Statistics and Its Interface Volume 14 (2021) 279–294

Community detection for statistical citation network by D-SCORE*

Tianchen Gao, Yan Zhang, Siyu Wang, Yuehan Yang, and Rui Pan†

With the wide application of statistics, it is important to identify research trends and the development of statistics. In this paper, we analyze a citation network of the top 4 statistical journals from 2001 to 2018, applying the directed spectral clustering on the ratio-of-eigenvectors (D-SCORE) method to detect the community structure of citation network. We find that statistical researchers are becoming more and more collaborative. The number of influential papers which account for the majority of citations is small. High betweenness centrality and high closeness centrality papers are concentrated in Annals of Statistics (AoS). Furthermore, we detect 4 communities and 11 sub-communities such as "High-dimensional Model", "Variable Selection", and "Covariance Matrix Analysis". Then, we compare the results of D-SCORE with three other methods and find that D-SCORE is more suitable for our citation network. Finally, we identify the dynamic nature of the communities. Our findings present trends and topological patterns of statistical papers, and the data set provides a fertile ground for future research on social networks.

Keywords and phrases: Community detection, D-SCORE, Citation network.

1. CITATION NETWORK

1.1 Statistical citation network of top 4 journals

In the era of big data, statistics has attracted more and more attentions. Accordingly, statistical methods are applied in a wide range of disciplines such as engineering [49], economics [1], biochemistry [13], ecology [22] and many others. Statistical analysis goes in many directions, such as variable selection [18, 19], causal inference [27, 14], Bayesian analysis [20, 16], non- and semiparametric models [58, 63, 68] and many others [5]. As a result, it is of great interest to understand research trends and the development of statistics. To be more specific, it is essential to identify research topics and key papers in a scientific community [30].

Citation network analysis is a powerful tool for identifying established and emerging research topics [17]. Some studies identified research topics through citation network analysis. [50] proposed a visualization technique for citation networks by applying the topic-based paper clustering. [43] proposed the citation-network topic model to jointly model research publications and their citation network. Their model offered substantial performance improvement over previous work in terms of model fitting and clustering evaluation. [65] used the infomap algorithm iteratively in order to extract topics from the directed citation network in the field of astrophysics and obtained 22 topics. Then, they generated a cognitive map of the field using a topic affinity network to highlight the relationships among the topics.

In citation network analysis, community detection is a prominent task. Communities are groups of nodes that are densely interconnected but sparely connected with the rest of the network [21, 66, 11, 70]. The goal of community detection is to identify such groups. Community detection helps to identify important research topics and understand the research problems of a subject. It can be applied to many fields such us physics [8], biology [26], statistics [30] and others. A common assumption made in most community detection methods is that the number of communities K is known, but in practice K is often unknown. Therefore, it is of great practical and theorectical importance to estimate K. Many methods have been proposed, including a recursive approach [73], information theory [59, 56], Bayesian inference [39, 47], sequential tests [6, 41], spectral methods [40, 45], network cross-validation [7] and likelihood-based methods [12, 62, 67, 29].

Nowadays, exploring the dynamic of the communities is becoming more and more important. [28] analyzed different time snapshots of the citation network provided by NEC Citeseer Database. By analyzing the community structure of different snapshot networks, tracking the evolution of communities over time. They found that the emergence of new communities involve a new research area. [54] analyzed the evolution of communities in collaboration network

*The research is supported by National Natural Science Foundation of China (NSFC, 11971504, 11631003, 12001557, 71771224), the Fundamental Research Funds for the Central University of Finance and Economics (QL18010), the Youth Talent Development Support Program (QYP1911), the Program for Innovation Research in Central University of Finance and Economics, and the disciplinary funding of Central University of Finance and Economics.
†Corresponding author.

Table 1. An example to show 7 variables of a published paper

Title	Journal	Publication Date	Author
Variable selection via nonconcave penalized likelihood and its oracle properties	JASA	2001	Fan, J. Q. Li, R. Z.
Keywords	**Abstract**	**Details of its References**	
Hard thresholding Nonnegative garrote Penalized likelihood Lasso...	Variable selection is fundamental to high-dimensional statistical modeling, including nonparametric regression...	Regularization of wavelet approximations Antoniadis, A and Fan, J. Q. Journal of the American Statistical Association 2001	

and phone-call network. They found that small communities are more stable, while large communities change drastically. [71] proposed a dynamic stochastic block model to detecting communities and the evolution of communities. [44] proposed a method named Persistent Community Detection (PCD) to detect the communities which show persistent behavior over time. [46] proposed a model that combines a static stochastic block model for its static part with independent markov chains for the evolution of the communities over time which can detect the communities in discrete time dynamic networks. [72] proposed a approach for detecting common modules in the time-varying gene regulation network, which can be applied to other time-varying networks.

