



Studying the linkage patterns and incremental evolution of domain knowledge structure: a perspective of structure deconstruction

Qi Wang¹ · Bentao Zou¹ · Jialin Jin³ · Yuefen Wang^{1,2} 

Received: 2 November 2023 / Accepted: 16 May 2024 / Published online: 14 June 2024
© Akadémiai Kiadó, Budapest, Hungary 2024

Abstract

The knowledge structure is the result of continuous evolution of the forces intertwined with knowledge linkages and structural patterns. However, the dynamics along this path are still not fully understood. This study aims to investigate the linkage patterns and incremental evolution of domain knowledge structure from the perspective of structure deconstruction. To this end, we proposed a novel framework that integrates the incremental update mechanism of knowledge network construction, subgraph enumeration, and knowledge combination. The proposed integrative framework enables us to embed time-related node attributes into identified subgraphs and to deconstruct specific types of decomposable structure into exiting knowledge combinations and potential knowledge combinations. Results from our case studies, the IIoT and the Metaverse fields, confirmed that the proposed framework is applicable to reveal the underlying knowledge linkage patterns and relative evolution strength. The identified decomposable structures suggest that the path toward knowledge linkages mainly follows a mixed strategy (e.g., high impact knowledge elements are more likely to be linked with elements of middle/low level of impact). The framework designed in this study, together with findings from two fields, elucidates specific evolutionary dynamics through a combined analysis of motifs and structural deconstruction. These findings hold implications for practitioners and policymakers seeking to develop a nuanced understanding of the field.

Keywords Knowledge structure · Dynamic keywords co-occurrence networks · Decomposable structures · Linkage patterns

✉ Yuefen Wang
yuefen163@163.com

¹ School of Economics & Management, Nanjing University of Science & Technology, Nanjing, China

² Institute for Big Data Science, Tianjin Normal University, Tianjin, China

³ School of Management and Economics, North China University of Water Resources and Electric Power, Zhengzhou, China

Introduction

Studying knowledge structure has become increasingly crucial for both researchers and managers (Khasseh et al., 2017; Cheng et al., 2020). For the most part, researchers have concerned on representing and applying knowledge structures (Katsurai & Ono, 2019; Hosseini et al., 2021). To explore such knowledge structures' evolutionary characteristics, networks comprising of knowledge elements as nodes and interrelationships as edges become an appropriate fit. Pairwise connections, i.e., edges, are non-decomposable in such networks, which substantively hinders the understanding of knowledge structure. Taking into account dynamic linkages of collective knowledge elements (Fortunato et al., 2018), combined with knowledge elements' attributes, enables the construction of scientific knowledge networks. While the literature extensively covers the study of global properties of knowledge structure (Castillo-Vergara et al., 2018; Cho, 2020), the investigation of decomposable structures has received limited attention. Given the central role decomposable structures play in shaping knowledge structures, tracking linkage patterns of re-occurring and significant decomposable structures has the potential to provide important and distinctive insights into the evolution of knowledge structures. To this end, decomposable motif structures such as 3-node subgraphs (i.e., triads), 4-node subgraphs (i.e., tetrads), etc., provide a range of possible structures as the output value (i.e., knowledge structure) for further investigation.

In statistical bibliography or bibliometrics, existing studies have characterized the knowledge structure of a field or discipline through two network-based methods (Choudhury et al., 2020). The first method involves employing co-citation networks to explore the structure of scientific communication based on the relationships between various cited and citing documents (González-Valiente et al., 2021). The second method entails utilizing word co-occurrence or co-word networks (Lee & Lee, 2021). In its essence, both methods capture the interactions of a domain-specific knowledge system by representing metadata information (such as authors, institutions, citations, keywords, etc.) from scientific publications as nodes, and their associations (co-citation or co-occurrence relationships) as edges connecting them. The term co-occurrence was first coined by Harris (1954), and the techniques of word co-occurrence analysis was then developed by Callon et al. (Callon et al., 1991). When two words co-occur within one paper, it suggests a relationship between the topics they represent. In most cases, authors select keywords for their articles to summarize the research topics. Therefore, word co-occurrence networks are often constructed using author-selected keywords, such that they become keywords co-occurrence networks (KCNs). The growth of KCNs is basically contributed by the increase of publications and, more specifically, by the addition of a complete N-graph (N fully connected nodes) from a new paper's keywords list. Owing to the continuous input of such complete N-graphs, a range of different types and sizes of subgraphs were generated, resulting in an increase in the scale and complexity of the network structure.

On the other hand, scientific knowledge creation is in nature a dynamic process and benefits from the scholarly communication within and across fields. This implies that keywords with different properties are connected simultaneously or successively. This inspiration led us to collectively consider the properties of knowledge elements, specific decomposable structures, and the dynamics of knowledge networks, thereby revealing patterns and the evolution process of knowledge combinations. Figure 1 presents the overall framework behind the rationale. In particular, the dynamic KCNs are generated by progressively adding complete N-graphs, during which the nodes and topology are

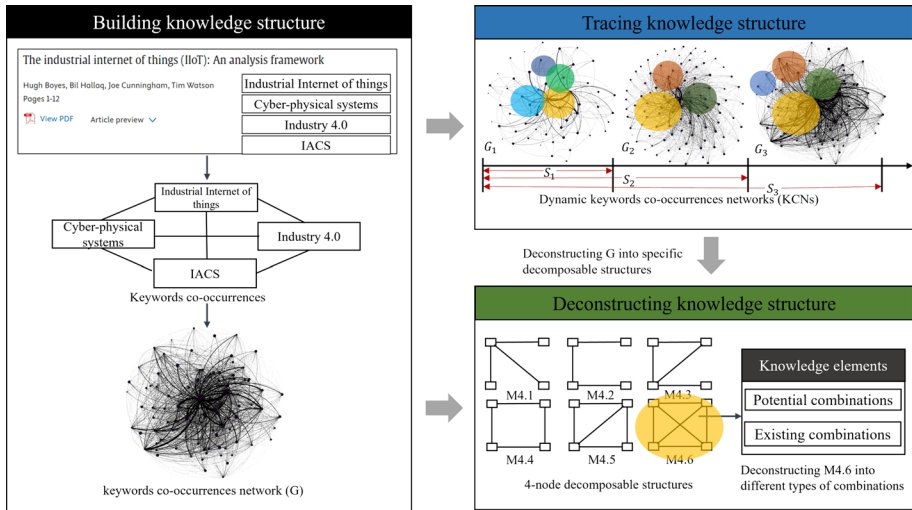


Fig. 1 Overview of the proposed research framework

incrementally updated in tandem with new papers published. This method enables us to view papers as a variable that controls the generation of knowledge networks. Then, we break the network into specific decomposable structures based on a motif detection algorithm (Wernicke, 2005), and investigate implicit combination patterns via time-related node properties. Lastly, leveraging the output decomposable structures, we innovatively investigate the evolution of knowledge structures by measuring the relative variation of different types of output N-graphs.

In this paper, we are going to present answers to the following two questions: (1) what rules underlie the linkage patterns of knowledge elements that we see in decomposable structures? (2) taking the literature growth and keywords properties into account, how do we quantify existing or "potential" knowledge combinations, and further understand the evolution of knowledge structures? The remainder of this paper is organized as follows. Sect. “**Related Work**” briefly reviews related work to our study, Sect. “**Data and Methods**” shows the data and methods, Sect. “**Results**” presents two case studies from IIoT and Metaverse, respectively, Sect. “**Discussion**” discusses our findings, and Sect. “**Conclusion**” concludes this paper.

Related work

As a specific kind of knowledge representation techniques, knowledge networks of varying types have been employed to understand the organization and evolution of scientific knowledge. Therefore, in what follows, we first briefly reviewed major methods and results in analyzing knowledge structure through the use of KCN, which are determined by the structure to be represented. Then, we can move on to the decomposable structures in the literature, which have been mainly studied either from the theory or application aspects.

Knowledge structure and dynamic KCNs

Since scientific knowledge is formed by concepts and relations embodied in various scholarly artifacts, it is best described as network with complex topology (Boccaletti et al., 2006; Fortunato et al., 2018). KCNs have been extensively applied to the elucidation and mapping of knowledge structure, because of the promising effectiveness in expressing knowledge components and knowledge structure (Haunschild et al., 2019; Cho, 2020). To analyze the structure of scientific knowledge, a series of statistical metrics are utilized to shed light on exploit the structural properties at the basis of real networks, for instance node degree, degree distribution, network diameter, clustering coefficient, shortest path length, etc. The main result has been the discerning of the connectivity, sparsity, and aggregation of knowledge structure (Zhang et al. 2016; Cheng et al., 2020). Moreover, several unifying principles are revealed, such as the degree correlations, scale-free distribution, relatively small characteristic path lengths, and the presence of community structure (La & Chai, 2021).

There is also a spectrum of literature that uses dynamic KCNs to reproduce the growth of network and capture the dynamic nature of knowledge structure. In dynamic KCNs, new knowledge elements and interactions are added over time (Katsurai & Ono, 2019). In extant works, dynamic KCNs are defined as a set of temporal networks $\{G_1, G_2, \dots, G_t\}$ (Balili et al., 2020). Although dynamic KCNs have received enormous interests, they still present two major drawbacks: (1) temporal networks have limited ability to accurately represent the growth of scientific knowledge, as 't' is usually divided into equally spaced time intervals (often on an annual basis); (2) while temporal networks can control the division of time intervals, it also precludes the possibility of controlling the division of papers. In response to the above issues, we propose a simple, yet powerful method based on graph operations for constructing dynamic KCNs. This method adheres to the principle that literature growth drives the growth of networks, providing flexible control over the input subgraphs.

Decomposable structure and knowledge combinations

Decomposable structures are small, connected, non-isomorphic subgraphs, also known as network motifs. Therefore, the motif discovery techniques are able to find significantly over-represented decomposable structures in specific networks. After Milo et al. (2002) introduced the concept of network motifs, the use of motifs to capture interactions among hidden basic units has been well established. Motifs have been proved to be a powerful graph analysis tool in various fields. Sporns and Kötter (2004) performed an analysis of structural and functional motifs in various brain networks, leading to the discovery that highly evolved neural architectures are structured to optimize functional repertoires. Then, Krumov et al. (2011) measured the correlation between motifs and citations in co-authorship networks. More recently, Zou et al. (2023) connected motifs with collaboration networks to investigate the knowledge transmitting functions in identified scientific teams. Furthermore, motifs have been applied in a variety of tasks, from defining which communities that vertices belong to (Arenas et al., 2008) to improving network clustering accuracy (Benson et al., 2016), and to optimizing link prediction performance (Wang et al., 2020).

