview

To analyze the overall core-periphery structures in author citation networks, we first employ the connection strength perspective to identify the various interplay patterns between different authors.

Fig. 2 presents the distribution of the authors' degree (sum of in- and out-degree) in the particular three disciplines. This is visualized by smoothing the curves using logarithmic bins and taking logarithms for both the vertical and horizontal axes. The degree distribution of authors exhibits deviations from the theoretical power-law distribution, as evidenced by the observed "hook" shape on the left side of the curve (Milojević, 2010). This deviation among authors with lower degrees indicates that the author citation network is not scale-free.

Given the unbalanced distribution of authors' degrees, the differentiation between rich and ordinary authors is determined using the cumulative degree of the authors (Methods). To examine the core-periphery structure in author citation networks from a holistic perspective, we consider the sum of the edge weights between rich-rich, rich-ordinary, and ordinary-ordinary author pairs in the three disciplines of Nano, Chem. Phys., and HCI. The results are normalized using a random network as a null model. This normalization allows for insights into citation intensity and hierarchical organization within scholarly communities, highlighting relative connection strengths among different groups of authors. This approach offers an understanding of verified connection patterns across core nodes (rich authors) and periphery nods (ordinary authors). Table 2 describes the normalized connection strength in these disciplines.

In all three disciplines, the connection strength between rich-rich authors exceeds 1, even reaching up to 4, indicating denser connectivity among the affluent authors. On the other hand, the connection strength among ordinary-ordinary authors remains around 0.9, which is <1. This suggests that connections among ordinary authors are weaker compared to what is observed in randomized networks.

We shift our research focus to the formation of core-periphery structures in author citation networks by constructing separate networks of author citation relationships up to a specific year and analyzing the trends in this structure over time.

Fig. 3 shows the trend of rich authors. The curve representing both the percentage and absolute number of rich authors shows that, although the number of rich authors has risen rapidly, the percentage has decreased. This trend highlights a concentration of rich authors within a smaller community, alongside a substantial increase in the ordinary number of authors on a larger scale. Despite the rising number of rich authors, their relative proportion in the larger authorship community has diminished, indicating a concentration of influence among a smaller elite group. The landscape is becoming increasingly polarized, with a small number of authors garnering a disproportionate amount of recognition and resources compared to their colleagues. The pronounced spike in the curve for nanoscience, observed in the early 1980s, can be attributed to the relatively small network size during the initial years of this emerging field. As nanoscience is a relatively new discipline that experienced rapid growth in the early twenty-first century (Hulla et al., 2015), the initial sample size was smaller, leading to greater fluctuations in the percentage of rich authors.

We analyzed the normalized connection strengths for multiple cohorts to investigate the trends in evolution, as shown in Fig. 4 and Fig. 5. The results are similar across nanoscience, chemical physics, and human-computer interaction disciplines, with a clear trend toward increasing the strength of rich-rich connections over time, a slight increase in the strength of rich-ordinary connections, and a slight shrinking of the strength of ordinary-ordinary connections. The connection strength among the rich authors gradually intensifies, suggesting that the phenomenon of "clustering" among the rich authors becomes more pronounced. The observed core-

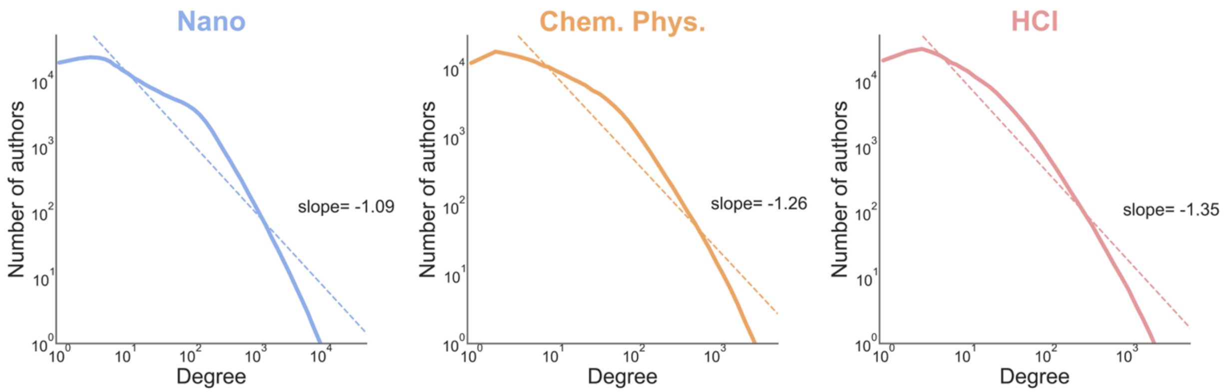

Fig. 2. The degree distribution of author citation networks. The dotted line represents the fitting result of a power-law distribution. Across the three disciplines, the scale-free phenomenon is not observed.

Table 2
The connection strength between authors in the three disciplines. The connection strength metrics are normalized using null models. In these three disciplines, a denser citation relationship is observed among rich authors, while a sparser citation relationship is observed among ordinary authors.

Connection strength	Nano	Chem. Phys.	HCI
Rich-Rich	4.834	4.314	4.328
Rich-Ordinary	1.290	1.271	1.254
Ordinary- Ordinary	0.905	0.910	0.912

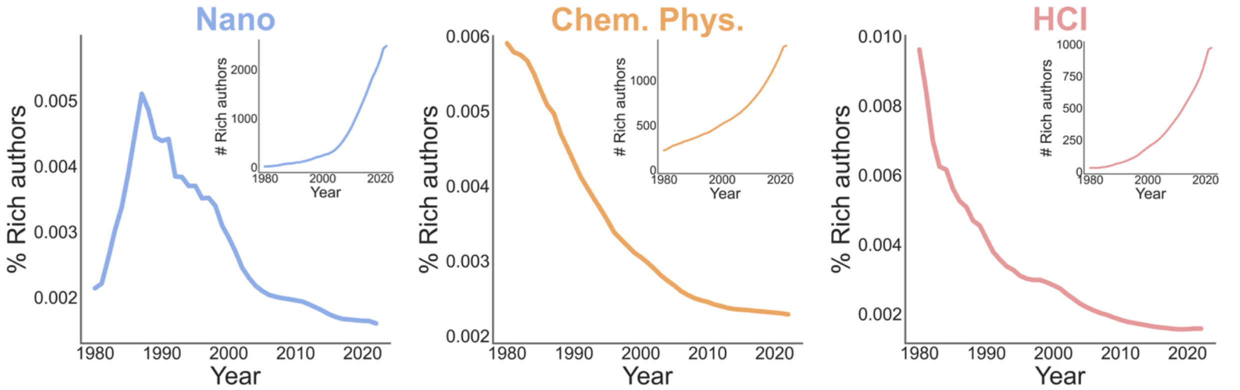

Fig. 3. The number (inner) and percentage (outer) of rich authors in different years. The percentage of rich authors declines over time despite a rapid increase in their absolute numbers.