In this paper, we analyze a citation network for statistical publications. The data are collected from the "Web of Science" (http://apps.webofknowledge.com/). We initially collect 6,497 papers. After data cleaning, we remove 751 isolated nodes whose in-degree and out-degree are both zero. The rest of the data consists of 5,746 papers from 2001 to 2018 published in the top 4 statistical journals, i.e., *Annals of Statistics* (AoS), *Biometrika*, *Journal of American Statistical Association* (JASA) and *Journal of Royal Statistical Society (Series B)* (JRSS-B). For each paper, the following variables are obtained: title, authors, abstract, keywords, publisher, publication date and its reference list. Table 1 presents an example of a published paper. It is worth noting that there exist some other prominent statistical journals, such as those mentioned in [64] and [2], which should be analyzed in the future. Due to the data constraint, we focus on the top 4 statistical journals in this work.

1.2 Descriptive analysis of the citation network

There are 5,746 nodes in citation network. Especially, there are 1,723 with zero in-degree which may be newly published papers and 967 nodes with zero out-degree. These nodes may be papers published earlier, and their citations are not in our dataset. There are 5,538 different authors in citation network, and thus each author publishes an average of 1.04 paper. We faced some challenges in data cleaning such as ambiguous author names. Different authors may have the same name, and one author may have different initials. For example, author "Fan Jianqing" may also be written as "Fan, J." or "Fan, J. Q.". To address this issue, we adopt the following methods. Firstly, some authors have Open Researcher and Contributor ID (ORCID). We can uniquely identify authors according to their ORCIDs. For authors without ORCID, we use institutions and addresses of authors to identify them.

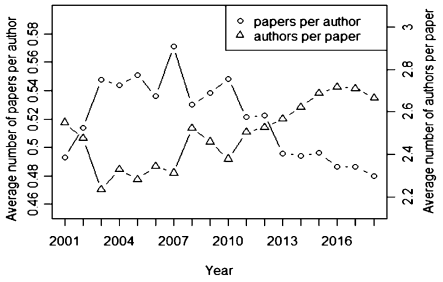


Figure 1. Yearly average productivity per author (left y-axis) and yearly average authors per paper (right y-axis).

The left y-axis of Figure 1 shows the yearly average productivity per author, and the right y-axis shows the yearly average authors per paper. In most cases, when the yearly average productivity per author goes up, the yearly average authors per paper goes down. In a few years (2011-2012, 2013-2015 and 2017-2018), the yearly average productivity per author and the yearly average authors per paper show the same trend, but the changes of the two indicators are small. As one can see, from 2001 to 2018, the number of papers per author shows a trend of rising first and then declining, while the number of authors per paper is moving in the opposite direction. It indicates that the amount of statisticians has greatly increased in recent years, and the amount of collaboration among authors has increased as well.

Consider a directed citation network structure. This structure contains 5,746 nodes corresponding to the papers and the directed edges describe the citation relationships among nodes. Without loss of generality, nodes are not allowed to be self-related. There are 23,737 edges in the citation network. Let $d = m/(N^2 - N)$ be the density of a directed network, where m denotes the number of edges and N denotes the number of nodes. The density of the citation network is 0.00072. We use the in-degree, the number of citations received by each paper, to measure the importance

Table 2. Top 10 high in-degree papers

ID	Title	Journal	Year	In-degree
1	Variable selection via nonconcave penalized likelihood and its oracle properties	JASA	2001	335
2	The adaptive lasso and its oracle properties	JASA	2006	208
3	Least angle regression	AoS	2004	177
4	High-dimensional graphs and variable selection with the lasso	AoS	2006	172
5	The dantzig selector: statistical estimation when p is much larger than n	AoS	2007	144
6	Sure independence screening for ultrahigh dimensional feature space	JRSS-B	2008	137
7	Simultaneous analysis of lasso and dantzig selector	AoS	2006	129
8	Nearly unbiased variable selection under minimax concave penalty	AoS	2010	110
9	Regularization and variable selection via the elastic net	JRSS-B	2005	108
10	Model selection and estimation in regression with grouped variables	JRSS-B	2006	108

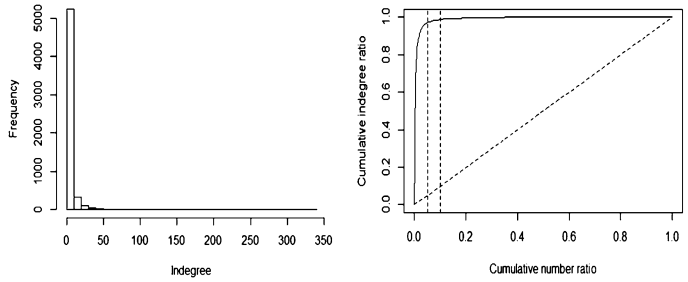


Figure 2. Histogram of in-degree distribution (left) and Lorenz curve (right).

of papers. Nowadays, There are many literatures to describe the distribution of degrees [69]. Actually, many degree distributions follow the power law distribution, a phenomenon which is also found in [4, 51, 30]. In our citation network, we use the method proposed by [10] to test the in-degree distribution, and find that it doesn't follow the power law distribution. The left-hand panel of Figure 2 shows the histogram of in-degree distribution in the citation network. As one can see, it is highly right-skewed. Most of the nodes have very low in-degree. The Gini coefficient is 0.708 [25], suggesting that the in-degree is highly dispersed. The right-hand panel of Figure 2 shows the Lorenz curve of in-degrees, confirming that the distribution of in-degrees is highly skewed. For example, the top 10% highly cited papers receive about more than 95% of all citation counts. It indicates that the highly cited papers play an irreplaceable role in the development of statistics.

