Statistics and Its Interface Volume 17 (2024) 145–158

Community detection in temporal citation network via a tensor-based approach*

Tianchen Gao, Rui Pan, Junfei Zhang†, and Hansheng Wang

In the era of big data, network analysis has attracted widespread attention. Detecting and tracking community evolution in temporal networks can uncover important and interesting behaviors. In this paper, we analyze a temporal citation network constructed by publications collected from 44 statistical journals between 2001 and 2018. We propose an approach named Tensor-based Directed Spectral Clustering On Ratios of Eigenvectors (TD-SCORE) which can correct for degree heterogeneity to detect the community structure of the temporal citation network. We first explore the characteristics of the temporal network via in-degree distribution and visualization of different snapshots, and we find that both the community structure and the key nodes change over time. Then, we apply the TD-SCORE method to the core network of our temporal citation network. Seven communities are identified, including variable selection, Bayesian analysis, functional data analysis, and many others. Finally, we track the evolution of the above communities and reach some conclusions.

Keywords and phrases: Community detection, Temporal citation network, TD-SCORE, Community evolution, CP decomposition.

1. INTRODUCTION

Network analysis is an important statistical tool arising in many disciplines [43, 39, 56]. A network consists of nodes and edges. According to the meaning of the nodes and edges, abundant networks can be defined and studied, such as social networks [32], cooperation networks [55], electrical networks [39] and many others. A citation network is one of the typical networks, where nodes are treated as papers and edges are the citation relationships among them. Statistical analysis on citation networks proceeds in many directions, including academic evaluation indicators [46, 57], citation network visualization [40, 29] and community detection [24]. The aim of community detection is to identify groups of nodes that are densely interconnected [16]. Detecting communities in citation networks can identify research topics and key papers in a certain discipline [8, 24, 23].

According to whether the network changes over time, algorithms for community detection can be roughly divided into static community detection and dynamic community detection. Existing algorithms for static community detection can be divided into two types, i.e., traditional algorithms and modularity-based algorithms [21]. Traditional algorithms include but are not limited to spectral clustering algorithms [51, 25], hierarchical clustering algorithms [45, 1] and partitional clustering algorithms [3, 4]. Modularity-based algorithms include extremal optimization algorithms [35, 5], greedy optimization algorithms [11] and many others. Among them, the D-SCORE algorithm proposed by [24] is a spectral clustering algorithm, which is used for detecting the community structure in a citation network. However, in practice, a citation network can be seen as temporal, which is constantly evolving over time [6]. It is then important to explore the development of communities and trace their evolution.

Algorithms on dynamic network community detection can be divided into three classes [41]. The first class focuses on instant optimal community detection, which can be seen as a natural extension of static community detection. The basic idea is to apply a static community detection algorithm to find communities and then use similarity indicators to match communities at adjacent moments [9, 18, 20]. The second class is temporal trade-off community detection, where the incremental algorithm is extensively used [58, 44, 31]. This class of approaches adjusts the community structure according to the increase and decrease of nodes and edges in different steps to optimize the loss function. It is suitable for long-term networks with relatively drastic changes [41]. The third class is cross-network community detection [22, 38, 50], where different steps involved in network development are not treated independently. In this kind of method, tensor decomposition has attracted the attention of many researchers [53, 2, 37, 36, 26].

The temporal network can be regarded as a tensor when it has a multilayer network structure [38]. To see why, we use the dataset in our research for explanation. The

* The research of Rui Pan is supported by National Natural Science Foundation of China (No. 11971504). Hansheng Wang's research is partially supported by National Natural Science Foundation of China (No. 11831008) and also partially supported by the Open Research Fund of Key Laboratory of Advanced Theory and Application in Statistics and Data Science (KLATASDS-MOE-ECNU-KLATASDS2101). Rui Pan and Junfei Zhang are both supported by the disciplinary funding and the Emerging Interdisciplinary Project of Central University of Finance and Economics.

†Corresponding author.

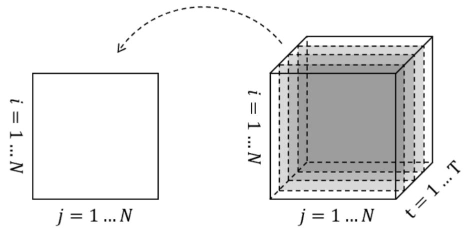


Figure 1. An adjacency matrix and a third-order tensor formed by T adjacency matrices.

dataset is collected from the "Web of Science"(*http://apps.webofknowledge.com/*). Specifically, we collect 59,690 papers published from 2001 to 2018 in 44 statistical journals. The adjacency matrix is constructed by year and arranged to form a third-order tensor, where nodes are papers and edges are the citation relationships. Figure 1 shows the diagrammatic sketch of an adjacency matrix and a third-order tensor.

At present, tensor-based methods are widely used to identify community structure. [42, 28] apply the implicit slice canonical decomposition method, which is a type of CP decomposition, to obtain the compilation features. These features are further used to form clusters by k-means. [17] investigate nonnegative tensor decomposition to extract the community activity structure of temporal multilayer networks. This method is applied to a network of students' social interactions at school. However, when the layers are very sparse or when there are many connections among communities in each layer, this method cannot achieve the desired effect. Aiming at this defect, [34] propose an improved algorithm for nonnegative tensor decomposition. The authors reduce the number of layers by merging strong correlation layers before decomposition. [26] provides statistical guarantees for community detection methods in multilayer networks based on tensor decomposition. From the perspective of application, [37] use tensor-based methods to identify communities in brain networks using rs-fMRI data, which leads to stable and accurate results.

Considering the temporal citation network as a third-order tensor, we focus on community detection based on tensor decomposition and propose a method called Tensor-based Directed Spectral Clustering On Ratios of Eigenvectors (TD-SCORE) to explore the dynamic nature and community evolution events of the citation network. Compared to the tensor-based methods mentioned above, the main innovation of our work is to use the entry-wise ratios of eigenvectors as feature matrices for clustering [24], which can correct for degree heterogeneity. Furthermore, previous studies mostly focus on multilayer networks. Our work treats the temporal network as a tensor to detect the community structure. Finally, the dataset used in the work is on a large scale, containing 59,690 papers over 18 years published in 44 statistical journals. This paper expands the scope of related work, and the method proposed in this paper provides new orientations for dynamic network community detection.

The rest of the paper is organized as follows. Section 2 presents the temporal citation network and its descriptive analysis. Section 3 illustrates CP decomposition with ALS and the newly proposed TD-SCORE method. Section 4 detects the dynamic communities of the citation network. We explore the dynamic nature of the citation network and analyze the evolution events. Section 5 concludes the paper.

2. DATA DESCRIPTION

2.1 Statistical citation network of 44 journals

The papers in our citation network are collected from 44 statistical journals. Table 1 provides the list of the 44 journals, including *Annals of Statistics* (AoS), *Journal of American Statistical Association* (JASA), *Journal of the Royal Statistical Society Series B* (JRSS-B) and *Biometrika*. For each paper, the following variables are obtained: title, authors, abstract, keywords, publisher, published date and the reference list. Table 2 presents an example of a published paper.

2.2 Descriptive analysis of the citation network

In this section, a descriptive analysis of the citation network is given. In our citation network, there are 59,690 nodes and 265,038 edges. The period of the citation network is from 2001 to 2018, where the network can be constructed yearly. The edges in the network are directed, which describe the citation relationships among nodes. Without loss of generality, nodes are not allowed to be self-related. Fixed time t, let $d_t = n_t/(N_t^2 - N_t)$ be the density of a directed network, where n_t denotes the number of edges at time t and N_t denotes the number of nodes at time t. The density of our citation network in 2018 is 0.00744 %, which indicates that the citation network is extremely sparse. We also notice that 12,660 papers in our citation network are written by individual authors.

Degree is an index to measure the importance of nodes in the network. In a practical sense, the in-degree of a node means the number of papers that have cited this paper. It is a natural measurement of the importance of a paper. On the other hand, out-degree of a node means the number of references cited by this paper. There are 9,399 nodes with zero in-degree, which may be newly published papers. While 19,362 nodes are with zero out-degree. These nodes may be papers published earlier, and their citations are not observed in our dataset. Figure 2 shows the log-log in-degree distribution, from which we can see that the in-degree distribution approximately obeys a power-law distribution. That is, the in-degree of most of nodes are small, and a few nodes have a large in-degree. The Gini coefficient of in-degree is 0.753, which means that the in-degree is highly dispersed.

