Postprint: Knowledge Network Structural Relationship Extraction Based on Eigen-decomposition

Authors: Luan Yu, Teng Guangqing, tranquility, Han Shangxuan

Date: 2023-07-26T00:00:00+00:00

Abstract

[Purpose/Significance] Effective identification and extraction of structural relationships in knowledge networks facilitates the detection of topological structures and evolution patterns from complex data. [Method/Process] This paper proposes a method for extracting structural relationships in knowledge networks based on eigen decomposition of adjacency matrices. Based on real-world data, comparative analyses are conducted between the eigen decomposition method and the traditional co-occurrence frequency method from two perspectives: static structural relationship extraction and dynamic structural evolution, with additional comparison to the Pathfinder algorithm. The effectiveness of the eigen decomposition method for extracting knowledge network structural relationships is validated. [Results/Conclusion] Research results demonstrate that the eigen decomposition method can identify principal component information in the original knowledge network, accurately identify low-frequency relationships that are important for the overall topological structure of the network, and the extraction method is flexible and unconstrained.

Full Text

Preamble

Vol. 63, No. 7, April 2019
ChinaXiv Cooperative Journal

Extraction of Structural Relationships in Knowledge Networks Based on Eigen Decomposition

Luan Yu, Teng Guangqing, An Ning, Han Shangxuan
School of Information Science and Technology, Northeast Normal University, Changchun 130117

Abstract

[**Purpose/Significance**] Effective identification and extraction of structural relationships in knowledge networks helps detect the topological structure and evolution patterns of knowledge networks from complex data. [**Method/Process**] This paper proposes a method for extracting structural relationships in knowledge networks based on eigen decomposition of adjacency matrices. Using real data, the eigen decomposition method and traditional correlation frequency method are compared and analyzed from both static structural relationship extraction and dynamic structural evolution perspectives, and compared with the Pathfinder algorithm. The validity of extracting structural relationships in knowledge networks based on the eigen decomposition method is verified. [**Result/Conclusion**] Research results show that the eigen decomposition method can identify the main component information in original knowledge networks, accurately identify low-frequency correlations that are important to the overall network topology, and the extraction method is flexible and free.

Keywords: knowledge network; eigen network; eigen decomposition; structural relationship
Classification Number: G255.76
DOI: 10.13266/j.issn.0252-3116.2019.07.012

Knowledge network analysis methods have become a new research paradigm in library and information science for exploring association relationships and structural patterns among knowledge units. Researchers construct knowledge networks such as keyword networks and patent networks by using knowledge units (e.g., keywords, patent technologies) as network nodes and association relationships between knowledge units as edges. By analyzing the statistical characteristics and topological structure of knowledge networks, issues such as hierarchical structure and evolution mechanisms of knowledge can be studied. With the application of big data analytics in scientific research, knowledge network studies have exhibited characteristics of large-scale data and diversity. However, while focusing on high-value data analysis, knowledge network research must confront the low value density problem in big data analysis. Therefore, while ensuring high-frequency significance of data information, extracting effective structural relationships from knowledge networks has become an urgent problem to be solved.

2 Related Research

With the publication of several important research achievements in network science at the end of the 20th century, network analysis has attracted increasing attention from academia as a new research paradigm. Network thinking is particularly suitable for describing and analyzing components and their relationships in complex systems, abstracting system components as network nodes and relationships between components as network edges. As a general abstraction and description of these complex systems, networks emphasize the topological structure and statistical characteristics of complex systems. The combination of network science and library and information science has also changed the traditional descriptive statistics-based research paradigm in the field.

Researchers have introduced network analysis theory into patent analysis, citation analysis, hotspot identification, knowledge emergence, important scholar identification, and knowledge network structure, better revealing the structural relationships of research objects. J. Liebowitz pointed out that network analysis methods help better explain the association relationships between knowledge and the hierarchical structure of knowledge, playing an increasingly prominent role in knowledge management.

While conducting analysis and mining based on data, researchers must face the problem of low value density in big data analysis as data volume increases. Early node frequency methods first selected knowledge units (e.g., keywords) with high frequency in the domain as network nodes, then established association relationships between nodes. This method emphasizes the high frequency of individual knowledge units (nodes) but pays insufficient attention to knowledge association relationships. Currently, the more commonly used method is the correlation frequency method, which first selects high-frequency associations (node pairs) and builds knowledge networks based on these high-frequency associations. Knowledge networks extracted at specific correlation frequency levels, called level networks, highlight the importance of associations in network analysis compared to node frequency methods. This method for extracting network information is simple, intuitive, and computationally efficient, and has been widely applied in library and information science research, achieving corresponding results in related studies such as associated topic extraction and knowledge network structure identification.