Table 2 lists the title, journal, published year and in-degree of the top 10 cited papers. All of them were published before 2010. It is natural since recent papers do not have enough time to accrue citations. The top 10 papers are all in a specific sub-area of high-dimensional variable selection. The results suggest that variable selection is a "hot" area of statistical research.

Furthermore, we use centralities, i.e., betweenness centrality and closeness centrality, to identify the influential papers in the citation network. Centrality is an informative approach to identify the most "important" papers of a citation network. There are many different measures of centrality. Both betweenness centrality and closeness centrality are standard measures: 1) Betweenness centrality measures the extent to which a node "standing" between other nodes [23]; 2) Closeness centrality, calculated as the average length of the shortest paths between the node and all other nodes, is used to reflect the proximity of a node to other nodes [61].

Table 3 shows the top 10 high betweenness centrality papers of the citation network. The high betweenness centrality papers are mainly published in JRSS-B and AoS, around 2008 and 2013. The values of the top 15 high betweenness centrality papers are relatively small. It means that these papers cannot be seen as the irreplaceable bridges in the statistical citation network. Table 4 presents the top 10 high closeness centrality papers, which are mainly published in JASA and AoS around 2001 and 2007. A node with high closeness centrality is in a central position of a network for it is close to all other nodes. Paper "Variable selection via nonconcave penalized likelihood and its oracle properties" has the highest closeness centrality. This paper can be seen as one of the most central nodes in the citation network.

1.3 Exploratory analysis of community structure

In this part, we try to explore the community structure of citation network. Figure 3 shows a small group of the citation network, which are composed of 27 closely connected nodes. The edges within this group are denser than those without the group. The density of this small group is 0.151, which is much larger than that of the citation network (i.e., 0.00072). We find that the nodes in this group are all involved in the area of "Variable Selection", including some important papers such as "Variable selection via nonconcave penalized likelihood and its oracle properties", "The adaptive lasso and its oracle properties" and so on. Therefore, it is reasonable to speculate that there exist group structures in the citation network, namely, the community.

Figure 4 shows another small group of the citation network with 123 closely connected nodes, which are involved in the area of "False Discovery Rate". The density of the

Table 3. Top 10 high betweenness centrality papers

ID	Title	Journal	Year	Betweenness Centrality
1	Sure independence screening for ultrahigh dimensional feature space	JRSS-B	2008	0.00063
2	On asymptotically optimal confidence regions and tests for high-dimensional models	AoS	2014	0.00052
3	Sure independence screening in generalized linear models with np-dimensionality	AoS	2010	0.00033
4	Large covariance estimation by thresholding principal orthogonal complements	JRSS-B	2013	0.00033
5	Confidence sets in sparse regression	AoS	2013	0.00031
6	Multiscale change point inference	JRSS-B	2014	0.00030
7	Innovated higher criticism for detecting sparse signals in correlated noise	AoS	2010	0.00030
8	Maximum likelihood estimation in semiparametric regression models with censored data	JRSS-B	2007	0.00025
9	One-step sparse estimates in nonconcave penalized likelihood models	AoS	2008	0.00025
10	Optimal detection of sparse principal components in high dimension	AoS	2013	0.00024

Table 4. Top 10 high closeness centrality papers

ID	Title	Journal	Year	Closeness centrality
1	Variable selection via nonconcave penalized likelihood and its oracle properties	JASA	2001	0.1125
2	Regularization of wavelet approximations	JASA	2001	0.0939
3	Least angle regression	AoS	2004	0.0911
4	Nonconcave penalized likelihood with a diverging number of parameters	AoS	2004	0.0817
5	Adaptive model selection	JASA	2002	0.0802
6	High-dimensional graphs and variable selection with the Lasso	AoS	2006	0.0794
7	The adaptive lasso and its oracle properties	JASA	2006	0.0758
8	Generalized likelihood ratio statistics and Wilks phenomenon	AoS	2001	0.0744
9	Greedy function approximation: A gradient boosting machine	AoS	2001	0.0720
10	Variable selection for Cox's proportional hazards model and frailty model	AoS	2002	0.0702

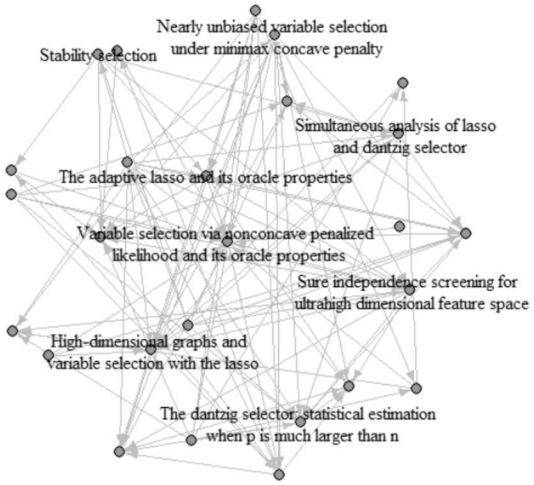


Figure 3. Citation network shows community structure.

