Journal of Informetrics 16 (2022) 101345

Contents lists available at ScienceDirect

Journal of Informetrics

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

A directed collaboration network for exploring the order of scientific collaboration

Li Zhai [a], Xiangbin Yan [b,*]

[a] School of Statistics, Jilin University of Finance and Economics, Changchun, 130117, China
[b] School of Economics and Management, University of Science and Technology Beijing, Beijing, 100083, China

ARTICLE INFO

Keywords:
Collaboration networks
Collaboration order
Directed networks
Collaboration diversity

ABSTRACT

The rapid growth of scientific collaboration and its significant role in promoting academic productivity has attracted increasing scientific community attention. The collaboration networks have become a powerful tool for studying scientific collaboration. Collaboration networks commonly used in research treat the collaborators as equal in status. However, the roles and contributions of different collaborators are not the same. Those differences are usually reflected through the signature order of academic achievements. This paper expands the construction of scientific collaboration networks with a directed collaboration network (DCN) to describe the different roles of collaborators and the connectivity and strength of collaborations. We analyzed the theoretical properties of the DCN and constructed evaluation indexes describing the diversity of collaboration order. Based on a case study of published papers in the business field, we discuss the value of the DCN in the characterization and evaluation of scientific collaboration and compare the DCN with two other collaboration networks. We found that the DCN provides a powerful new approach for investigating collaboration laws and patterns.

1. Introduction

The rapid development of information technology has enhanced scholars' ability to communicate and interact with each other, which significantly promotes scientific collaboration. The number of scientific collaborations in various fields is growing rapidly (Ribeiro, Rapini, Silva, & Albuquerque, 2018; Zhai, Yan, Shibchurn, & Song, 2014). Moreover, modern science has entered the era of big science, and its comprehensiveness and complexity require scientific collaboration. Thus, it has become the mainstream approach in scientific development (Cugmas, Mali, & Ziberna, 2020; Milojević, 2010).

Scientific collaboration affects scientific productivity in various ways, such as improving the research efficiency of scholars and the quantity and quality of research achievements (J. Guan, Yan, & Zhang, 2017; Lee & Bozeman, 2005; Scarazzati & Wang, 2019), increasing the possibility of publication (Laband & Tollison, 2000; Lopaciuk-Gonczaryk, 2016), reducing the cost of scientific research (Laband & Tollison, 2000), improving academic reputation (Floyd & Finn, 1994), promoting the dissemination and creation of knowledge (Wuchty, Jones, & Uzzi, 2007), and facilitating the sharing of resources and the cultivation of young scholars (C. He, Wu, & Zhang, 2021). With the development of social networks and complex networks, constructing scientific collaboration networks (Newman, 2001a, 2001b) and studying their structural characteristics have become an important method and research hotspot (Zeng, et al., 2017).

Based on scientific collaboration networks, scholars have conducted extensive and in-depth research on scientific collaboration among academic entities, such as scholars, institutions, countries, regions, and fields of study. The research contents include the for-

* Corresponding author
E-mail address: xbyan@ustb.edu.cn (X. Yan).

https://doi.org/10.1016/j.joi.2022.101345
Received 16 January 2022; Received in revised form 25 September 2022; Accepted 13 October 2022
Available online 28 October 2022

mation and dynamic evolution laws of collaboration (Abbasi, Hossain, & Leydesdorff, 2012; Purwitasari, Fatichah, Sumpeno, Steglich, & Purnomo, 2020; Wagner, Whetsell, & Leydesdorff, 2017), the measurement of collaboration ability and characteristics (C. He, et al., 2021; Schubert, 2012; X. Yan, Zhai, & Fan, 2013), the identification of outstanding scholars and the impact of collaboration on scientific research performance (Abbasi, Wigand, & Hossain, 2014; Khalid, Ibrahim, Selamat, & Kadir, 2017; Liao, 2017), knowledge creation and dissemination (Wang, 2016), etc. For example, a collaboration network of scholars is a social network composed of scholars and the collaborative relationship between them. In this network, scholars are represented by nodes and the collaborative relationships between scholars are expressed as links (Newman, 2001a, 2001b). Since co-authorship is the standard way to measure scientific collaboration, these links usually express the co-authorship between scholars.

Much research is based on the unweighted network. The unweighted network only considers the existence of collaboration, while the weighted network further expresses the strength of collaboration by the link weight. The definition of link weight is mostly based on the number of collaborations (Newman, 2001a) or cumulative contributions (Newman, 2001b). If two scholars published n papers together, there exists a link between them, and the link weight is n. In the setting of link weight by cumulative contributions, if an article has k authors, there are links between k scholars, and all the link weights are $1/(k\text{-}1)$. The link weight will accumulate with increased collaborations. Although research based on the above collaboration networks can reflect collaborator information and collaboration frequency, such network construction methods cannot meet the increasing research needs with various purposes (X. Liu, Bollen, Nelson, & Van, 2005; Zeng, et al., 2017). The above collaboration networks constructed by co-authorship assume that all authors made the same contribution to an article, and the roles of all co-authors are treated as equal. They only consider the existence or strength of the collaboration and ignore information about the collaboration roles. Treating all scholars as the same is likely to lead to deviation of academic ranking and prediction, which will fail to find deep collaboration laws (Zhou, Zeng, Fan, & Di, 2018).

In general academic practice in most fields, the signature order (author order) of an article directly represents the positions of participants in the collaboration, with different roles and contributions (B. He, Ding, & Yan, 2012; Jennings & El-Adaway, 2012; Waltman, 2012). It can also be called the collaboration order. Some scholars have established collaboration networks based on collaboration order. Kim and Diesner (2015) proposed a framework to conceptualize the co-authorship network as a directed, weighted, and self-looped graph according to author order, which can measure the academic contributions of scholars. In a directed network, a directed link (arc) from a node to itself is defined as a self-loop. Here the self-looped graph means each node has an arc that starts from itself and points to itself. In the network, the lower ranked scholars direct to the higher ranked scholars, and the network-based allocation model (Kim & Diesner, 2014) of co-authorship credit is used to construct the weight of links (Kim & Diesner, 2015). Some studies only consider the relationship between the first author and the non-first authors. Yoshikane, Nozawa, and Tsuji (2006) believe that the first author is the leader and the other co-authors are the followers, and they establish a directed collaboration network to identify essential researchers. The arc of the network is constructed by the non-first authors directing to the first author, and the weight of the arc is measured by the number of collaborations. Based on a case study, the topology characteristics of the directed collaboration network for the first author and the correlations of the betweenness centrality, PageRank, HITS algorithm, and the importance of nodes in SIR model are discussed (Zhou, et al., 2018). The SIR model is one of the classic epidemic models. In the SIR model, all nodes in the network belong to one of three states: susceptible (S), infected (I), or recovered (R). It could be extended to measure the spreading effect of nodes in the network according to random simulation under the SIR model (Lü, et al., 2016). The directed collaboration network for the first-author only describes the collaboration between the non-first authors and the first author, but lacks description of the relationship between the non-first authors and ignores the influence of these relationships on the importance of scholars. Although the existing studies have improved the construction of collaboration networks in the setting of weights and direction of links to reflect the collaboration order fully or partially, they are still in the exploratory stage.

Unlike previous studies, in this paper, we construct a new directed collaboration network (DCN), which can reflect the connectivity, order, and intensity of collaborative relationships simultaneously. The proposed construction method will describe the collaborative relationship in more detail and identify patterns and laws of orderly collaboration, providing a new analysis method for scientific collaboration research. Without loss of generality, this paper takes the co-authors network of scholars as an example, and the construction and analysis methods of the network are also applicable to other types of collaboration, such as those among institutions, countries, or regions.

In detail, the orderly collaborative relationship among all collaborators is considered, and the collaboration order is represented by an arc, which points from the lower ranking scholar to the higher ranking scholar, and the arc weight is defined as the number of collaborations. This approach is more conducive to analyzing and discovering patterns in collaboration networks. Since the purpose of the network is to discover the collaboration rules and patterns among collaborators, without considering each scholar's independent work, there is no self-loop in the network, that is, there is no arc from a node to itself, which is different from the study of Kim and Diesner (2015). The directed, weighted network explicitly expresses the order of collaborators, with the first author coming before the second author, the second author before the third, and so on.