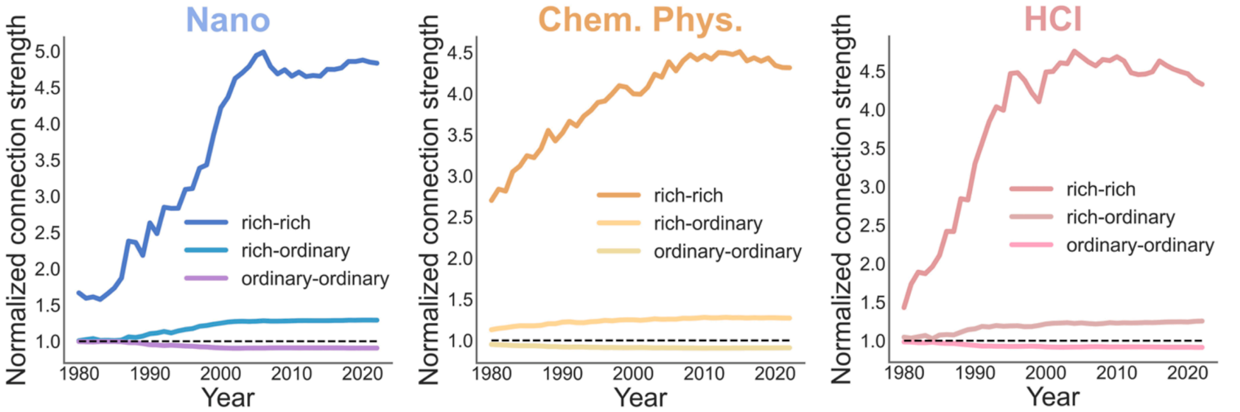

Fig. 4. Temporal changes in connection strength between rich and ordinary authors across the three disciplines. The normalized connection strength metric, calculated using the null model, is presented for each year.

periphery structure in these disciplines has important implications for understanding scientific communication and information flows, as highly connected core authors may display significant influence over research directions and trends. Our finding aligns with the conclusion that global citation inequality is increasing, particularly among the rich (Nielsen & Andersen, 2021).

In Fig. 4, the heightened normalized connection strength among the rich authors obscures the nuanced variations between rich-ordinary and ordinary-ordinary authors. This dominance of strong connections within the rich category can overshadow the subtler but equally significant dynamics relevant to ordinary authors, potentially impacting our understanding of the overall network structure. To emphasize these patterns over time, we have narrowed the range of the vertical axis in Fig. 5 to delve into the details.

Over the years under investigation, the total normalized interconnection edge weights between rich and ordinary authors in author citation networks demonstrate a tendency to increase. Conversely, the total of normalized interconnection edge weights among ordinary authors show a decreasing trend. We observe a clear rise in connections between the affluent and ordinary authors within the scientific ecosystem, alongside a decline in mutual attention and recognition within the ordinary group. The scientific community has experienced exponential growth in recent decades, leading to a substantial increase in the number of ordinary authors. These ordinary authors exhibit a more loosely connected pattern of interaction, which is characteristic of the periphery in the core-periphery structure. The influx of ordinary authors, coupled with their peripheral positioning, underscores the challenges they face in gaining visibility and

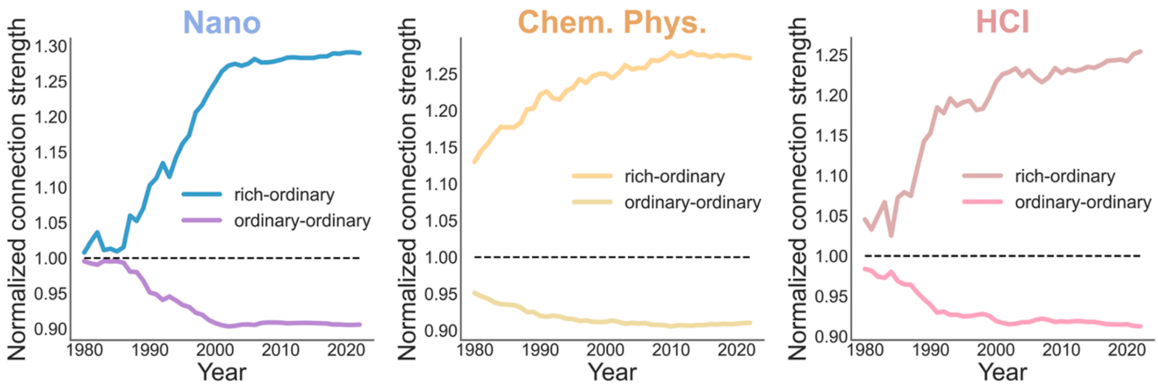

Fig. 5. Temporal changes in connection strength between rich and ordinary authors across the three disciplines. The normalized connection strength metric, calculated using the null model, is presented for each year (zoomed in from Fig. 4).

establishing robust collaborative networks.

5.2. Robustness checks

As observed in the previous section, the connection strength between rich and ordinary authors exhibits distinct evolving patterns. To further substantiate our findings, we employ several additional robustness check approaches. The methods we adopted here are three commonly used approaches to investigate the core-periphery structure identified in author citation networks.

5.2.1. The rich club perspective

Rich clubs describe a phenomenon in networks in which hubs are closely connected (Zhou & Mondragon, 2004). This substructure of the network also could be interpreted as the tendency of nodes with high degrees to form densely interconnected communities (Colizza et al., 2006). Highly influential or well-connected nodes are more likely to establish robust connections with each other. For instance, in the context of the human brain, core brain hubs function as a high-capacity backbone for communication of neurons (Van den Heuvel et al., 2012). Although the concept of a rich club differs from that of a core-periphery structure, both concepts involve a set of closely connected high-degree nodes and can be understood in terms of a two-class partitioning of the network (Ma & Mondragón, 2015). The densely connected rich club can be interpreted as the core of the core-periphery structure. In addition, the presence of a rich club is also commonly associated with various network structural characteristics, such as clustering coefficient and assortativity, which are also relevant to the core-periphery structure (Mondragon, 2018). For author citation networks, the rich club coefficient is a common metric to describe the core-periphery structure of a network (Wang et al., 2022), and the normalized rich club coefficient is obtained by randomizing the real author citation network, providing a way to assess whether the observed rich club structure is statistically significant. Fig. 6 provides an example of the normalized rich club coefficient curves for author citation networks in the field of nanoscience up until the end of 2022. The results indicate a significant rich club structure among highly cited authors, with a steep increase in the coefficient for nodes with higher degree.

The standardized rich club coefficient curve shows a smooth increase followed by a sharp rise. We then adopt the two afore-

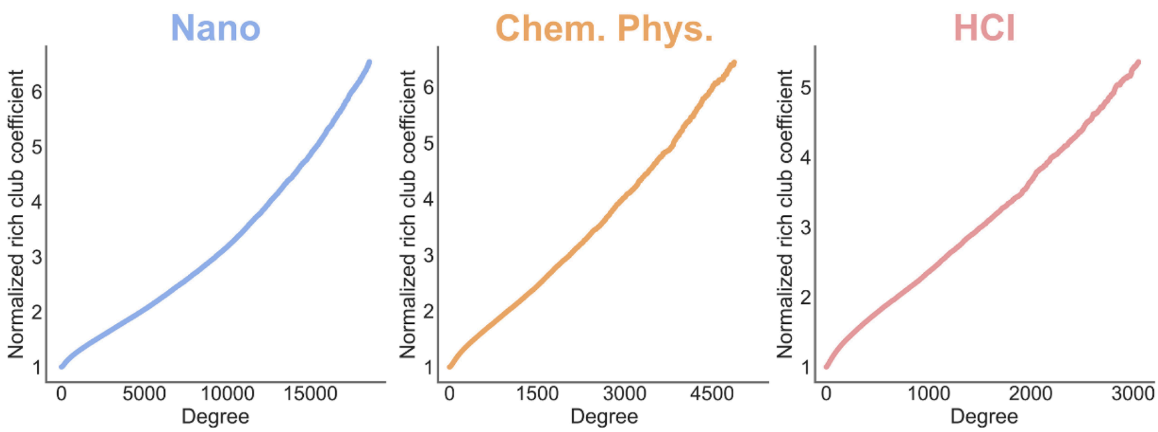

Fig. 6. An upward-sloping pattern exists in the normalized rich club coefficient curves for all three disciplines. This figure encompasses the entire time range of our dataset. A steep increase in the normalized rich club coefficient among high-degree authors is observed across all three disciplines.