Figure 4. Sub-network shows community structure.

group is 0.031. The edges within this group are densely connected as well. By further researching on this small community, we find that this community can be further divided into two sub-communities. As shown in Figure 4, we can find two closely connected groups. Therefore, it is reasonable to speculate that there are some communities in sub-network.

Results show that the keywords of papers in the same area have obvious changes, which means that there may exists a dynamic nature of the communities. For instance, in the area of variable selection, the keywords of papers published between 2001 and 2005 are mostly "model selection" and "variable selection". After 2006, we find that the keyword

"lasso" appears more and more frequently. That is because the lasso appears in the area of variable selection since 2006, where paper "The adaptive lasso and its oracle properties" [74] and "High-dimensional graphs and variable selection with the Lasso" [48] are representative papers of the lasso. Therefore, it is reasonable to speculate that there may be a dynamic nature of the communities in the citation network.

To summarize, we have the following three questions to address.

- We find that there exist some closely connected groups, namely communities in the network, such as "Variable Selection" community. So can we explore different communities in the citation network?
- We find that a closely connected community with a similar topic can also be divided into several sub-communities. Are there sub-communities in the communities identified in the previous question? If so, what are their sub-communities?
- We find that there is a dynamic nature of the communities. Can we detect this phenomenon by community detection methods? How have the characteristics of the communities changed over time?

2. LITERATURE REVIEW

As shown above, we can discover some research topics such as "Variable Selection" and "False Discovery Rate" by identifying the groups in the citation network. To identify the communities, we conduct community detection on the citation network. It is helpful for us to identify "hot" research topics and understand the research problems of statistics in 4 journals.

The methods of community detection can be classified into two categories according to whether the network is directed or not. In an undirected network, there are many classic algorithms. [9] presented a faster greedy algorithm for the problem of modularity maximization based on a hierarchical clustering approach. [31] proposed an approach to community detection called the Spectral Clustering On Ratios-of-Eigenvectors (SCORE), which could remove the effect of degree heterogeneity. SCORE is a conceptually simple and flexible idea. It can be extended to other areas of research, such as topic estimation in text mining [37] and state aggregation in control systems and reinforcement learning [15]. In the network study, many approaches based on SCORE have been proposed. [32] proposed Mixed-SCORE to membership, which accommodated the settings where almost all nodes may be mixed and where they allow severe degree heterogeneity. [33] proposed SCORE+ as a refinement of SCORE to conduct community detection, especially for networks with weak signals. [36] proposed Tensor-SCORE for hypergraph community detection. This method firstly used a tensor power iteration technique for community detection and accommodated degree heterogeneity. In recent years, [55] proposed a community detection algorithm (Fluid Communities) based on the propagation methodology, which performed very well in terms of computational cost and scalability.

In the directed network, the direction of an edge contains important information such as asymmetric influence or information flow. Therefore, any kind of community detection approach may fail to detect the communities correctly if the direction of the edge is not considered properly [38, 42]. Many researchers have made profound contributions to this field. [53] used the machinery of probabilistic mixture models and the expectation-maximization algorithm to detect types of structure in networks without prior knowledge. However, the approach fails to detect obvious community structures if there are some nodes with zero out-degree or in-degree [57]. Later, [60] used the probability flow of random walks on a network and decomposed the network into modules to reveal community structure in weighted and directed networks. [42] extended the spectral modularity methods [52] for undirected networks to directed networks, based on spectral optimization of the modularity. However, there are some limitations to this method. [38] found that it failed to properly distinguish the directions of the edges and could not detect communities representing directionality patterns among the nodes. [30] proposed D-SCORE for community detection of the citation network. D-SCORE is an adaption of SCORE [31] to directed networks. They also compared D-SCORE with the method proposed by [42] and got the same conclusion with [38].

Newly proposed D-SCORE method [30] computes faster and is easy to implement, thus is suitable for large citation network. We apply it to conduct community detection in our citation network. Our study adopts different nodes from the study of [30]. The nodes in our citation network represent papers, while in the [30]'s citation network, the nodes represent authors. Our study focuses on the interactions among papers rather than authors. We detect 4 communities and 11 sub-communities, which are to be presented later.