With regard to the specific application of motifs in depicting knowledge structure, prior research has mostly focused on examining how the concentration of various building blocks

affects structural stability. Feng et al. (2021) used network motifs to discover the underlying information structure and evolution rules of tagged knowledge networks. Based on this, Wang et al. (2022) further utilized network motifs to analysis the structure and development of knowledge label network. It has shown that the network motifs theory plays an essential role in characterizing the mesoscopic layer of knowledge structure. These studies follow the same strict norm, the significance characteristic per se, which originates from the motif definition. However, what behind such building blocks is still unclear. It is safe to say that less attention has been given to deconstruct knowledge structure, especially the part in connecting motif theory with knowledge linkage patterns.

When new knowledge elements are successfully combined with established ones, it often violates expectations and leads to the creation of novel ideas with high impact (Larivière et al., 2015). Many networks have complex and highly non-linear structures. While motifs alone only reveal significant subgraph patterns in knowledge networks, node attributes however are needed to understand how and why certain subgraphs are significant patterns. Towards this, colored motifs and types graphlets are the branches of studies most closely related to ours (Qian et al., 2011; Ribeiro & Silva, 2014; Rossi et al., 2021). These studies typically extend the purely structural motifs by integrating color or type information into nodes or edges. Building upon this inspiration, we explore how the scientific knowledge, originating from different time periods or influenced by different factors, is organized and evolves from a knowledge element properties combination view. Also, the relationship between innovation and knowledge combinations has been widely discussed, and it is generally accepted that effective knowledge combinations are more likely to lead to innovative breakthroughs (Kogut & Zander, 1992; Nerkar, 2003; Tolstoy, 2010; Ji et al., 2020; Han et al., 2020). That being said, not every knowledge combination has the same level of power to support knowledge creation and contributes identically to the evolution of knowledge structure (Katila & Ahuja, 2002; Kuo et al., 2019). Thus, it further demands deconstruction on specific structures to identify both existing and potential knowledge combinations by controlling over the input subgraphs.

Data and methods

Data and pre-processing

We limit this study to illustrate evolutionary knowledge structures in specific domains, where knowledge combinations are assumed to be frequently updated, and the evolution of knowledge structure is more observable. The first domain is related to the Industrial Internet of Things (IIoT).¹ With the rapid advancement of Artificial Intelligence and related technologies, the IIoT has emerged as a highly demanding field that involves the integration of physical machines, advanced analytics, and the internet to improve industrial processes, increase efficiency, and enhance productivity (Serror et al., 2021). The second domain revolves around the Metaverse¹, conceptualized as enduring and immersive digital

¹ The search strategies were defined as TS=("industrial internet" OR "industrial IoT" OR "IIoT") for IIoT, and TS=(Metaverse OR "Virtual World") for Metaverse, with English articles published between 1999 and 2022 being included. Additionally, searches were conducted in other databases such as Scopus and ScienceDirect. While there were no notable discrepancies between them, WoS offered more comprehensive metadata for articles.

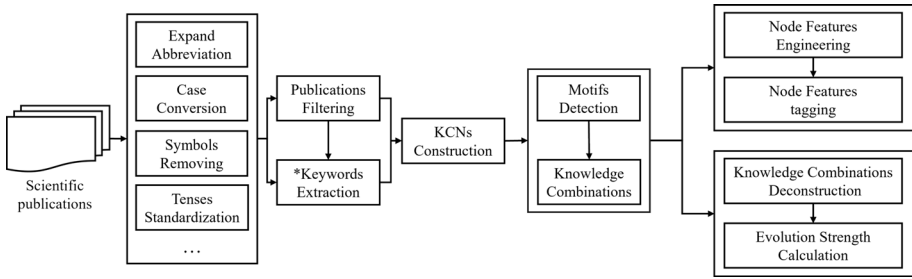


Fig. 2 The workflow of data processing

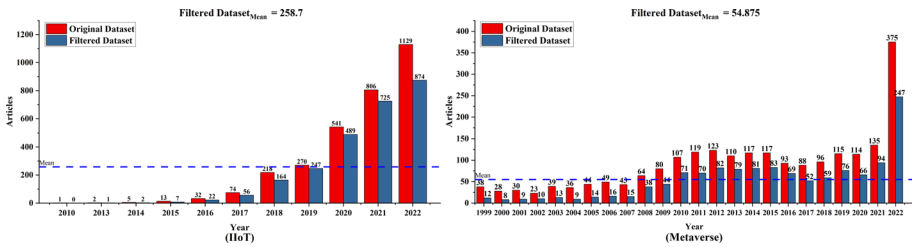


Fig. 3 Distribution of articles in IIoT (left) and Metaverse (right)

environments that leverage various enabling technologies, facilitating the creation, dissemination, and utilization of digital assets (Gao et al., 2023). Article metadata information includes title (TI), abstract (AB), author-selected keywords (DE), year of publication (PY), publication date (PD), and WoS number (UT). The following constraints were considered during the search: (1) articles published in English journals from Jan. 1999 to Dec. 2022, (2) articles indexed in the Web of Science Core Collection, including SCI-E, SSCI, and CPCI-S, and (3) articles were sorted by date in ascending order (oldest first) prior to downloading the metadata. To save space, we will use the abbreviations D_I and D_M to refer to the datasets related to IIoT and Metaverse, respectively, throughout the rest of this article.

The construction of KCN starts with the extraction and filtering of author-selected keywords. Firstly, we removed missing values instead of filling them with keywords extracted from other fields (such as titles, abstracts, or introduction). The reason behind is that we focused exclusively on the author-selected keywords that the authors recognized as relevant to the article. Next, we move on to the text processing phase. Apart from a few commonly used abbreviations, we manually standardized abbreviations into their full forms. With the help of the NLTK text preprocessing tool, we removed symbols like hyphens and performed lemmatization. Figure 2 illustrates the workflow of data processing.

Typically, scholars select a “highly relevant” subset of keywords for co-word analysis, among which keyword frequency is a common criterion. Therefore, we extracted author-selected keywords that appeared in more than one article and labeled them as “*keywords” (Choudhury et al., 2020). Keywords that did not receive the minimum attention from the research community were considered irrelevant and discarded. Likewise, articles that lack more than one “*keywords” are deemed unrelated and removed from consideration. As a result, the *duration of D_I was reduced to the year 2013. Figure 3 shows yearly statistics of articles across various fields, with the filtered dataset comprising highly relevant articles identified in this study. For IIoT, there has been a significant upward trend in article

Table 1 Basic statistics of the scholarly dataset

Dataset	*Articles	*Keywords	*Ratio (%)	*Duration
D_I	2587 (83.69% of original dataset)	1551	72.72	2013–2022
D_M	1320 (57.97% of original dataset)	1019	52.60	1999–2022

numbers since 2017, with an increase of 100–200 articles annually compared to the previous year, indicating a continual rise in academic and industrial interest in this domain. In contrast, research on Metaverse has been relatively limited, with fewer than 100 articles published annually before 2021. However, starting from 2022, the number of articles on Metaverse surged to 247.

At the end of this pre-processing, the number of those cleaned and transformed keywords were identified as unique and were used in further analyses. Basic dataset statistics are presented in Table 1. Note that *Ratio refers to the frequency accumulation ratio of keywords between the filtered and original dataset.

Methods

Constructing dynamic KCNs

In graph theory, a network is commonly represented by a graph, which can undergo various operations, such as union, intersection, join, etc. *Igraph* is a particularly useful R library providing a set of data types and graph operations for network analysis. In this study, **graph.full()** is used to create a full N-graph, $g_i=(v_i, e_i)$. The graph vertices (v_i) are keywords and edges (e_i) between vertices are the relations of co-occurrence for each article i . Then, with a new article j recorded in WoS (relative to article i), we use **graph.union()** to merge g_i and g_j into a new graph G . By default, **graph.union()** keeps the attributes of both graphs. It can lead to name clashes if an attribute is presented in multiple graphs. In this case, suffixes are used for clarity, for example, `weight_1` and `weight_2` represent the edge weight for g_i and g_j respectively. It should be noted that the edge weight between non-identical or non-overlapping vertices will be marked as null values in G . As such, we assign 0 to the null value of `weight_1` and `weight_2` to address the absent edge weight after the operation of graph union.

In this study, dynamic KCNs are a set of graphs constructed by graph operations from multiple snapshots. For a dataset D with n articles, dynamic KCNs are denoted by the series $G_D = \{G_{D(1)}, G_{D(2)}, \dots, G_{D(n)}\}$, which can be obtained by incrementally collapsing subgraphs (g_i). Each snapshot G_i can be obtained as follows:

$$G_i = \begin{cases} g_i, & i = 1 \\ g_i \cup G_{i-1}, & i > 1 \end{cases} \tag{1}$$

The construction of dynamic KCNs has been extensively studied in the literature, with perspectives from both sliding-window and aggregate-window approaches, addressing short-term as well as long-term dynamics (Balili et al., 2020). In line with these two approaches, we generate two types of KCNs: G'_{D_x} from the sliding-windows manner (described by $G'_{D_x} = \{G'_{D_x(1)}, G'_{D_x(2)}, G'_{D_x(3)}, \dots, G'_{D_x(S)}\}$, $x = I/M$ for IIoT/Metaverse), and G_{D_x} from the aggregate-windows manner (described by $G_{D_x} = \{G_{D_x(1)}, G_{D_x(2)}, G_{D_x(3)}, \dots, G_{D_x(S)}\}$). The differences between these two forms are

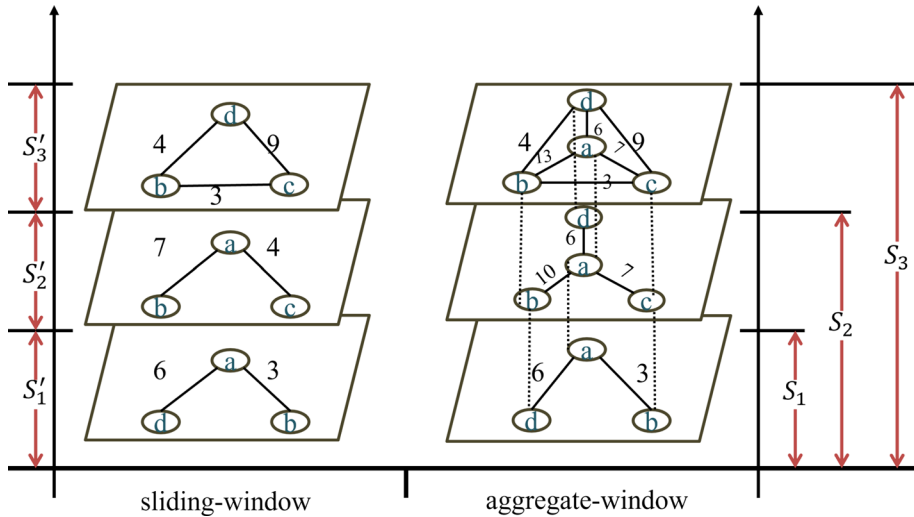


Fig. 4 Taking snapshots based on sliding-window or aggregate-window way

illustrated in Fig. 4, which was adapted from (Balili et al., 2020). Note that snapshots constructed in the sliding-window manner only accounts for the co-occurrences within subsets.