Table 1. The list of 44 famous statistical journals

ID	Journals	ID	Journals
1	ADVANCES IN DATA ANALYSIS AND CLASSIFICATION	23	JOURNAL OF NONPARAMETRIC STATISTICS
2	AMERICAN STATISTICIAN	24	JOURNAL OF STATISTICAL COMPUTATION AND SIMULATION
3	ANNALS OF APPLIED STATISTICS	25	JOURNAL OF STATISTICAL PLANNING AND INFERENCE
4	ANNALS OF STATISTICS	26	JOURNAL OF STATISTICAL SOFTWARE
5	ANNALS OF THE INSTITUTE OF STATISTICAL MATHEMATICS	27	JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
6	ANNUAL REVIEW OF STATISTICS AND ITS APPLICATION	28	JOURNAL OF THE ROYAL STATISTICAL SOCIETY SERIES A-STATISTICS IN SOCIETY
7	BAYESIAN ANALYSIS	29	JOURNAL OF THE ROYAL STATISTICAL SOCIETY SERIES B-STATISTICAL METHODOLOGY
8	BERNOULLI	30	JOURNAL OF THE ROYAL STATISTICAL SOCIETY SERIES C-APPLIED STATISTICS
9	BIOINFORMATICS	31	JOURNAL OF THE ROYAL STATISTICAL SOCIETY SERIES D-THE STATISTICIAN
10	BIOMETRICS	32	JOURNAL OF TIME SERIES ANALYSIS
11	BIOMETRIKA	33	R JOURNAL
12	BIOSTATISTICS	34	SCANDINAVIAN JOURNAL OF STATISTICS
13	COMMUNICATIONS IN STATISTICS-SIMULATION AND COMPUTATION	35	SPATIAL STATISTICS
14	COMMUNICATIONS IN STATISTICS-THEORY AND METHODS	36	STATISTICA SINICA
15	COMPUTATIONAL STATISTICS	37	STATISTICAL ANALYSIS AND DATA MINING
16	COMPUTATIONAL STATISTICS & DATA ANALYSIS	38	STATISTICAL METHODOLOGY
17	ELECTRONIC JOURNAL OF STATISTICS	39	STATISTICAL METHODS AND APPLICATIONS
18	INTERNATIONAL STATISTICAL REVIEW	40	STATISTICAL MODELLING
19	JOURNAL OF APPLIED STATISTICS	41	STATISTICAL SCIENCE
20	JOURNAL OF BUSINESS & ECONOMIC STATISTICS	42	STATISTICS
21	JOURNAL OF COMPUTATIONAL AND GRAPHICAL STATISTICS	43	STATISTICS & PROBABILITY LETTERS
22	JOURNAL OF MULTIVARIATE ANALYSIS	44	STATISTICS AND COMPUTING

Table 2. An example to show 7 variables of a published paper

Title	Journal	Publication Date	Author
The adaptive lasso and its oracle properties	JASA	2006	Zou, Hui
Keywords	**Abstract**	**Details of its References**	
asymptotic normality lasso minimax oracle inequality variable selection	The lasso is a popular technique for simultaneous estimation and variable selection. Lasso variable selection has been shown...	Regularization of wavelet approximations Antoniadis, A and Fan, J. Q. Journal of the American Statistical Association 2001	

To comprehend the changes in high in-degree papers over time, we divide the temporal citation network into three snapshots, i.e., 2001–2006, 2001–2012 and 2001–2018. Each snapshot retains only the nodes and edges that exist in the period. Table 3 shows the top 5 high in-degree papers in each snapshot.

From Table 3, we can see that a majority of the high in-degree papers are published in JASA, JRSS-B, and AoS. In the period of 2001–2006, the top 5 in-degree papers reveal three important research topics in the field of statistics, i.e., "Bayesian Analysis", "Biostatistics" and "Hypothesis Testing". In the period of 2001–2012, many papers on variable selection appeared in the top 5 high in-degree papers. Bayesian analysis is also a very popular research topic. As we can see from the last snapshot, variable selection becomes one of the "hottest" topics.

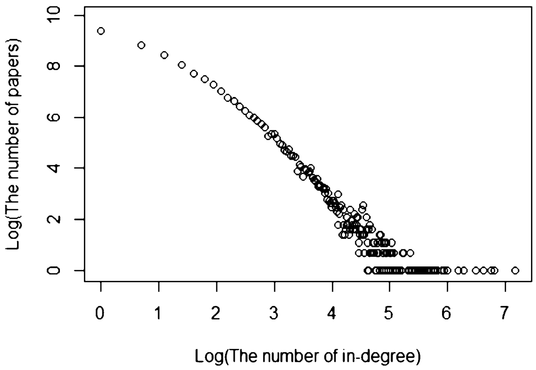
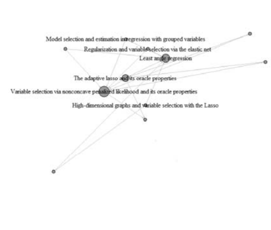
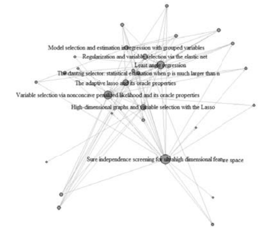


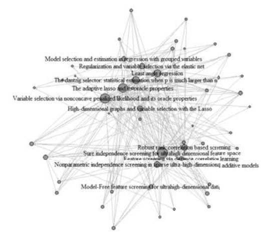
Figure 2. log-log in-degree distribution.



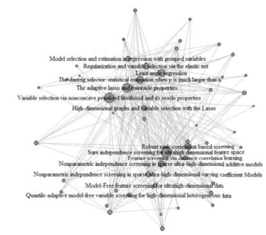
(a) Network from 2001 to 2006



(b) Network from 2001 to 2008



(c) Network from 2001 to 2012



(d) Network from 2001 to 2018

Figure 3. Visualization of networks composed of papers related to variable selection at different snapshots. We can see that the citation network has a community structure and that the communities are constantly evolving.

2.3 Visualization of the community structure

To initially explore the community structure of the citation network, we visualize some subnetworks. Papers related to "variable selection" are selected to explore the community structure and dynamic nature of the citation network. Figure 3 shows the visualization of networks at different snapshots, i.e., 2001–2006, 2001–2008, 2001–2012, and 2001–2018.

Figure 3 indicates that the citation network has a community structure and that the communities are constantly evolving. The variable selection community continues to grow, which is reflected in the increasing number of papers in this community. In addition, the variable selection community gradually splits into two communities, namely the variable selection community and the ultrahigh-dimensional community. The former mainly studies traditional variable selection methods, such as lasso and elastic net, while the latter focuses on variable selection under ultrahigh-dimensional data in the era of big data, such as screening-based methods.

Table 3. Top 5 high in-degree papers in three snapshots.

Period	Title	Journal	Year	In-degree
2001–2006 ($t = 1 \sim 6$)	Empirical Bayes analysis of a microarray experiment	JASA	2001	98
	Bayesian measures of model complexity and fit	JRSS-B	2002	95
	Comparison of discrimination methods for the classification of tumors using gene expression data	JASA	2002	84
	Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments	SS	2002	74
	A direct approach to false discovery rates	JRSS-B	2002	74
2001–2012 ($t = 1 \sim 12$)	Bayesian measures of model complexity and fit	JRSS-B	2002	443
	Variable selection via nonconcave penalized likelihood and its oracle properties	JASA	2001	407
	Least angle regression	AoS	2004	354
	The adaptive lasso and its oracle properties	JASA	2006	294
	Empirical Bayes analysis of a microarray experiment	JASA	2001	249
2001–2018 ($t = 1 \sim 18$)	Variable selection via nonconcave penalized likelihood and its oracle properties	JASA	2001	1254
	The adaptive lasso and its oracle properties	JASA	2006	885
	Bayesian measures of model complexity and fit	JRSS-B	2002	837
	Least angle regression	AoS	2004	731
	Regularization and variable selection via the elastic net	JRSS-B	2005	635

*The journals are abbreviated as follows: *Annals of Statistics* (AoS), *Journal of American Statistical Association* (JASA), *Journal of Royal Statistical Society (Series B)* (JRSS-B), *Statistica Sinica* (SS)

3. METHODOLOGY

It is well known that CANDECOMP/PARAFAC (CP) decomposition [30] and Tucker decomposition [48] are two prevailing tensor decomposition methods. Tucker decomposition is commonly implemented through high-order SVD (HOSVD) [13] or high-order orthogonal iteration (HOOI) [14].

The Tucker decomposition is a form of higher-order principal component analysis (PCA) [33]. The main difference between CP decomposition and Tucker decomposition is that there is a core tensor by using Tucker decomposition, which can be compared to the principal component factor in PCA and can reflect most of the properties of the original tensor. Each factor matrices in Tucker decomposition corresponds only to a transformation (multiplication) operation of a different mode. CP decomposes a tensor into a sum of component rank-one tensors, which can expressed as a multilinear product with a diagonal core [10]. The factor matrices in CP decomposition contains more information of the original tensor.

Typically, the CP decomposition is used for decomposing data into easy to interpret components, such as rank-one tensors. And the Tucker decomposition is more often used to reduce data into a small size core tensor. Tucker decomposition is widely used in computer vision and signal processing. Our method is a generalization of D-SCORE to use factor matrices for clustering. The main purpose is to use the SCORE normalization and intersection-with-attachment approach on the factor matrix to correct for degree heterogeneity. Therefore, CP decomposition with ALS is used in the TD-SCORE method.