3. COMMUNITY DETECTION

In this section, we apply the D-SCORE [30] to analyze the citation network. D-SCORE is a detection method and is preferred for large scale data since it is fast and simple. We first give a brief introduction to this method. Consider a directed citation network. Let A be the adjacency matrix with elements denoting the indicators of whether or not the jth paper is cited by the ith paper. Assume there are K communities. D-SCORE applies Singular Value Decomposition (SVD) on the adjacent matrix A to obtain the first K left singular vectors $\hat{u}_1, \hat{u}_2, ..., \hat{u}_K$, and the first K right singular vectors $\hat{v}_1, \hat{v}_2, ..., \hat{v}_K$ of A, where $\hat{u}_k \in \mathbb{R}^n$, $\hat{v}_k \in \mathbb{R}^n$, and $k = 1, 2, ..., K$. Let $\mathcal{N}_1$ be the collection of indexes where $\hat{u}_1 = (\hat{u}_{11}, \hat{u}_{12}, ..., \hat{u}_{1n})^{\mathrm{T}} \in \mathbb{R}^n, |\hat{u}_{1i}| \neq 0$. $\mathcal{N}_1 = \{i : |\hat{u}_{1i}| \neq 0\}$. Let $\mathcal{N}_2$ be the collection of indexes where

$\hat{v}_1 = (\hat{v}_{11}, \hat{v}_{12}, ..., \hat{v}_{1n})^{\mathrm{T}} \in \mathbb{R}^n, |\hat{v}_{1i}| \neq 0$. $\mathcal{N}_2 = \{i : |\hat{v}_{1i}| \neq 0\}$. Then we have all nodes split into four disjoint subsets,

$$\mathcal{N} = (\mathcal{N}_1 \cap \mathcal{N}_2) \cup (\mathcal{N}_1 \backslash \mathcal{N}_2) \cup (\mathcal{N}_2 \backslash \mathcal{N}_1) \cup (\mathcal{N} \backslash (\mathcal{N}_1 \cup \mathcal{N}_2)).$$

Define two $n \times (K-1)$ matrices $\hat{R}^{(l)}$ and $\hat{R}^{(r)}$ as follows,

$$\hat{R}^{(l)}(i,k) = \begin{cases} \operatorname{sgn}\left(\frac{\hat{u}_{k+1}(i)}{\hat{u}_1(i)}\right) \cdot \min\left\{\left|\frac{\hat{u}_{k+1}(i)}{\hat{u}_1(i)}\right|, \log(n)\right\}, & i \in \mathcal{N}_1, \\ 0, & i \notin \mathcal{N}_1, \end{cases}$$

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

where $1 \leq k \leq K-1$ 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 $\hat{R}^{(l)}$ and $\hat{R}^{(r)}$.

Step 1: Restrict the rows of $\hat{R}^{(l)}$ and $\hat{R}^{(r)}$ to the set $\mathcal{N}_1 \cap \mathcal{N}_2$ and get two new matrices $\tilde{R}^{(l)} \in \mathbb{R}^{m \times (K-1)}$ and $\tilde{R}^{(r)} \in \mathbb{R}^{m \times (K-1)}$ where $m = |\mathcal{N}_1 \cap \mathcal{N}_2|$ is the size of $\mathcal{N}_1 \cap \mathcal{N}_2$. Assume there are K communities, and apply k-means to the columns of $B = (\tilde{R}^{(l)}, \tilde{R}^{(r)}) \in \mathbb{R}^{m \times 2(K-1)}$, so that nodes in $\mathcal{N}_1 \cap \mathcal{N}_2$ are divided into K communities.

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

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

Step 4: For a node i in $\mathcal{N} \backslash (\mathcal{N}_1 \cup \mathcal{N}_2)$, 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.

We consider the weakly connected citation network, where two nodes are connected by an undirected edge if one has cited the other [3]. There are 57 components for the original citation network in the associated weakly connected network. The giant component has 5601 nodes and all other components have no more than 6 nodes. There is no the largest strong connected component because citations between papers are one-way. In what follows, we restrict our attention to the weakly connected giant component. For simplicity of expression, we denote it as the citation network for the rest of the paper instead of the weakly connected giant component of the original citation network.

3.1 Community detection of the citation network

The scree plot of the citation network is shown in Figure 5. As can be seen, it suggests 2 or more communities in $\mathcal{N}$. We initially try 2, 3, 4 and 5 communities. The results show that it is best to divide the citation network into 4 communities. When we divide the network into 2 or 3 communities, some important research topics are not detected, such as "Functional Data Analysis" and "False Discovery Rate". One community may contain various research topics. When we divide the network into 5 communities, there are some common research topics among communities. For example, three different communities all involve "Variable Seletion" area. Assuming $K = 4$, applying D-SCORE, we have $|\mathcal{N}_1 \cap \mathcal{N}_2| = 4{,}070$, $|\mathcal{N}_1 \backslash \mathcal{N}_2| = 895$, $|\mathcal{N}_2 \backslash \mathcal{N}_1| = 605$, and $|\mathcal{N} \backslash (\mathcal{N}_1 \cup \mathcal{N}_2)| = 31$. D-SCORE identifies four communities: (1)"Variable Selection" community containing 2074 nodes; (2) "Sparse Covariance Matrix" community containing 886 nodes; (3)"Functional Data Analysis & Dimension Reduction" community containing 1939 nodes; (4) "False Discovery Rate" community containing 702 nodes. Table 5 reports the top 5 high in-degree papers in each community. We also separately count the keywords of the papers in each community to identify communities accurately. Table 6 shows the top 3 keywords of each community.