In addition to extraction methods based on node association frequency, there is another class of path-based extraction methods represented by the Pathfinder algorithm. In the Pathfinder algorithm, if the distance D between two nodes is less than the weight of the edge directly connecting them, that edge is removed from the network. Therefore, the main idea of the Pathfinder algorithm is to compare and judge the importance of edges directly connecting nodes by calculating the weights of different paths between nodes. The most typical application of this algorithm in knowledge network path research and simplified analysis is in the literature analysis software CiteSpace. Chen Chaomei introduced the Pathfinder algorithm into knowledge network analysis, using its completeness to highlight important link features while simplifying networks. Due to the uniqueness of the Pathfinder algorithm in main path identification, this method has also been widely applied in community knowledge structure, member structure, knowledge exchange paths, and key literature identification.

With the publication and accumulation of achievements in network topology structure research, the above-mentioned correlation frequency method and Pathfinder algorithm, while highlighting association frequency or link importance, have shown deficiencies in extracting specific network topology information, potentially causing researchers to miss important structural information in networks. In response to these limitations of traditional methods such as correlation frequency in extracting network information, this paper proposes a knowledge network structural relationship extraction method that considers both network association frequency and topological information, and verifies the effectiveness of this method through static and dynamic knowledge network structural relationship extraction.

Traditional methods for extracting knowledge network structural relationships mainly include node frequency method, correlation frequency method, and Pathfinder algorithm. The node frequency method prioritizes the importance of individual knowledge units (network nodes). Li Gang et al. pointed out that using high-frequency words to represent the overall research direction of a field has inherent defects, while low-frequency words help obtain information about some implicit or forward-looking topics. The correlation frequency method, when obtaining knowledge networks at a certain correlation frequency threshold level, only examines the association frequency at the individual relationship (node pair) level, failing to consider the overall network topology, and may to some extent ignore or omit important low-frequency knowledge associations. The Pathfinder algorithm can only remove edges between nodes when there exists a path with lower weight between them, so the calculation result of the Pathfinder algorithm is a connected network, and the node scale of the original network will not be reduced. In knowledge network research, the association frequency between knowledge nodes (such as keywords) is only local information of the knowledge network. Judging the importance of knowledge nodes and their associations requires not only reference to local information but also consideration from the overall network topology. In the process of domain knowledge evolution, even low-frequency associations (node pairs) sometimes play very important roles.

3 Theory and Methods

3.1 Related Theory of Knowledge Network Structural Relationships

Knowledge units are generally considered the smallest units of domain knowledge, but the attributes of individual knowledge units cannot characterize the overall attributes of domain knowledge. As J. Gleick, author of *The Information: A History, A Theory, A Flood*, pointed out, the connectivity between knowledge is more important than knowledge units themselves. Numerous individual knowledge units in a domain form knowledge networks based on certain association relationships. Research on knowledge networks shows that many knowledge units gather together through specific association relationships, forming specific topological structure features, and then macro-level patterns and laws emerge in knowledge networks.

The adjacency matrix is the mathematical representation of a knowledge network. Eigen decomposition of the adjacency matrix can focus on local attributes of knowledge units and associations while considering the overall attributes of the knowledge network topology. Therefore, this paper proposes an information extraction method for knowledge network structural relationships based on adjacency matrix eigen decomposition. Eigen decomposition refers to decomposing a matrix into a product represented by its eigenvalues and eigenvectors. The adjacency matrix A of the original knowledge network G(N, L) is a real symmetric square matrix, so it can be eigen-decomposed. A matrix can be understood as a description of a linear transformation in its linear space. Through eigen decomposition, the eigenvalues and eigenvectors of the matrix can be discovered, from which the transformation form described by the matrix can be derived, and structural relationship information corresponding to the main eigenvalues of the matrix can be extracted. The adjacency matrix A carries all connection information of the original knowledge network G, where element values are association frequencies between knowledge nodes, and eigenvalues correspond to topological structure information in the network. Extracting the main eigenvalues of adjacency matrix A is equivalent to extracting topological structure information of network G after incorporating the PageRank algorithm idea. Based on this, the eigen network required for research can be generated.

3.2 Method and Process

The adjacency matrix A constructed from real data is a multi-valued matrix, where element values represent association frequencies between knowledge nodes. For adjacency matrix A, if an element value is 0, there is no association between the corresponding knowledge nodes; if the element value is greater than 0, there is an association between the corresponding knowledge nodes. For an N×N adjacency matrix A, it has N real eigenvalues, denoted in descending order as λ_i (i = 1, 2, 3, …, N), and correspondingly has N linearly independent eigenvectors, denoted as q_i (i = 1, 2, 3, …, N). Eigen decomposition of adjacency matrix A yields:

$$A = Q\Lambda Q^T$$

where Q is an orthogonal matrix composed of eigenvectors, the i-th column of Q equals q_i^T, and Λ is a diagonal matrix with diagonal elements being the descending eigenvalues λ_i (i = 1, 2, 3, …, N).