The goal of this paper is to explore the laws and patterns of scientific collaboration based on the collaboration order, focusing on the discovery of collaboration laws, rather than determining the contributions of scholars. In the majority of disciplines and fields, it is generally accepted that authors are ranked in descending order of importance and relative contribution (Hagen, 2013; Hu, Rousseau, & Chen, 2010; Marui, Bonjak, & Jeroni, 2011; Savitz, 1999; Wager, 2009), where the collaboration order is equivalent to the contribution order. However, there are exceptions in which the collaboration order is not the same as contribution order. First, in some disciplines or fields, there are convention or cunstoms in which author order is alphabetical (Joanis & Patil, 2021; Levitt & Thelwall, 2013; Waltman, 2012). The use of alphabetical authorship has a continuing downward trend, which hides the relative contributions of co-authors and is detrimental to scientific collaboration (Joanis & Patil, 2021; Kim & Diesner, 2015; Xie, Zhang, & Song, 2021). Second, the corresponding author is considered to have an important contribution and leadership in collaboration, and the collaboration order does not always reflect this, especially when the corresponding author is the last author, as is the convention of some disciplines (Bao & Zhai, 2017; Ueda, Nishizaki, Homma, Devos, & Sanada, 2021). In some fields, the last author is also important and deserves special respect. This person might be the research group leader or senior expert who has made great contributions to the research design (Du & Tang, 2013; Galam, 2011; Hagen, 2010). Some articles mark co-first authors (equal or joint first authors) or co-corresponding authors (equal or joint corresponding authors). In this way, these authors share the identity of the first author or corresponding author and are considered to have equal contributions (Akhabue & Lautenbach, 2010; Du & Tang, 2013; Hu, 2009; Marek & Kosmulski, 2012). Determining the contribution order in these cases is complex and varies by country, discipline, and field (Marui, et al., 2011; Nishizaki, Homma, Ueda, Devos, & Sanada, 2020; Ueda, et al., 2021).

Using the author order as the collaboration order has the following rationale. First, the determination of the author order is complicated and affected by many factors. However, regardless of the principles for determining the author order and the reasons behind them, the author order is agreed to and confirmed by all authors. So, the author order is an important basis for analyzing the collaboration patterns and laws among scholars. Second, this paper focuses on exploring the laws and patterns of orderly collaboration and does not assume that the author order is the contribution order of authors. Therefore, although the author order may be inconsistent with the contribution order, it does not affect the rationality of using the author order to analyze the laws and patterns of orderly collaboration. And, in addition to corresponding authors and other authors with special important status, scholars with higher collaboration ranking will usually be accorded greater perceived contribution and advantages in terms of tenure, professional recognition, salary return, and productivity (Du & Tang, 2013; Wager, 2009), even in fields where the author order is determined by alphabetization (Einav & Yariv, 2006; Joanis & Patil, 2021; Rudd, 1977). Therefore, the author order is important and worthy of further study.

It should be noted that the directed network method proposed in this paper can be improved by some adjustment methods, such as rearrangement of the author order (Kim & Diesner, 2014; X. Z. Liu & Fang, 2012), so as to study the contribution order of scholars. In the Conclusions and Discussion section, we will discuss this and provide a brief processing framework.

The following sections will analyze the DCN theoretically and empirically, and are organized as follows. In Section 2, the construction of the DCN is described in detail, and the differences between it and two other collaboration networks are analyzed theoretically. Section 3 discusses the theoretical properties of the DCN and constructs an evaluation method to handle the diversity of orderly collaboration. Based on a case in the business field, Section 4 provides a detailed analysis of the DCN, including network topology, evaluation of collaboration diversity, and how it reflects collaboration order. Conclusions and discussion are in Section 5.

2. Conceptualization of the directed collaboration network

We extract the orderly collaborative relationships between scholars from the co-authorships. First, take an article with four authors as an example to demonstrate the construction of a directed collaboration network (DCN), as shown in Figure 1. Scholars A1-A4 are the four authors of the article. If only considering the existence of collaboration between scholars A1-A4, a global coupling network of four scholars can be obtained. It is an undirected collaboration network (UCN), where the four scholars are connected to each other. If only the collaborations between the first author and non-first authors are considered, a directed collaboration network for the first author (DCN-F) can be constructed, where the arcs are from the non-first authors to the first author of this article. In the DCN-F, the collaborations between the non-first authors are ignored. In the DCN, the four scholars are connected to each other, and the arcs point from the lower ranking scholar to the higher ranking scholar based on the author order. In this way, the DCN can not only reflect the complete collaborative relationship between scholars, but also reflect the position of the author in the collaboration.

For a set S of articles, collaboration networks formed by each article of set S can be obtained according to the construction method shown in Figure 1. By combining these networks, the complete collaboration network constructed by set S is obtained. In the merging process, we adopt the most used and effective practice. If there are multiple arcs between two scholars (multiple arcs are arcs that have the same source node and target node), these multiple arcs will be merged into a weighted arc, whose weight is the number of multiple arcs, that is, the number of collaborations between the two scholars. Thus, the UCN, DCN-F, and DCN of set S can be constructed, which are all weighted networks with the same set of nodes and different sets of links.

As an example, the UCN, DCN-F, and DCN generated from five articles are shown in Figure 2. They have different structures. The DCN contains the most information and measures the existence, order, and strength of collaborative relationships, providing a framework for analyzing scientific collaboration.

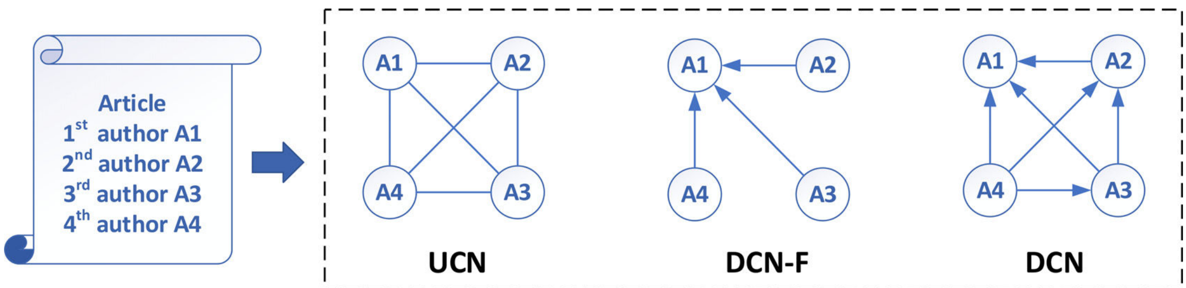


Figure 1. The schematic diagram of three forms of collaboration networks for an article.

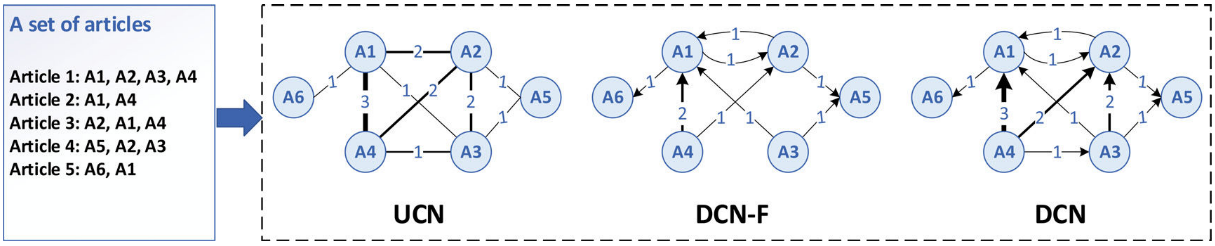


Figure 2. An example of network constructions. (The width of each link is proportional to its weight, and the number on each link is its weight.)

In the following, the relationships among the UCN, DCN-F, and DCN are analyzed. It is assumed that the three networks are constructed from the same article set.

For simplicity, we use an adjacency matrix to represent each network. The adjacency matrix is a square matrix, whose element a_{ij} in row *i* and column *j* represents the weight of the link from source node *i* to target node *j*. If there is no link from node *i* to node *j*, the element $a_{ij} = 0$. The adjacency matrices of the UCN, DCN-F, and DCN are denoted as A_{UCN}, $A_{DCN\text{-}F}$, and A_{DCN}, respectively. According to whether the collaborative relationship is directed, the adjacency matrix A_{UCN} is a symmetric matrix, while $A_{DCN\text{-}F}$ and A_{DCN} are asymmetric. In the UCN, all collaborations are reciprocal, and $a_{ij} = a_{ji}$ for any node *i* and *j*. Since there are no self-loops, the main diagonal elements a_{ii} of these adjacency matrices are all 0 for any node *i*.

By the definitions of the UCN, DCN-F, and DCN, the following conclusions can be drawn.

Conclusion 1. It holds that $A_{UCN} = A_{DCN} + A'_{DCN}$, and A_{UCN} can be uniquely determined by A_{DCN}, and vice versa, where A'_{DCN} is the transposition of A_{DCN}.

Proof: According to the network constructions, the number of collaborations between two scholars *i* and *j* is equal to the sum of the number of collaborations where *i* is sorted before *j* and the number of collaborations where *i* is sorted after *j*, so $a_{ij} = b_{ij} + b_{ji}$ for any *i* and *j*, and thus $A_{UCN} = A_{DCN} + A'_{DCN}$. Here a_{ij} and b_{ij} are elements of A_{UCN} and A_{DCN}, respectively. Therefore, if the A_{DCN} is determined, its transposed matrix A'_{DCN} is uniquely determined. Since $A_{UCN} = A_{DCN} + A'_{DCN}$, A_{UCN} can be uniquely determined. On the other hand, given A_{UCN}, the equation $A_{UCN} = A_{DCN} + A'_{DCN}$ does not uniquely determine A_{DCN}. So we finish the proof.