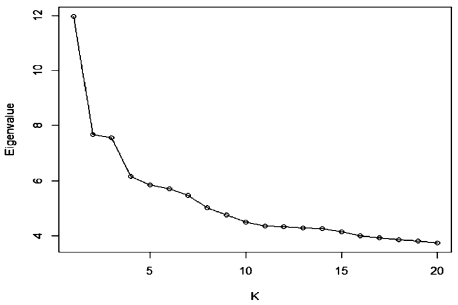


Figure 5. The scree plot of the citation network.

The first community is the largest community in citation network. Papers in this community focus on variable selection. The top 3 keywords of this community are "variable selection", "lasso" and "sparsity". It indicates that variable selection is a very "hot" topic in statistics. The second community is mainly relative to the sparse covariance matrix. Nevertheless, this community contains some other topic related papers, making the research structure unclear. It is necessary to get more detection and discussion for this community. The third community involves two main research topics, functional data analysis and dimension reduction. This community seems to contain some substructures and we expect to distinguish those for better interpretation and identification. The last community is the smallest and is detected clearly. This community studies issues such as false discovery rate, multiple testing and so on. For further information, we restrict the networks to their respective communities and detect the sub-communities in section 3.2.

Futhermore, we analyze the differences between the communities in the four journals. For AoS, about 41% papers are

Table 5. The result of community detection by D-SCORE

Community	Title	Journal	Year	In-degree
Variable Selection	Variable selection via nonconcave penalized likelihood and its oracle properties	JASA	2001	335
	The adaptive lasso and its oracle properties	JASA	2006	208
	Least angle regression	AoS	2004	177
	High-dimensional graphs and variable selection with the Lasso	AoS	2006	172
	Sure independence screening for ultrahigh dimensional feature space	JRSS-B	2008	137
Sparse Covariance Matrix	On the distribution of the largest eigenvalue in principal components analysis	AoS	2001	80
	On consistency and sparsity for principal components analysis in high dimensions	JASA	2009	48
	Operator norm consistent estimation of large dimensional sparse covariance matrices	AoS	2008	41
	Optimal rates of convergence for covariance matrix estimation	AoS	2010	40
	Adaptive thresholding for sparse covariance matrix estimation	JASA	2011	34
Functional Data Analysis & Dimension Reduction	Functional data analysis for sparse longitudinal data	JASA	2005	82
	Bayesian measures of model complexity and fit	JRSS-B	2002	75
	An adaptive estimation of dimension reduction space	JRSS-B	2002	65
	Gibbs sampling methods for stick-breaking priors	JASA	2001	64
	Generalized likelihood ratio statistics and Wilks phenomenon	AoS	2001	61
False Discovery Rate	A direct approach to false discovery rates	JRSS-B	2002	90
	Empirical Bayes analysis of a microarray experiment	JASA	2001	80
	Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach	JRSS-B	2004	62
	A stochastic process approach to false discovery control	AoS	2004	54
	Large-scale simultaneous hypothesis testing: The choice of a null hypothesis	JASA	2004	52

Table 6. Top 3 keywords of each community

Community	Keywords	Frequency
Variable Selection	Variable selection	117
	Lasso	105
	Sparsity	101
Sparse Covariance Matrix	Sparsity	27
	Principal component analysis	24
	Markov Chain Monte Carlo	23
Functional Data Analysis & Dimension Reduction	Functional data analysis	70
	Markov Chain Monte Carlo	68
	Nonparametric regression	66
False Discovery Rate	False discovery rate	66
	Multiple testing	55
	Multiple comparisons	24

Table 7. Two sub-communities from "Variable Selection" community

Community	Title	Journal	Year	In-degree
High-dimensional Model	Model selection and estimation in the Gaussian graphical model	Biometria	2007	54
	Regularized estimation of large covariance matrices	AoS	2008	40
	High-dimensional generalized linear models and the lasso	AoS	2008	38
	on asymptotically optimal confidence regions and tests for high-dimensional models	AoS	2014	35
	A constrained l_1 minimization approach to sparse precision matrix estimation	JASA	2011	32
Variable Selection	Variable selection via nonconcave penalized likelihood and its oracle properties	JASA	2001	279
	The adaptive lasso and its oracle properties	JASA	2006	183
	Least angle regression	AoS	2004	153
	High-dimensional graphs and variable selection with the Lasso	AoS	2006	141
	The Dantzig selector: Statistical estimation when p is much larger than n	AoS	2008	123

in "Variable Selection" community, which is larger than the other three communities (19%, 26%, 14%). For JASA and JRSS-B, around 40% and 41% papers in these two journals are in "Functional Data Analysis & Dimension Reduction" community, respectively. For Biometrika, 35% and 34% papers are in "Variable Selection" community and "Functional Data Analysis & Dimension Reduction" community, respectively. Around 10% papers in the four journals belong to the "False Discovery Rate" community, which is the smallest community in the citation network.