Meanwhile, the adjacency matrix A is binarized to obtain binary matrix C, which is used to determine the existence of association relationships between knowledge nodes. According to research needs, several eigenvalues that can represent the main transformation forms of adjacency matrix A are selected. By assigning 0 to other eigenvalues in diagonal matrix Λ except those to be retained, Λ_B is obtained, and matrix B is constructed as:

$$B = Q\Lambda_B Q^T$$

Matrix B is an extraction of information from adjacency matrix A, containing transformation information corresponding to specific eigenvalue combinations and their eigenvectors. The element values in matrix B are the weights of network edges under specific eigenvalue combinations. Two factors affect the element values in matrix B: first, the association frequency between knowledge nodes; second, the network topology information corresponding to specific eigenvalues. Furthermore, matrix B is compared with binary matrix C. If an association does not exist in binary matrix C, matrix B is corrected by setting the corresponding edge weight to 0. The corrected matrix B corresponds to a knowledge network with the same topology as the original knowledge network, but the edge weights of the knowledge network are reassigned. Unlike the fixed weights in traditional correlation frequency methods, edge weights in eigen decomposition methods are dynamic and relative. Different selected eigenvalues result in different extracted topology information, and the weight value of each edge in the network changes relatively. Edge weights defined in this way consider both local network information and overall topology information.

Next, based on edge weights, a threshold is set to discard edges with weights below the threshold in the reassigned knowledge network, and isolated nodes are removed to generate eigen network E, completing the extraction of knowledge network structural relationship information. The entire process is shown in Figure 1 [Figure 1: see original paper].

It should be noted that adjacency matrix A constructed from real data is often sparse, resulting in a few large eigenvalues and many small eigenvalues. Therefore, only a small number of eigenvalue and eigenvector combinations are needed to describe the main transformation content of adjacency matrix A. For the original knowledge network G, these few eigenvalues and eigenvectors correspond to important network topologies. The selection of eigenvalues determines which topology information will be extracted from the network; any single or multiple continuous or discontinuous eigenvalues at any ordinal position can be selected. Additionally, element values in matrix B are assigned as edge weights to the edge set L of the original knowledge network G, while the edge weight threshold is used to determine and extract edges with relatively high importance to the selected topology. Eigenvalue selection determines the topology information extracted from the network, and the edge weight threshold determines the scale of the final extracted eigen network. If more network information needs to be extracted, the number of eigenvalues can be increased and the weight threshold appropriately lowered; conversely, the number of eigenvalues can be reduced and the weight threshold appropriately raised. In the eigen decomposition method, edge weights in the selected eigenvalue combination's corresponding network topology are enhanced, while other edge weights are relatively reduced, allowing the method to highlight the importance of local structures according to different focus points. Because of this characteristic, the eigen decomposition method can not only simplify the network as a whole but also identify and extract important local information from the network.

3.3 Eigen Decomposition Method vs. Correlation Frequency Method

In previous studies using the correlation frequency method, threshold setting often depends on researchers' personal experience, and threshold selection is limited to the positive integer domain. In terms of operations on adjacency matrix A, this means reassigning elements below the threshold to 0. Through in-depth comparative analysis of the processes of eigen decomposition method and correlation frequency method for extracting network information, it can be found that the correlation frequency method is actually a special case of the eigen decomposition method. In the eigen decomposition method, when all eigenvalues of matrix A are retained, Λ_B equals Λ, and matrix B equals matrix A. If the edge weight threshold is set to the same positive integer as the correlation frequency threshold, the eigen decomposition method returns the same result as the correlation frequency method. At this point, both extraction methods operate on the same matrix A with the same threshold.

Clearly, the correlation frequency method is an eigen decomposition method that selects all eigenvalues of the adjacency matrix and limits edge weights to the positive integer domain. From the perspective of eigen decomposition, the correlation frequency method is scientifically reasonable as it selects all eigenvalues of the adjacency matrix based on a global network perspective and simplifies the network with a positive integer threshold. However, as a special case of the eigen decomposition method, the correlation frequency method ignores the sparsity problem of adjacency matrices in real-world networks and indiscriminately selects all eigenvalues, making it unable to identify and extract specific local network information components. Meanwhile, the threshold of the correlation frequency method is naturally limited to the positive integer domain. The correlation frequency method can only extract the most network information at the threshold level of 2, which is clearly not conducive to in-depth analysis of knowledge networks by researchers. The eigen decomposition method is a general extension of the correlation frequency method, breaking through the limitations of the correlation frequency method in eigenvalue and threshold selection, and enabling more flexible and accurate extraction of network information.