Conclusion 1 shows that the UCN can be derived from the DCN, but the reverse is not true. Therefore, the DCN contains more information and is more important than the UCN.

Compared with A_{DCN}, $A_{DCN\text{-}F}$ lacks the relationship and order of collaborations among the non-first authors, so the following conclusion holds.

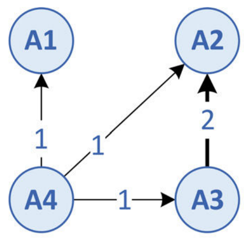
Conclusion 2. For any nodes *i* and *j*, it holds that $c_{ij} \leq b_{ij}$, where c_{ij} and b_{ij} are the elements of $A_{DCN\text{-}F}$ and A_{DCN}, respectively, and the matrix $A_{DCN} - A_{DCN-F}$ is the adjacency matrix of the directed collaboration network for the non-first authors, formed by co-authorship among the non-first authors of articles.

Conclusion 2 shows that, for any *i* and *j*, $b_{ij} - c_{ij}$ measures the times that *i* follows *j* where *j* is not the first author. Conclusion 2 can be easily verified by the collaboration network shown in Figure 2. The collaborations for the non-first authors only exist among nodes A1-A4. The network and adjacency matrix of the directed collaboration network for the non-first authors are shown in Figure 3.

Conclusion 3. For the UCN, DCN-F, and DCN built from the same set of articles, the weak connectivity is consistent.

According to the definitions of the UCN, DCN-F, and DCN, Conclusion 3 is obviously true. If the nodes are divided according to the weak connectivity, the partitions of the three networks are consistent.

Based on the above conclusions, the DCN is the best according to the richness of information; the UCN can be generated by the DCN. Compared with the DCN, the DCN-F lacks the collaboration order and relationship between the non-first authors, and it cannot completely depict collaborations among all authors. The DCN-F can be regarded as a subgraph of DCN. By combining DCN and DCN-F, the collaboration network for the non-first authors can be obtained, which provides a tool for analyzing their collaborations. Therefore, the DCN provides a new framework for the study of scientific collaboration.



$$A_{DCN} - A_{DCN-F} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 2 & 0 & 0 & 1 & 0 \\ 3 & 2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} - \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Figure 3. The directed collaboration subnetwork for the non-first authors in the network shown in Figure 2. (The width of each arc is proportional to its weight.)

3. Analysis of the collaboration order based on the DCN

In this section, we will use the DCN to analyze the characteristics of the collaboration order of scholars. The node degree (Freeman, 1979; Nieminen, 1974; Shaw, 1954) and node strength (Barrat, Barthélemy, Pastor-Satorras, & Vespignani, 2004) are essential network metrics to measure the power and status of nodes. According to the DCN construction method, the arc between two scholars represents the frequency and order of their collaboration, so the meanings of node degree and node strength in the DCN are as follows.

Property 1. There are two node degree properties for the DCN.

(1) The out-degree of a node measures the number of different scholars who have collaborated with it and ranked ahead of it. The in-degree of a node measures the number of scholars who have collaborated with it and ranked behind it.
(2) As the sum of its out-degree and in-degree, the node degree of a node is the number of its orderly collaborators. Here, orderly collaborator refers to the collaborator with a fixed order. If a scholar has collaborated with another scholar twice in a different order, it is recorded as two orderly collaborators.

Property 2. There are two node strength properties for the DCN.

(1) Out-strength of a node represents its number of collaborations in which it is ranked behind, while in-strength represents its number of collaborations in which it is ranked ahead.
(2) As the sum of out-strength and in-strength, the node strength is the total number of collaborations between it and others.

By comparing the node degree and node strength in networks DCN, UCN, and DCN-F, the following corollaries hold. Supposing that three networks are generated from the same set of articles, d_{in}, d_{out}, and d denote the in-degree, out-degree, and degree of a node, respectively, and s_{in}, s_{out}, and s denote the in-strength, out-strength, and strength of a node, respectively. From the conclusions given in Section 2 and the constructions of the UCN, DCN-F, and DCN, it is obvious that the following three corollaries are true.

Corollary 1. For any node i, $d_{in}^{DCN}(i) \geq d_{in}^{DCN-F}(i)$, $d_{out}^{DCN}(i) \geq d_{out}^{DCN-F}(i)$, and $d^{DCN}(i) \geq d^{DCN-F}(i)$, and this is also true for its in-strength, out-strength, and strength.

Corollary 2. For any node i, $d^{DCN}(i) \geq d^{UCN}(i)$, and "=" holds if and only if there is only one arc between node i and each of its collaborators, that is, node i has a fixed collaboration order with each collaborator.

Corollary 3. For any node i, $s^{DCN}(i) = s^{UCN}(i) \geq s^{DCN-F}(i)$. And $s^{UCN}(i) = s^{DCN-F}(i)$ if and only if node i has never been the non-first author of an article co-authored by more than two scholars.

The node degree and node strength in the DCN provide intuitive, effective, and diverse analytical methods and tools for describing collaboration patterns of scholars. The node degree and node strength can be analyzed visually. As shown in Figure 4, for any node i in the DCN, it corresponds to a point $(d_{in}(i), d_{out}(i))$ in the two-dimensional plane rectangular coordinate system with in-degree and out-degree as coordinate axes, which must be in the first quadrant and its boundary.

According to Property 1(1), if the point $(d_{in}(i), d_{out}(i))$ falls on the diagonal with equation $d_{in} = d_{out}$, the number of collaborators of node i ahead of i is equal to the number of collaborators behind it. The closer the point $(d_{in}(i), d_{out}(i))$ is to the diagonal, the closer the in-degree and out-degree of node i are to each other. If node i is below the diagonal, then the in-degree of the node is greater than the out-degree, that is, the collaboration order is closer to the front in author order. In contrast, if a node is above the diagonal, then it is farther back in author order. A point on the horizontal axis has an out-degree of 0, indicating that it has always been the first author in collaborations, while a point on the vertical axis has an in-degree of 0, indicating that it has always been in the last place in collaborations. With the increase of node degree, the number of orderly collaborators increases, and the scope of collaboration

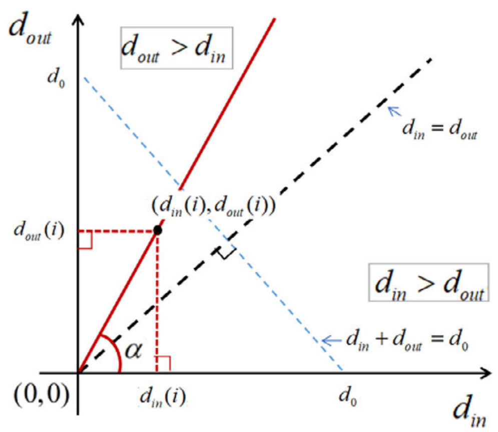


Figure 4. Two-dimensional representation of node degree in the DCN.

increases. As shown in Figure 4, the nodes on the blue dotted line segment all have the same node degree d_0, and they can vary greatly in the distribution of the in-degree and out-degree. As the node degree d_0 increases, the line segment will move parallel to the upper right.

Similarly, a two-dimensional plane rectangular coordinate system with a node's in-strength and out-strength as the coordinate axes can be constructed. Based on the position of the point $(s_{in}(i), s_{out}(i))$ in the coordinate system, the number of times that node i ranks behind other scholars and the number of times that it ranks ahead of other scholars in its collaborations can be analyzed.

Therefore, we can conduct visual analysis for scholars based on the number of collaborators and the number of collaborations according to the two coordinate systems above. Based on these, some derivative indexes can also be built, such as the node degree $d = d_{in} + d_{out}$ and the node strength $s = s_{in} + s_{out}$, which measure the total size of the orderly collaborators and the number of collaborations of a scholar, but these cannot describe collaboration order. In the following, we will build a p^d_{in} index and a p^s_{in} index based on (d_{in}, d_{out}) and (s_{in}, s_{out}) to quantify the diversity of the collaboration order of scholars, respectively.

Definition 1. For any node i in the DCN, we define the p^d_{in} index as $\frac{d_{in}(i)}{d(i)}$, and p^d_{out} index as $\frac{d_{out}(i)}{d(i)}$, where $d(i) = d_{in}(i) + d_{out}(i)$.

Definition 2. For any node i in the DCN, we define the p^s_{in} index as $\frac{s_{in}(i)}{s(i)}$ and p^s_{out} index as $\frac{s_{out}(i)}{s(i)}$, where $s(i) = s_{in}(i) + s_{out}(i)$.

The p^d_{in} index and p^d_{out} index are the proportions of in-degree and out-degree in the node degree, respectively, and the p^s_{in} index and p^s_{out} index are the proportions of in-strength and out-strength in the node strength, respectively. By Definition 1 and Definition 2, the following properties hold.

Property 3. For any node i in the DCN, it holds $p^d_{in}(i) \in [0, 1]$,$p^d_{out}(i) \in [0, 1]$, and $p^d_{in}(i) + p^d_{out}(i) = 1$. In the same way, $p^s_{in}(i) \in [0, 1]$, $p^s_{out}(i) \in [0, 1]$, and $p^s_{in}(i) + p^s_{out}(i)= 1$.