Combining motifs and node properties

The motif detection is designed to find frequently re-occurred patterns in real networks. According to Ribeiro (2011), the criteria of network motif depend on four parameters, $\{P, U, D, N\}$, where P is the probability threshold that indicates the frequency of a subgraph in the real network is less than that in N randomized networks (preserving the same node degree distribution as the real network), and U defines the minimum frequency a subgraph in the real network should have, and D is the minimum frequency deviation that ensures sufficient difference between the real network and the random networks. Therefore, the mathematical definition that a subgraph could be viewed as a network motif is formulated as follows:

$$Probability(\bar{f}_{rand}(g_{k,i}) > f_{real}(g_{k,i})) \leq P$$

$$f_{real}(g_{k,i}) \geq U$$

$$f_{real}(g_{k,i}) - \bar{f}_{rand}(g_{k,i}) > D \times \bar{f}_{rand}(g_{k,i})$$

where, $g_{k,i}$ is a subgraph (i) with k nodes, $\bar{f}_{rand}(g_{k,i})$ is the average frequency over all random networks of $g_{k,i}$, $f_{real}(g_{k,i})$ is the frequency of $g_{k,i}$ in the real network.

Based on the aforementioned conditions, a subgraph can be identified as a motif through the following three main procedures: (1) conducting the subgraph census; (2) generating randomized networks; (3) computing subgraphs’ significance. The goal of

the subgraph census is to enumerate all possible subgraphs of size k . This is a time-consuming process, and we adopt an efficient algorithm proposed by Wernicke (2005) for full subgraph census. Then, we generated 100 randomized networks as null models, preserving the same number of nodes and edges hold by the real network. Two metrics were involved in computing subgraphs' significance, Z-score ($Z(g_{k,i})$, Eq. (2)), and significance profile ($SP_{k,i}$, Eq. (3)). Z-score is a key metric to reflect the statistical significance of motifs and is related to the network size. To make comparisons among networks of different sizes, we use Eq. (3) to normalize and get a significance profile value that ranges from - 1 to 1. Besides, we use the concentration value, Eq. (4) to calculate the relative concentration of detected subgraphs with the same size k ranging from 0 to 1.

$$Z(g_{k,i}) = \frac{(f_{real}(g_{k,i}) - \bar{f}_{rand}(g_{k,i}))}{\sigma_{g_{k,i}}^{rand}} \tag{2}$$

$$SP_{k,i} = \frac{z_{k,i}}{\sqrt{\sum_{i=1}^n z_{k,i}^2}} \tag{3}$$

$$C(g_{k,i}) = \frac{f_{real}(g_{k,i})}{\sum_{j=1}^n f_{real}(g_{k,j})} \tag{4}$$

As mentioned before, motifs alone only reveal significant subgraph patterns in knowledge networks. Not limited to the pattern significance, we further investigate the implicit connected patterns of knowledge elements, with the extraction of two time-related node properties, the age and the impact of a node. To capture the rules that underlie the organization of decomposable structures, the task of node feature engineering can be considered as a classification problem. In other words, we label nodes with its properties for all possible subgraphs. With the growing development of a field, some keywords are repeatedly used, and some might be used because of the emerge of new concepts. Thus, the age of a keyword intrinsically relates to its position in knowledge networks. For each *keywords, the age is computed by the difference between the first appearance timestamp (denoted by τ) and the observation timestamp (denoted by δ). If $\delta - \tau \leq \theta$, then *keywords is classified as 'New knowledge element' (N); otherwise, it belongs to 'Existing knowledge element' (E). The threshold $\theta = 3$ is a measurement time window. Another critical dimension is the impact of a keyword. The relative importance of keywords in a snapshot G_i is estimated by computing their PageRank values (Cheng et al., 2020). In this paper, keywords are classified into three types: (1) 'High-level impact' (H) if the PageRank value falls within the upper quartile; (2) 'Low-level impact' (L) if the PageRank value falls within the lower quartile; and (3) 'Medium-level impact' (M) for the remaining keywords.

As such, we conduct subgraph censuses, and then record all occurrences for a given structure of size k . It should be noted that counting k -motifs is an expensive operation, as their number grows exponentially with k increases. In this study, a total of 1551 and 1019 *keywords (i.e., vertices) are collected from D_I and D_M , respectively. We focus only on 4-node subgraphs, as this process is of highly complexity in both time and space and higher-order subgraphs are also composed of lower-order subgraphs (Zou et al., 2023). In Fig. 5, we enumerate all types of four-node motifs.

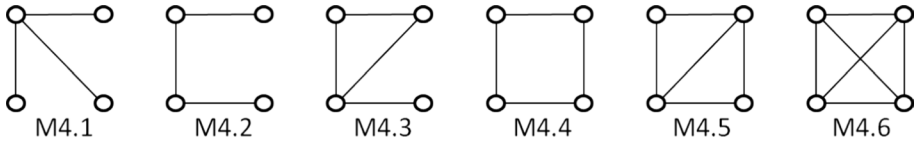
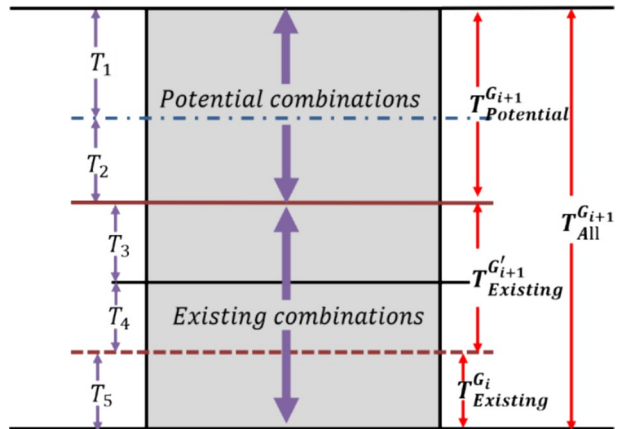


Fig. 5 All four-node motifs

Fig. 6 The composition of M4.6 in snapshot G_{i+1}



Measuring the evolution of knowledge structure

The task of measuring the evolution of knowledge structure can be simplified and tracked by modeling the system as dynamic KCNs. Viewing the network growth at the finest-grained article-level, the inclusion of a new article can lead to fluctuations of the knowledge structure. To magnify these fluctuations, the overall evolution process can be divided into snapshots (G_i) with each containing a fixed number of articles. The subgraph M4.6 (see Fig. 5) has the highest connectivity among all 4-node subgraphs and is able to be deconstructed into any of the other 5 subgraphs. Based on such properties, we define two types of change events to capture the dynamics between two consecutive snapshots, intra-change and inter-change events. The intra-change reveals the knowledge linkages between nodes across or within literature, while inter-change indicates a shift in motifs from one type to another. Among those changes, again, we give special focus to the knowledge combinations that lead to the formation of a fully connected structure, i.e., M4.6 in Fig. 6. It could be further categorized into two types, the existing combinations and the potential combinations. The existing combinations arise from the incremental intra-change, indicating all nodes co-appeared in one successfully published literature. The potential combinations are the result of inter-change shifts, which suggest that linkages have been constructed among all nodes but have not yet been confirmed. These unconfirmed combinations, however, reveal beliefs about which elements of knowledge are most likely to work well together and should be preferentially considered. Followed by this, we propose a new approach to examine the relative change of knowledge structure, innovatively revealing changes among the combinations of knowledge elements. This measurement takes into account the intra-change and inter-change events in dynamic KCNs. Figure 6 illustrates the composition of M4.6 in a snapshot G_{i+1} , and we coded different types of combinations

Table 2 Descriptive statistics of dynamic KCNs properties ($D_I = \text{IIoT}$ and $D_I = \text{Metaverse}$)

Network properties	G'_{D_I}			G_{D_I}			G'_{D_M}			G_{D_M}		
	Mini map	Min	Max	Mini map	Min	Max	Mini map	Min	Max	Mini map	Min	Max
Articles		199	199		199	2587		100	100		100	1300
Nodes (N)		345	499		345	1550		181	252		181	1003
Edges (E)		1217	3893		1217	23317		385	825		390	5469
Avg. Clustering Coefficient		0.1420	0.224		0.105	0.143		0.1690	0.404		0.0970	0.269
Avg. Path Length		2.4512	2.808		2.4782	2.808		2.8273	3.821		3.0403	3.571
Network Density		0.1890	0.033		0.0140	0.021		0.0180	0.031		0.0100	0.024
α of the formula: $p(x) = Cx^\alpha$		2.5522	2.786		2.1392	2.736		2.7773	3.419		2.6022	2.963
Modularity		0.3700	0.520		0.2580	0.501		0.5340	0.748		0.4180	0.681

as $X = \{\text{All, Existing, Potential}\}$, and the form of $T_X^{G_i}$ is the frequency of a specific type of combinations X in a prior snapshot G_i . T_2 or T_4 is the intersection of existing combinations between T^{G_i} and $T^{G'_{i+1}}$. $T_{Existing}^{G'_{i+1}}$ is the frequency of M4.6 in snapshot G'_{i+1} , which is added to G_{i+1} as an intra-change.