4 Case Analysis

4.1 Research Data

Research data were collected from the China National Knowledge Infrastructure (CNKI) and Wanfang Data knowledge service platforms. Using "knowledge management" as the search theme and library and information science as the discipline domain, all papers published in CSSCI journals between 1999-2017 were retrieved, extracting paper titles, keywords, publication dates, and other relevant information. To better compare and analyze knowledge structure relationship extraction methods from the perspective of associations between different knowledge communities, further searches were conducted using "social network," "complex network," "social network analysis," "network analysis," and "network science" as keywords. On CNKI, the search domains were "natural science theory and methods," "social science theory and methods," "mathematics," and "nonlinear science and systems science"; on Wanfang Data, the domains were "natural science general," "social science general," and "mathematics and physics." All papers published in CSSCI journals between 1999-2017 were retrieved, extracting paper titles, keywords, and publication dates. All collected data were aggregated, with duplicate papers, conference notices, journal announcements, and other invalid data removed, resulting in a total of 1,842 journal papers and 3,018 keywords.

Starting from 1999-2001, a 3-year window was used as a time window with a 1-year step, smoothly moving to 2015-2017. Data for each time window (3 years) were treated as current occurrence values. To avoid the impact of drastic data fluctuations on domain knowledge development, a "smoothing" process was applied by moving one step (1 year) forward in the time series. Through cross-platform combinations of multiple search themes and keywords (CNKI and Wanfang Data) and moving smoothed time window divisions, the inherent development patterns of knowledge units and associations could be highlighted while showing cross-aggregation evolution phenomena between the "knowledge management" knowledge community and "social network," "complex network" knowledge communities over time.

Basic statistics for the 17 time windows are shown in Table 1 .

Table 1 Number of Papers and Keywords in Each Time Window

Time Window	Papers	Keywords	Time Window	Papers	Keywords

4.2 Extraction Process

First, original adjacency matrices and original networks were constructed based on the data in Table 1. The co-occurrence relationships and frequencies of keywords in literature under time windows t_1-t_{17} were statistically analyzed, and multi-valued adjacency matrices A_1-A_{17} for each time window were established. Based on these 17 adjacency matrices, 17 domain knowledge networks G_1-G_{17} were generated according to keyword co-occurrence relationships. Second, structural relationship information of knowledge networks was extracted using the conventional correlation frequency method. In the correlation frequency method, the smaller the threshold, the more detailed the extracted knowledge network structural relationship information. In this study, 2 was used as the frequency threshold to filter out associations with co-occurrence frequencies less than 2 in the original network, i.e., removing elements less than 2 in adjacency matrices A_1-A_{17}, thereby obtaining knowledge networks F_1-F_{17} at the corresponding frequency level. Third, structural relationship information of original networks was extracted using the eigen decomposition method. Adjacency matrices A_1-A_{17} were eigen-decomposed, matrix eigenvalues were extracted in different combinations, knowledge network structural relationship information was extracted according to set thresholds, and eigen networks E_1-E_{17} under corresponding eigenvalues were generated. Finally, eigen networks extracted based on eigen decomposition (E_1-E_{17}) were compared and analyzed with original networks (G_1-G_{17}) and level networks extracted by the correlation frequency method (F_1-F_{17}) to verify the function and characteristics of the eigen decomposition method in identifying and extracting main structural relationship information from networks.

Taking the knowledge network G_{17} under time window t_{17} as an example, its adjacency matrix A_{17} was eigen-decomposed. Adjacency matrix A_{17} has a total of 748 eigenvalues. The eigenvalues of adjacency matrix A_{17} were sorted in descending order and numbered, with the ordinal number as the horizontal axis and eigenvalues as the vertical axis. The arrangement of 748 eigenvalues of matrix A_{17} is shown in Figure 2 [Figure 2: see original paper].

Figure 2 [Figure 2: see original paper] Eigenvalues of Adjacency Matrix A_{17}

5 Analysis Results

5.1 Static Network Structural Relationship Extraction

In the eigen decomposition method, determining the combination of eigenvalues to be extracted determines the structural relationship information of the relevant network to be extracted. The edge weight threshold determines the scale of the sub-network finally extracted from the original network. Taking time window t_{17} as an example, A_{17} was eigen-decomposed, extracting 22 eigenvalues with relatively large absolute values (the 1st-11th and 738th-748th eigenvalues). Edge weight thresholds were set to 0.5, 1.0, 1.1, and 1.2, respectively. Based on the network structural relationship information of these 22 eigenvalues, eigen networks under different edge weight thresholds were constructed and compared with the original knowledge network G_{17} and the level network F_{17} extracted by the correlation frequency method. The results are shown in Figure 4 [Figure 4: see original paper]. Networks in Figure 4 were drawn using the F-R layout algorithm proposed by T.M.J. Fruchterman and E.M. Reingold and a self-developed network visualization program.