Property 4. For any node i in the DCN, it holds $p^d_{in}(i) = 0 \Leftrightarrow p^s_{in}(i) = 0$ and $p^d_{in}(i) = 1 \Leftrightarrow p^s_{in}(i) = 1$. For the out-direction, the conclusion still holds.

Proof: By Definition 1 and Definition 2, $p^d_{in}(i) = 0 \Leftrightarrow d_{in}(i) = 0$ and $p^s_{in}(i) = 0 \Leftrightarrow s_{in}(i) = 0$. In the DCN, if $d_{in}(i) = 0$, then node i has no in-directed collaborator, so there is no in-directed collaboration, and thus $s_{in}(i) = 0$. And vice versa. So $p^d_{in}(i) = 0 \Leftrightarrow d_{in}(i) = 0 \Leftrightarrow s_{in}(i) = 0 \Leftrightarrow p^s_{in}(i) = 0$. In the same way, $p^d_{out}(i) = 0 \Leftrightarrow p^s_{out}(i) = 0$. By Property 3, $p^d_{in}(i) = 1 \Leftrightarrow p^d_{out}(i) = 1 - p^d_{in}(i) = 0$ and $p^s_{in}(i) = 1 \Leftrightarrow p^s_{out}(i) = 1 - p^s_{in}(i) = 0$. So from $p^d_{out}(i) = 0 \Leftrightarrow p^s_{out}(i) = 0$, we have $p^d_{in}(i) = 1 \Leftrightarrow p^s_{in}(i) = 1$. In the same way, $p^d_{out}(i) = 1 \Leftrightarrow p^s_{out}(i) = 1$. So we can finish the proof.

Note that the p^d_{in} index of 0 means that the node is always the last author in collaborations, and the p^d_{in} index of 1 means that the node is only the first author in collaborations. The same is true of the p^s_{in} index. In extreme cases where the p^d_{in} index (p^s_{in} index) of a scholar equals 0 or 1, the collaboration order of the scholar is constant, and the scholar has no diversity in collaboration order.

Property 5. For any node i in the DCN:

(1) $p^d_{in}(i) = \frac{1}{2} \Leftrightarrow d_{in}(i) = d_{out}(i)$, $p^d_{in}(i) > \frac{1}{2} \Leftrightarrow d_{in}(i) > d_{out}(i)$, and $p^d_{in}(i) < \frac{1}{2} \Leftrightarrow d_{in}(i) < d_{out}(i)$;

(2) $p^s_{in}(i) = \frac{1}{2} \Leftrightarrow s_{in}(i) = s_{out}(i)$, $p^s_{in}(i) > \frac{1}{2} \Leftrightarrow s_{in}(i) > s_{out}(i)$, and $p^s_{in}(i) < \frac{1}{2} \Leftrightarrow s_{in}(i) < s_{out}(i)$.

Property 5 shows that if the p^d_{in} index of a scholar is equal to 1/2, it means that the scholar's out-degree is equal to the in-degree, and then the scholar is in the middle position in collaborators. If it is greater than 1/2, then the scholar is more inclined to be in the top ranking, otherwise the scholar is in the bottom ranking. Furthermore, the relationship between the p^s_{in} index and 1/2 gives a scholar's position in collaborations in terms of the number of collaborations.

Property 6. In the two-dimensional plane with the coordinates of a node's out-degree and in-degree, for any node i in the DCN with coordinate $(d_{in}(i), d_{out}(i))$, the slope of the line that passing through the point $(d_{in}(i), d_{out}(i))$ and the origin is $\frac{1}{p^d_{in}(i)} - 1$. In the same way, in the two-dimensional plane with the coordinates of a node's out-strength and in-strength, the slope of the line that passing through the point $(s_{in}(i), s_{out}(i))$ and the origin is $\frac{1}{p^s_{in}(i)} - 1$.

Proof: By Definition 1 and Property 3, the slope is $\tan(\alpha) = \frac{d_{out}(i)}{d_{in}(i)} = \frac{p^d_{out}(i)}{p^d_{in}(i)} = \frac{1}{p^d_{in}(i)} - 1$ for the line passing through point $(d_{in}(i), d_{out}(i))$ and the origin. The proof of the conclusion for node strength is similar. So we finish the proof.

Property 6 establishes the geometric representations of the p^d_{in} index and p^s_{in} index. By Property 6, if the in-degree of a node is 0, then $\frac{1}{0}$ is regarded as infinity, and the conclusion is still valid.

The p^d_{in} index of a scholar describes the standardization of his/her average ranking among collaborators based on the local network characteristics of the scholar in the DCN. And the p^s_{in} index can describe the standardization of his/her average ranking in their collaborations. The two indexes are conducive to the evaluation of the collaboration status of scholars at a finer granularity.

According to Definition 1 and Definition 2, the p^d_{in} index and p^s_{in} index of nodes A1-A6 in the DCN of Figure 2 are shown in Table 1. According to the two indexes, node A1 and node A2 have the same diversity. The p^d_{in} indexes of A5 and A6 are 1, ranking first in collaborations all the time. However, from the in-degree and in-strength, the number of collaborators and the number of collaborations are different between A5 and A6, which cannot be reflected by the p^d_{in} index and p^s_{in} index. Node A4 is also extreme, ranking last in all collaborations, since its in-degree and in-strength are 0. Thus, A4 has a constant collaboration order, so it lacks diversity in collaboration order. A3 has collaboration diversity and repeated collaborations in the out-direction. Both the p^d_{in} index and p^s_{in} index of A3 are less than 1/2, indicating that scholar A3 mainly ranks lower in collaborations. The p^d_{in} index and p^s_{in} index can

Table 1
Collaboration metrics of nodes in the DCN in Figure 2.

Node	d_{in}	d_{out}	p_{in}^{d}	s_{in}	s_{out}	p_{in}^{s}
A1	3	2	0.600	5	2	0.714
A2	3	2	0.600	5	2	0.714
A3	1	3	0.250	1	4	0.200
A4	0	3	0.000	0	6	0.000
A5	2	0	1.000	2	0	1.000
A6	1	0	1.000	1	0	1.000

Note: d_{in}, d_{out}, s_{in}, s_{out}, p_{in}^{d}, and p_{in}^{s} represent the in-degree, out-degree, in-strength, out-strength, p_{in}^{d} index, and p_{in}^{s} index, respectively.

provide an evaluation perspective of collaboration diversity. In practice, they can be used in combination with other metrics, such as directed degree and directed strength, to provide a more comprehensive analysis of scientific collaboration.

4. An example

4.1. Data

We take an example from the business field to systematically analyze the collaboration networks and compare different construction methods. We choose 24 leading business journals as the data source, which are used in the UTD Top 100 Business School Research Rankings[1]. These journals are an important reference for ranking the research strength of business schools around the world and are highly recognized internationally. The Naveen Jindal School of Management at the University of Texas at Dallas (UTD) uses publications in these journals to provide the top 100 business school rankings based on the total contributions of faculty. The data from the 24 journals for the published period of 2000-2009 are retrieved from the Web of Science database, with 11099 articles (with 12419 authors) recorded by "Article" downloaded in December 2012. There were 9079 (81.8%) co-authored articles and 2.318 authors per article on average.

Based on the co-authorships of these articles, the UCN, DCN-F, and DCN networks are constructed. The networks all contain 12419 nodes, of which 596 (4.799%) are isolated nodes. The number of links in the UCN, DCN-F, and DCN networks are 20098, 12889, and 23347, respectively. The DCN has 3249 (16.166%) more arcs than the UCN. The DCN can depict the collaborations; there are at least two collaborations between two scholars with different author orders. In the partition of weakly connected components, the three networks are consistent, which is consistent with Conclusion 3 in Section 2. In 1821 weakly connected components, the giant connected component contains 8004 (72.11%) nodes. In order to remove the influence of network disconnection on the analysis results, the following analysis is carried out in the giant connected components of each network.

4.2. Results and analysis

4.2.1. Topology characteristics of collaboration networks

First, we explore the similarities and differences in topology characteristics of the DCN, UCN, and DCN-F networks. Table 2 shows that the DCN is different from the UCN and DCN-F in network structure. These differences are consistent with the previous theoretical analysis and are caused by the differences in network construction. Because orderly collaborative relationship is expressed by directed links, the DCN is larger than the UCN and DCN-F in terms of the number of links, network diameter, average path length, average node degree, and network density. The DCN's clustering coefficient lies between those of the UCN and DCN-F. The nodes in the DCN have orderly collaboration with 4.212 scholars on average, where different collaboration orders are counted separately, which is different from the UCN and DCN-F. Like Corollary 3, the DCN and UCN networks have the same average node strength, which is greater than that of the DCN-F.