Define $\mathbb{X} \in \mathbb{R}^{P_1\times P_2\times P_3}$ as a third-order tensor, where P_1, P_2, and P_3 are integers. Column (mode-1), row (mode-2) and tube (mode-3) fibers exist in a third-order tensor, and fibers are always assumed to be column vectors. Figure 4 shows the fibers of a third-order tensor. The tensor can be unfolded into a matrix by the mode, which arranges the mode-q fibers to be the columns of the matrix. Denote the mode-q matricization of a tensor $\mathbb{X}$ as $X_{(q)}$, $q = 1, 2, 3$. For example, the tensor $\mathbb{X} \in \mathbb{R}^{P_1\times P_2\times P_3}$ can be mode-1 unfolded into a matrix $X_{(1)} \in \mathbb{R}^{P_1\times(P_2P_3)}$. Slices refer to two-dimensional sections of a tensor defined by fixing all but two indices. For third-order tensors, there are horizontal, lateral and frontal slides, as shown in Figure 4.

Consider a temporal citation network, which can be represented by a third-order adjacency tensor $\mathbb{A} \in \mathbb{R}^{N\times N\times T}$, where N is the number of nodes, and T is the time span of the temporal network. Let $A^{(t)} = (a_{ijt}) \in \mathbb{R}^{N\times N}, t = 1, \dots, T$ be the adjacency matrix with elements denoting the indicators of whether the jth paper is cited by the ith paper in snapshot t. If the ith paper cites the jth paper at time t and before, then $a_{ijt} = 1$; otherwise, $a_{ijt} = 0$. In order to match each snapshot, we set each snapshot $A^{(t)}$ to be $N \times N$- dimensional matrix. It is possible that some snapshots contain nodes appeared in the future, so we don't consider the edges appearing after snapshot t in $A(t)$. Denote the collection of papers that appeared by time t as S_t, where $|S_t| = N_t$, and $N_1 \le N_2 \le \cdots \le N_T = N$. Specifically, we can see that $S_t \backslash S_{t-1}$ $(t = 2, \dots, T)$ is the collection of papers published in the period from $t-1$ to t. It is observed that $A^{(t)}$ $(t = 1, \dots, T)$ are the frontal slides of tensor $\mathbb{A}$ corresponding to the temporal citation network.

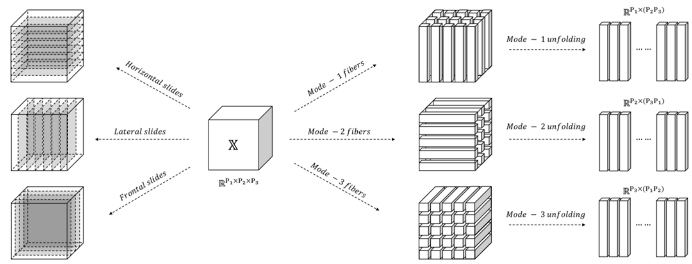


Figure 4. The fibers and slides of a third-order tensor: $\mathbb{X} \in \mathbb{R}^{P_1\times P_2\times P_3}$. Column (mode-1), row (mode-2) and tube (mode-3) fibers and horizontal, lateral and frontal slides exist in a third-order tensor. The tensor can be unfolded into a matrix by the mode.

Remark 1. *For convenience, we omit the superscript of tensor $\mathbb{A}$ in time T. In fact, the third-order adjacency tensor $\mathbb{A} \in \mathbb{R}^{N\times N\times T}$ should be written as $\mathbb{A}^{(T)}$. In addition, for each time t, the tensor can be written as $\mathbb{A}^{(t)} \in \mathbb{R}^{N_t\times N_t\times t}, t = 1, \dots, T$.*

3.1 CP decomposition with ALS

We use the alternating least squares (ALS) method [19, 7, 33] to compute the CP decomposition. CP decomposition with ALS is an approximate algorithm. The goal of the algorithm is to minimize the modulus of the difference between the true value $\mathbb{A}$ and estimation $\hat{\mathbb{A}}$; i.e.,

$$\min_{\hat{\mathbb{A}}} \|\mathbb{A} - \hat{\mathbb{A}}\|_F, \tag{1}$$

where $\hat{\mathbb{A}}$ is defined as

$$\hat{\mathbb{A}} = \sum_{r=1}^{R} \lambda_r \mathrm{m}_r \circ \mathrm{p}_r \circ \mathrm{q}_r = [\![\lambda; \mathrm{M}, \mathrm{P}, \mathrm{Q}]\!] \tag{2}$$

where $\lambda_r \in \mathbb{R}, \mathrm{m}_r \in \mathbb{R}^N$, $\mathrm{p}_r \in \mathbb{R}^N$, $\mathrm{q}_r \in \mathbb{R}^T$, for $r = 1, \dots, R$, where R is the number of components. Symbol $\circ$ represents the vector outer product, which means that each element of the tensor is the product of the corresponding vector elements. In addition, $\mathrm{M} = [\mathrm{m}_1, \mathrm{m}_2, \cdots, \mathrm{m}_R] \in \mathbb{R}^{N\times R}$, $\mathrm{P} = [\mathrm{p}_1, \mathrm{p}_2, \cdots, \mathrm{p}_R] \in \mathbb{R}^{N\times R}$, and $\mathrm{Q} = [\mathrm{q}_1, \mathrm{q}_2, \cdots, \mathrm{q}_R] \in \mathbb{R}^{T\times R}$ are the factor matrices, which are related to the combination of the vectors from the rank-one components. Furthermore, $\lambda = (\lambda_1, \dots, \lambda_R)^\top \in \mathbb{R}^R$ is the vector that absorbs the weights from the factor matrices. According to the matricization of a tensor, equation (2) can be rewritten as

$$
\begin{aligned}
\mathrm{A}_{(1)} &\approx \mathrm{M}(\mathrm{Q} \odot \mathrm{P})^{\top}, \\
\mathrm{A}_{(2)} &\approx \mathrm{P}(\mathrm{Q} \odot \mathrm{M})^{\top}, \\
\mathrm{A}_{(3)} &\approx \mathrm{Q}(\mathrm{P} \odot \mathrm{M})^{\top}.
\end{aligned} \tag{3}
$$

where the symbol $\odot$ denotes the Khatri-Rao product [47], which is the "matching columnwise" Kronecker product.

Assuming the number of components R is fixed, we use the ALS method to compute the CP decomposition, which is one of the most popularly used algorithms. This method simplifies the problem of computing the CP decomposition approximate solution to a linear least square problem. We need only to compute the pseudoinverse of an $R \times R$ matrix instead of an $NT \times R$ matrix. The following is the process of CP decomposition with ALS.

Algorithm 1 (*CP decomposition with ALS procedure*). Input: an adjacency tensor $\mathbb{A}$ and the number of components R.

Step 1: Initialization: initialize M, P and Q randomly.

Step 2: Iteration: fix P and Q to solve M, then fix M and Q to solve P, then fix M and P to solve Q, and continue to repeat the entire procedure until some convergence criterion is satisfied. The convergence criterion we set is that there is little or no improvement in the loss function or the algorithm reaches the maximum number of iterations. Denote M, P, and Q after convergence as $\hat{\mathrm{M}}$, $\hat{\mathrm{P}}$, and $\hat{\mathrm{Q}}$, respectively.

Step 3: Normalization: normalize the columns of $\hat{\mathrm{M}}$ to obtain M. We obtain the decomposition result M, P and Q in the end.

For each iteration, because we have fixed all but one matrix, the problem can be reduced to a linear least squares problem. For example, we suppose that P and Q are fixed. Then, we can rewrite the above minimization problem 1 in matrix form as $\min \left\| \mathrm{A}_{(1)} - \hat{\mathrm{M}}(\mathrm{P} \odot \mathrm{Q})^{\top} \right\|_F$, where $\hat{\mathrm{M}} = \mathrm{M} \cdot \mathrm{diag}(\lambda)$. The optimal solution is $\hat{\mathrm{M}} = \mathrm{A}_{(1)} \left[(\mathrm{P} \odot \mathrm{Q})^{\mathrm{T}} \right]^{\dagger}$, where $\dagger$ represents the conjugate transpose of the matrix. According to the properties proposed by [49, 47], the Khatri-Rao product pseudoinverse has a special form. Finally, the optimal solution can be rewritten as $\hat{\mathrm{M}} = \mathrm{A}_{(1)}(\mathrm{Q} \odot \mathrm{P}) \left(\mathrm{Q}^{\top}\mathrm{Q}\mathrm{P}^{\top}\mathrm{P} \right)^{\dagger}$.