Note: KM-Knowledge Management, SN-Social Network, CN-Complex Network, SNA-Social Network Analysis, KG-Knowledge Graph, IS-Information Science, L-Library

Figure 4 [Figure 4: see original paper] Original Network, Eigen Network, and Level Network

In Figure 4, (a) is the original knowledge network G_{17}, (b) is the eigen network with threshold = 0.5, (c) is the eigen network with threshold = 1.0, (d) is the eigen network with threshold = 1.1, (e) is the eigen network with threshold = 1.2, and (f) is the level knowledge network F_{17}. Meanwhile, basic statistical features of eigen network E $_{17}$, original network G $_{17}$, and level network F $_{17}$ generated based on the eigen decomposition method are shown in Table 2 .

Table 2 Network Statistical Features

Network	Nodes	Edges	Density	Clustering Coefficient	Characteristic Pat
Original Network G $_{17}$	748	2054	0.0074	0.1664	3.3612
Eigen Network E $_{17}$ (threshold=0.5)	591	930	0.0053	0.0743	3.1906
Eigen Network E $_{17}$ (threshold=1.0)	147	183	0.0171	0.0262	3.1083
Eigen Network E $_{17}$ (threshold=1.1)	56	68	0.0442	0.0800	3.8617
Eigen Network E $_{17}$ (threshold=1.2)	51	60	0.0471	0.0750	4.3388
Level Network F $_{17}$	51	57	0.0447	0.0610	2.6267

Based on the results in Figure 4 and Table 2, it can be found that in the eigen decomposition method, after determining the eigenvalues of the adjacency matrix to be extracted, the lower the edge weight threshold, the closer the eigen network is to the original network G $_{17}$; the higher the weight threshold, the more streamlined the eigen network becomes while retaining the main structural features of the original network. In this example, when the edge weight threshold is set to 1.2, the obtained eigen network has approximately the same number of nodes and edges as the level network F $_{17}$ with correlation frequency threshold = 2. However, Figure 4(e) shows a connected network, while Figure 4(f) shows a non-connected network. Obviously, the eigen network tends to retain the main structural relationships of the network, while the level network focuses on the statistical significance of association frequencies between network nodes. This characteristic of eigen networks will help reveal the growth or decline of knowledge association relationships in dynamic network analysis of domain knowledge evolution.

In addition, the edge weight threshold in the eigen decomposition method can take either integer or decimal values, offering far richer choices than the correlation frequency method. In the correlation frequency method, the minimum selectable frequency threshold is the integer 2, meaning information can only be extracted from the original network at the threshold = 2 level (the smaller the frequency threshold, the richer the information extracted from the original network). The eigen decomposition method can freely extract and scale overall or local topological structure features from the original network through flexible combinations of eigenvalues and thresholds, not limited to fixed thresholds.

5.2 Dynamic Evolution of Structural Relationships

By extracting eigenvalues with relatively large absolute values, the main structural relationship information in the original knowledge network can be identified. Static analysis of knowledge networks reveals that under the same edge weight conditions, sub-networks such as Knowledge Management (KM), Social

Network (SN), and Complex Network (CN) are relatively large in scale, consistent with the data collection scheme. On the other hand, the emergence of knowledge association relationships can affect the evolution of knowledge network topology. Therefore, the eigen decomposition method was further used to extract eigenvalues corresponding to knowledge management, social network, and complex network from original knowledge networks (G_1-G_{17}) between 1999-2017, combining them to generate time-series eigen networks (E_1-E_{17}). The cross-association status of three knowledge communities in the network was tracked and compared with time-series level knowledge networks (F_1-F_{17}) obtained by the correlation frequency method.

The eigen decomposition method was used to extract structural relationship information from original networks under 17 time windows. First, eigenvalues related to knowledge management, social network, and complex network were extracted from the adjacency matrices of 17 time windows. Then, using 0.5 as the weight threshold, relevant eigenvalue combinations from the original knowledge networks under the 17 time windows were extracted to generate corresponding eigen networks. Using the same F-R layout algorithm, time-series eigen networks with threshold = 0.5 are shown in Figure 5 [Figure 5: see original paper].

Figure 5 [Figure 5: see original paper] Time-Series Eigen Networks (threshold = 0.5)

The time-series eigen networks in Figure 5 show that in the initial stage of the time axis (t_1, t_2), the social network (SN) and knowledge management (KM) knowledge communities under investigation were not yet connected and were in their respective development stages. Complex network (CN) related knowledge had not yet appeared in the investigation objects. By time window t_3, the knowledge management and social network knowledge communities were connected for the first time. In time window t_4, the complex network knowledge community appeared for the first time but was in a non-connected state with knowledge management and social network knowledge communities. In time window t_5, the previously connected knowledge management and social network knowledge communities were disconnected again, with all three knowledge communities being non-connected with each other. In time window t_6, social network and complex network knowledge communities were connected with the knowledge management knowledge community respectively, meaning social network and complex network could only establish connections through knowledge management. Starting from time window t_7, the three knowledge communities of social network, complex network, and knowledge management remained mutually connected through multiple knowledge nodes or other small-scale communities. In subsequent time windows, the three knowledge nodes of social network, complex network, and knowledge management relatively stably formed three peaks in the network that were interconnected (or directly associated) through bridge points, with the association relationships among the three being clearly presented.