As shown in Figure 5, the distributions of degree, in-degree, and out-degree all have power-law characteristics in the DCN. The distributions of degree, in-degree, and out-degree in DCN-F all have a larger power exponent in absolute value, and the distributions drop faster. The difference in distributions of node degree between the UCN and DCN is very small, and the fit results of the power-law distribution are the same. Since 0 is the minimum value for both in-degree and out-degree of nodes, $\log(k+1)$ is used to replace $\log(k)$ when exploring the cumulative distributions of in-degree and out-degree of nodes in log-log coordinates. This transformation does not change the original distribution shape.

Assortativity analysis is a standard tool in analyzing network structure (Newman, 2002), which quantifies the assortative or disassortative tendency of a network. In the assortative undirected networks, a node with high node degree tends to link to other nodes with high node degree. Conversely, in the disassortative undirected networks, a node with high node degree tends to link to other nodes with low node degree. The assortativity affects crucial structural and dynamic properties of real-world networks, such as

[1] https://jsom.utdallas.edu/the-utd-top-100-business-school-research-rankings/index.php

Table 2
The topology characteristics of the giant connected components of the DCN, UCN, and DCN-F.

Topology	DCN	UCN	DCN-F
N	8004	8004	8004
E	16856	16354	10201
<l>	12.628	9.554	12.275
D	46	31	33
<d>	4.212	4.086	2.549
<s>	4.823	4.823	2.941
<c>	0.284	0.638	0.033
Density	2.631E-4	1.277E-4	1.593E-4

Note: *N, E, <l>, D, <d>, <s>, <c>*, and *Density* represent the network size, the number of links, average path length, network diameter, average node degree, average node strength, average clustering coefficient, and network density, respectively.

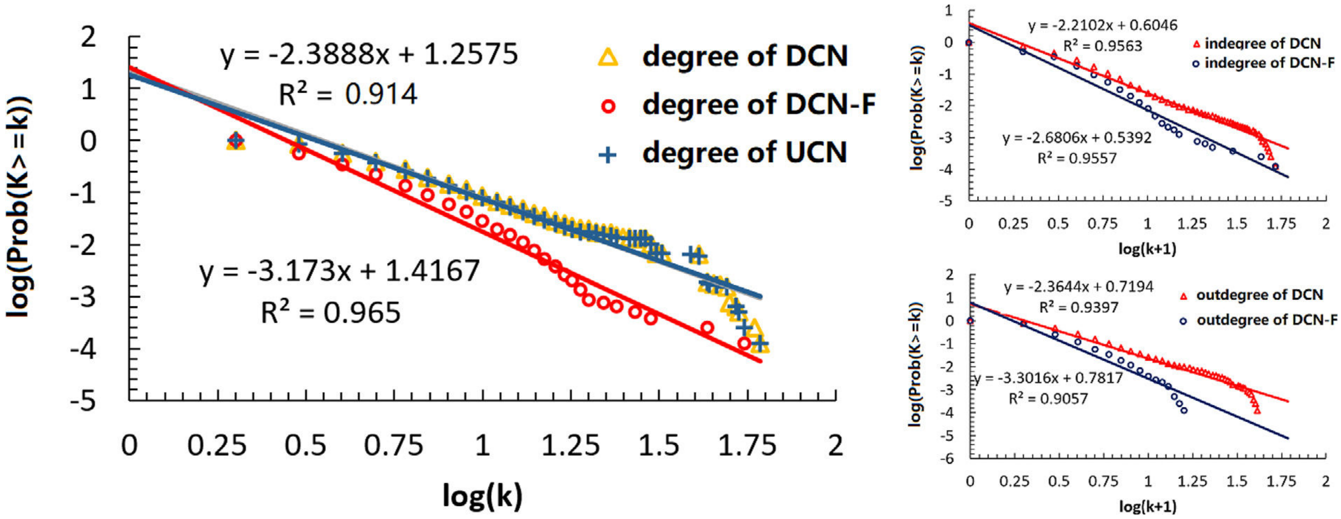
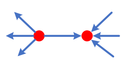
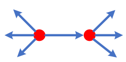
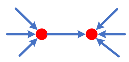
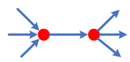
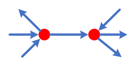


Figure 5. Distributions of node degree in the DCN, UCN, and DCN-F.

Table 3
The assortativity coefficients of the DCN, UCN, and DCN-F.

	r(*out, in*)	*r*(*out, out*)	*r*(*in, in*)	*r*(*in, out*)	*r*(*d, d*)
illustration					
DCN	0.516	0.464	0.369	0.197	0.704
UCN	-	-	-	-	0.729
DCN-F	-0.115	-0.084	-0.161	0.039	-0.095

error tolerance or epidemic spreading (Foster, Foster, Grassberger, & Paczuski, 2010). The assortativity of a network is measured by the assortativity coefficient, which is defined as the Pearson correlation between node degrees of the two nodes connected by a link (Newman, 2002). If assortativity coefficient $r > 0$, then the network is assortative, else if $r < 0$, then the network is disassortative. Foster, et al. (2010) proposed four directed assortativity measures based on in-degree and out-degree of a node in directed networks, which are *r*(*out, in*), *r*(*out, out*), *r*(*in, in*), and *r*(*in, out*) in Table 3. Taking *r*(*out, in*) as an example, it is defined as the Pearson correlation of out-degree of the source node and in-degree of the target node for all arcs. That is, $r(out, in) = \frac{E^{-1}\sum_i (d_{out}^i - \bar{d}_{out})(d_{in}^i - \bar{d}_{in})}{\sigma_{out}\sigma_{in}}$, where E is the number of arcs in the network, d_{out}^i and d_{in}^i are out-degree of the source node and in-degree of the target node for arc i; $\bar{d}_{out} = \frac{1}{E}\sum_i d_{out}^i$, $\sigma_{out} = \sqrt{\frac{1}{E}\sum_i (d_{out}^i - \bar{d}_{out})^2}$, $\bar{d}_{in}$, and σ_{in} are similarly defined. On the basis of the four kinds of assortativity measures (Foster, et al., 2010), we can define the *r*(*d, d*) in a similar way to measure the assortativity between the degrees of source node and target node.

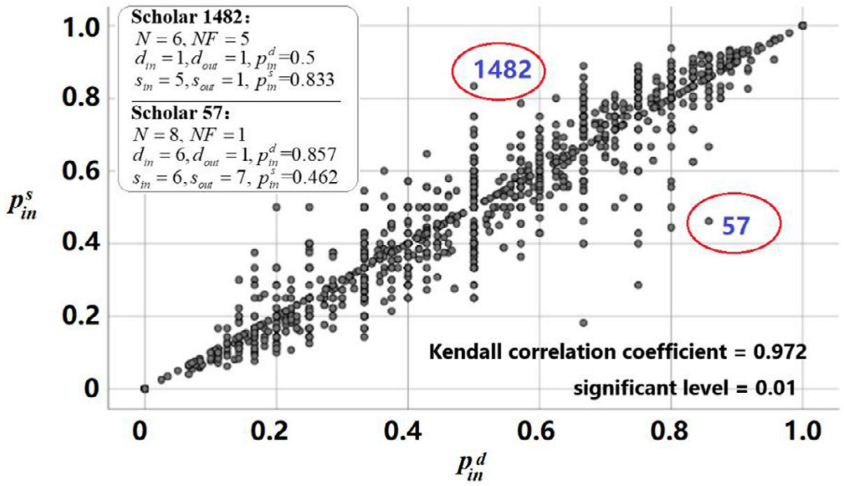


Figure 6. The scatter diagram of the p_{in}^d index and p_{in}^s index. (*N* and *NF* represent the number of articles and the number of articles as the first author, respectively.)

We calculate the five kinds of assortativity measures for the DCN, UCN, and DCN-F, respectively. Table 3 shows that the assortativity coefficients of the DCN are different from those of the DCN-F and UCN. In the DCN, the assortativity coefficients are at least 0.197, which means that the network is assortative. The *r*(*out, in*) is 0.516 in the DCN, indicating that source nodes with high out-degree tend to link to target nodes with high in-degree, that is, scholars who follow many scholars tend to follow scholars who are followed by many others. Both the DCN and UCN have high degree correlation (0.704 and 0.729, respectively), showing that nodes with large node degree tend to link to nodes with large node degree. This implies that scholars with more collaborators tend to cooperate with scholars with more collaborators. Different from the DCN, the DCN-F shows weak disassortative tendency. In the DCN-F, the out-degree of a node refers to the number of articles where it is not the first author, while the in-degree of a node refers to the number of collaborators where it is the first author. Except for *r*(*in, out*), the directed degree of the source node is negatively correlated with that of the target node in DCN-F, where the *r*(*out, out*) is -0.084 and *r*(*in, in*) is -0.161, which is consistent with the conclusions in Zhou, et al. (2018).

4.2.2. Evaluation of orderly collaboration based on the DCN

The DCN provides a new tool for evaluating collaboration patterns. First, the orderly collaboration pattern of scholars will be evaluated according to the proposed p_{in}^d index and p_{in}^s index in Definition 1 and Definition 2. They describe the orderly collaboration of scholars based on the number of orderly collaborators and the number of orderly collaborations, respectively.