mentioned measurements, α and s, to separately quantify the relative level of the cluster phenomenon in the part of ordinary and rich authors for each year. By calculating the slope of the curve in the section of the degree interval corresponding to the ordinary authors, denoted as α, and comparing the change in the intensity of the rich club within the ordinary authors in the network in each of the three disciplines, we conclude that the strength of the core-periphery structure decreases over time in the ordinary author group. As shown in Fig. 7, the diminishing attention toward ordinary authors reflects the fact that the scientific output of these researchers is not receiving sufficient recognition within the scientific community.

When considering the rich authors within author citation networks and using the s indicator to measure the 'density' of connections among the affluent, the results indicate that, while there were some fluctuations before 2000, the density of connections among the rich notably elevated after 2000. As shown in Fig. 8, within the scientific ecosystem, it is evident that upper-class authors are more inclined to follow and cite papers produced specifically by the affluent community themselves.

The rich club approach offers further empirical evidence of changes in connection strength both among the rich and among the ordinary, supporting the phenomenon of increased connections among the rich and weakened connections among the ordinary in author citation networks. In the context of the scientific ecosystem, these findings highlight the dynamic shifts in scholarly interactions, where established experts tend to reinforce their ties while emerging voices may encounter challenges in fostering robust connections. We present an effective method to compare rich club curves across different time periods, focusing on both rich and ordinary author communities. This offers a novel approach to investigate the temporal characteristics of the rich club phenomenon (Pedreschi et al., 2022).

5.2.2. The assortativity perspective

Fig. 9 shows the temporal evolution of assortativity of the author citation network in each of the three disciplines.

We observe that, in all three disciplines, the assortativity generally increases and then decreases over time, but the range of assortativity values remains relatively stable over time. The overall assortativity of the citation networks decreases slightly over time, indicating a weaker tendency for authors with similar degrees to connect with each other. We have uncovered different connection patterns between the rich and ordinary authors, which together create a complex and dynamic core-periphery structure within the scientific ecosystem. Essentially, this complex network configuration reflects the intricate dynamics of knowledge flows and influence among scientists.

5.2.3. Network percolation perspective

To further enhance the validation of the complex core-periphery structure in the network, we employ the network percolation method. This technique gradually removes rich authors from the author citation network, allowing for a thorough analysis of the core-periphery dynamics within the residual network. The rationale behind employing this approach lies in its ability to unveil the underlying core-periphery dynamics among the less prominent authors by reducing the influence of the dominant, high-degree authors.

The network percolation method involves several steps. Initially, authors with the highest degrees are identified within specific percentile thresholds (top 0.1 %, 0.2 %, 0.3 %, …, up to 1 % in our analysis) in each discipline. The core idea of network percolation is to sequentially remove these authors to eliminate their impact on the citation network, thereby allowing analysis of the residual network structure in stages. After removing each percentile of high-degree authors and their connections, the remaining authors in the network are re-categorized as either rich or ordinary based on their cumulative degree in the modified, partially depleted network, using the same method as previously described. The core-periphery structure of the residual network is then analyzed subsequently. By focusing on interactions and connections among the remaining authors, we can reveal underlying core-periphery dynamics that may be obscured by the presence of highly influential authors. This process aids in understanding the structural integrity and connectivity among ordinary authors, thereby uncovering hierarchical structures among relatively ordinary authors.

We create a scenario where the influence of dominant figures diminishes, allowing us to focus on the interactions and relationships obscured by the rich authors. The top 0.1 %, 0.2 %, 0.3 %, …, and 1 % degrees of the overall authors in each discipline were removed,

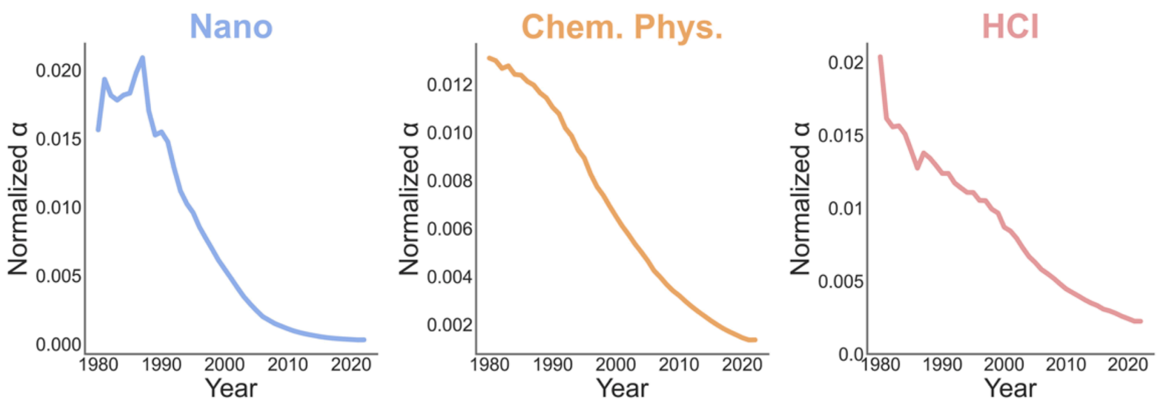

Fig. 7. The changes of normalized α over time. α depicts the change in the normalized rich club coefficient per unit increase in author degree for the ordinary authors (see Methods).

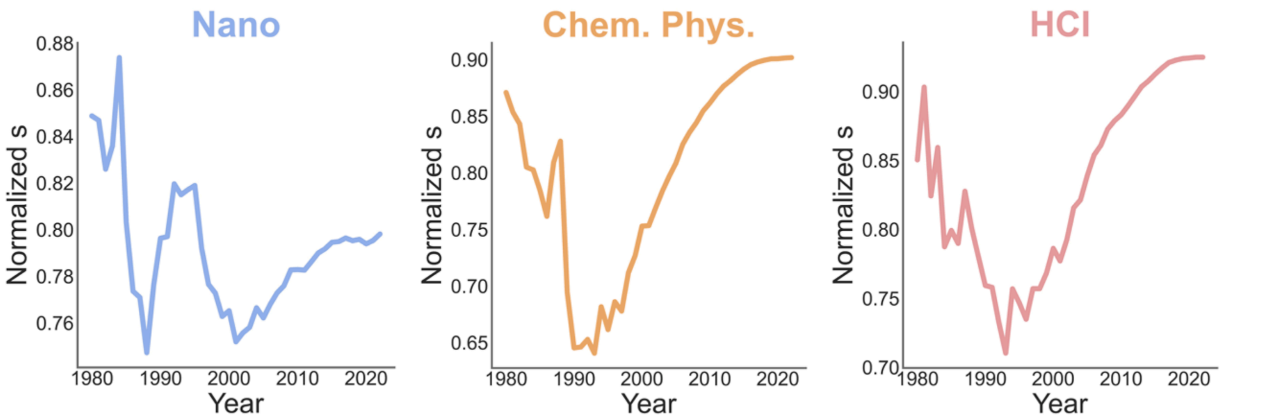

Fig. 8. The changes of normalized **s** over time. **s** depicts the relative intensity of the rich club phenomenon for the rich authors (see Methods).