3.2 Tensor directed-spectral clustering on ratios of eigenvectors

For dynamic community detection of directed temporal networks, we proposed a method: Tensor-based Directed Spectral Clustering On Ratios of Eigenvectors (TD-SCORE). Let $\mathbb{A}^{(t)} \in \mathbb{R}^{N_t \times N_t \times t}$ be the third-order tensor corresponding to each snapshot t. Assume there are K_t communities in snapshot t, $t = 1, \ldots, T$. TD-SCORE applies CP decomposition with ALS on the adjacent tensor $\mathbb{A}^{(t)}$ to obtain the Mode-1 singular matrices $\hat{\mathrm{M}}^{(t)} = (\hat{\mathrm{m}}_1^{(t)}, \ldots, \hat{\mathrm{m}}_R^{(t)}) \in \mathbb{R}^{N_t \times R}$ and Mode-2 singular matrices $\hat{\mathrm{P}}^{(t)} = (\hat{\mathrm{p}}_1^{(t)}, \ldots, \hat{\mathrm{p}}_R^{(t)}) \in \mathbb{R}^{N_t \times R}$.

Denote $\mathcal{N}^{(t)}$ as the collection of nodes corresponding to $\mathbb{A}^{(t)}$ at time t. Let $\mathcal{N}_1^{(t)} = \{i : \left|\hat{\mathrm{m}}_{1i}^{(t)}\right| \neq 0\}$ be the collection of nodes at time t, where $1 \leq i \leq N_t$, $1 \leq t \leq T$, and $\hat{\mathrm{m}}_1^{(t)} = (\hat{\mathrm{m}}_{11}^{(t)}, \hat{\mathrm{m}}_{12}^{(t)}, \ldots, \hat{\mathrm{m}}_{1N_t}^{(t)})^{\top} \in \mathbb{R}^{N_t}$. Let $\mathcal{N}_2^{(t)} = \{i : \left|\hat{\mathrm{p}}_{1i}^{(t)}\right| \neq 0\}$ be the collection of nodes at time t, where $1 \leq i \leq N_t$, $1 \leq t \leq T$, $\hat{\mathrm{p}}_1^{(t)} = (\hat{\mathrm{p}}_{11}^{(t)}, \hat{\mathrm{p}}_{12}^{(t)}, \ldots, \hat{\mathrm{p}}_{1N_t}^{(t)})^{\top} \in \mathbb{R}^{N_t}$. We define the sets $\mathcal{N}_1^{(t)}$ and $\mathcal{N}_2^{(t)}$ are based on the intersection-with-attachment approach proposed by [24]. The nodes in $\mathcal{N}_1^{(t)} \cap \mathcal{N}_2^{(t)}$ are much more separable than the other nodes. The nodes in the intersection set extract the core of the network, and other nodes act as the periphery and mislead the clustering result. It is worth noting that there are very few nodes that belong to neither. Thus the intersection-with-attachment technique works by taking the clustering results for the intersection and attaching the periphery nodes to these clusters using the edges in the original network, and hence yields better performance [54].

Then, we have all nodes of temporal networks at time t split into four disjoint subsets,

$$
\mathcal{N}^{(t)} = \left(\mathcal{N}_1^{(t)} \cap \mathcal{N}_2^{(t)}\right) \cup \left(\mathcal{N}_1^{(t)} \backslash \mathcal{N}_2^{(t)}\right) \cup \left(\mathcal{N}_2^{(t)} \backslash \mathcal{N}_1^{(t)}\right) \cup \left(\mathcal{N}^{(t)} \backslash \left(\mathcal{N}_1^{(t)} \cup \mathcal{N}_2^{(t)}\right)\right)
$$

For each adjacency tensor $\mathbb{A}^{(t)} \in \mathbb{R}^{N_t \times N_t \times t}$, we define two $N_t \times (K_t - 1)$ matrices $\hat{R}_m^{(t)}$ and $\hat{R}_p^{(t)}$ as follows:

$$
\hat{R}_m^{(t)}(i,k) = \begin{cases} \operatorname{sgn}\left(\hat{\mathrm{m}}_{k+1}^{(t)}(i) / \hat{\mathrm{m}}_1^{(t)}(i)\right) \cdot \min\left\{ \left| \frac{\hat{\mathrm{m}}_{k+1}^{(t)}(i)}{\hat{\mathrm{m}}_1^{(t)}(i)} \right|, \log(N_t) \right\}, & i \in \mathcal{N}_1^{(t)}, \\ 0, & i \notin \mathcal{N}_1^{(t)}, \end{cases}
$$

$$
\hat{R}_p^{(t)}(i,k) = \begin{cases} \operatorname{sgn}\left(\hat{\mathrm{p}}_{k+1}^{(t)}(i) / \hat{\mathrm{p}}_1^{(t)}(i)\right) \cdot \min\left\{ \left| \frac{\hat{\mathrm{p}}_{k+1}^{(t)}(i)}{\hat{\mathrm{p}}_1^{(t)}(i)} \right|, \log(N_t) \right\}, & i \in \mathcal{N}_2^{(t)}, \\ 0, & i \notin \mathcal{N}_2^{(t)}, \end{cases}
$$

where $1 \leq k \leq K_t - 1$ and $\operatorname{sgn}(x)$ stands for the sign function satisfying $\operatorname{sgn}(x) = 1$ when $x > 0$, $\operatorname{sgn}(0) = 0$, and $\operatorname{sgn}(x) = -1$ when $x < 0$. Then, we cluster nodes in the above four subsets separately using $\hat{R}_m^{(t)}$ and $\hat{R}_p^{(t)}$, which is called the SCORE normalization [27]. This treatment can correct for degree heterogeneity.

Algorithm 2 (*Tensor directed-spectral clustering on ratios of eigenvector procedure*). Input: the matrices $\hat{R}_m^{(t)}$ and $\hat{R}_p^{(t)}$, the number of communities K_t, and the node sets $\mathbb{N}_1^{(t)}$ and $\mathbb{N}_2^{(t)}$.

Step 1: Restrict the rows of $\hat{R}_m^{(t)}$ and $\hat{R}_p^{(t)}$ to the set $\mathbb{N}_1^{(t)} \cap \mathbb{N}_2^{(t)}$ and obtain two new matrices $\tilde{R}_m^{(t)} \in \mathbb{R}^{s \times (K_t - 1)}$

and $\tilde{R}_p^{(t)} \in \mathbb{R}^{s \times (K_t-1)}$, where $s = \left|\mathbb{N}_1^{(t)} \cap \mathbb{N}_2^{(t)}\right|$ is the size of $\mathbb{N}_1^{(t)} \cap \mathbb{N}_2^{(t)}$. Assume there are K_t communities, and apply k-means to the columns of $B = (\tilde{R}_m^{(t)}, \tilde{R}_p^{(t)}) \in \mathbb{R}^{s \times 2(K_t-1)}$, so that nodes in $\mathbb{N}_1^{(t)} \cap \mathbb{N}_2^{(t)}$ are divided into K_t communities.

Step 2: Compute the mean of the row vectors of $\tilde{R}_m^{(t)}$ in each community, and take them as the community center. For a node i in $\mathbb{N}_1^{(t)} \backslash \mathbb{N}_2^{(t)}$, classify it to one of the communities whose community center is the closest to the ith row of $\hat{R}_m^{(t)}$.

Step 3: Compute the mean of the row vectors of $\tilde{R}_p^{(t)}$ in each community, and take them as the community center. For a node i in $\mathbb{N}_2^{(t)} \backslash \mathbb{N}_1^{(t)}$, classify it to one of the communities whose community center is the closest to the ith row of $\hat{R}_p^{(t)}$.

Step 4: For a node i in $\mathbb{N}_t \backslash \left(\mathbb{N}_1^{(t)} \cup \mathbb{N}_2^{(t)}\right)$, compute the numbers of edges (ignore directions) it has with nodes in each community and then classify it to the community with which it has the largest number of edges.

4. DYNAMIC COMMUNITY DETECTION AND COMMUNITY EVOLUTION

4.1 Dynamic community detection

Many real-world networks show a core-periphery structure, and citation networks are more likely to exhibit such phenomena. Therefore, this paper focuses on the core subnetwork of the citation network and extracts the core subnetwork according to the method proposed by [52]. For one iteration, we delete nodes with degrees less than 10 and their connected edges. Then, the above process is repeated until the network no longer changes. In this way, a core subnetwork composed of 4,101 papers is constructed. The core subnetwork has 48,409 edges, and its density is 0.288%.