We calculate the p_{in}^d index and p_{in}^s index for all scholars. There is a significant positive correlation between the p_{in}^d index and p_{in}^s index shown in Figure 6, and the Kendall correlation coefficient is 0.972. There is also a significant difference between them, which provides a possibility for evaluating collaboration patterns. If two scholars have the same p_{in}^d index, then the scholar with the larger p_{in}^s index will collaborate more frequently in the higher ranking position while maintaining a certain order with the collaborators. For example, scholar 1482 has the largest p_{in}^s index of 0.833 among all scholars with p_{in}^d=0.5. This scholar has 6 articles, in which he/she is the first author of 5 articles, however, he/she works with only one scholar, whose out-degree and in-degree are both 1. In contrast, if two scholars have the same p_{in}^s index, the larger p_{in}^d index of a scholar means that more collaborators follow this scholar, and there are more scattered collaborations between the scholar and the collaborators following him/her. For example, scholar 57 has 8 articles and is the first author of one. This scholar has 7 collaborators, and the in-degree is 6, the out-degree is 1, the in-strength is 6, and the out-strength is 7. So, the p_{in}^s index is 0.462, and the p_{in}^d index is 0.857. Therefore, the p_{in}^d index and p_{in}^s index in the evaluation have a complementary role and cannot replace each other.

In order to investigate whether using the DCN to describe scientific collaboration can synthesize and reflect the academic performance and orderly collaboration of scholars, we will calculate the correlations of the DCN-based network metrics with collaboration order indicators and academic performance of scholar. The collaboration order indicators include the number of articles as the first author, the proportion of articles as the first author, and the average author order of scholars. The academic performance of a scholar is measured by the number of articles and the total number of citations obtained by the articles. These are the most fundamental indexes to measure the academic performance of a scholar (Moldwin & Liemohn, 2018; Onodera & Yoshikane, 2015), measuring the productivity and academic influence of scholars, respectively. They are calculated from the data used in this paper, with the citations data obtained from Web of Science in December 2012.

According to the theory of social capital, scholars accumulate social capital while participating in scientific collaboration. The social capital refers to all the resources obtained through social networks (Abbasi, et al., 2014; Colladon, D'Angelo, & Gloor, 2020), which can bring advantages to enhance the influence of scientific research (Abbasi, et al., 2014; Li, Liao, & Yen, 2013; E. Yan & Ding, 2009; Y. Yan & Guan, 2018). Collaboration networks have become a standard method to measure the social capital accumulated by scholars in collaboration, which reflects the structural dimension of social capital. Existing studies are based on the UCN, without

Table 4
Results of Kendall rank correlation test for nodes in the DCN.

DCN-based metrics	*NF*	*PF*	*AO*	*N*	*C*
in-degree	0.583**	0.410**	-0.360**	0.429**	0.257**
out-degree	-0.244**	-0.505**	0.663**	0.330**	0.240**
degree	0.226**	-0.035**	0.176**	0.595**	0.370**
in-strength	**0.606****	0.410**	-0.360**	0.463**	0.274**
out-strength	-0.231**	-0.496**	0.636**	0.372**	0.263**
strength	0.254**	-0.020*	0.152**	**0.668****	**0.402****
PageRank	0.570**	0.454**	-0.407**	0.366**	0.219**
p_{in}^{d}-index	0.518**	0.738**	-0.735**	0.053**	0.003
p_{in}^{s}-index	0.523**	**0.746****	**-0.741****	0.051**	0.001

Note: *NF, PF, AO, N,* and *C* represent the number of articles as the first author, the proportion of articles as the first author, the average of author order, the number of articles, and the number of citations of articles. The correlation coefficients with largest absolute value are shown in bold. ** $p<0.01$. * $p<0.05$.

considering the collaboration order. The node centralities in the collaboration networks can reflect position and status of scholars in the collaboration, can be used to measure the social capital of scholars (Colladon, et al., 2020; Khalid, et al., 2017), and can also predict the academic influence of scholars (Bertsimas, Brynjolfsson, Reichman, & Silberhoz, 2013; J. C. Guan & Pang, 2018; Khalid, et al., 2017). Existing studies based on the UCN have found that node centralities such as node degree, node strength, and PageRank have a positive correlation with academic performance (Abbasi & Altmann, 2011; Stefano & Zaccarin, 2016; E. Yan & Ding, 2009). Therefore, in the following correlation analysis, we choose the directed node degree, directed node strength, and PageRank (Brin & Page, 1998) as the centrality measures. In addition, the DCN-based network metrics include the p_{in}^{d} index and p_{in}^{s} index. Table 4 shows the Kendall rank correlation test results in the giant connected component of the DCN.

The p_{in}^{d} index and p_{in}^{s} index have absolute advantages for describing orderly collaboration. In detail, the number of articles as the first author has positive correlations with the in-directed metrics of a node. The in-strength has the largest correlation with the number of articles as the first author, followed by in-degree, PageRank, p_{in}^{s} index, and p_{in}^{d} index. These correlations are all greater than 0.5, showing moderate correlation. For the proportion of articles as the first author, the correlation is the strongest between the p_{in}^{s} index and it, reaching 0.746, followed by the p_{in}^{d} index with 0.738. The node out-degree and out-strength are negatively correlated with it, and the absolute values of the correlation coefficients are moderate. The average of author order of a scholar is negatively correlated with the p_{in}^{s} index and has the largest absolute value in all network metrics, followed by the p_{in}^{d} index, whose correlation also reaches -0.735. Both the p_{in}^{d} index and p_{in}^{s} index well reflect the average of author order. In addition, the node's out-degree and out-strength all have a positive correlation with the average of author order, and the values are between 0.6 and 0.7. In general, in measuring collaboration order, the p_{in}^{s} index and p_{in}^{d} index have a good complementary effect.

The node centralities are significantly positively correlated with the academic performance indicators, and the node strength has the strongest positive correlation, which can be seen in Table 4. The node degree also has a relatively greater correlation with academic performance. Therefore, increasing collaboration and collaborators is conducive to improving academic performance, which is consistent with the existing studies (Abbasi & Altmann, 2011; Stefano & Zaccarin, 2016; E. Yan & Ding, 2009). In terms of correlation with academic performance, in-degree is better than out-degree, and in-strength is better than out-strength. This shows that in-directed indicators are better than out-directed ones, indicating that the higher the ranking of a scholar is, the better the academic performance of the scholar will be. This is consistent with the existing research conclusion that scholars with higher ranking have advantages in terms of professional recognition and productivity (Du & Tang, 2013; Einav & Yariv, 2006; Joanis & Patil, 2021). Because the p_{in}^{d} index and p_{in}^{s} index are two structural relative indicators, which just measure the proportion of in-degree in node degree and the proportion of in-strength in node strength for a node, respectively, their ability to reflect the absolute position advantage of nodes in the network will be weak. So, as can be seen from Table 4, the p_{in}^{d} index and p_{in}^{s} index have significant weak positive correlations with the number of articles and insignificant weak positive correlations with the number of citations. Although the correlation is weak, the positive correlation still implies that the above conclusion is tenable, and it is more beneficial to have a higher collaboration ranking. These findings further illustrate the necessity of studying orderly collaboration. The above analysis enlightens the DCN-based network metrics that have a certain predictive effect on academic performance.

5. Conclusions and discussion

Scientific collaboration networks are powerful tools for scientific collaboration research. This paper constructs a directed collaboration network (DCN) to avoid the deficiency that the scientific collaboration network ignores the collaboration order, which reflects the hierarchical relationship of orderly collaboration. Through theoretical analysis, the relationships and differences between the DCN and undirected collaboration network (UCN) and directed collaboration network for the first author (DCN-F) are clarified. The DCN can be degenerated into the UCN. The directed collaboration network for the non-first authors can be derived from the DCN and DCN-F networks. In the DCN, the directed node degree and directed node strength well describe the orderly collaboration pattern of scholars. On this basis, we propose the p_{in}^{d} index and p_{in}^{s} index to measure and evaluate the diversity of orderly collaboration and further discuss their properties and geometric representation.

Based on a practical case constructed by data from 24 business journals over 10 years, we investigate and compare the characteristics and outcomes of the DCN, UCN, and DCN-F. The analysis of network topologies shows that the DCN, UCN, and DCN-F have similar but different network structures, such as having the same number of nodes and the same weak connected component partitions, but having different number of links, shortest path, diameter, average node degree, clustering coefficient, and network assortativity coefficients. The three networks have similar degree distributions, and they all have power-law characteristics. The DCN is assortative, and scholars with more collaborators tend to link to scholars with more collaborators. Among the five kinds of assortativity coefficients, $r(out, in)$ is the largest, indicating that scholars who follow many scholars are most inclined to follow scholars who are followed by many scholars. The diversity evaluation of collaboration order based on the DCN shows that the p_{in}^{d} index and p_{in}^{s} index can identify features of orderly collaboration of scholars from different perspectives and can comprehensively evaluate the pattern of orderly collaboration. For collaboration order, the p_{in}^{d} index and p_{in}^{s} index have a good complementary effect. Node centralities based on the DCN can provide scholars with a more detailed measurement of social capital in social structure than the UCN. The study found that the correlations of in-degree and in-strength with academic performance are higher than the corresponding out-directed indicators, which indicates that the top ranking is better than the bottom ranking in collaboration.