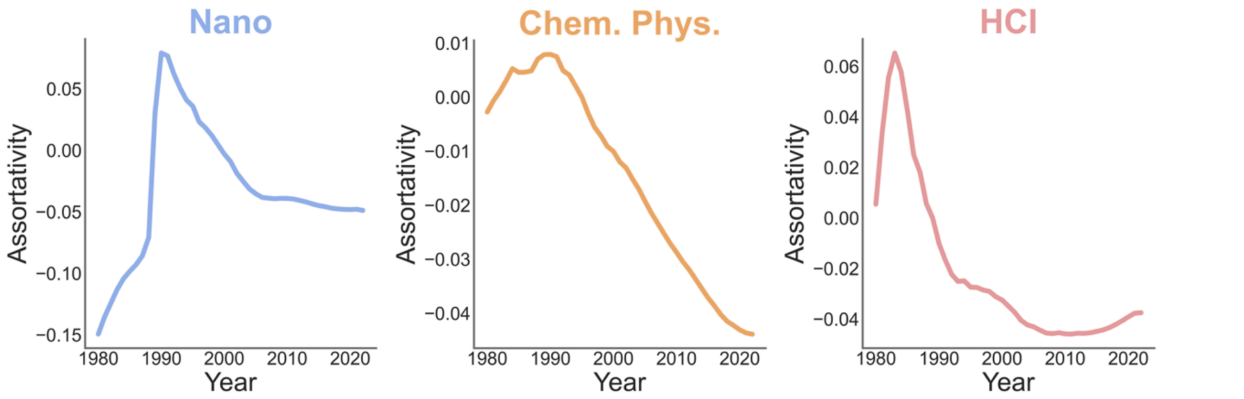

Fig. 9. Assortativity of author citation network over time. The networks exhibit a decreasing trend in the connectivity of nodes with similar degrees.

along with their connected edges. The remaining authors were then categorized as either rich or ordinary based on the cumulated degree in the author's partly removed network, using the same method as previously described. The connection strength among newly divided rich authors is shown in Fig. 10.

The overall trend of all three curves is decreasing, and the connection strength among the sub-richest group in the remaining network after removing the richest group in the network decreases with the percentage of the richest authors removed, indicating that the connection strength among ordinary authors in these three disciplinary author citation networks is lower than the connection

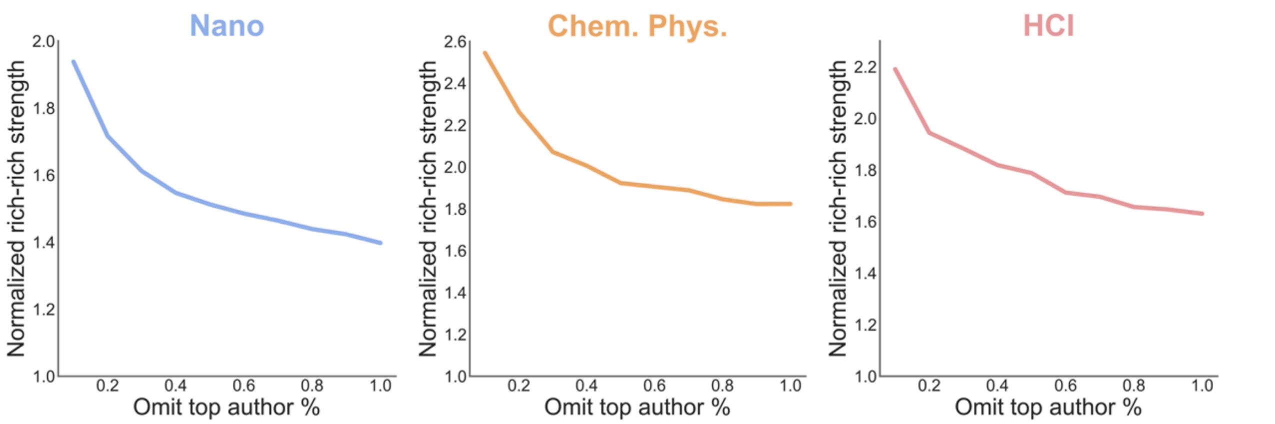

Fig. 10. Effects of removing top degree authors on the connection strength among remaining relatively rich authors. Sequentially, the top 0.1 %, 0.2 %, 0.3 %, …, and 1 % degrees of the highest-degree authors in each discipline are removed to reveal the connection patterns within the residual network.

strength among rich authors. This suggests that connections among ordinary authors are not as intensive, consistent with the conclusions obtained in the previous section.

Using the same network percolation method, we shifted our focus to examining the key role among relatively ordinary authors in the author citation network. Specifically, we investigated the proportion of connection strength within the remaining sub-rich group of authors, which offers a suitable perspective on the interaction among these authors, highlighting their interactions and contributions distinct from the dominant top-degree authors. The proportion of connection strength within the sub-rich group was calculated by dividing the connection strength among the sub-rich authors by the sum of the connection strength among the sub-rich authors and the connection strength between the sub-rich authors and the more ordinary authors. The results are shown in Fig. 11.

As we sequentially removed the top-degree authors, we observed a decrease in the proportion of connections within the sub-rich group relative to the entire remaining network. This finding suggests an increasing trend in citation relationships between the sub-rich group and the most ordinary authors in the network. This method provides insight into the multilayered and complex relationships within the author citation network, complementing our understanding of the core-periphery structure within the network.

6. Discussion

The Matthew Effect has been extensively discussed by researchers across different levels in the scientific community, including institutions, authors, publications, etc. (Merton, 1968; Price & Beaver, 1966; Price, 1976). The Matthew Effect, which is a form of inequality, illustrates the scenario where those who are already privileged continue to get more advantages, while those with fewer advantages face increasing challenges. This pattern of inequality is readily apparent in academia, where elite scientists receive preferential attention. One possible explanation for the phenomenon of the rich getting richer is the process of self-reinforcement (Nielsen & Andersen, 2021). More successful individuals are more likely to receive additional resources and attention from academia, which further promotes their development within the academic community. It has been demonstrated that the fame of an author or the familiarity of their name influences the citing behavior of others (Brogaard et al., 2020).

6.1. Theoretical and practical implications

The phenomenon of inequality within the scientific ecosystem carries several profound implications for scientists themselves as well as for the scientific ecosystem as a whole. In the network of the scientific ecosystem, it is typical for authors in elevated positions within the scientific hierarchy to naturally attract more attention. This elevated visibility often translates into increased citations, access to funding opportunities, and enhanced prospects for career advancement, which are vital components of a thriving scientific community (Koh et al., 2016; Rehrl et al., 2014). However, it is essential to recognize that, while a degree of inequality can be expected and may even be beneficial, excessive concentration of attention and disparity can be detrimental. The challenge lies in distinguishing between attention genuinely driven by the quality and impact of research and that influenced disproportionately by non-substantive factors. In our approach, employing a null model partially offers insights into citation patterns influenced more by author reputation and fame rather than strictly merit-based recognition. This is because the null model captures some fundamental yet latent information regarding the ordering of merit degrees.