We apply the TD-SCORE method to detect dynamic network communities in the citation network from 2010 to 2018. For real data analysis, how to choose the tensor rank is very important. [33] list some typical ranks for third-order tensors. In this paper, we use a simple and commonly used method, that is, we try to set the number of components from 1 to 20 respectively. According to the computing performance and actual community detection results, We finally set the number of components in our adjacency tensor decomposition to 7. As the development of the research community in citation networks is relatively slow, to accumulate a certain amount of information, we detect the community structure of citation networks in time $t = 10, \ldots, 18$. Then, to use the TD-SCORE, we need to determine the number of communities K, which is unknown. Assume that there are 5, 5, 6, 6, 6, 6, 7, 7, and 7 communities in each snapshot, respectively. We then apply the TD-SCORE algorithm 9 separate times for $t = 10, \ldots, 18$ to obtain the community structure in our temporal citation network. We determine

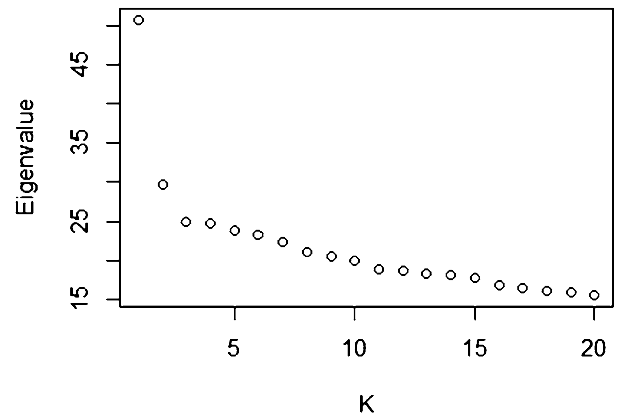


Figure 5. Scree plot at $T = 18$.

that the above K is the most reasonable choice in the following way. Take $t = 18$ as an example. Firstly, we use the scree plot to determine the range of K as 3 to 8. And we believe that the number of communities continues to grow slowly because the field of statistical research continues to expand. Hence, we also set the K of the current period to be no less than the previous period. Secondly, for each K we implement TD-SCORE and inspect the effect by checking the key papers and keywords of each communities. The criteria is that papers in the same community should be in the same or similar research fields and main research fields vary between communities.

We use 2018 as an example to explore the research interests of various communities. For each of the communities in 2018, we count the frequency of keywords and select the top 5 keywords. Table 4 shows the top 5 keywords and the number of nodes in each community, from which we can comprehend the research topics of each community. For example, the main research interests of the testing community are false discovery rate, multiple testing, multiple comparisons and so on. It is worth noting that that we initially set up 7 communities in 2018, but we find that two communities are very similar, both mainly involved in the field of dimensional reduction, so we merged them. Therefore, only 6 communities appear in Table 4. The details of community evolution will be discussed in the next section.

4.2 Community evolution

Figure 6 shows the evolution of communities in the citation core subnetwork between 2010 and 2018. There are many rectangles of different colors in the figure, among which different colors represent different years, from blue to purple representing 2010 to 2018, respectively. The length of each rectangle in the figure represents the number of nodes in the community. The larger the length of the rectangle, the larger the scale of the community. The rectangle in the uppermost area of each year represents the set of nodes that appeared in and after that year, namely, the “new” area.

Table 4. Top 5 keywords and the number of nodes in each community in 2018.

Community	Keywords	Nodes
Variable Selection	Variable selection, Lasso, Sparsity, Model selection, Oracle property	2089
Functional Data Analysis	Functional data, Functional principal components, Functional linear regression, Longitudinal data, Smoothing	555
Testing	False discovery rate, Multiple testing, Multiple comparisons, Familywise error rate, P-value	173
Bayesian Analysis	Markov chain Monte Carlo, Bayesian inference, Bayesian nonparametrics, Dirichlet process, Kriging	600
Dimension Reduction	Dimension reduction, Sliced inverse regression, Central subspace, Sufficient dimension reduction, Single-index model	250
Covariance Matrix	Sparsity, Covariance matrix, High-dimensional data, Gaussian graphical model, Precision matrix	434

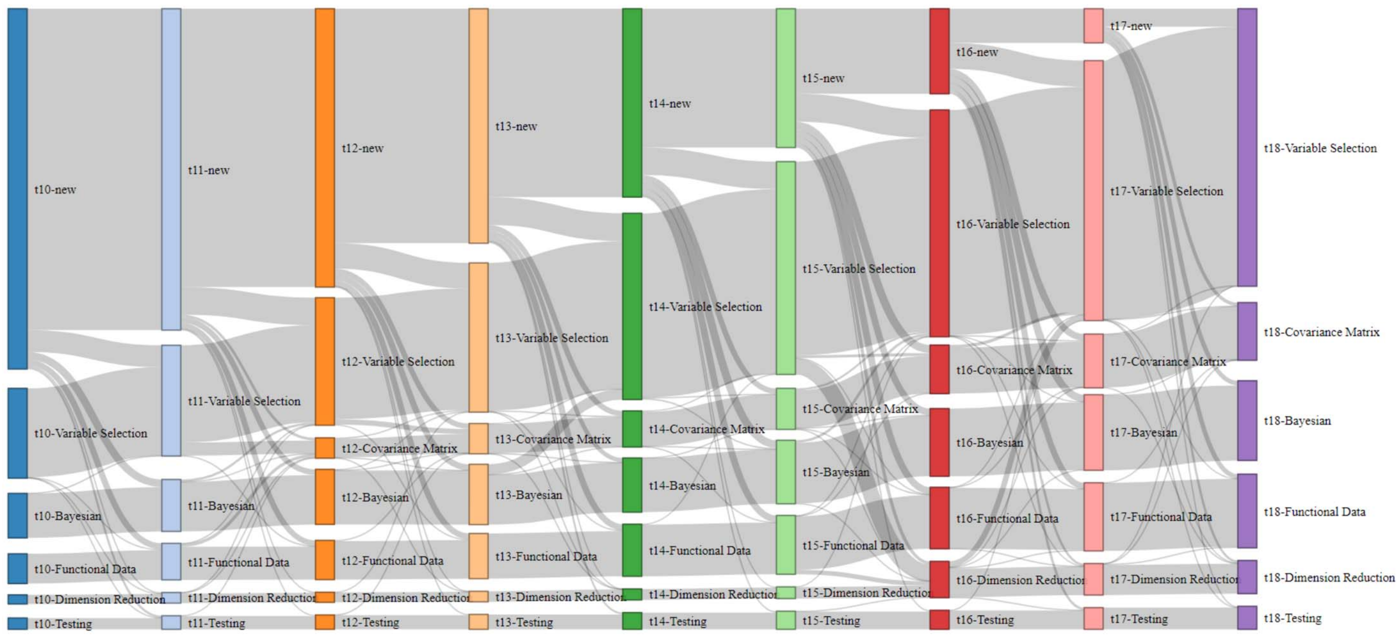


Figure 6. The evolution of communities in the citation core subnetwork between 2010 and 2018. The rectangles represent different communities, and the colors represent different years. The length of each rectangle indicates the number of nodes in each community. The "new" area represents the set of nodes that appeared in and after that year.

Every year, the nodes that appear in this year flow from the "new" area to other communities, and the width of the flow represents the number of nodes that flow into the community.

In general, the number of nodes in the core subnetwork of the statistical citation network increased from 1,389 in 2010 to 4,101 in 2018. From Figure 6, we can see that nearly half the new nodes flow into the variable selection community each year, which means that variable selection is the main research in the field of statistics. Denote the number of nodes in 2018 as the size of each community. We can see that the variable selection community has the largest community size, followed by the Bayesian analysis community and functional data analysis community. The size of the testing community and dimension reduction community is relatively small. In addition, we notice that only a small number of papers change communities every year, which means that our method is stable.

Denote $R = (\prod_{i=1}^{t}(1+p_i))^{1/t} - 1$ as the growth rate of each community, where p_i is the growth rate at time i. Figure 7 indicates the growth rate of each community. For newly emerging communities, we calculate growth rates from the second year of their appearance. The covariance matrix community is the fastest-growing community among all communities, with a growth rate of 19.0%. The dimension reduction community is the second fastest-growing community because of the artificial merging of communities, so this growth rate needs to be treated with more caution. The variable selection community is the third fastest-growing community, whose growth rate is 15.1%. The functional data analysis,

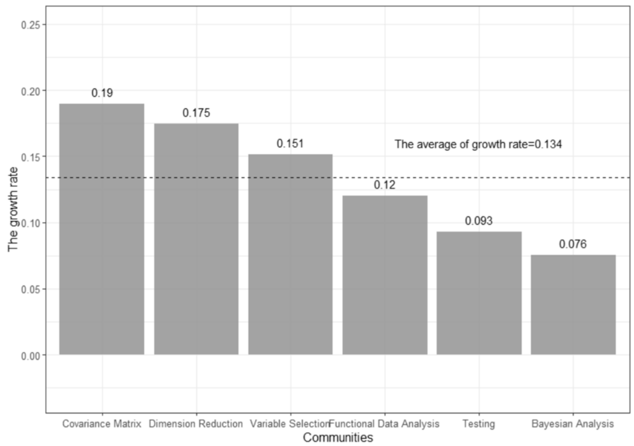


Figure 7. The growth rate of each community.