This paper proposes and demonstrates the necessity of constructing DCNs for scientific collaboration research. The DCN enriches scientific collaboration research tools. It refines and expands the UCN to reflect the existence and strength of collaboration as well as the order of collaboration. In the DCN, the directed node degree, directed node strength, and the indexes for collaboration order diversity can evaluate the orderly collaboration patterns of scholars and help to explore the laws of orderly collaboration. Although the analysis is carried out by taking the co-authors network as an example, the method is applicable to the research of orderly collaboration among any scientific entities, such as between institutions and countries or regions.

Here, we also distinguish and discuss the value of the DCN in discovering collaboration laws and identifying contributions of scholars. First, the main goal of the DCN is to find collaboration patterns based on orderly collaboration, rather than determining the contributions of scholars. The construction method and properties of the DCN proposed in this paper are based on the author order, regardless of the method of determining author order. Second, from the aspect of identifying author contributions, the DCN can express the contribution order if the author order is completely determined by the contributions of scholars. In most fields, the author order is determined by the contribution order (Hu, et al., 2010; Wager, 2009; Waltman, 2012), and the scope of disciplines and fields in which this specification is adopted is gradually expanding (B. He, et al., 2012). However, there are still special cases in which it is difficult to determine the contribution order, such as alphabetical order and corresponding author. When the author order is not determined by contribution, we can adjust the DCN's construction method to reflect this. We provide a preliminary framework containing two methods, using (Kim & Diesner, 2014; X. Z. Liu & Fang, 2012) for reference. One is to set up bi-directed arcs to deal with the situation where multiple scholars have the same contribution, such as alphabetical order, co-first authors, and co-corresponding authors. The second is to rearrange the author order to correct the deviation of the author order of some scholars, such as the corresponding author and the last author in some disciplines.

Generally, the alphabetical order of authors hides the contribution of authors, and it is often simply treated as if the contributions of the authors were equal (Zhen & Hui, 2012). For scholars with equal contributions, a complete bi-directed relationship can be established between them, that is, there are two opposite arcs between any two scholars pointing to each other. In order to maintain the property that the arc weight can reflect the number of collaborations, we can set up the pair of bi-directed arcs equally by dividing the original weight of 1. This setting can keep the node strength of scholars unchanged in the adjusted DCN and the original DCN. Similarly, the above method is also applicable to the establishment of network relationships between co-first authors or co-corresponding authors. However, in the case of co-first authors or co-corresponding authors, there may be other authors with contributions different from theirs. In such cases, we can use the following rearrangement strategy of author order.

Taking the case of corresponding author as an example, we introduce the rearrangement strategy. In an article, if there is only one corresponding author and the author's order is inconsistent with his/her contribution, the author order can be adjusted in combination with the recognition of the contribution of the corresponding author (Kim & Diesner, 2014; X. Z. Liu & Fang, 2012). Most studies believe that the contribution of corresponding authors is equal to that of the first author (Bao & Zhai, 2017; Kim & Diesner, 2014; Kim & Kim, 2015; X. Z. Liu & Fang, 2012; Xie, et al., 2021). In this case, the first author and the corresponding author can be adjusted to rank first. We can establish the bi-directed arcs between them according to the above-mentioned treatment of the authors with equal contributions. At the same time, other authors point to them, and the relationship between other authors can be established by the original DCN according to their author order. The contribution of corresponding authors may be greater than that of the first author, and may be less than that of the first author but greater than other authors (Kim & Diesner, 2014). Then, the author order can be rearranged according to their contribution order, and the DCN can be constructed according to the original method. For the case that the last author is the research team leader or senior expert in some fields, the author order can be rearranged by referring to the processing method of the corresponding author. It should be noted that the above rearrangement of author order should be based on the facts and perceptions of the contribution in different disciplines and fields to the first author, the corresponding author, the last author, and other authors (Du & Tang, 2013; Retzer & Jurasinski, 2009; Tscharntke, Hochberg, Rand, Resh, & Krauss, 2007).

To sum up, the complex situation of author order is discussed here, and a preliminary adjustment scheme for the DCN is proposed to adapt to the research of contribution order. More properties and conclusions of the adjusted DCN should be explored in future research. The research will include more practical cases in different fields and explore the patterns and mechanisms of orderly collaboration.

Acknowledgement

This work was supported by the National Natural Science Foundation of China [grant numbers 71971096, 72025101, 72001103], the National Social Science Foundation of China [grant number 20BTJ062], Jilin Province education science planning project [grant number ZD21055] and the Key project of Jilin University of Finance and Economics [grant number 2019Z03].

References

Abbasi, A., & Altmann, J. (2011). On the correlation between research performance and social network analysis measures applied to research collaboration networks. The 44th Hawaii International Conference on System Science (HICSS-44). *Hawaii, USA*.

Abbasi, A., Hossain, L., & Leydesdorff, L. (2012). Betweenness centrality as a driver of preferential attachment in the evolution of research collaboration networks. *Journal of Informetrics, 6*(3), 403–412.

Abbasi, A., Wigand, R. T., & Hossain, L. (2014). Measuring social capital through network analysis and its influence on individual performance. *Library & Information Science Research, 36*(1), 66–73.

Akhabue, E., & Lautenbach, E. (2010). "Equal" contributions and credit: an emerging trend in the characterization of authorship. *Annals of Epidemiology, 20*, 868–871.

Bao, P., & Zhai, C. (2017). Dynamic credit allocation in scientific literature. *Scientometrics, 112*(1), 1–12.

Barrat, A., Barthélemy, M., Pastor-Satorras, R., & Vespignani, A. (2004). The architecture of complex weighted networks. In *Proceedings of the National Academy of Sciences of The United States of America: 101* (pp. 3747–3752).

Bertsimas, D., Brynjolfsson, E., Reichman, S., & Silberhoz, J. (2013). Network analysis for predicting academic impact. *Thirty Fourth International Conference on Information Systems (ICIS)*. Milan.

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems, 30*, 107–117.

Colladon, A. F., D'Angelo, C. A., & Gloor, P. A (2020). Predicting the future success of scientific publications through social network and semantic analysis. *Scientometrics, 124*(1), 357–377.

Cugmas, M., Mali, F., & Ziberna, A. (2020). Scientific collaboration of researchers and organizations: a two-level blockmodeling approach. *Scientometrics, 125*(3), 2471–2489.

Du, J., & Tang, X. (2013). Perceptions of author order versus contribution among researchers with different professional ranks and the potential of harmonic counts for encouraging ethical co-authorship practices. *Scientometrics, 96*(1), 277–295.

Einav, L., & Yariv, L. (2006). What's in a Surname? The Effects of Surname Initials on Academic Success. *Journal of Economic Perspectives, 20*(1), 175–188.

Floyd, S. W., & Finn, D. M. (1994). "Only If I'm First Author": conflict over credit in management scholarship. *Academy of Management Journal, 37*(3), 734–747.

Foster, J. G., Foster, R. V., Grassberger, R., & Paczuski, R. (2010). Edge direction and the structure of networks. In *Proceedings of the National Academy Of Sciences Of The United States Of America: 107* (pp. 10815–10820).

Freeman, L. C. (1979). Centrality in social networks: conceptual clarification. *Social Networks, 1*(3), 215–239.

Galam, S. (2011). Tailor based allocations for multiple authorship: a fractional gh-index. *Scientometrics, 89*(1), 365–379.

Guan, J., Yan, Y., & Zhang, J. J. (2017). The impact of collaboration and knowledge networks on citations. *Journal of Informetrics, 11*(2), 407–422.

Guan, J. C., & Pang, L. X. (2018). Bidirectional relationship between network position and knowledge creation in Scientometrics. *Scientometrics, 115*(1), 201–222.

Hagen, N. T. (2010). Harmonic publication and citation counting: sharing authorship credit equitably – not equally, geometrically or arithmetically. *Scientometrics, 84*(3), 785–793.

Hagen, N. T. (2013). Harmonic coauthor credit: A parsimonious quantification of the byline hierarchy. *Journal of Informetrics, 7*(4), 784–791.

He, B., Ding, Y., & Yan, E. (2012). Mining patterns of author orders in scientific publications. *Journal of Informetrics, 6*(3), 359–367.

He, C., Wu, J., & Zhang, Q. (2021). Characterizing research leadership on geographically weighted collaboration network. *Scientometrics, 126*(5), 4005–4037.

Hu, X. (2009). Loads of special authorship functions: Linear growth in the percentage of "equal first authors" and corresponding authors. *Journal of the American Society for Information Science & Technology, 60*(11), 2378–2381.