When a disproportionate share of recognition and resources is concentrated among a limited group of scientists, there is a heightened risk of neglecting diverse voices within the scientific community, which can ultimately lead to a stifling of diversity and innovation within the ecosystem (Painter et al., 2021). This imbalance reverberates throughout the scientific landscape. Talented younger and lesser-known scholars, who constitute the majority of ordinary authors within the scientific community, may encounter formidable obstacles on their path to prominence. These obstacles can result in inhibiting the vital infusion of innovative spirits into scientific advancements (Chu & Evans, 2021). Building on this understanding, it is essential to explore how scientific attention inequality specifically impacts the development of science. The denser connection among rich authors indicates a concentration of influence and resources among a select group of high-degree authors. Theoretically, the concept of cumulative advantage (Merton, 1968) suggests that initial recognition can disproportionately amplify the visibility and impact of already prominent scientists, thereby perpetuating existing inequalities. This can skew scientific attention toward well-established topics influenced by elite clubs, potentially overlooking novel or interdisciplinary fields that may require more nurturing. Our analysis of ordinary authors provides positive insights into this issue. While prominent authors exhibit a strong tendency to form elite research clubs characterized by concentrated citations within this community, we have identified a mitigating trend in which sub-rich authors direct attention toward relatively ordinary authors. This indicates that pathways for inclusivity and recognition do exist. We understand that the challenge of fully comprehending attention inequality can cast doubt on the effectiveness of these efforts. However, it is important to recognize that fostering inclusivity is not merely a hopeful aspiration; it is a strategic necessity for the long-term vitality of the scientific community. To sustain this inclusive practice, it is crucial to cultivate an equitable environment that recognizes the contributions of all scholars, regardless of their seniority or prominence, and actively nurtures and amplifies these voices.

Our study offers a multifaceted measurement of the core-periphery structure within the scientific community, examining the interconnectivity between different author groups from a social stratification perspective. However, it is worth further exploration to understand how the increasing density of associations among established scholars and the decreasing density among less prominent individuals will impact the development of science. A key motivation behind citation behavior is the reinforcement of strong social connections (Vinkler, 1987). Academic social capital gained through collaborative relationships may influence citation practices, making it essential to examine how attributes of the collaboration network impact the citation network. Given that the author collaboration networks are closely intertwined with author citation networks, investigating the influence of the changing dynamics in

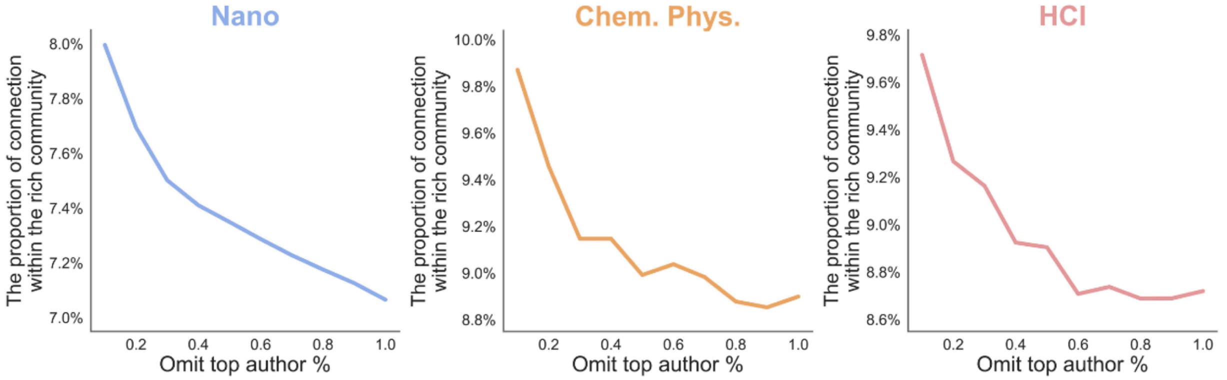

Fig. 11. Effects of removing top-degree authors on the proportion of connection strength within the remaining relatively rich authors. Sequentially, the top 0.1 %, 0.2 %, 0.3 %, …, and 1 % degrees of the highest-degree authors in each discipline are removed. As the sub-richest groups in the author citation network become more ordinary, their connections with other ordinary authors increase.

the complex core-periphery structure on the collaboration and citation behaviors of young scholars holds significant importance for scientific innovation. This analysis can reveal how shifts in the core-periphery balance influence the opportunities for emerging researchers to engage in collaborative projects and receive citations, thereby impacting their career progression and the overall diversity of ideas within the scientific ecosystem. By examining the dynamics of core-periphery structures, we aim to gain insights into how these structures might facilitate or hinder the dissemination of work by emerging scholars. We expect that such insights could help inform strategies to support the career development of these researchers and contribute to fostering a more inclusive and diverse scientific community. While this remains speculative, ensuring that emerging scholars have equitable opportunities for collaboration and citation is an important consideration for maintaining the diversity of ideas and promoting the long-term sustainability of the scientific ecosystem.

6.2. Limitations and future directions

While our study provides valuable insights into the phenomenon of attention inequality within scientific ecosystems, it is essential to acknowledge several limitations that may impact interpretation of the research results. One notable limitation of this study is related to the construction of the author citation network. We exclude the citation records across different disciplines, aiming to obtain a citation network specific to our target field. This choice might underestimate the significance of interdisciplinary publications. Another potential limitation relates to our study's comparison benchmark. To establish a baseline for assessment, we randomize the actual author citation network to create a null model for comparison. The methodology and the frequency of randomization iterations can affect the credibility of the null model as a valid comparison benchmark. Our findings may require further validation to ensure robustness. We attempted to assess the core-periphery structure from multiple angles using several indicators, including connection strength, rich club, and assortativity perspectives, aiming to provide a more detailed and nuanced understanding of the core-periphery dynamics. However, we acknowledge that our approach may not capture every aspect necessary for a truly comprehensive measurement of core-periphery structures.

Finally, we conclude our study by signifying some promising directions for future research. A crucial aspect of understanding inequality within citation networks lies in exploring the complex relationship between prior collaboration dynamics and citation patterns. Collaboration among researchers plays a pivotal role in shaping citation relationships, as collaborative efforts often result in co-authored publications that garner citations. However, disparities in collaboration opportunities and visibility can exacerbate existing inequalities within the scientific community.

In addition, the potential existence of multiple cores within author citation networks requires further analysis. The specialization of science, as outlined by Chubin (1976), may lead to the fragmentation of these networks into smaller, densely connected subgroups, each representing a specialized area of research. This fragmentation suggests that authors from different time periods or sub-disciplinary cohorts might exhibit varying degrees of connectivity, indicating a dynamic and layered network structure. Specifically, investigating how these multiple cores interact and influence information flow within the network could provide deeper insights into knowledge dissemination patterns. The concept of core-periphery structure requires that the core be well-connected. However, the specific level of connectivity is not explicitly defined. Considering the potential for multiple cores to exist in the author citation network, future analysis of the connectivity of these cores is crucial for yielding meaningful results. Connectivity not only reinforces the position of core authors but also shapes the interactions between different sub-networks. Investigating these connections further could reveal whether certain cores act as isolated knowledge clusters or as hubs that facilitate broader dissemination.

Another promising avenue for future research is the exploration of proportional effects in citation relationships. Distinguishing between authors who cite a substantial proportion of another author's body of work versus those who cite a smaller fraction could yield deeper insights into citation dynamics. Our current study primarily focuses on the core-periphery structure of the citation network by analyzing the absolute number of citation links. However, incorporating the proportionality of citations may uncover additional layers of scholarly attention and influence. This approach could enrich our understanding of the patterns of attention inequality.