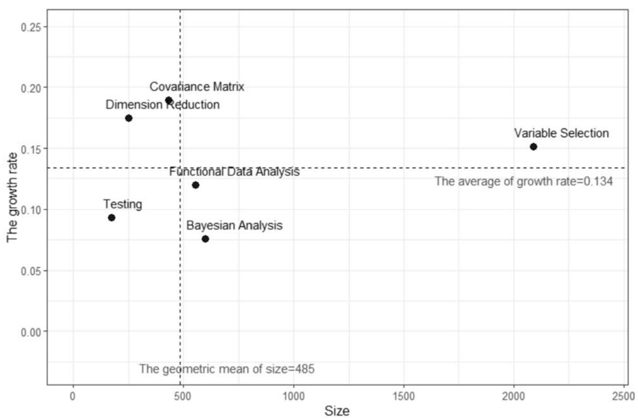


Figure 8. Scatterplot of the size and growth rate of communities.

testing and Bayesian analysis communities are growing, increasing by 12.0%, 9.3% and 7.6%, respectively. However, the above three communities are all lower than the growth rate of the nodes in the core subnetwork.

Combining the size and growth rate of each community, Figure 8 shows the scatterplot of the size and the growth rate of communities. We can divide the communities into three classifications by the size and growth rate as follows:

- Community with high size and high growth rate: variable selection community.
- Community with high size and low growth rate: Bayesian analysis community and functional data analysis community
- Community with low size and high growth rate: covariance matrix community and dimension reduction community.
- Community with low size and low growth rate: testing community.

4.3 Variable selection community evolution

The variable selection community mainly studies high-dimensional data analysis, which is widely used in many fields, such as economics, engineering, genomics, and medical science. High-dimensional variable selection is the frontier field of statistical development. Different from conventional studies, the number of parameters p is larger than the number of observations n in high-dimensional analysis, such as high-frequency financial data, proteomics data, high-resolution images and many others. In the period of 2001–2018, many studies were dedicated to exploring the performance of different variable selection methods. Table 5 shows the top 10 high in-degree papers in the variable selection community. These papers mainly discuss theoretical questions about variable selection methods, such as dimensionality limits and optimality characterization. The above papers are mainstream papers in the field of variable selection.

Table 6 shows the top 10 high growth rate papers in the variable selection community. We note that the current development direction of the field of variable selection includes but is not limited to penalized likelihood estimation methods (e.g., lasso, SCAD, MCP, elastic net), ultrahigh-dimensional variable selection (e.g., independence screening, feature screening), penalized least squares sampling properties (e.g., persistence, consistency, oracle property) and many others. Especially in the past ten years, because of the proliferation of data dimensions, ultrahigh-dimensional

Table 5. Top 10 high in-degree papers in variable selection community.

ID	Title	Journal	Year	In-degree
1	Variable selection via nonconcave penalized likelihood and its oracle properties	JASA	2001	1254
2	The adaptive lasso and its oracle properties	JASA	2006	885
3	Least angle regression	AoS	2004	731
4	Regularization and variable selection via the elastic net	JRSS-B	2005	635
5	Model selection and estimation in regression with grouped variables	JRSS-B	2006	522
6	High-dimensional graphs and variable selection with the Lasso	AoS	2006	454
7	The Dantzig selector: Statistical estimation when p is much larger than n	AoS	2007	379
8	Regularization paths for generalized linear models via coordinate descent	JSS	2010	372
9	Sure independence screening for ultrahigh dimensional feature space	JRSS-B	2008	366
10	Nearly unbiased variable selection under minimax concave penalty	AoS	2010	324

Table 6. Top 10 high growth rate papers in variable selection community.

ID	Title	Journal	Year	Growth rate
1	A selective overview of variable selection in high dimensional feature space	SS	2010	1.684
2	Regularization paths for generalized linear models via coordinate descent	JSS	2010	1.679
3	Coordinate descent algorithms for nonconvex penalized regression, with applications to biological feature selection	AoAS	2011	1.454
4	Nearly unbiased variable selection under minimax concave penalty	AoS	2010	1.338
5	New efficient estimation and variable selection methods for semiparametric varying-coefficient partially linear models	AoS	2011	1.254
6	Nonparametric independence screening in sparse ultra-high-dimensional additive models	JASA	2011	1.197
7	Variable selection in nonparametric additive models	AoS	2010	1.194
8	Stability selection	JRSS-B	2010	1.012
9	Bayes and empirical-bayes multiplicity adjustment in the variable-selection problem	AoS	2010	0.974
10	Measuring and testing dependence by correlation of distances	AoS	2007	0.848

*The journal is abbreviated as follows: *Journal of Statistical Software* (JSS), *Annals of Applied Statistics* (AoAS)

Table 7. Top 10 high in-degree papers in functional data analysis community.

ID	Title	Journal	Year	In-degree
1	Functional data analysis for sparse longitudinal data	JASA	2005	243
2	Selecting the number of knots for penalized splines	JCGS	2002	154
3	Generalized functional linear models	AoS	2005	149
4	Sequential Monte Carlo samplers	JRSS-B	2006	146
5	Varying-coefficient models and basis function approximations for the analysis of repeated measurements	Biometrika	2002	140
6	Functional linear regression analysis for longitudinal data	AoS	2005	139
7	Prediction in functional linear regression	AoS	2006	127
8	Spline estimators for the functional linear model	SS	2003	126
9	Methodology and convergence rates for functional linear regression	AoS	2007	112
10	Nonparametric mixed effects models for unequally sampled noisy curves	Biometrics	2001	110

*The journal is abbreviated as follows: *Journal of Computational and Graphical Statistics* (JCGS)

analysis has attracted the attention of many scholars. There are also many theoretical and applied papers based on independent screening methods.

The first paper [15] in Table 6 is a paper review, which systematically summarizes the development of theories, methods and implementations of high-dimensional variable selection. The main conclusions drawn in this chapter are consistent with this. Through the above analysis, we notice that high-dimensional variable selection is still the most important and frequent research field in statistics and that its development can help us better handle and solve various challenges in scientific research. Moreover, combining with other disciplines can further advance scientific development.

4.4 Functional data analysis community evolution

Functional data analysis is a major and incumbent research field in statistics, and its research objects are functions rather than numbers or vectors. Because the key idea of functional data analysis is to hope that the object of its research is a smoothing curve, the raw data require a preliminary treatment by basis representation for dimension reduction or noise removal [12]. The calcium curves in experimental cardiology and magnetic field intensity recorded by geomagnetometer can be expressed as functional data.

Table 7 and Table 8 show the top 10 high in-degree and high growth rate papers in the functional data analysis community, respectively. From these two tables, we can see that the current development direction of the field of functional data analysis includes but is not limited to functional regression (e.g., scalar response, functional response, functional nonparametric regression), functional classification (e.g., functional classification, unsupervised functional classification), functional dimension reduction (e.g., functional principal components, partial least squares, variable selection) and functional bootstrap methods. In addition, we note that functional additive mixed models are a current hot research direction.

The conclusions drawn in this chapter are consistent with the results of the first review paper [12] in Table 8. Functional data analysis can analyze the essential characteristics of complex data and obtain a more reasonable and intuitive data interpretation. The development of functional data analysis is vital to statistics and even science.

Table 8. Top 10 high growth rate papers in functional data analysis community.

ID	Title	Journal	Year	Growth rate
1	A partial overview of the theory of statistics with functional data	JSPI	2014	2.622
2	Functional Additive Mixed Models	JCGS	2015	1.756
3	Functional Regression	ARSIA	2015	1.414
4	Single and multiple index functional regression models with nonparametric link	AoS	2011	1.357
5	A reproducing kernel Hilbert space approach to functional linear regression	AoS	2010	1.108
6	Nonparametric time series prediction: A semi-functional partial linear modeling	JMVA	2008	1.029
7	Weakly dependent functional data	AoS	2010	0.865
8	k-Nearest Neighbour method in functional nonparametric regression	JNS	2009	0.857
9	Flexible Bayesian quantile regression for independent and clustered data	Biostatistics	2010	0.838
10	Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models	JRSS-B	2011	0.820

*The journal is abbreviated as follows: *Journal of Statistical Planning and Inference* (JSPI), *Annual Review of Statistics and Its Application* (ARSIA), *Journal of Multivariate Analysis* (JMVA), *Journal of Nonparametric Statistics* (JNS)

4.5 Comparison of community detection methods

In this part, we compare the dynamic community detection results of D-SCORE by [24]. We mainly compare the algorithms from the actual effect.

D-SCORE is a directed spectral clustering method based on ratios of eigenvectors. Firstly, the first K left and right singular vectors of the adjacency matrix are obtained by singular value decomposition. Secondly, the entry-wise ratios between the first leading vector and each of the other leading vectors is calculated for clustering, which can remove the effect of degree heterogeneity effectively. Finally, the nodes into four disjoint subsets are split and the k-means algorithm is used to detect the communities. We apply the D-SCORE method to detect dynamic network communities in the citation network from 2010 to 2018. By using the scree plot and the actual effect, we assume that there are 3, 3, 3, 4, 4, 4, 4, 5, and 5 communities in each snapshot, respectively.