Hu, X., Rousseau, R., & Chen, J. (2010). In those fields where multiple authorship is the rule, the h-index should be supplemented by role-based h-indices. *Journal of Information Science, 36*(1), 73–85.

Jennings, M. M., & El-Adaway, I. H. (2012). Ethical issues in multiple-authored and mentor-supervised publications. *Journal of Professional Issues in Engineering Education & Practice, 138*(1), 37–47.

Joanis, S. T., & Patil, V. H. (2021). Alphabetical ordering of author surnames in academic publishing: A detriment to teamwork. *PLoS One, 16*(5), Article e0251176.

Khalid, N. H., Ibrahim, R., Selamat, A., & Kadir, M. R. A. (2017). Collaboration patterns of researchers using social network analysis approach. In *IEEE International Conference on Systems, Man, and Cybernetics* (pp. 001632–001637).

Kim, J., & Diesner, J. (2014). A network-based approach to coauthorship credit allocation. *Scientometrics, 101*(1), 587–602.

Kim, J., & Diesner, J. (2015). Coauthorship networks: A directed network approach considering the order and number of coauthors. *Journal of the Association for Information Science & Technology, 66*(12), 2685–2696.

Kim, J., & Kim, J. (2015). Rethinking the comparison of coauthorship credit allocation schemes. *Journal of Informetrics, 9*(3), 667–673.

Laband, D. N., & Tollison, R. D. (2000). Intellectual Collaboration. *Journal of Political Economy, 108*(3), 632–662.

Lee, S., & Bozeman, B. (2005). The impact of research collaboration on scientific productivity. *Social Studies of Science, 35*(5), 673–702.

Levitt, J. M., & Thelwall, M. (2013). Alphabetization and the skewing of first authorship towards last names early in the alphabet. *Journal of Informetrics, 7*(3), 575–582.

Li, E. Y., Liao, C. H., & Yen, H. R. (2013). Co-authorship networks and research impact: A social capital perspective. *Research Policy, 42*(9), 1515–1530.

Liao, C. H. (2017). Exploring the social effect of outstanding scholars on future research accomplishments. *Journal of the Association for Information Science & Technology, 68*(10), 2449–2459.

Liu, X., Bollen, J., Nelson, M. L., & Van, D. (2005). Co-authorship networks in the digital library research community. *Information Processing and Management, 41*(6), 1462–1480.

Liu, X. Z., & Fang, H. (2012). Fairly sharing the credit of multi-authored papers and its application in the modification of h-index and g-index. *Scientometrics, 91*(1), 37–49.

Lopaciuk-Gonczaryk, B. (2016). Collaboration strategies for publishing articles in international journals – A study of Polish scientists in economics. *Social Networks, 44*, 50–63.

Lü, L., Chen, D., Ren, X. L., Zhang, Q. M., Zhang, Y. C., & Zhou, T. (2016). Vital nodes identification in complex networks. *Physics Reports, 650*, 1–63.

Marek, & Kosmulski. (2012). The order in the lists of authors in multi-author papers revisited. *Journal of Informetrics, 6*(4), 639–644.

Marui, A., Bonjak, L., & Jeroni, A. (2011). A systematic review of research on the meaning, ethics and practices of authorship across scholarly disciplines. *PLoS One, 6*(9), e23477.

Milojević, S. (2010). Modes of collaboration in modern science: Beyond power laws and preferential attachment. *Journal of the American Society for Information Science and Technology, 61*(7), 1410–1423.

Moldwin, M. B., & Liemohn, M. W. (2018). High-citation papers in space physics: Examination of gender, country, and paper characteristics. *Journal of Geophysical Research-Space Physics, 123*(4), 2557–2565.

Newman, M. E. J. (2001a). Scientific collaboration networks. I. Network construction and fundamental results. *Physical Review E Statistical Nonlinear & Soft Matter Physics, 64*(1), Article 016131.

Newman, M. E. J. (2001b). Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical Review E Statistical Nonlinear & Soft Matter Physics, 64*(2), Article 016132.

Newman, M. E. J. (2002). Assortative mixing in networks. *Physical Review Letters, 89*(20), Article 208701.

Nieminen, J. (1974). On the centrality in a graph. *Scandinavian journal of psychology, 15*(1), 332–336.
Nishizaki, Y., Homma, Y., Ueda, R., Devos, P., & Sanada, S. (2020). Differences in author ranking: Development of SIGAPS scoring system for Japanese version. *Journal of General and Family Medicine, 21*(5), 208–209.
Onodera, N., & Yoshikane, F. (2015). Factors affecting citation rates of research articles. *Journal of the Association for Information Science and Technology, 66*(4), 739–764.
Purwitasari, D., Fatichah, C., Sumpeno, S., Steglich, C., & Purnomo, M. H. (2020). Identifying collaboration dynamics of bipartite author-topic networks with the influences of interest changes. *Scientometrics, 122*(3), 1407–1443.
Retzer, V., & Jurasinski, G. (2009). Towards objectivity in research evaluation using bibliometric indicators—a protocol for incorporating complexity. *Basic and Applied Ecology, 10*, 393–400.
Ribeiro, L. C., Rapini, M. S., Silva, L. A., & Albuquerque, E. M. (2018). Growth patterns of the network of international collaboration in science. *Scientometrics, 114*(1), 159–179.
Rudd, E. (1977). The effect of alphabetical order of author listing on the careers of scientists. *Social Studies of Science, 7*(2), 268–269.
Savitz, D. A. (1999). What can we infer from author order in epidemiology? *American Journal of Epidemiology, 149*(5), 401–403.
Scarazzati, S., & Wang, L. L. (2019). The effect of collaborations on scientific research output: the case of nanoscience in Chinese regions. *Scientometrics, 121*(2), 839–868.
Schubert, A. (2012). A Hirsch-type index of co-author partnership ability. *Scientometrics, 91*(1), 303–308.
Shaw, M. E. (1954). Group structure and the behavior of individuals in small groups. *Journal of Psychology: interdisciplinary and applied, 38*(1), 139–149.
Stefano, D. D., & Zaccarin, S. (2016). Co-authorship networks and scientific performance: an empirical analysis using the generalized extreme value distribution. *Journal of Applied Statistics, 43*(1), 262–279.
Tscharntke, T., Hochberg, M. E., Rand, T. A., Resh, V. H., & Krauss, J. (2007). Author sequence and credit for contributions in multiauthored publications. *PLoS Biology, 5*(1), e18.
Ueda, R., Nishizaki, Y., Homma, Y., Devos, P., & Sanada, S. (2021). The relationship between contributions of authors and author order. *Journal of General and Family Medicine, 22*(6), 361–362.
Wager, E. (2009). Recognition, reward and responsibility: Why the authorship of scientific papers matters. *Maturitas, 62*(2), 109–112.
Wagner, C. S., Whetsell, T. A., & Leydesdorff, L. (2017). Growth of international collaboration in science: revisiting six specialties. *Scientometrics, 110*(3), 1633–1652.
Waltman, L. (2012). An empirical analysis of the use of alphabetical authorship in scientific publishing. *Journal of Informetrics, 6*(4), 700–711.
Wang, J (2016). Knowledge creation in collaboration networks: Effects of tie configuration. *Research Policy, 45*(1), 68–80.
Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science, 316*(5827), 1036–1039.
Xie, Q., Zhang, X., & Song, M. (2021). A network embedding-based scholar assessment indicator considering four facets: Research topic, author credit allocation, field-normalized journal impact, and published time. *Journal of Informetrics, 15*(4), Article 101201.
Yan, E., & Ding, Y. (2009). Applying centrality measures to impact analysis: A coauthorship network analysis. *Journal of the American Society for Information Science and Technology, 60*(10), 2107–2118.
Yan, X., Zhai, L., & Fan, W. (2013). C-index: A weighted network node centrality measure for collaboration competence. *Journal of Informetrics, 7*(1), 223–239.
Yan, Y., & Guan, J. C. (2018). Social capital, exploitative and exploratory innovations: The mediating roles of ego-network dynamics. *Technological Forecasting and Social Change, 126*, 244–258.
Yoshikane, F., Nozawa, T., & Tsuji, K. (2006). Comparative analysis of co-authorship networks considering authors' roles in collaboration: Differences between the theoretical and application areas. *Scientometrics, 68*(3), 643–655.
Zeng, A., Shen, Z. S., Zhou, J. L., Wu, J. S., Fan, Y., Wang, Y. G., & Stanley, H. E. (2017). The science of science: From the perspective of complex systems. *Physics Reports, 714-715*, 1–73.
Zhai, L., Yan, X., Shibchurn, J., & Song, X. (2014). Evolutionary analysis of international collaboration network of Chinese scholars in management research. *Scientometrics, 98*(2), 1435–1454.
Zhen, L. X., & Hui, F. (2012). Modifying h-index by allocating credit of multi-authored papers whose author names rank based on contribution. *Journal of Informetrics, 6*(4), 557–565.
Zhou, J., Zeng, A., Fan, Y., & Di, Z. (2018). Identifying important scholars via directed scientific collaboration networks. *Scientometrics, 114*(3), 1327–1343.