In our approach, the relative change (denoted by $\xi_X^{G_i \rightarrow G_{i+1}}$) of all combinations, existing combinations, or potential combinations across consecutive snapshots is calculated **based on deconstructing M4.6**. We defined this change as the knowledge structure evolution strength and, $\xi_X^{G_i \rightarrow G_{i+1}}$ can be computed by Eq. (5).

$$\xi_X^{G_i \rightarrow G_{i+1}} = \frac{T_X^{G_{i+1}} - T_X^{G_i}}{T_X^{G_i}} \tag{5}$$

Results

Network properties of the dynamic KCNs

Adhering to the workflow introduced in Sect. “Data and pre-processing”, two sets of dynamic KCNs were built for D_I and D_M , (D_x hereafter, for simplicity). With regard to the yearly statistics of articles in IIoT and Metaverse, D_x is divided into $S = 13$ equally spaced intervals.² Each subset of D_I comprises 199 papers, while each subset of D_M consists of

² We assumed that the states of the current time window are all evolved based on the state development of the previous time window. As a result of that, the subset associated with multiples of the initial observation values can all be covered.

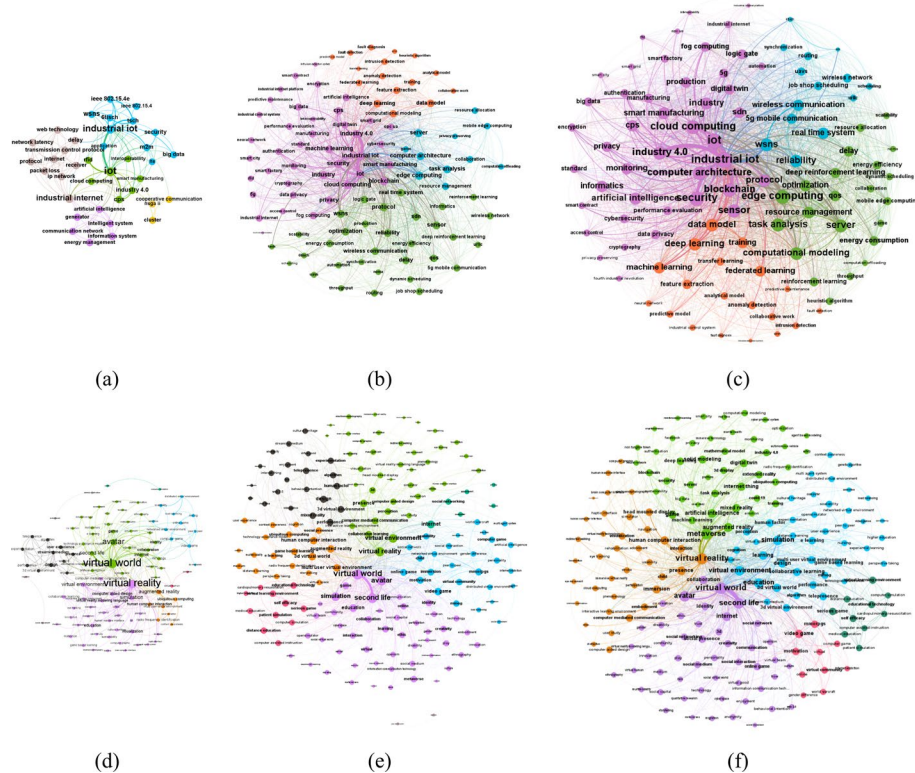


Fig. 7 Network snapshots have been arbitrarily picked from the subset of G_{D_I} and G_{D_M} . All network snapshots are divided into different colors according to the community detection results. The size of the node represents keyword's degree. The edge thickness represents corresponding edge weight. From these network snapshots, many small, originally isolated groups evolve into larger ones and become major communities

100 papers.³ This enabled us to explore the evolution of knowledge structure. Table 2 presents the descriptive statistics of all snapshots in G'_{D_x} and G_{D_x} , respectively. In the short-term, based on the derived metrics, there are noticeable differences in the structural properties between snapshots of G'_{D_x} , despite the input subsets is divided into equal intervals. This suggests that the volume of scientific literature is not equivalent to the scope of scientific ideas (Fortunato et al., 2018). In the long-term, with the expansion of scientific literature, the snapshots of G_{D_x} become increasingly nonlinear and complex. In particular, for both scenarios (G_{D_I} and G_{D_M}), all α fall into the range of (2, 3), indicating the network growth follows a **preferential attachment process** (Barabási & Albert, 1999). Also, in G_{D_x} both the *Network density* and the *Average Clustering Coefficient* experienced a decrease and then followed by an increase. In contrast, the *Average Path Length* and the *Modularity* undergone a decrease trend during the whole process. Generally, in practice, the value of modularity greater than approximately 0.3 appears to indicate a high quality of clustering

³ For comparison with D_I , we excluded 20 articles that were last published. Thus, the *keywords of D_M was reduced to 1003.

(Newman, 2004). It is evident that the boundaries of communities are becoming blurry with more knowledge elements and their interrelationships added to the network.

As depicted in Fig. 7, network snapshots have been arbitrarily picked from the subset of G_{D_i} (a–c) and G_{D_M} (d–f). On the one hand, with the continuous growth of literature, the development of IIoT has entered the deep cultivation stage of practice from the concept popularization, and keywords had gradually evolved from concept definition and bottom technology development to industry application, service quality management, energy consumption management, etc. The keywords related to IIoT research can be categorized into five types: manufacturing industry development (e.g., *smart manufacturing*, *industry 4.0*), core technology (*data privacy*, *blockchain*, *edge computing*, etc.), platform infrastructure (e.g., *WSNs*, *cyber security*), application scenarios (e.g., *smart factory*, *resource management*, *energy consumption*), and Function (e.g., *authentication*, *feature extraction*, *optimization*, *task analysis*). For snapshots in the Metaverse domain (d–f), although the research papers published on the metaverse is limited, significant advancements in technologies such as *virtual reality*, *augmented reality*, *digital twins*, and *blockchain* have led to substantial transformations. Moving from its initial stage, primarily focused on gaming and social interactions, known as metaverse 1.0, to the current era of metaverse 2.0, the technology has evolved significantly (Gao et al., 2023).

Selecting appropriate keywords is not always straightforward for authors, yet this process is crucial for the searchability and impact of an article (Choudhury et al., 2020). Therefore, certain representative keywords that are more likely to be preferentially chosen owing to their high popularity and relevance in a specific field. For example, in the networks of IIoT, it is noticeable that the representative keywords acquire more connections with more papers published and emerge as hotpots of the respective research domain, such as *CPS*, *industry 4.0*, *blockchain*, *WSNs*, *QoS*, *cloud computing*, *virtual reality*. Furthermore, while new keywords initially co-occur with less-known ones, they eventually secure new connections with representative topics, corresponding to the process of **preferential attachment**. For instance, *digital twins* was found to be linked with *keywords *CPS* and *smart manufacturing* in Fig. 7a. In Fig. 7b–c, it was further connected with keywords like *smart manufacturing*, *cloud computing*, and achieved endorsement from the representative keywords (e.g., *industrial IoT*, *industry 4.0*) in later snapshots. This phenomenon, indeed, represents a crucial feature in many relevant circumstances. In addition, a similar situation occurs in the knowledge network of Metaverse.

In summary, the above descriptive statistics and visualization characterize the connection of knowledge elements and the evolution of knowledge structure in a certain field. However, knowledge structure evolution is embodied in node evolution and structural evolution. Nodes with different properties are combined into specific patterns simultaneously or successively, and undergo pattern shifts from one type to another. The subsequent stages of this section are driven by the expectation that comprehending the decomposable nature of knowledge structures would facilitate a better grasp of their organization and enable more accurate measurement of their evolution.

Identified knowledge linkage patterns

In total, we found 6 different types of tetradic motifs among all snapshots in G_{D_x} (except for M4.4 in $G_{D_i(1)}$, $G_{D_M(1)}$ and $G_{D_M(2)}$ with $SP < 0$).

For each snapshot in G_{D_x} , its frequency and concentration value of a subgraph type are represented in Fig. 8. In particular, red lines represent the counts of identified significant

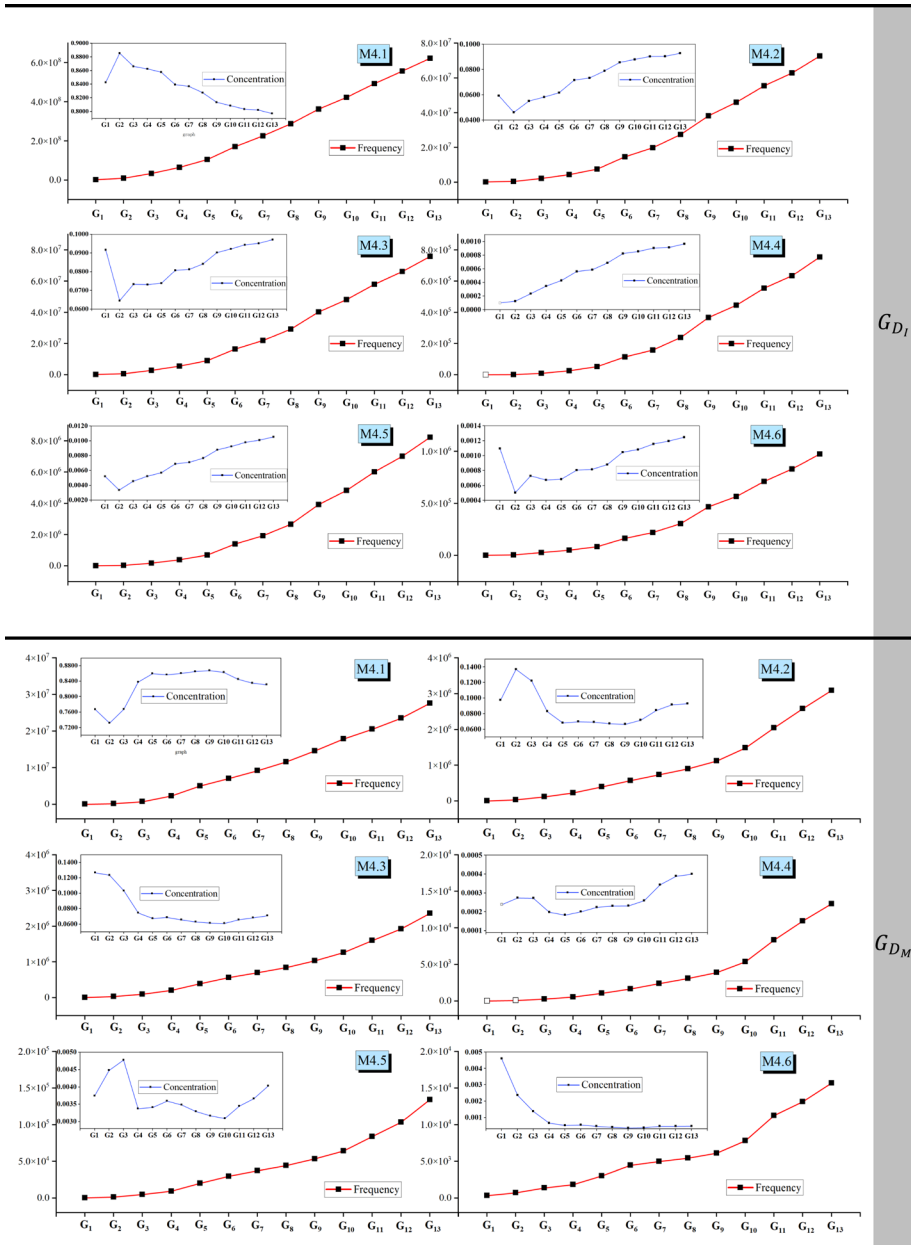


Fig. 8 Incremental changes in the frequency and concentration of G_{D_i} and G_{D_M}

and re-occurring subgraph patterns, and blue lines in the inset plots provide detailed information on the concentrations of these detected structures.