To gain a clearer understanding of the extraction effects of the eigen decomposition method under different edge weights, eigen networks were further generated using 1.0 as the edge weight threshold, as shown in Figure 6 [Figure 6: see original paper].

Figure 6 [Figure 6: see original paper] Time-Series Eigen Networks (threshold = 1.0)

The eigenvalues extracted in Figure 6 are the same as in Figure 5, with the difference being that the edge weight threshold is set to 1.0, based on which time-series eigen networks are generated. Comparative analysis between Figure 6 and Figure 5 reveals that due to the increased edge weight, some edges below the threshold are discarded, and corresponding isolated nodes are eliminated. This results in a corresponding reduction in the scale of the three knowledge communities, and some structural relationships visible under low weight thresholds are correspondingly hidden, such as in time windows t_3, t_4, t_6, t_7, t_8, t_9, etc. Nevertheless, Figure 6 generally reveals the structural relationships among knowledge management, social network, and complex network knowledge communities consistently with Figure 5, forming three relatively distinct peaks interconnected through bridge points (or directly associated) in the later stage of the time series.

For comparative validation purposes, node relationship information in knowledge networks G_1-G_{17} was further extracted using the general correlation frequency method. Using 2.0 as the correlation frequency threshold, association relationships and corresponding node pairs were extracted from the original networks to generate level networks F_1-F_{17} at the correlation frequency level of 2.0, as shown in Figure 7 [Figure 7: see original paper].

Figure 7 [Figure 7: see original paper] Time-Series Level Networks (threshold = 2.0)

A correlation frequency threshold of 2.0 in the correlation frequency method means that, apart from the original network (equivalent to a correlation frequency threshold of 1.0), this is the smallest threshold, generating level networks closest to the original network. The resulting level networks (see Figure 7) have significantly more branch structures (non-connected fragments) than eigen networks generated by the eigen decomposition method (see Figures 5 and 6). In all time windows of Figure 7, the level networks generated by the correlation frequency method contain non-connected fragments, indicating that some important structural relationships are rejected by the correlation frequency threshold. In the first half of the time axis (t_1-t_9), the three knowledge communities of knowledge management, social network, and complex network never established connections. In the second half of the time axis, in time windows t_{10}, t_{12}, t_{13}, t_{15}, and t_{16}, the complex network could only establish connections with the knowledge management knowledge community through the social network; in time window t_{11}, knowledge management was connected with social network, while complex network was non-connected with both; in time window t_{17}, social network was connected with complex network, while knowledge management was non-connected with both. Only time window t_{14} shows the three peaks interconnected through bridge points (or directly associated) that are prominently displayed in the second half of the time axis in the eigen decomposition method.

In fact, cross-fusion of knowledge is a universal law in knowledge development. Since the revival of network science at the end of the 20th century, network analysis as a new research paradigm has been introduced into many aspects of knowledge management research, with cross-combinations of social network, complex network, and knowledge management appearing in numerous literature. In eigen networks generated by eigen decomposition method-based structural relationship extraction, the specific formation process of cross-associations among the three knowledge communities of social network, complex network, and knowledge management can be observed. The correlation frequency method only considers the association frequency between nodes when setting edge weights, without considering the overall topological structure properties of the network, and cannot identify edge information that is below the threshold in frequency but in a relatively important position in the overall topology (specifically manifested as non-connection in multiple time windows and loss of bridging relationships between "peaks" in Figure 7). Therefore, in level networks generated by edges (node pairs) meeting the frequency threshold, important low-frequency association relationships in the knowledge network evolution process will be lost.

5.3 Comparison Between Eigen Decomposition Method and Pathfinder Algorithm

Based on the analysis of differences between the eigen decomposition method and correlation frequency method, the eigen decomposition method was further compared with another path-based network simplification algorithm, the Pathfinder algorithm. First, the Pathfinder algorithm was used to simplify knowledge networks under time windows t_2, t_7, t_{12}, and t_{17}. Parameter q was set to n-1, meaning all real paths between nodes were considered. Parameter r was set to 1 and infinity (Inf) respectively to observe the network simplification effects under different parameter r values.