Future research could also consider expanding the dataset to include more comprehensive data from the period before 2000. This would provide a clearer picture of earlier network dynamics and help determine whether the trends observed in more recent years are consistent with past citation patterns. Additionally, it would be beneficial to examine the role of evolving publication and collaboration practices, particularly how digitalization and open-access publishing have influenced author connectivity.

Lastly, it would be beneficial to explore the longitudinal effects of unequal attention dynamics on scientific careers. Long-term studies tracking the trajectories of researchers over time can reveal how disparities in attention accumulation manifest throughout researchers' careers and their subsequent impact on scientific productivity, career advancement, and retention within the scientific community.

CRediT authorship contribution statement

Haoyang Wang: Writing – original draft, Visualization, Methodology, Investigation, Formal analysis. **Win-bin Huang:** Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Formal analysis, Conceptualization. **Yi Bu:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis.

Acknowledgements

The authors acknowledge the financial support by the National Natural Science Foundation of China (#72104007, #72474009, and #72174016) and by the Fundamental Research Funds for the Central Universities, Peking University. The authors are grateful to the fruitful discussion with Fan Meng, Xingkun Liang, all other members in the knowledge discovery group at the Department of Information Management, Peking University, and anonymous reviewers for their constructive comments.

APPENDIX. Pseudocode of the rich author classification process.

The algorithm ranks authors by their degrees in descending order and calculates the cumulative degree. Authors are sequentially added to the 'rich' author set until cumulative degree exceeds a threshold defined as 10 % of the total author degree.

```
Algorithm: Finding rich authors in an author citation network
Input: author citation network; rich threshold (10 % in the later analyses)
Output: rich author set
Sort the authors by their degrees in descending order
Define a rich author set R and cumulative degree CD
While CD <= sum of total author degree × rich threshold:
    categorize the first author A in degree descending order as a rich author
    add author A to set R and delete author A from the original author list
    CD = CD + degree of author A
Return rich author set R
```

Data availability

Data will be made available on request.

References

Ahuja, G., Soda, G., & Zaheer, A. (2012). Introduction to the special issue: The genesis and dynamics of organizational networks. *Organization Science (New York, N.Y.), 23*(2), 434–448. https://doi.org/10.1287/orsc.1110.0695

Alba, R. D., & Moore, G. (1978). Elite Social Circles. *Sociological Methods (San Diego, Calif.) & Research; A Journal of Science and its Applications, 7*(2), 167–188. https://doi.org/10.1177/004912417800700204

Allison, P. D., Long, J. S., & Krauze, T. K. (1982). Cumulative advantage and inequality in science. *American Sociological Review, 47*(5), 615–625. https://doi.org/10.2307/2095162

Almeida-Neto, M., Guimarães, P., Guimarães, P. R., Jr, Loyola, R. D., & Ulrich, W. (2008). A consistent metric for nestedness analysis in ecological systems: Reconciling concept and measurement. *Oikos (Copenhagen, Denmark), 117*(8), 1227–1239. https://doi.org/10.1111/j.0030-1299.2008.16644.x

Alstott, J., Panzarasa, P., Rubinov, M., Bullmore, E. T., & Vértes, P. E. (2014). A unifying framework for measuring weighted rich clubs. *Scientific Reports, 4*(1), 1. https://doi.org/10.1038/srep07258. Article.

Alvarez-Hamelin, J.I., Dall'Asta, L., Barrat, A., & Vespignani, A. (2005). *k-core decomposition: A tool for the visualization of large scale networks* (arXiv:Cs/0504107). arXiv. https://doi.org/10.48550/arXiv.cs/0504107.

Atmar, W., & Patterson, B. (1993). The measure of order and disorder in the distribution of species in fragmented habitat. *Oecologia, 96*, 373–382. https://doi.org/10.1007/BF00317508

Azoulay, P., Stuart, T., & Wang, Y. (2014). Matthew: Effect or fable? *Management Science (New York, N.Y.), 60*(1), 92–109. https://doi.org/10.1287/mnsc.2013.1755