3.2 Detection of sub-networks

To get more information from each community, we ignore all the edges and nodes outside the four communities. We apply D-SCORE to the four communities for community detection, respectively. Figure 6 shows their scree plots, suggesting that there are 2, 2, 5 and 2 communities in the four sub-networks respectively.

For "Variable Selection" community, assuming $K = 2$, we obtain two sub-communities: "High-dimensional Model" containing 636 nodes, "Variable Selection" containing 1215 nodes. Table 7 reports 5 key nodes (papers) from each sub-community. The top 5 papers of this sub-community are also the Top 5 high in-degree papers of the original citation network. From the original citation network to the sub-community, their in-degrees decrease differently. Paper "Variable selection via nonconcave penalized likelihood and its oracle properties" decreases the most, from 335 to 279, indicating that many papers from other sub-communities have cited this paper. In addition, all the other 4 papers decrease about 25 citation counts. We suggest they have a similar influence on other sub-communities. In the other sub-community, "High-dimensional Model", the second high in-degree paper "Regularized estimation of large covariance matrices" is the only one of the five that is also the Top 20 high in-degree papers of the original citation network. Its citation count is reduced from 99 to 40. Note that the twentieth high in-degree paper of the original citation network has 78 citations. It may mean that papers in the "High-dimensional Model" sub-community do not have much influence on other communities.

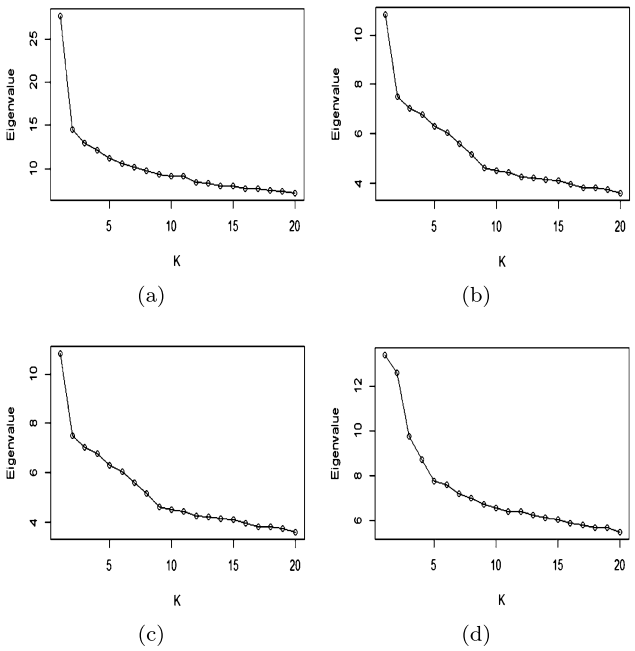


Figure 6. Scree Plots of four sub-networks. (a):Variable Selection Community. (b):Sparse Covariance Matrix Community. (c):Functional Data Analysis & Dimension Reduction Community. (d):False Discovery Rate Community.

Table 8. Two sub-communities from "Sparse Covariance Matrix" community

Community	Title	Journal	Year	In-degree
Covariance Matrix Analysis	On the distribution of the largest eigenvalue in principal components analysis	AoS	2001	52
	Operator Norm Consistent Estimation of Large dimensional Sparse Covariance Matrices	AoS	2008	24
	Generalized Thresholding of Large Covariance Matrices	JASA	2009	22
	A tale of two time scales: Determining integrated volatility with noisy high-frequency data	JASA	2005	21
	Optimal Rates of Convergence for Covariance Matrix Estimation	AoS	2010	21
Principal Component Analysis & Network Analysis	On Consistency and Sparsity for Principal Components Analysis in High Dimensions	JASA	2009	33
	Spectral clustering and the high-dimensional stochastic blockmodel	AoS	2011	18
	High-dimensional analysis of semidefinite relaxations for sparse principal components	AoS	2009	18
	Sparse principal component analysis and iterative thresholding	AoS	2013	17
	Minimax bounds for sparse pca with noisy high-dimensional data	AoS	2013	16

For "Sparse Covariance Matrix" community, assuming $K = 2$, we obtain two sub-communities: "Covariance Matrix Analysis" containing 352 nodes, "Principal Component Analysis & Network Analysis" containing 219 nodes. The latter is related to the former, but covered by the former in the previous section. Table 8 reports 5 key nodes (papers) from each sub-community. Most of them were published in AoS.