It should be noted that in order to achieve the most reasonable actual performance, the setting on number of communities may be different for the TD-SCORE and SCORE. From the process of choosing the optimal K, we mention that we first determine a range of K, and then traverse each K to obtain the most reasonable actual performance. Hence, the number of communities used in D-SCORE is different from the TD-SCORE. Table 9 shows the community detection results of the D-SCORE. We enumerate the communities we detected, and the number after the community represents the number of papers in that community. The number after the keywords is the frequency.

By comparing the community detection results of D-SCORE and TD-SCORE, we find that the results of D-SCORE mainly have two disadvantages. The first point is that the appearance time of some communities does not match the facts. we notice that the Bayesian community and the Dimension Reduction community emerge in 2017 by D-SCORE, which does not match the facts. In the community detection results by TD-SCORE, these two communities have always existed from 2010 to 2018. The second point is that some communities detected by D-SCORE are involved in two fields, and the algorithms are not well separated, such as the Testing & Covariance Matrix communities from 2016 to 2018. By calculating the keywords in Testing & Covariance Matrix communities, we notice that "false discovery rate" and "multiple testing" are mainly "Testing" fields, while "sparsity" and "covariance matrix" are mainly related to "Covariance Matrix" fields.

5. CONCLUSION

In this study, we analyze the temporal citation network constructed by papers collected from 44 famous statistical journals between 2001 and 2018. We proposed an approach named Tensor-based Directed Spectral Clustering On Ratios of Eigenvectors (TD-SCORE) to detect the community structure of a temporal citation network. We initially explore the characteristics of the temporal network via in-degree distribution and visualization of different snapshots. We find that both the community structure and the key nodes change over time. Then, we apply the TD-SCORE method to the core network of our temporal citation network. Seven "hot" communities are identified, including variable selection, Bayesian analysis, functional data analysis communities and many others. Through the size and growth rate, we study the evolution of each community. Variable selection has been a long-established and well-developed community in recent years, among which ultrahigh-dimensional data analysis is an emerging hot research issue. The Bayesian analysis community is an incumbent community but is relatively slow.

For the dynamic community detection analysis, the stable of the methods is very important. Our results are stable

Table 9. The results of dynamic community detection by D-SCORE.

Period	**Community**	**Keywords**
2010	Functional Data(652)	functional data analysis(64), mcmc(51)
	Variable Selection(564)	variable selection(100), lasso(98)
	Testing(171)	false discovery rate(44), multiple testing(36)
2011	Functional Data(837)	functional data analysis(77), mcmc(59)
	Variable Selection(656)	lasso(125), variable selection(124)
	Testing(187)	false discovery rate(53), multiple testing(42)
2012	Variable Selection(1370)	variable selection(166), lasso(134)
	Covariance Matrix & Functional Data(333)	sparsity(29), mcmc(23), functional data analysis(21)
	Testing(303)	false discovery rate(60), multiple testing(49)
2013	Variable Selection(911)	variable selection(197), lasso(164)
	Functional Data(903)	functional data analysis(94), longitudinal data(67)
	Covariance Matrix(341)	sparsity(29), covariance matrix(26)
	Testing(181)	false discovery rate(59), multiple testing(46)
2014	Variable Selection(1186)	variable selection(258), lasso(191)
	Functional Data(801)	functional data analysis(98), longitudinal data(71)
	Covariance Matrix(377)	covariance matrix(30), sparsity(29)
	Testing(319)	false discovery rate(72), multiple testing(52)
2015	Variable Selection(1315)	variable selection(296), lasso(210)
	Functional Data(894)	functional data analysis(117), functional data(79)
	Testing(469)	false discovery rate(78), multiple testing(54)
	Covariance Matrix(378)	covariance matrix(36), sparsity(35)
2016	Variable Selection(1584)	variable selection(350), lasso(234)
	Functional Data(873)	functional data analysis(128), functional data(82)
	Testing & Covariance Matrix(570)	false discovery rate(81), multiple testing(57), sparsity(48), covariance matrix(42)
	Bayesian(434)	dirichlet process(49), bayesian nonparametrics(41)
2017	Variable Selection(1867)	variable selection(393), lasso(270)
	Testing & Covariance Matrix(523)	false discovery rate(78), multiple testing(54), covariance matrix(38), sparsity(30)
	Functional Data(508)	functional data analysis(132), functional data(87)
	Bayesian(507)	dirichlet process(56), bayesian nonparametrics(53)
	Dimension Reduction(439)	longitudinal data(60), dimension reduction(52)
2018	Variable Selection(2033)	variable selection(420), lasso(283)
	Bayesian(597)	dirichlet process(58), bayesian nonparametrics(55)
	Functional Data(544)	functional data analysis(147), functional data(95)
	Dimension Reduction(464)	longitudinal data(65), dimension reduction(54)
	Testing & Covariance Matrix(463)	false discovery rate(89), multiple testing(64), covariance matrix(33), sparsity(26)

for the following reasons. Firstly, we find that the core papers with high in-degree do not frequently change the community. Secondly, we also notice that the core papers with high in-degree of the statistical research groups are theoretical papers, with insufficient attention paid to applied and interdisciplinary papers. Theoretical papers tend not to cover too many fields, so this paper's approach of assigning each paper to a single community is stable. However, our method also has limitations and we cannot address the problem of mix-membership. Overlapping community detection on dynamic citation network may obtain some interesting findings. Hence, proposing a high-precision and stable overlapping community detection method is one of our future research directions.

Additionally, several directions for future research are possible. First, the number of communities K is estimated by the silhouette coefficient, scree plot and actual clustering effect of manual inspection. How to automatically and accurately estimate the number K is of great practical and theoretical importance. Second, the temporal network is from 2001 to 2018, which includes only part of the development process in the field of statistics. There are also many important statistical theories and applications published before 2000. However, because of the availability of papers in "Web of Science", these papers are not considered in our work. In future work, we can try to broaden the scope of research and analyze the citation network more comprehensively.

6. APPENDIX

6.1 The dynamic community detection results by D-SCORE

Table 9 shows the community detection results of the D-SCORE. We enumerate the communities we detected, and the number after the community represents the number of papers in that community. The number after the keywords is the frequency.

6.2 The determination of the number of components

How to choose the tensor rank (i.e. the number of components in CP decomposition) is very important. For our real data, the adjacency tensor is a $59,690 \times 59,690 \times 18$ tensor, which is a large scale tensor. The computation of CP decomposition is a challenge. We use the following process to determine the number of components. Firstly, due to the limitation of computing resources and time cost, We set the initial range of the number of components from 1 to 20. Secondly, for each components, we take CP decomposition on the tensor, and then use the scree plot to determine the range of the number of communities K. Thirdly, for each K, we compare the effects of community detection based on the key papers and keywords of communities. Finally, determine the number of components.

Received 23 November 2021

REFERENCES

[1] Ahn, Y.-Y., Bagrow, J. P. and Lehmann, S. (2010). Link communities reveal multiscale complexity in networks. *Nature* **466** 761–764.

[2] Al-Sharoa, E., Al-khassaweneh, M. and Aviyente, S. (2017). A tensor based framework for community detection in dynamic networks. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* 2312–2316. IEEE.

[3] Bai, L., Cheng, X., Liang, J. and Guo, Y. (2017). Fast graph clustering with a new description model for community detection. *Information Sciences* **388** 37–47.

[4] Bezdek, J. C. (2013). *Pattern recognition with fuzzy objective function algorithms*. Springer Science & Business Media. MR0631231

[5] Boettcher, S. and Percus, A. G. (2001). Extremal optimization for graph partitioning. *Physical Review E* **64** 026114. MR1993187

[6] Bommarito II, M. J., Katz, D. M., Zelner, J. L. and Fowler, J. H. (2010). Distance measures for dynamic citation networks. *Physica A: Statistical Mechanics and its Applications* **389** 4201–4208.

[7] Carroll, J. D. and Chang, J.-J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. *Psychometrika* **35** 283–319.

[8] Chen, P. and Redner, S. (2010). Community structure of the physical review citation network. *Journal of Informetrics* **4** 278–290.

[9] Chen, Z., Wilson, K. A., Jin, Y., Hendrix, W. and Samatova, N. F. (2010). Detecting and tracking community dynamics in evolutionary networks. In *2010 IEEE International Conference on Data Mining Workshops* 318–327. IEEE.

[10] Cichocki, A., Mandic, D., De Lathauwer, L., Zhou, G., Zhao, Q., Caiafa, C. and Phan, H. A. (2015). Tensor decompositions for signal processing applications: From two-way to multiway component analysis. *IEEE Signal Processing Magazine* **32** 145–163.

[11] Clauset, A., Newman, M. E. and Moore, C. (2004). Finding community structure in very large networks. *Physical Review E* **70** 066111.