It can be seen that the red lines in the main graphs exhibit a nonlinear increase as new subgraphs ($G'_{D_x(i+1)}$) are sequentially added to $G_{D_x(i)}$. This trend is found for every type of subgraph, suggesting that the complexity of knowledge structure gradually

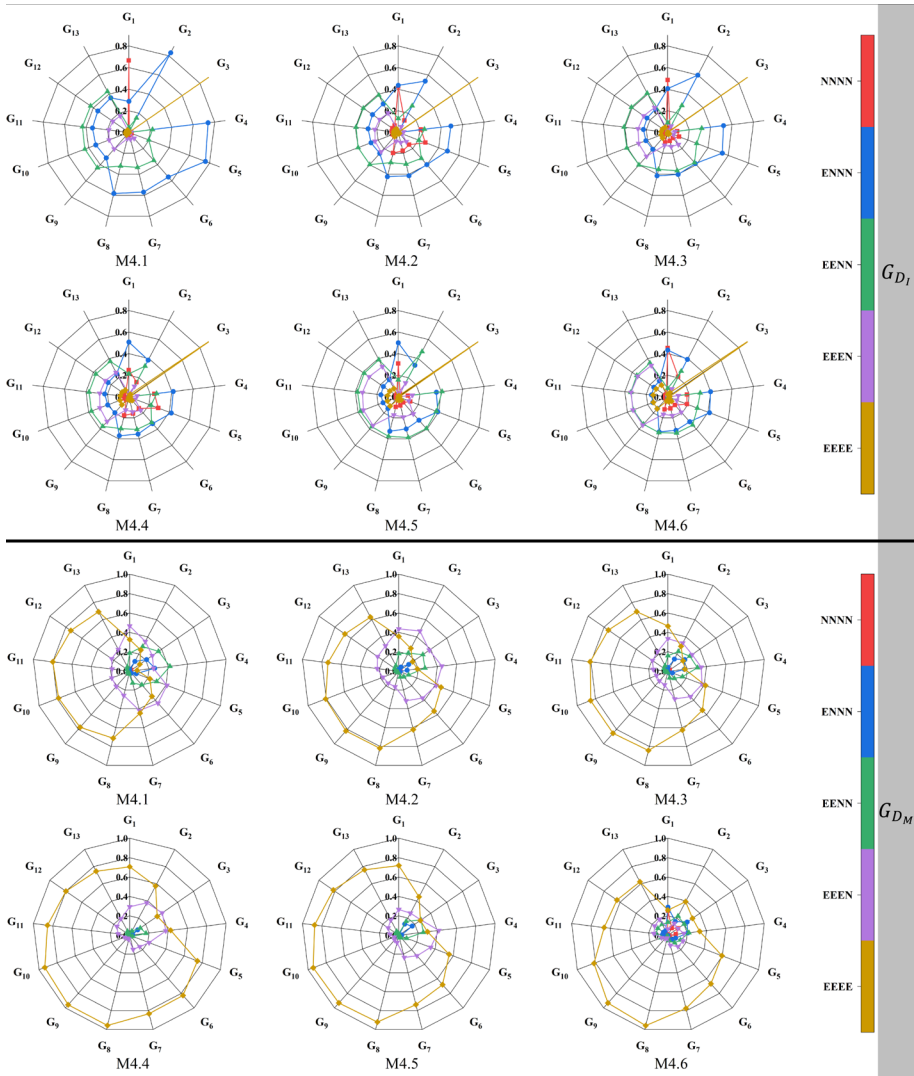


Fig. 9 The dynamic linkage results of keywords from different ages in G_{D_I} and G_{D_M}

increases over time. This finding is consistent with the growing need for knowledge linkage to facilitate scientific creation in the two fields. We emphasize that these decomposable structures serve as the “building blocks” for the formation and evolution of knowledge structure. In the context of G_{D_I} , the result shows that while the frequency of M4.1 is significantly higher than that of other subgraphs, its concentration value gradually decreases. Compared to other subgraphs, M4.1 has the lowest connectivity as it is characterized by a star-shaped structure. This implies that a shift from subgraphs with a low connectivity to those with a high connectivity, resulting in an increasing level of connectivity within the knowledge structure. For G_{D_M} , there is an interval where the blue line remains stable, ranging from $G_{D_M(5)}$ to $G_{D_M(10)}$. This indicates that the

increasing level of connectivity within the knowledge structure may primarily stem from the intra-change events.

To investigate the linkage patterns behind these “building blocks”, we further look at the node properties. Figure 9 displays the results of dynamic linkages of keywords from different ages. Since this study does not take into account the differences in node positions in the same type of subgraphs, we identified 5 different types of combinations ($Q = \{\text{NNNN}, \text{ENNN}, \text{EENN}, \text{EEEN}, \text{EEEE}\}$). The color of each line represents the concentration value (denoted by $C_{(g_{4,i})}^Q$) of a given “building block” of a specific subgraph type.

Considering G_{D_i} , as the network size increased (from $G_{D_i(1)}$ to $G_{D_i(13)}$), we found that knowledge elements undergo a rapid transition from new knowledge elements to existing knowledge elements. An interesting fact is that, as knowledge structures gradually mature, the concentration values of mono-type combinations (i.e., **NNNN** and **EEEE**) are significantly lower than other types of combinations (i.e., **ENNN**, **EENN**, and **EEEN**). In other words, a mixture of new and existing knowledge elements is the main path toward creating new knowledge (Kuo et al., 2019). The principles behind this path are, existing knowledge elements are more likely to increase the likelihood of the article being discovered and cited by other researchers, and new knowledge elements are used to highlight topics, techniques or methods that are different from previous research. Furthermore, as the concentration values of subgraphs containing new knowledge units and those without new knowledge units exhibit an inverse relationship. This implies that the increase of new elements is lower than the growth of existing elements.

It is important to recognize that knowledge evolution can differ greatly across various fields. As mentioned earlier, the research on the Metaverse has been relatively limited, resulting in connectivity changes within the knowledge structure from $G_{D_M(5)}$ to $G_{D_M(9)}$ (2012–2021) that primarily rely on the combination of existing knowledge elements. However, this changed in 2022 when there was a significant increase in investments and the number of papers in the field of Metaverse (Gao et al., 2023). During this period, corresponding to $G_{D_M(10)}$ to $G_{D_M(13)}$, a substantial influx of new knowledge elements entered, reflecting technological advancements or conceptual developments. Existing knowledge elements established linkages with these new elements, resulting in the increasing of hybrid combinations, while the concentration of singular types of combinations began to decline.

On the other hand, considering the impact scope of keywords, we identified 81 distinct combinations in G_{D_i} and 79 in G_{D_M} . In the same vein, these combinations are further typed into three categories: mono-type (**H-H**, **M-M**, and **L-L**), pairwise-type (**H-M**, **H-L**, and **M-L**), and mixed-type (**H-M-L**). In other words, a total of 7 different types of combinations ($R = \{\text{H-H}, \text{M-M}, \text{L-L}, \text{H-M}, \text{H-L}, \text{M-L}, \text{H-M-L}\}$) can be identified in G_{D_i} . Figure 10 provides the dynamic linkage results of keywords from different impact-scope. Being a crucial structure of knowledge structure, the combinations with higher concentration values (denoted by $C_{(g_{4,i})}^R$) in M4.1 dictates the form of knowledge element combination patterns. As can be seen in Fig. 10 (G_{D_i} , M4.1), the combinations of pairwise-type ($C_{(g_{4,1})}^{H-M} \in [0.471, 0.591]$ and $C_{(g_{4,1})}^{H-L} \in [0.081, 0.121]$) and mixed-type ($C_{(g_{4,1})}^{H-M-L} \in [0.305, 0.368]$) take the dominant positions among all patterns. This implies that highly important nodes are directly paired nodes of different types.

Subgraphs with low concentration values can still be relatively dominant subgraph types (Zou et al., 2023). As the connectivity degree grows (from M4.1 to M4.6), there are clear increases in the concentration values of **H-H**. This type also serves as the dominant mode