Table 3 Statistical Features of Pathfinder Subnets (PFNETs)

Network	r=1 Nodes	r=1 Edges	r=Inf Nodes	r=Inf Edges
PFNETs	202	526	725	2029
	839	2429	748	2044
	202	500	725	1938
	839	2323	748	1997

The Pathfinder algorithm retains at least one original path connecting two nodes during network simplification, thus not destroying the original connectivity of the network. The extracted subnet PFNETs maintain the same number of nodes as the original network, with no reduction in node scale. In terms of the number of edges, subnet PFNETs simplify the network to a lesser degree than the eigen decomposition method and correlation frequency method. The Pathfinder algorithm is suitable for solving certain problems about main network paths. The flexibility of the eigen decomposition method in identifying and extracting network topology information also lies in its ability to freely determine the network components and scale to be extracted through eigenvalue combination selection and threshold setting, which is something the Pathfinder algorithm cannot achieve.

6 Conclusion and Discussion

This paper proposes a method for extracting knowledge network structural relationships based on eigen decomposition of adjacency matrices. The study generates eigen networks through extracted eigenvalue combinations and compares and analyzes the eigen decomposition method and correlation frequency method from both static analysis and dynamic analysis perspectives. The analysis results show that knowledge network structural relationship extraction based on the eigen decomposition method has the following characteristics:

(1) The eigen decomposition method can identify main component information in knowledge networks. When extracting structural relationship information from knowledge networks, the importance of nodes and edges in the network should be judged, and relatively less important edges should be removed, in principle, while considering the overall topological structure properties of knowledge networks. Previous data processing methods that set thresholds based on node occurrence frequency or node pair association frequency determine the overall attributes of domain knowledge at the macro level based on local information of knowledge units or knowledge associations at the micro level. Their correctness and effectiveness cannot meet the needs of presenting global network structural relationships. These two traditional methods ignore the topological structure properties of networks, deviating to some extent from the original intention of network scientific thinking, and may cause researchers to miss important structural information in networks. Knowledge units or cross-associations between knowledge communities in their infancy (low frequency) are often ignored.

(2) The eigen decomposition method can accurately identify low-frequency but topologically important associations and is more sensitive to structural relationships between knowledge. In the process of domain knowledge growth and development, the addition of new knowledge and establishment of new relationships need to go through a process from weak to strong. Changes in network overall structure often come from repeated superposition of small changes, so low-frequency associations may also play a crucial role in changing domain knowledge network topology. The eigen

decomposition method retains important paths at the topological level and well presents three peaks of social network, complex network, and knowledge management that are interconnected (or directly associated) through bridge points (see Figures 5 and 6). The correlation frequency method clearly rejects low-frequency associations between complex network and knowledge management (see Figure 7). Obviously, focusing solely on association frequency reduces the presentation of key details at the network topology level to some extent, while the eigen decomposition method can observe the specific process of cross-associations between nodes corresponding to specified eigenvalues in the extracted network. Therefore, the eigen decomposition method is more sensitive to cross-associations between knowledge units or knowledge communities and is more helpful for studying the dynamic evolution mechanisms of knowledge.

(3) The eigen decomposition method can flexibly and freely extract knowledge network structural relationship information. The correlation frequency method uses association frequency between knowledge units as the basis for determining the importance of associations between knowledge units. The eigen decomposition method comprehensively considers both the overall topological structure properties of the network and local association relationships between nodes when setting weights and thresholds for knowledge unit associations, making it superior to the correlation frequency method. Additionally, through eigenvalue selection, the eigen decomposition method can not only generate eigen networks reflecting the original network topology using eigenvalues with relatively high absolute values but also generate eigen networks needed by researchers by selecting specified eigenvalues for research focus. This characteristic makes knowledge network structural relationship extraction based on the eigen decomposition method flexible and free, applicable to both overall network topology identification and analysis of subtle changes in specific knowledge associations.

Although the eigen decomposition method can accurately identify and extract specific information from knowledge networks, its implementation process is more complex compared to traditional network data processing methods. In subsequent research, we will continue to explore application patterns of the eigen decomposition method in knowledge network research, further standardizing and simplifying the specific steps of the eigen decomposition method. Meanwhile, in the process of eigen decomposition of adjacency matrices, it was preliminarily observed that the network topology structures corresponding to eigenvalues with relatively small absolute values also affect the statistical characteristics of knowledge networks to some extent. Exploring the meaning of these relatively small absolute value eigenvalues in knowledge networks is also part of future research work.