[12] Cuevas, A. (2014). A partial overview of the theory of statistics with functional data. *Journal of Statistical Planning and Inference* **147** 1–23. MR3151843

[13] De Lathauwer, L., De Moor, B. and Vandewalle, J. (2000). A multilinear singular value decomposition. *SIAM Journal on Matrix Analysis and Applications* **21** 1253–1278. MR1780272

[14] De Lathauwer, L., De Moor, B. and Vandewalle, J. (2000). On the best rank-1 and rank-(r 1, r 2,..., rn) approximation of higher-order tensors. *SIAM Journal on Matrix Analysis and Applications* **21** 1324–1342. MR1780276

[15] Fan, J. and Lv, J. (2010). A selective overview of variable selection in high dimensional feature space. *Statistica Sinica* **20** 101. MR2640659

[16] Fortunato, S. (2010). Community detection in graphs. *Physics Reports* **486** 75–174. MR2580414

[17] Gauvin, L., Panisson, A. and Cattuto, C. (2014). Detecting the community structure and activity patterns of temporal networks: a non-negative tensor factorization approach. *PloS One* **9** e86028.

[18] Goldberg, M., Magdon-Ismail, M., Nambirajan, S. and Thompson, J. (2011). Tracking and predicting evolution of social communities. In *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing* 780–783. IEEE.

[19] Harshman, R. A. et al. (1970). Foundations of the PARAFAC procedure: Models and conditions for an" explanatory" multimodal factor analysis.

[20] Hopcroft, J., Khan, O., Kulis, B. and Selman, B. (2004). Tracking evolving communities in large linked networks. *Proceedings of the National Academy of Sciences* **101** 5249–5253.

[21] Javed, M. A., Younis, M. S., Latif, S., Qadir, J. and Baig, A. (2018). Community detection in networks: A multidisciplinary review. *Journal of Network and Computer Applications* **108** 87–111.

[22] Jdidia, M. B., Robardet, C. and Fleury, E. (2007). Communities detection and analysis of their dynamics in collaborative networks. In *2007 2nd International Conference on Digital Information Management* **2** 744–749. IEEE.

[23] Ji, P., Jin, J., Ke, Z. T. and Li, W. (2021). Co-citation and Co-authorship Networks of Statisticians. *Journal of Business & Economic Statistics* 1–17. MR4410871

[24] Ji, P., Jin, J. et al. (2016). Coauthorship and citation networks for statisticians. *The Annals of Applied Statistics* **10** 1779–1812. MR3592033

[25] Jin, J. et al. (2015). Fast community detection by score. *Annals of Statistics* **43** 57–89. MR3285600

[26] Jing, B.-Y., Li, T., Lyu, Z. and Xia, D. (2021). Community detection on mixture multilayer networks via regularized tensor decomposition. *The Annals of Statistics* **49** 3181–3205. MR4352527

[27] Ke, Z. T. and Jin, J. (2022). The SCORE normalization, especially for highly heterogeneous network and text data. *arXiv preprint arXiv:2204.11097*.

[28] Ke, Z. T., Shi, F. and Xia, D. (2019). Community detection for hypergraph networks via regularized tensor power iteration. *arXiv preprint arXiv:1909.06503*.

[29] Khalid, A., Afzal, M. T. and Qadir, M. A. (2011). *Citation network visualization of CiteSeer dataset*. Citation network visualization of CiteSeer dataset.

[30] Kiers, H. A. (2000). Towards a standardized notation and terminology in multiway analysis. *Journal of Chemometrics* **14** 105–122.

[31] Kim, M.-S. and Han, J. (2009). A particle-and-density based evo-

lutionary clustering method for dynamic networks. *Proceedings of the VLDB Endowment* **2** 622–633.
[32] Knoke, D. and Yang, S. (2019). *Social Network Analysis*. Sage Publications.
[33] Kolda, T. G. and Bader, B. W. (2009). Tensor decompositions and applications. *SIAM Review* **51** 455–500. MR2535056
[34] Li-hu, C., You-fang, L., Zhi-hao, W. and Li-ping, J. (2017). A Community Detection Algorithm of Multiplex Networks with Layer Reduction. *Computer and Modernization* **6** 84.
[35] Li, L., Du, M., Liu, G., Hu, X. and Wu, G. (2014). Extremal optimization-based semi-supervised algorithm with conflict pairwise constraints for community detection. In *2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014)* 180–187. IEEE.
[36] Lyu, Z., Xia, D. and Zhang, Y. (2021). Latent Space Model for Higher-order Networks and Generalized Tensor Decomposition. *arXiv preprint arXiv:2106.16042*.
[37] Mirzaei, S. and Soltanian-Zadeh, H. (2019). Overlapping brain Community detection using Bayesian tensor decomposition. *Journal of Neuroscience Methods* **318** 47–55.
[38] Mucha, P. J., Richardson, T., Macon, K., Porter, M. A. and Onnela, J.-P. (2010). Community structure in time-dependent, multiscale, and multiplex networks. *Science* **328** 876–878. MR2662590
[39] Pagani, G. A. and Aiello, M. (2013). The power grid as a complex network: a survey. *Physica A: Statistical Mechanics and its Applications* **392** 2688–2700. MR3043171
[40] Portenoy, J. and West, J. D. (2016). Dynamic Visualization of Citation Networks Showing the Influence of Scholarly Fields over Time. In *International Workshop on Semantic, Analytics, Visualization*.
[41] Rossetti, G. and Cazabet, R. (2018). Community discovery in dynamic networks: a survey. *ACM Computing Surveys (CSUR)* **51** 1–37.
[42] Selee, T. M., Kolda, T. G., Kegelmeyer, W. P. and Griffin, J. D. (2007). Extracting clusters from large datasets with multiple similarity measures using IMSCAND. *CSRI Summer Proceedings* 87–103.
[43] Serrano, M. A. and Boguná, M. (2003). Topology of the world trade web. *Physical Review E* **68** 015101.
[44] Shang, J., Liu, L., Xie, F., Chen, Z., Miao, J., Fang, X. and Wu, C. (2014). A real-time detecting algorithm for tracking community structure of dynamic networks. *arXiv preprint arXiv:1407.2683*.
[45] Shen, H., Cheng, X., Cai, K. and Hu, M.-B. (2009). Detect overlapping and hierarchical community structure in networks. *Physica A: Statistical Mechanics and its Applications* **388** 1706–1712.
[46] Sidiropoulos, A., Katsaros, D. and Manolopoulos, Y. (2007). Generalized Hirsch h-index for disclosing latent facts in citation networks. *Scientometrics* **72** 253–280.
[47] Smilde, A., Bro, R. and Geladi, P. (2005). *Multi-way analysis: applications in the chemical sciences*. John Wiley & Sons.
[48] Tucker, L. R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika* **31** 279–311. MR0205395
[49] Van Loan, C. F. (2000). The ubiquitous Kronecker product. *Journal of Computational and Applied Mathematics* **123** 85–100. MR1798520
[50] Viard, T., Latapy, M. and Magnien, C. (2016). Computing maximal cliques in link streams. *Theoretical Computer Science* **609** 245–252. MR3426953
[51] Von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and Computing* **17** 395–416. MR2409803
[52] Wang, S. and Rohe, K. (2016). Discussion of "Coauthorship and citation networks for statisticians". *The Annals of Applied Statistics* **10** 1820–1826. MR3592035
[53] Wang, X., Che, M. and Wei, Y. (2020). Tensor neural network models for tensor singular value decompositions. *Computational Optimization and Applications* **75** 753–777. MR4078368
[54] Wang, Z., Liang, Y. and Ji, P. (2020). Spectral algorithms for community detection in directed networks. *arXiv preprint arXiv:2008.03820*. MR4209439
[55] Xie, X.-M., Zeng, S. and Tam, C. M. (2010). Overcoming barriers to innovation in SMEs in China: A perspective based cooperation network. *Innovation* **12** 298–310.
[56] Yang, B. and Li, J. (2020). Complex network analysis of three-way decision researches. *International Journal of Machine Learning and Cybernetics* 1–15.
[57] Ying, D., Yan, E., Frazho, A. and Caverlee, J. (2014). PageRank for ranking authors in co-citation networks. *Journal of the American Society for Information Science & Technology* **60** 2229–2243.
[58] Zhou, D., Councill, I., Zha, H. and Giles, C. L. (2007). Discovering temporal communities from social network documents. In *Seventh IEEE International Conference on Data Mining (ICDM 2007)* 745–750. IEEE.

Tianchen Gao
School of Economics
Xiamen University
Xiamen, Fujian
P. R. China
E-mail address: gaotc_cufe@126.com

Rui Pan
School of Statistics and Mathematics
Central University of Finance and Economics
Beijing
P. R. China
E-mail address: panrui_cufe@126.com

Junfei Zhang
School of Statistics and Mathematics
Central University of Finance and Economics
Beijing
P. R. China
E-mail address: junfeizhang_cufe@163.com

Hansheng Wang
Guanghua School of Management
Peking University
Beijing
P. R. China
E-mail address: hansheng@pku.edu.cn