Balland, P.-A., Suire, R., & Vicente, J. (2010). How do clusters/pipelines and core/periphery structures work together in knowledge processes? *Papers in Evolutionary Economic Geography (PEEG)*.
Barabási, A.-L. (2016). Network science. http://networksciencebook.com/.
Bassett, D. S., & Bullmore, E. T. (2009). Human brain networks in health and disease. *Current Opinion in Neurology, 22*(4), 340–347. https://doi.org/10.1097/WCO.0b013e32832d93dd
Bassett, D. S., Wymbs, N. F., Rombach, M. P., Porter, M. A., Mucha, P. J., & Grafton, S. T. (2013). Task-based Core-Periphery Organization of Human Brain Dynamics. *PLOS Computational Biology, 9*(9), Article e1003171. https://doi.org/10.1371/journal.pcbi.1003171
Bianchi, C., Galaso, P., & Palomeque, S. (2021). Patent collaboration networks in Latin America: Extra-regional orientation and core-periphery structure. *The Journal of Scientometric Research; A Journal of Science and its Applications, 10*(1s), s59–s70. https://doi.org/10.5530/jscires.10.1s.22
Borgatti, S. P., & Everett, M. G. (2000). Models of core/periphery structures. *Social Networks, 21*(4), 375–395. https://doi.org/10.1016/S0378-8733(99)00019-2
Bornmann, L. (2014). Assigning publications to multiple subject categories for bibliometric analysis: An empirical case study based on percentiles. *The Journal of Documentation, 70*(1), 52–61. https://doi.org/10.1108/JD-10-2012-0136
Brogaard, J., Engelberg, J., Eswar, S., & Van Wesep, E. D. (2020). On the causal effect of fame on citations. *SSRN Electronic The Journal*. https://doi.org/10.2139/ssrn.3565487
Bruckner, S., Hüffner, F., & Komusiewicz, C. (2015). A graph modification approach for finding core–periphery structures in protein interaction networks. *Algorithms for Molecular Biology, 10*(1), 16. https://doi.org/10.1186/s13015-015-0043-7
Catanzaro, M., Caldarelli, G., & Pietronero, L. (2004). Assortative model for social networks. *Physical Review E (Norwalk, Conn.), 70*(3). https://doi.org/10.1103/PhysRevE.70.037101
Chen, Z., & Guan, J. (2016). The core-peripheral structure of international knowledge flows: Evidence from patent citation data. *R&D Management, 46*(1), 62–79. https://doi.org/10.1111/radm.12119
Chu, J. S. G., & Evans, J. A. (2021). Slowed canonical progress in large fields of science. *Proceedings of the National Academy of Sciences, 118*(41). https://doi.org/10.1073/pnas.2021636118
Chubin, D. E. (1976). The conceptualization of scientific specialties. *The Sociological Quarterly, 17*(4), 448–476.
Clough, J. R., Gollings, J., Loach, T. V., & Evans, T. S. (2015). Transitive reduction of citation networks. *The Journal of Complex Networks, 3*(2), 189–203. https://doi.org/10.1093/comnet/cnu039
Colizza, V., Flammini, A., Serrano, M. A., & Vespignani, A. (2006). Detecting rich-club ordering in complex networks. *Nature Physics, 2*(2), 110–115. https://doi.org/10.1038/nphys209
Corradini, E. (2024). Deconstructing cultural appropriation in online communities: A multilayer network analysis approach. *Information Processing & Management, 61*(3), Article 103662. https://doi.org/10.1016/j.ipm.2024.103662
Da Silva, M. R., Ma, H., & Zeng, A.-P. (2008). Centrality, network capacity, and modularity as parameters to analyze the core-periphery structure in metabolic networks. *Proceedings of the IEEE, 96*(8), 1411–1420. https://doi.org/10.1109/JPROC.2008.925418. Proceedings of the IEEE.
Eder, J. (2019). Innovation in the periphery: A critical survey and research agenda. *International Regional Science (New York, N.Y.) Review, 42*(2), 119–146. https://doi.org/10.1177/0160017618764279
Fontana, M., Iori, M., Montobbio, F., & Sinatra, R. (2020). New and atypical combinations: An assessment of novelty and interdisciplinarity. *Research; A Journal of Science and its Applications Policy, 49*(7), Article 104063. https://doi.org/10.1016/j.respol.2020.104063
Foster, J. G., Foster, D. V., Grassberger, P., & Paczuski, M. (2010). Edge direction and the structure of networks. *Proceedings of the National Academy of Sciences, 107*(24), 10815–10820. https://doi.org/10.1073/pnas.0912671107
Freeman, L. C. (1984). The impact of computer based communication on the social structure of an emerging scientific specialty. *Social Networks, 6*(3), 201–221. https://doi.org/10.1016/0378-8733(84)90011-X
Gallagher, R. J., Young, J.-G., & Welles, B. F. (2021). A clarified typology of core-periphery structure in networks. *Science (New York, N.Y.) Advances, 7*(12), eabc9800. https://doi.org/10.1126/sciadv.abc9800
Garfield, E. (1965). Can citation indexing be automated?. In *Statistical Association Methods for Mechanized Documentation, Symposium Proceedings* (pp. 189–192).
Granovetter, M. (1978). Networks of Collective Action: A perspective on Community Influence Systems. Edward O. Laumann, Franz U. Pappi. *American The Journal of Sociology, 83*(6), 1538–1542. https://doi.org/10.1086/226720
Guo, C., Milojević, S., & Liu, X. (2015). How are academic articles cited over time? *iConference 2015 Proceedings*. https://hdl.handle.net/2142/73747.
Hagberg, A., Swart, P. J., & Schult, D. A. (2008). *Exploring network structure, dynamics, and function using networkx (LA-UR-08-05495; LA-UR-08-5495)*. Los Alamos, NM (United States): Los Alamos National Laboratory (LANL). https://www.osti.gov/biblio/960616.
Hao, B., & Kovács, I. A. (2024). Proper network randomization is key to assessing social balance. *Science (New York, N.Y.) Advances, 10*(18), eadj0104. https://doi.org/10.1126/sciadv.adj0104
Hessey, R., & Willett, P. (2013). Quantifying the value of knowledge exports from librarianship and information science research. *The Journal of Information Science (New York, N.Y.), 39*(1), 141–150. https://doi.org/10.1177/0165551512442476
Hiebert, E. N. (1996). Discipline identification in chemistry and physics. *Science (New York, N.Y.) in Context, 9*(2), 93–119. https://doi.org/10.1017/S0269889700002362
Hojman, D. A., & Szeidl, A. (2008). Core and periphery in networks. *The Journal of Economic Theory, 139*(1), 295–309. https://doi.org/10.1016/j.jet.2007.07.007
Huang, Y., Lu, W., Liu, J., Cheng, Q., & Bu, Y. (2022). Towards transdisciplinary impact of scientific publications: A longitudinal, comprehensive, and large-scale analysis on Microsoft Academic Graph. *Information Processing & Management, 59*(2), Article 102859. https://doi.org/10.1016/j.ipm.2021.102859
Hulla, J., Sahu, S., & Hayes, A. (2015). Nanotechnology: History and future. *Human & Experimental Toxicology, 34*(12), 1318–1321. https://doi.org/10.1177/0960327115603588
Kahneman, D. (1975). Attention and effort. *The American The Journal of Psychology (Savannah, Ga.), 88*(2), 339. https://doi.org/10.2307/1421603
Khanna, R., & Guler, I. (2022). Degree assortativity in collaboration networks and invention performance. *Strategic Management The Journal, 43*(7), 1402–1430. https://doi.org/10.1002/smj.3367
Khelfaoui, M., & Gingras, Y. (2019). Physical Review: From the periphery to the center of physics. *Physics in Perspective, 21*(1), 23–42. https://doi.org/10.1007/s00016-019-00235-y
Klamer, A., & Dalen, H. P.van (2002). Attention and the art of scientific publishing. *The Journal of Economic Methodology, 9*(3), 289–315. https://doi.org/10.1080/1350178022000015104
Koh, E. J., Dunwoody, S., Brossard, D., & Allgaier, J. (2016). Mapping neuroscientists' Perceptions of the nature and effects of public visibility. *Science (New York, N.Y.) Communication, 38*(2), 170–196. https://doi.org/10.1177/1075547016635180
Kojaku, S., & Masuda, N. (2017). Finding multiple core-periphery pairs in networks. *Physical Review E (Norwalk, Conn.), 96*(5), Article 052313. https://doi.org/10.1103/PhysRevE.96.052313
Krugman, P. R. (1996). *The self-organizing economy*. Blackwell Publishers.
Kunnath, S. N., Herrmannova, D., Pride, D., & Knoth, P. (2021). A meta-analysis of semantic classification of citations. *Quantitative Science (New York, N.Y.) Studies, 2*(4), 1170–1215. https://doi.org/10.1162/qss_a_00159
Lee, S. H., Cucuringu, M., & Porter, M. (2014). Density-based and transport-based core-periphery structures in networks. *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics, 89*, Article 032810. https://doi.org/10.1103/PhysRevE.89.032810
Liao, C. H. (2021). The Matthew effect and the halo effect in research funding. *The Journal of Informetrics, 15*(1), Article 101108. https://doi.org/10.1016/j.joi.2020.101108
Ma, A., & Mondragón, R. J. (2015). Rich-cores in networks. *PloS one, 10*(3), Article e0119678. https://doi.org/10.1371/journal.pone.0119678
Merton, R. K. (1968). The Matthew Effect in science. *Science (New York, N.Y.), 159*(3810), 56–63. https://doi.org/10.1126/science.159.3810.56

Milojević, S. (2010). Modes of collaboration in modern science: Beyond power laws and preferential attachment. *The Journal of the American Society for Information Science (New York, N.Y.) and Technology (Elmsford, N.Y.), 61*(7), 1410–1423. https://doi.org/10.1002/asi.21331

Mondragon, R.J. (2018). *Beyond the rich-club: Properties of networks related to the better connected nodes* (arXiv:1810.12328). arXiv. https://doi.org/10.48550/arXiv.1810.12328.