References

[1] LEWIS T G. Network Science: Principles and Applications[M]. Translated by Chen Xiangyang, Ju Xiulian, et al. Beijing: China Machine Press, 2011.
[2] ZHENG Jinlian, DI Zengru. Complex Network Research and Complex Phenomena[J]. Journal of Systems Dialectics, 2005, 13(4): 8-13.
[3] KRAFFT J, QUATRARO F, SAVIOTTI P P. The knowledge-base evolution in biotechnology: a social network analysis[J]. Economics of innovation and new technology, 2011, 20(5): 445-475.
[4] CHEN C, IBEKWE-SANJUAN F, HOU J. The structure and dynamics of cocitation clusters: a multiple-perspective cocitation analysis[J]. Journal of the Association for Information Science and Technology, 2010, 61(7): 1386-1409.
[5] ZHAO Rongying, WANG Ju. Knowledge Graph Analysis of Library Science[J]. Journal of Library Science in China, 2011, 37(2): 40-50.
[6] AN Ning, TENG Guangqing, BAI Shuchun, et al. Research on Domain Core Knowledge Emergence Based on Network Hub[J]. Library and Information Service, 2017, 61(18): 98-106.
[7] QIU Junping, ZHANG Xiaopei. Author Co-citation Analysis of Domestic Knowledge Management Field Based on CSSCI[J]. Information Science, 2011, 29(10): 1441-1445.
[8] LIU Xiang, MA Feicheng, WANG Xiaoguang. Structure and Process Models of Knowledge Networks[J]. Systems Engineering - Theory & Practice, 2013, 33(7): 1836-1844.
[9] LIEBOWITZ J. Linking social network analysis with the analytic hierarchy process for knowledge mapping in organizations[J]. Journal of knowledge management, 2005, 9(1): 76-86.
[10] KATAKIS I, TSOUMAKAS G, VLAHAVAS I. Multilabel text classification for automated tag suggestion[EB/OL]. [2018-07-01]. http://lpis.csd.auth.gr/publications/katakis_{{{ecmlkdd(
[11] GONZALEZ-ALCAIDE G, CASTELLO-COGOLLOS L, NAVARRO-MOLINA C, et al. Library and information science research areas: analysis of journal articles in LISA[J]. Journal of the American Society for Information Science and Technology, 2008, 59(1): 150-154.
[12] ZHANG J, XIE J, HOU W, et al. Mapping the knowledge structure of research on patient adherence: knowledge domain visualization based on co-word analysis and social network analysis[J]. PloS one, 2012, 7(4): e34497.
[13] TENG Guangqing, BAI Shuchun, HAN Shangxuan, et al. Hierarchical Knowledge Network Principle Analysis Based on Scale-Free and Fractal Theory[J]. Library and Information Service, 2017, 61(14): 132-140.
[14] SCHVANEVELDT R W, DURSO F T, DEARHOLT D W. Network structures in proximity data[J]. The psychology of learning and motivation, 1989, 24: 249-284.
[15] CHEN C, MORRIS S. Visualizing evolving networks: minimum spanning trees versus pathfinder networks[C]//IEEE conference on information visualization. Washington DC: IEEE Computer Society, 2003: 67-74.
[16] HAN Yi, TONG Ying, XIA Hui. Comparison of Main Path Methods and Highly Cited Paper Methods for Domain Evolution Structure Identification[J].

Library and Information Service, 2013, 57(3): 11-17.
[17] AMITH M, CUNNINGHAM R, SAVAS L S, et al. Using Pathfinder networks to discover alignment between expert and consumer conceptual knowledge from online vaccine content[J]. Journal of biomedical informatics, 2017, 74: 33-45.
[18] SANCHEZ-FRANCO M J, MUÑOZ-EXPÓSITO M, VILLAREJO-RAMOS F. A knowledge structure exploration on social network sites[J]. Kybernetes, 2017, 46(5): 818-839.
[19] MA Ruimin, ZHANG Xin. Research on Domain Knowledge Exchange Main Path Discovery Based on Pathfinder Algorithm[J]. Journal of the China Society for Scientific and Technical Information, 2016, 35(8): 856-863.
[20] ZHANG Yun, HUA Weina, YUAN Shunbo. Exploration of Methods for Identifying Key Literature in a Field Using Citations[J]. Library and Information Service, 2016, 60(1): 66-73, 82.
[21] GLEICK J. The Information: A History, A Theory, A Flood[M]. Translated by GAO Bo. Beijing: Posts & Telecom Press, 2013: 409-421.
[22] LI Gang, BA Zhichao. Research on Several Issues in Co-word Analysis Process[J]. Journal of Library Science in China, 2017, 43(4): 93-113.
[23] BRIN S, PAGE L. The anatomy of a large-scale hypertextual web search engine[J]. Computer networks and ISDN systems, 1998, 30(1-7): 107-117.
[24] FRUCHTERMAN T M J, REINGOLD E M. Graph drawing by force-directed placement[J]. Software: practice and experience, 1991, 21(11): 1129-1164.
[25] TENG Guangqing. Emergence of Domain Knowledge Association Relationships Based on Frequency Evolution[J]. Journal of Library Science in China, 2018, 44(3): 79-95.
[26] MINTZ B, SCHWARTZ M. The power structure of American business[M]. Chicago: University of Chicago Press, 1987: 224-284.

Author Contributions

Luan Yu: Data collection and analysis, paper writing;
Teng Guangqing: Proposed research ideas, designed research scheme, paper writing and revision;
An Ning: Data analysis;
Han Shangxuan: Paper revision.

Note: Figure translations are in progress. See original paper for figures.