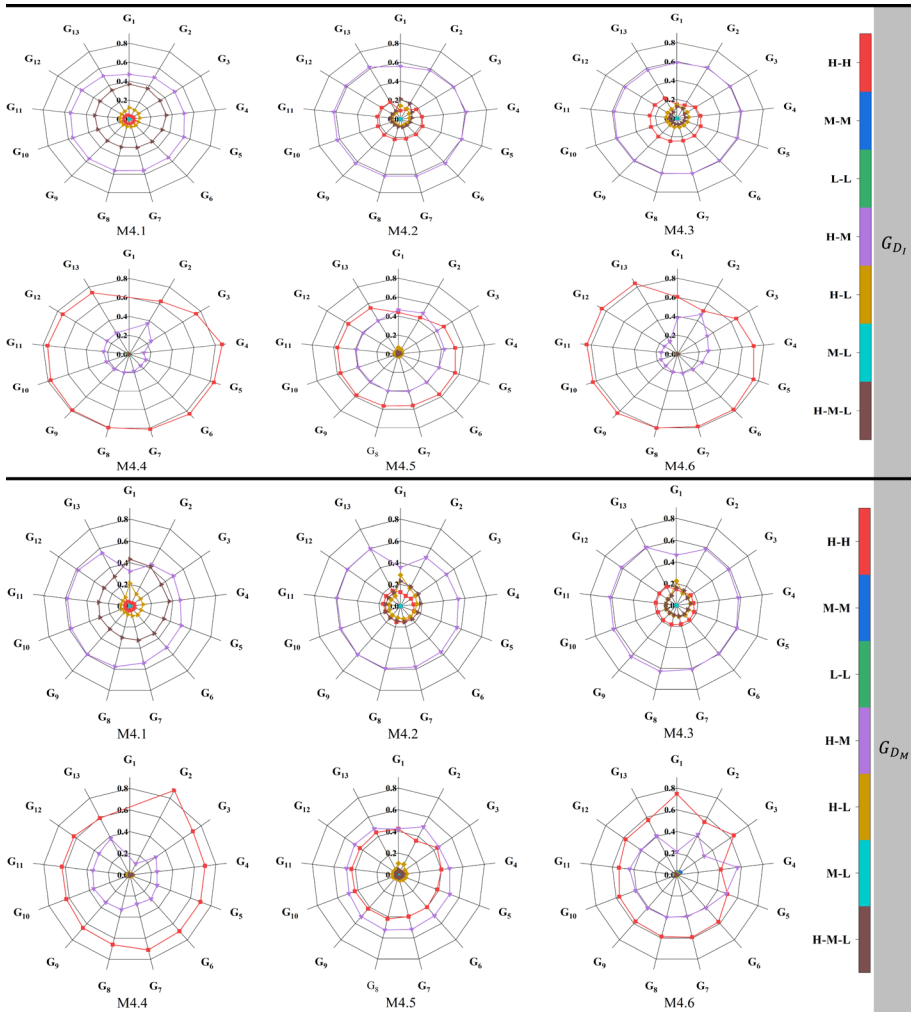


Fig. 10 The dynamic combination results of keywords from different impacts in G_{D_I} and G_{D_M}

of M4.4, M4.5, and M4.6. This means that direct links between high-impact nodes are favored owing to their high frequency of co-occurrence and thematic relevance, wherein such direct connections enhance information propagation efficiency and the structure stability of the knowledge. Likewise, in G_{D_M} , the combinations of pairwise-type and mixed-type similarly dominate among all patterns. Furthermore, with the increase in connectivity degree (from M4.1 to M4.6), there is also a noticeable increase in the concentration values of **H–H**. This suggests that the knowledge linking patterns found in G_{D_I} are not exceptional.

In previous studies, it has been shown that keywords with higher impact are general vocabularies with broad semantics (Cheng et al., 2020), which underlie the core structure of domain knowledge. This suggests that scientific creations are primarily grounded in the core topics of prior works. It should be noted that new keywords secure connections from representative keywords directly, instead of co-occurring with less-known keywords at

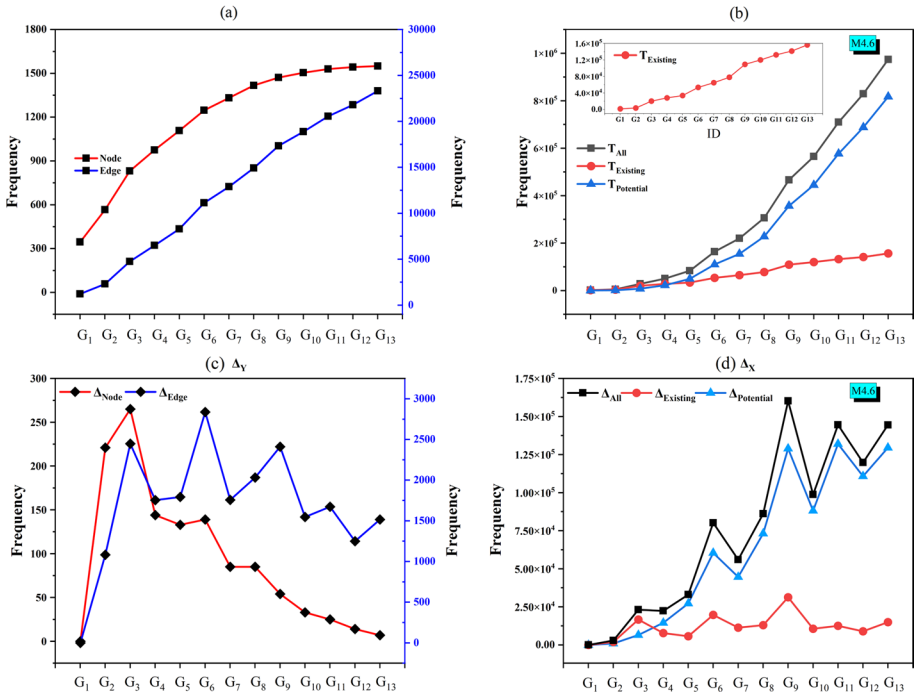


Fig. 11 The frequency distribution and relative changes of nodes, edges, and different types of combinations in G_{D_t}

their first timestamps. This finding challenges the conventional understandings described in Sect. “Network properties of the dynamic KCNs” and contributes to a more nuanced comprehension of domain knowledge structures.

The incremental changes of knowledge structure

Since the proposed approach to gauge knowledge structure evolution strength takes into consideration both knowledge content and structural information, it is beneficial to compare our approach with that only considers knowledge content or structural information.

Figure 11a–b displays the frequency distribution of nodes, edges, and different combination types across consecutive snapshots for G_{D_t} . The inset plots detailed frequency distribution of existing combinations. It can be seen that all lines in the main graphs increased over time, which shows a consistency with the continuous growth of scientific creations in the IIoT field.

Also, we calculated the incremental changes of different types of combinations, as shown in Fig. 11c–d. The addition of new nodes (Δ_{Node} , Eq. (6)) is gradually decreasing, while the addition of edges (Δ_{Edge} , Eq. (7)) continues to grow at a relative high level. In relation to the incremental changes of nodes or edges, the fluctuations of Δ_{Edge} are consistent with the fluctuations of Δ_X , suggesting the preservation, extension, and combination of existing simpler subgraphs. From a quantitative perspective, the substantial increase in fully connected subgraphs significantly outpaces the modest growth in edges and nodes, signifying a profound transition in the network’s topological patterns. In other words, previously disconnected

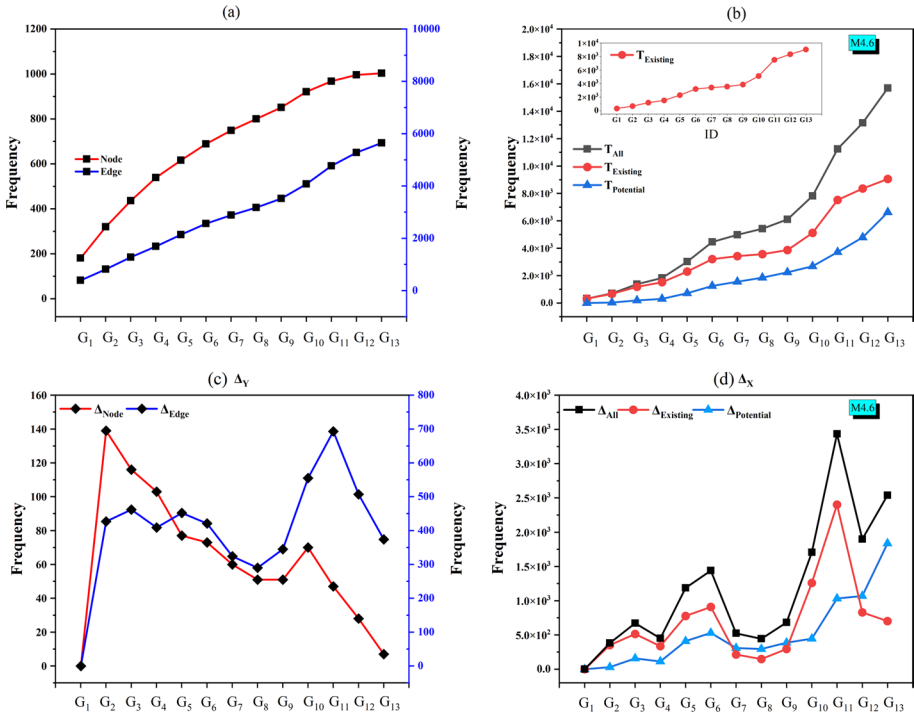


Fig. 12 The frequency distribution and relative changes of nodes, edges, and different types of combinations in G_{D_M}

ideas and resources are successfully combined, and this can be confirmed from the addition of edges. Another remarkable feature is that the gap between the size of $\Delta_{Existing}$ (see Eq. (8)) and $\Delta_{Potential}$ (see Eq. (9)) became greater over time, accompanied by a decrease in Δ_{Node} . This finding indicates that the knowledge structure of IIoT is gradually stabilizing, and scientific knowledge creation is primarily driven by the recombination of existing knowledge.

$$\Delta_{Node} = N_{G_{D_x(i+1)}} - N_{G_{D_x(i)}} \quad (6)$$

$$\Delta_{Edge} = E_{G_{D_x(i+1)}} - E_{G_{D_x(i)}} \quad (7)$$

$$\Delta_{Existing} = T_{Existing}^{G_{D_x(i+1)}} - T_{Existing}^{G_{D_x(i)}} \quad (8)$$

$$\Delta_{Potential} = T_{Potential}^{G_{D_x(i+1)}} - T_{Potential}^{G_{D_x(i)}} \quad (9)$$

For comparison, we conducted similar work on G_{D_M} . The upward trend of all lines in Fig. 12 (a-b) indicates a consistent pattern with the ongoing expansion of scientific developments within the Metaverse domain. One noticeable point is that the value of $T_{Existing}$ consistently surpasses $T_{Potential}$, completely contrasting with the situation observed in G_{D_I} . From $G_{D_M(10)}$ to $G_{D_M(13)}$, the burgeoning interest in the Metaverse led to a rapid reversal of $\Delta_{Potential}$ over $\Delta_{Existing}$. Meanwhile, the Δ_{Node} underwent a significant decline during this