Mullins, N. C., Hargens, L. L., Hecht, P. K., & Kick, E. L. (1977). The group structure of cocitation clusters: A comparative study. *American Sociological Review, 42*(4), 552–562. https://doi.org/10.2307/2094554

Newman, M. E. J. (2002). Assortative mixing in networks. *Physical Review Letters, 89*(20), Article 208701. https://doi.org/10.1103/PhysRevLett.89.208701

Nielsen, M. W., & Andersen, J. P. (2021). Global citation inequality is on the rise. *Proceedings of the National Academy of Sciences, 118*(7), Article e2012208118. https://doi.org/10.1073/pnas.2012208118

Noldus, R., & Van Mieghem, P. (2015). Assortativity in complex networks. *The Journal of Complex Networks, 3*(4), 507–542. https://doi.org/10.1093/comnet/cnv005

Oppenheim, C., & Renn, S. P. (1978). Highly cited old papers and the reasons why they continue to be cited. *The Journal of the American Society for Information Science (New York, N.Y.), 29*(5), 225–231. https://doi.org/10.1002/asi.4630290504

Painter, D. T., Daniels, B. C., & Laubichler, M. D. (2021). Innovations are disproportionately likely in the periphery of a scientific network. *Theory in Biosciences, 140*(4), 391–399. https://doi.org/10.1007/s12064-021-00359-1

Paszcza, B. (2016). *Comparison of microsoft academic (Graph) with web of science, scopus and google scholar [Master's thesis*. University of Southampton.] https://eprints.soton.ac.uk/408647/.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research; A Journal of Science and its Applications, 12*(85), 2825–2830.

Pedreschi, N., Battaglia, D., & Barrat, A. (2022). The temporal rich club phenomenon. *Nature Physics*, 1–8. https://doi.org/10.1038/s41567-022-01634-8

Price, D. J., & de, S. (1976). A general theory of bibliometric and other cumulative advantage processes. *The Journal of the American Society for Information Science (New York, N.Y.), 27*(5), 292–306. https://doi.org/10.1002/asi.4630270505

Price, D. J., de, S., & Beaver, D. (1966). Collaboration in an invisible college. *The American Psychologist, 21*, 1011–1018. https://doi.org/10.1037/h0024051

Rehrl, M., Palonen, T., Lehtinen, E., & Gruber, H. (2014). Experts in science: Visibility in research communities. *Talent Development (Cambridge, England) and Excellence, 6*, 31–45.

Rossa, F. D., Dercole, F., & Piccardi, C. (2013). Profiling core-periphery network structure by random walkers. *Scientific Reports, 3*(1), 1. https://doi.org/10.1038/srep01467. Article.

Safadi, H., Johnson, S. L., & Faraj, S. (2021). Who contributes knowledge? Core-periphery tension in online innovation communities. *Organization Science (New York, N.Y.), 32*(3), 752–775. https://doi.org/10.1287/orsc.2020.1364

Schneider, C. M., Moreira, A. A., Andrade, J. S., Havlin, S., & Herrmann, H. J. (2011). Mitigation of malicious attacks on networks. *Proceedings of the National Academy of Sciences of the United States of America, 108*(10), 3838–3841. https://doi.org/10.1073/pnas.1009440108

Sedita, S. R., Caloffi, A., & Lazzeretti, L. (2020). The invisible college of cluster research: A bibliometric core–periphery analysis of the literature. *Industry and Innovation, 27*(5), 562–584. https://doi.org/10.1080/13662716.2018.1538872

Sinha, A., Shen, Z., Song, Y., Ma, H., Eide, D., Hsu, B.-J.(Paul), & Wang, K. (2015). An overview of Microsoft academic service (MAS) and applications. In *Proceedings of the 24th International Conference on World Wide Web* (pp. 243–246). https://doi.org/10.1145/2740908.2742839

Sinha, G., Shahi, R., & Shankar, M. (2010). Human computer Interaction. *2010 3rd International Conference on Emerging Trends in Engineering and Technology (Elmsford, N.Y.)*, 1–4. https://doi.org/10.1109/ICETET.2010.85

Tunç, B., & Verma, R. (2015). Unifying inference of meso-scale structures in networks. *PloS one, 10*(11), Article e0143133. https://doi.org/10.1371/journal.pone.0143133

Van den Heuvel, M. P., Kahn, R. S., Goñi, J., & Sporns, O. (2012). High-cost, high-capacity backbone for global brain communication. *Proceedings of the National Academy of Sciences, 109*(28), 11372–11377. https://doi.org/10.1073/pnas.1203593109

Vinkler, P. (1987). A quasi-quantitative citation model. *Scientometrics, 12*(1), 47–72. https://doi.org/10.1007/BF02016689

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods (San Diego, Calif.), 17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

Waltman, L. (2016). A review of the literature on citation impact indicators. *The Journal of Informetrics, 10*(2), 365–391. https://doi.org/10.1016/j.joi.2016.02.007

Waltman, L., Calero-Medina, C., Kosten, J., Noyons, E. C. M., Tijssen, R. J. W., van Eck, N. J., van Leeuwen, T. N., van Raan, A. F. J., Visser, M. S., & Wouters, P. (2012). The Leiden ranking 2011/2012: Data collection, indicators, and interpretation. *The Journal of the American Society for Information Science (New York, N.Y.) and Technology (Elmsford, N.Y.), 63*(12), 2419–2432. https://doi.org/10.1002/asi.22708

Wang, H., Chen, H., Meng, F., & Bu, Y. (2022). Towards "rich clubs" in scientific publications: A preliminary exploration. In *26th International Conference on Science, Technology and Innovation Indicators (STI 2022)*. https://doi.org/10.5281/zenodo.6951730

Wang, K., Shen, Z., Huang, C., Wu, C.-H., Dong, Y., & Kanakia, A. (2020). Microsoft Academic Graph: When experts are not enough. *Quantitative Science (New York, N.Y.) Studies, 1*(1), 396–413. https://doi.org/10.1162/qss_a_00021

Wedell, E., Park, M., Korobskiy, D., Warnow, T., & Chacko, G. (2022). Center–periphery structure in research communities. *Quantitative Science (New York, N.Y.) Studies, 3*(1), 289–314. https://doi.org/10.1162/qss_a_00184

Weng, C., & Daim, T. U. (2012). Structural differentiation and its implications—Core/periphery structure of the technological network. *The Journal of the Knowledge Economy, 3*(4), 327–342. https://doi.org/10.1007/s13132-011-0048-5

Willis, C. L., & McNamee, S. J. (1990). Social Networks of science and patterns of publication in leading Sociology Journals, 1960 to 1985. *Knowledge, 11*(4), 363–381. https://doi.org/10.1177/107554709001100401

Wu, Z.-X., & Holme, P. (2011). Onion structure and network robustness. *Physical Review E (Norwalk, Conn.), 84*(2), Article 026106. https://doi.org/10.1103/PhysRevE.84.026106

Zelnio, R. (2012). Identifying the global core-periphery structure of science. *Scientometrics, 91*(2), 601–615. https://doi.org/10.1007/s11192-011-0598-0

Zhai, Y., Ding, Y., & Wang, F. (2018). Measuring the diffusion of an innovation: A citation analysis. *The Journal of the Association for Information Science (New York, N.Y.) and Technology (Elmsford, N.Y.), 69*(3), 368–379. https://doi.org/10.1002/asi.23898

Zhang, Y., & Zhang, C. (2015). Citation networks and the emergence of knowledge core. *IEEE Transactions on Knowledge and Data Engineering, 27*(12), 3203–3216. https://doi.org/10.1109/TKDE.2015.2454512. IEEE Transactions on Knowledge and Data Engineering.

Zhao, Z., Bu, Y., & Li, J. (2021). Characterizing scientists leaving science before their time: Evidence from mathematics. *Information Processing & Management, 58*(5), Article 102661. https://doi.org/10.1016/j.ipm.2021.102661

Zhou, S., & Mondragon, R. J. (2004). The rich-club phenomenon in the internet topology. *IEEE Communications Letters, 8*(3), 180–182. https://doi.org/10.1109/LCOMM.2004.823426. IEEE Communications Letters.