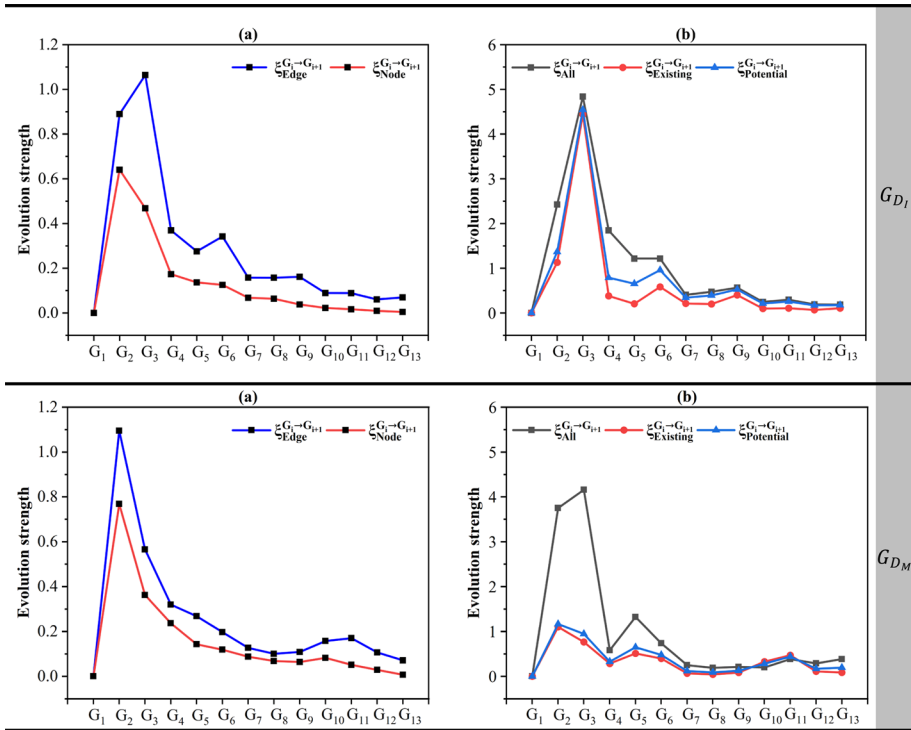


Fig. 13 The degree of knowledge structure evolution based on different strategies in G_{D_I} and in G_{D_M}

stage. This phenomenon suggests that a reshaping of the metaverse is being remodeled, with the abundant introduction of new knowledge elements generating more opportunities for scientific knowledge creation.

To investigate more in-depth how the knowledge structure evolves, we calculated the knowledge structure evolution strength based on different strategies. It should be considered that different types of evaluation strategies yield different results. Figure 13 shows the results for two types of strategies in relation to the evolution strength of knowledge structure. With the increase of nodes and edges, all lines are consistently above zero and reach a peak at a snapshot $G_{D_x(2)}$ or $G_{D_x(3)}$ and then shrink to a low level. This similar evolutionary trend suggests that the proposed approaches have been proven to be effective in quantifying the intensity of knowledge network evolution. Based on relative alterations in nodes or edges, dramatic shifts in the knowledge structure were found around the snapshots $G_{D_x(2)}$ or $G_{D_x(3)}$, for both the IIoT and Metaverse. However, this inference could be challenged when perspectives shift to subgraph evolution. More specifically, at snapshot G_3 , even though the value of knowledge structure evolution strength reached a peak, which predominantly propelled by existing knowledge components with high impact (see Figs. 9 and 10). At this phase, knowledge structure is undergoing a process of internalization, wherein established and new knowledge elements lack ample interactions. Yet, as the research field progressively unfolds, an escalating amount of new knowledge begins to intertwine with the existing knowledge, enriching the semantic of the domain knowledge and leading to a more profound evolution of the domain knowledge structure.

Discussion

Knowledge structure is the result of the continuous evolution of the forces that formed it, during which the node properties and structural patterns inherently affect the system's function. The development of knowledge structure is a complex process influenced by various factors, such as global network properties (Castillo-Vergara et al., 2018; La & Chai, 2021), meso-level community structures (Cho, 2020), and so on.

This study goes beyond macro global properties and pure structural motifs, and the goal is to uncover linkage patterns and incremental evolution of knowledge structure, thereby facilitating the process of knowledge recombination and generating new knowledge. We proposed a novel framework that integrates the incremental update mechanism of knowledge network construction, subgraph enumeration, and knowledge combination. On the one hand, node attributes are applied to the analysis of linkage patterns and incremental evolution of knowledge combinations. On the other hand, with the aid of incremental network methods, specific types of motifs are deconstructed to reveal potential knowledge combinations.

Key findings

The key findings arise from the application of the proposed framework and how linkage patterns and knowledge structure evolve in specific domains, in our case studies, the IIoT and the Metaverse fields.

Firstly, the proposed framework proves to be effective in capturing the dynamics of domain knowledge structure. With the integrated elements, it identifies the existence of intra-change and inter-change events that further associated with knowledge combinations.

Secondly, from a quantitative perspective, the star-like knowledge structure (i.e., M4.1) exhibits a significantly contrasting evolutionary trend compared to the behavior observed in other structures, and it partly explains how knowledge creation utilizes the recombination of existing knowledge elements (nodes not linked in M4.1) (Fortunato et al., 2018).

Thirdly, after incorporating node attributes into the combination structure, the distribution of knowledge linkage patterns of different types of decomposable structures, largely differ from each other. In a sense, the mixed strategy (e.g., high impact knowledge elements are more likely to be linked with elements of middle/low level of impact) is the main path toward knowledge linkages. The rationale might be that node attributes (here the node age and node impact) play a role in the evolution of knowledge structure (Ribeiro & Silva, 2014).

Theoretical and practical implications

This study contributes to innovation research and practice in several ways. Theoretically, taking the structure of knowledge networks as proxies for knowledge structure (Cheng et al., 2020; Cho, 2020), the novel application of motifs in the paper can identify decomposable structures of knowledge in a certain field, which is conducive to revealing the principles behind knowledge linkage and quantifying the evolution of knowledge structure. It is difficult to understand the evolution of knowledge structure when only involving global network properties (Castillo-Vergara et al., 2018; La & Chai, 2021) or purely structural motifs (Feng et al., 2021; Wang et al., 2022). Our research offers a much more comprehensive understanding of the formation and evolution of knowledge structure. The introduction

of node attributes enables researchers to investigate the rules that underlie the organization of decomposable structures. Also, the view of identifying existing or potential knowledge combinations from specific decomposable structures enriches the scenarios of employing network motif theory and methods.

Practically, beyond the reported results, the methodologies and findings of our study can serve as useful resources for stakeholders. Firstly, the proposed integrative framework could be applied into other domains, enabling researchers or policymakers in the short-term or long-term understanding on any topics, and the overall landscape of the field being studied (Balili et al., 2020). Secondly, the dynamics of knowledge structure found in this study, pave the way for firms to search for new potential applications around their prior existing knowledge combinations, and also to exploit toward more value from their established knowledge base (Kuo et al., 2019; Tolstoy, 2010).

Limitations and future work

Nevertheless, this work has several limitations, which suggest possibilities for future research. Firstly, it is crucial to select and extract “highly relevant” keywords when constructing knowledge networks to represent the domain-specific knowledge structure. To cope with the lack of author-selected keywords in some articles, we plan to formulate some extraction rules for collecting keywords from title, abstract, and full-text information (Ba & Liang, 2021). Secondly, the parameter used in defining the age property should be further validated, although using the default setting ($\theta = 3$) generates satisfactory results in the two fields. As a related point, other types of properties, such as functional, disciplinary, and semantic information, should be taken into account to further investigate the rules underlie the organization of higher-order structures. Thirdly, there are multi-level and multi-type relationships between keywords (e.g., co-occurrence, citations, co-citation, and semantic), as such, the co-occurrence of keywords cannot fully capture the topic or content correlation of a research field. In our future work, it is necessary to introduce multi-level or multi-type of relationships in building a simplex or multiplex knowledge network and analyzing the domain-specific knowledge structure (Boccaletti et al., 2014).

Conclusion

This study proposed an integrative framework to study the linkage patterns and evolutionary domain knowledge structure. The findings indicate the feasibility of utilizing motif analysis to uncover linkage patterns in knowledge structures, and the proposed methodology demonstrates effectiveness in quantifying the rate of knowledge structure evolution within specific domains. Guided by the framework, we investigated the evolution of domain knowledge structure based on incrementally updated dynamic KCNs. Node properties and motifs were combined to quantify knowledge combinations, upon which the evolution strength of specific structures was studied. Experiments from two fields demonstrate the framework’s capability in identifying favored knowledge combinations and elucidating the evolutionary trajectories of these knowledge structures. The insights generated constitute significant contributions to the research community and policymakers, bearing implications for advancing theoretical comprehension and facilitating practical applications.

Appendix

The whole procedure to construct dynamic KCNs is shown in pseudo codes described in Algorithm 1. The workflow of motifs detection and subgraph enumeration is shown in the Algorithm 2.

Algorithm 1: How to construct dynamic KCNs

```

Data: The final scholarly dataset,  $D_1$ 
Result: Dynamic KCNs,  $G_i$ 
1 begin
2   match DE within *keywords,  $D_2$ ;
3    $HK_1 = \text{GetDEfield}(D_2)$ ;
4    $N = \text{length}(HK_1)$ ;
5   for  $hk$  to  $HK_1$  do
6      $g = \text{makeFullGraph}(hk)$ ;
7      $\text{EdgeWeight}(g) = 1$ ;
8      $\text{VerticeName}(g) = hk$ ;
9      $P_{hk} = \text{SaveGraphToList}(g)$ ;
10  end
11   $\text{TempG} = g_1$ ;
12  for  $i \rightarrow 2$  to  $N$  do
13     $\text{TempG} = \text{unionGraphs}(\text{TempG}, g)$ ;
14    if  $\text{EdgeWeight}_1(\text{TempG}) = \text{null}$  then
15       $\text{EdgeWeight}_1(\text{TempG}) \leftarrow 0$ 
16    end
17    if  $\text{EdgeWeight}_2(\text{TempG}) = \text{null}$  then
18       $\text{EdgeWeight}_2(\text{TempG}) \leftarrow 0$ 
19    end
20     $\text{EdgeWeight}(\text{TempG}) = \text{EdgeWeight}_1(\text{TempG}) + \text{EdgeWeight}_2(\text{TempG})$ ;
21     $G_i = \text{SaveGraphToList}(\text{TempG})$ ;
22  end
23 end

```

Algorithm 2: The workflow of motifs detection and subgraph enumeration

```

Input: The real network,  $G_r$ 
Input: The number of random networks to generate,  $N$ 
Output: Frequencies and node combinations of all k-node subgraphs from the real
          network, the corresponding average frequencies standard deviations from
          random networks
1 begin
2    $\text{SubgraphCensus}(G_r)$ ;
3    $\text{SubgraphPrint}(G_r)$ ;
4   Generate a set of  $N$  random networks,  $\{R_1, R_2, \dots, R_N\}$ ;
5   for  $i$  in  $N$  do
6      $\text{SubgraphCensus}(R_i)$ 
7   end
8   Compute the significance for each detected k-node subgraph;
9   Verify the accuracy of the  $\text{subgraphCensus}(G_r)$  and  $\text{SubgraphPrint}(G_r)$  ;
10 end

```

Acknowledgements The authors are grateful to anonymous referees and editors for their invaluable and insightful comments and thank for the support by the National Social Science Fund of China (16ZDA224); the Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX22_0552). Qi Wang would like to thank Dr. Shi Xiong and LiuHong Wang for their invaluable discussions and suggestions.

Author contributions Qi Wang: Conceived and designed the analysis; Collected the data; Contributed data or analysis tools; Performed the analysis; Wrote the paper. Bentao Zou: Conceived and designed the analysis; Wrote the paper. Jialin Jin: Conceived and designed the analysis; Contributed data or analysis tools. Yuefen Wang: Conceived and designed the analysis.

Data availability The data used in this study is restricted by Web of Science Clarivate, and we do not have the right to distribute it. However, the search strategies are provided and could be used to recollect the data used. Additionally, the code will be made available upon request.

Declarations

Competing interests The authors have no relevant financial or non-financial interests to disclose. The authors have no conflicts of interest to declare that are relevant to the content of this article.

References

- Arenas, A., Fernández, A., Fortunato, S., & Gómez, S. (2008). Motif-based communities in complex networks. *Journal of Physics A: Mathematical and Theoretical*, *41*(22), 224001.
- Ba, Z., & Liang, Z. (2021). A novel approach to measuring science-technology linkage: From the perspective of knowledge network coupling. *Journal of Informetrics*, *15*(3), 101167.
- Balili, C., Lee, U., Segev, A., Kim, J., & Ko, M. (2020). TermBall: Tracking and predicting evolution types of research topics by using knowledge structures in scholarly big data. *IEEE Access*, *8*, 108514–108529.
- Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*(5439), 509–512.
- Benson, A. R., Gleich, D. F., & Leskovec, J. (2016). Higher-order organization of complex networks. *Science*, *353*(6295), 163–166.
- Boccaletti, S., Bianconi, G., Criado, R., del Genio, C. I., Gómez-Gardeñes, J., Romance, M., Sendiña-Nadal, I., Wang, Z., & Zanin, M. (2014). The structure and dynamics of multilayer networks. *Physics Reports*, *544*, 1–122.
- Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., & Hwang, D.-U. (2006). Complex networks: Structure and dynamics. *Physics Reports*, *424*(4), 175–308.
- Callon, M., Courtial, J. P., & Laville, F. (1991). Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemistry. *Scientometrics*, *22*(1), 155–205.
- Castillo-Vergara, M., Alvarez-Marin, A., & Placencio-Hidalgo, D. (2018). A bibliometric analysis of creativity in the field of business economics. *Journal of Business Research*, *85*, 1–9.
- Cheng, Q., Wang, J., Lu, W., Huang, Y., & Bu, Y. (2020). Keyword-citation-keyword network: A new perspective of discipline knowledge structure analysis. *Scientometrics*, *124*(3), 1923–1943.
- Cho, J. (2020). Intellectual structure evolution of open access research observed through correlation index of keyword centrality. *Scientometrics*, *125*(3), 2617–2635.
- Choudhury, N., Faisal, F., & Khushi, M. (2020). Mining temporal evolution of knowledge graphs and genealogical features for literature-based discovery prediction. *Journal of Informetrics*, *14*(3), 101057.
- Feng, X., Li, L., Li, J., Cui, M., Sun, L., & Wu, Y. (2021). Temporal evolution of tagging subnetwork features and motif under different activity levels – take the Q&A community Zhihu as an example. *Information Discovery and Delivery*, *49*(2), 151–161.
- Fortunato, S., Bergstrom, C. T., Börner, K., Evans, J. A., Helbing, D., Milojević, S., Petersen, A. M., Radicchi, F., Sinatra, R., Uzzi, B., Vespignani, A., Waltman, L., Wang, D., & Barabási, A.-L. (2018). Science of science. *Science*, *359*(6379), eaao0185.

- Gao, H., Chong, A. Y. L., & Bao, H. (2023). Metaverse: Literature review, synthesis and future research agenda. *Journal of Computer Information Systems*. <https://doi.org/10.1080/08874417.2023.2233455>
- González-Valiente, C. L., León Santos, M., Arencibia-Jorge, R., Noyons, E., & Costas, R. (2021). Mapping the evolution of intellectual structure in information management using author co-citation analysis. *Mobile Networks and Applications*, 26(6), 2374–2388.
- Han, J., Teng, X., Tang, X., Cai, X., & Liang, H. (2020). Discovering knowledge combinations in multidimensional collaboration network: A method based on trust link prediction and knowledge similarity. *Knowledge-Based Systems*, 195, 105701.
- Harris, Z. S. (1954). Distributional structure. *WORD*, 10(2–3), 146–162.
- Haunschild, R., Leydesdorff, L., Bornmann, L., Hellsten, I., & Marx, W. (2019). Does the public discuss other topics on climate change than researchers? A comparison of explorative networks based on author keywords and hashtags. *Journal of Informetrics*, 13(2), 695–707.
- Hosseini, S., Baziyaad, H., Norouzi, R., Jabbedari Khiabani, S., Gidófalvi, G., Albadvi, A., Alimohammadi, A., & Seyedabrishami, S. (2021). Mapping the intellectual structure of GIS-T field (2008–2019): A dynamic co-word analysis. *Scientometrics*, 126(4), 2667–2688.
- Ji, L., Yuan, C., Feng, T., & Wang, C. (2020). Achieving the environmental profits of green supplier integration: The roles of supply chain resilience and knowledge combination. *Sustainable Development*, 28(4), 978–989.
- Katila, R., & Ahuja, G. (2002). Something old, something new: A longitudinal study of search behavior and new product introduction. *Academy of Management Journal*, 45(6), 1183–1194.
- Katsurai, M., & Ono, S. (2019). TrendNets: Mapping emerging research trends from dynamic co-word networks via sparse representation. *Scientometrics*, 121(3), 1583–1598.
- Khasseh, A. A., Soheili, F., Moghaddam, H. S., & Chelak, A. M. (2017). Intellectual structure of knowledge in iMetrics: A co-word analysis. *Information Processing & Management*, 53(3), 705–720.
- Kogut, B., & Zander, U. (1992). Knowledge of the firm, combinative capabilities, and the replication of technology. *Organization Science*, 3(3), 383–397.
- Krumov, L., Fretter, C., Müller-Hannemann, M., Weihe, K., & Hütt, M.-T. (2011). Motifs in co-authorship networks and their relation to the impact of scientific publications. *The European Physical Journal B*, 84(4), 535–540.
- Kuo, C.-I., Wu, C.-H., & Lin, B.-W. (2019). Gaining from scientific knowledge: The role of knowledge accumulation and knowledge combination. *R&D Management*, 49(2), 252–263.
- La, Z., & Chai, L. H. (2021). Comprehensive study of evolution of global environmental quality research using informetric co-word network. *Journal of Environmental Informatics*, 38(2), 2.
- Larivière, V., Haustein, S., & Börner, K. (2015). Long-distance interdisciplinarity leads to higher scientific impact. *PLoS ONE*, 10(3), e0122565.
- Lee, K., & Lee, S. (2021). Knowledge structure of the application of high-performance computing: A co-word analysis. *Sustainability*, 13(20), 20. <https://doi.org/10.3390/su132011249>
- Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., & Alon, U. (2002). Network motifs: Simple building blocks of complex networks. *Science*, 298(5594), 824–827.
- Nerkar, A. (2003). Old is gold? The value of temporal exploration in the creation of new knowledge. *Management Science*, 49(2), 211–229.
- Newman, M. E. J. (2004). Fast algorithm for detecting community structure in networks. *Physical Review E, Statistical, Nonlinear, and Soft Matter Physics*, 69(6), 066133.
- Qian, J., Hintze, A., & Adami, C. (2011). Colored motifs reveal computational building blocks in the *C. elegans* brain. *PLoS ONE*, 6(3), e17013.
- Ribeiro, P. (2011). Efficient and scalable algorithms for network motifs discovery (Ph.D. thesis), University of Porto.
- Ribeiro, P., & Silva, F. (2014). Discovering colored network motifs. *Complex Networks V* (pp. 107–118). Springer.
- Rossi, R. A., Ahmed, N. K., Carranza, A., Arbour, D., Rao, A., Kim, S., & Koh, E. (2021). Heterogeneous graphlets. *ACM Transactions on Knowledge Discovery from Data*, 15(1), 1–43.
- Serror, M., Hack, S., Henze, M., Schuba, M., & Wehrle, K. (2021). Challenges and opportunities in securing the industrial internet of things. *IEEE Transactions on Industrial Informatics*, 17(5), 2985–2996.
- Sporns, O., & Kötter, R. (2004). Motifs in brain networks. *PLOS Biology*, 2(11), e369.
- Tolstoy, D. (2010). Network development and knowledge creation within the foreign market: A study of international entrepreneurial firms. *Entrepreneurship & Regional Development*, 22(5), 379–402.
- Wang, L., Ren, J., Xu, B., Li, J., Luo, W., & Xia, F. (2020). MODEL: Motif-Based deep feature learning for link prediction. *IEEE Transactions on Computational Social Systems*, 7(2), 503–516.

- Wang, X., Feng, X., & Guo, Y. (2022). Analysis of the structure and time-series evolution of knowledge label network from a complex perspective. *Aslib Journal of Information Management*, 75(6), 1056–1078.
- Wernicke, S. (2005). A faster algorithm for detecting network motifs international workshop on algorithms in bioinformatics. *Algorithms in Bioinformatics*. https://doi.org/10.1007/11557067_14
- Zhang, J., Yu, Q., Zheng, F., Long, C., Lu, Z., & Duan, Z. (2016). Comparing keywords plus of WOS and author keywords: A case study of patient adherence research: Comparing keywords plus of WOS and Author keywords. *Journal of the Association for Information Science and Technology*, 67(4), 967–972.
- Zou, B., Wang, Y., Kwoh, C. K., & Cen, Y. (2023). Directed collaboration patterns in funded teams: A perspective of knowledge flow. *Information Processing & Management*, 60(2), 103237.

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.