For "Functional Data Analysis & Dimension Reduction" community, assuming $K = 5$, we obtain five sub-communities: "Functional Data Analysis and Dimension Reduction" containing 524 nodes, "Markov Chain Monte Carlo" containing 419 nodes, "Dirichlet Process" containing 162 nodes, "Censored Data Analysis" containing 143 nodes, "Semiparametric & Nonparametric Statistics" containing 431 nodes. Table 9 reports 5 key nodes (papers) from each sub-community. It shows that "Dirichlet Process" mainly published in JASA; "Censored Data Analysis" mainly published in Biometrika, the other three sub-communities do not show the trend.

For "False Discovery Rate" community, assuming $K = 2$, we obtain two sub-communities: "False Discovery Rate" containing 258 nodes, "Testing" containing 251 nodes. Table 10 reports 5 key nodes (papers) from each sub-community. Among all the sub-communities, the paper citation counts of "Testing" is the least, while its number of nodes is not the least.

3.3 Comparison of community detection methods

In this part, we compare the community detection results of different methods, including SCORE by [31], Clauset-Newman-Moore greedy modularity maximization algorithm by [9], and the Fluid Communities (FluidC) algorithm by [55].

SCORE is a spectral clustering method based on ratios-of-eigenvectors. Firstly, obtain the K leading eigenvalues of the adjacency matrix. Secondly, calculate the entry-wise ratios between the first leading eigenvector and each of the other leading eigenvectors for clustering, which can remove the effect of degree heterogeneity effectively. [31]. Greedy modularity maximization algorithm initializes each node as a community. Then combine the pair of communities that most increases modularity into one community. Repeat this process until no such pair of communities exists [9]. FluidC initializes K communities, each of them initialized in different and random nodes. Then, the algorithm iterates over all nodes, updating the community of each node based on its own community and the communities of its neighbors. Repeat this process until no node changes the community it belongs to [55].

Considering the directed citation network as undirected network, we get the results of community detection by using the above three methods. Table 11 shows the top 3 keywords and the frequency of each community. We find three problems to the results of these methods. Firstly, there are many overlapping research topics among communities. For example, the second, third and fourth communities detected by SCORE all involve "model selection" and "Markov Chain Monte Carlo". Secondly, the differences among research topics within the same community are relatively large. For example, in the results of the SCORE and greedy modularity maximization algorithm, "model selection" and "Markov Chain Monte Carlo" are divided into the same community in most cases. Thirdly, the research topics found by these methods are not as comprehensive as those found by D-SCORE.

Table 9. Five sub-communities from "Functional Data Analysis & Dimension Reduction" community

Community	Title	Journal	Year	In-degree
Functional Data Analysis & Dimension Reduction	Functional data analysis for sparse longitudinal data	JASA	2005	68
	An adaptive estimation of dimension reduction space	JRSS-B	2002	55
	Dimension reduction for conditional mean in regression	AoS	2002	47
	Properties of principal component methods for functional and longitudinal data analysis	AoS	2006	42
	Generalized likelihood ratio statistics and Wilks phenomenon	JRSS-B	2001	42
Markov Chain Monte Carlo	Bayesian measures of model complexity and fit	JRSS-B	2002	53
	Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations	JRSS-B	2009	22
	Particle Markov chain Monte Carlo methods	JRSS-B	2010	14
	Marginal likelihood from the Metropolis-Hastings output	JASA	2001	13
	Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics	JASA	2004	13
Dirichlet Process	Gibbs sampling methods for stick-breaking priors	JASA	2001	50
	Order-based dependent Dirichlet processes	JASA	2006	21
	An ANOVA model for dependent random measures	JASA	2004	20
	A distributional approach for causal inference using propensity scores	JASA	2006	19
	Kernel stick-breaking processes	Biometrika	2008	15
Censored Data Analysis	Maximum likelihood estimation in semiparametric regression models with censored data	JRSS-B	2007	21
	Semiparametric analysis of transformation models with censored data	Biometrika	2002	18
	Efficient estimation of semiparametric transformation models for counting processes	Biometrika	2006	10
	A crossvalidation method for estimating conditional densities	Biometrika	2004	10
	On semiparametric transformation cure models	Biometrika	2004	9
Semiparametric &Nonparametric Statistics	Semiparametric and nonparametric regression analysis of longitudinal data	JASA	2001	27
	Marginal nonparametric kernel regression accounting for within-subject correlation	Biometrika	2003	23
	Efficient semiparametric marginal estimation for longitudinal/clustered data	JASA	2005	23
	Varying-coefficient models and basis function approximations for the analysis of repeated measurements	Biometrika	2002	21
	Semiparametric regression for clustered data using generalized estimating equations	JASA	2001	21

In addition to the topics such as "sparsity", "nonparametric regression" and "Markov Chain Monte Carlo" found by these three methods, D-SCORE also finds "principal component analysis", "functional data analysis", "multiple testing" and others. Overall, these three methods are not as suitable as D-SCORE in community detection for citation network.

4. THE DYNAMIC NATURE OF THE COMMUNITIES

To explore the dynamic nature of the communities, we divide the time into three periods: 2001 to 2006, 2001 to 2012, 2001 to 2018. For each period, we ignore all the edges and nodes outside the period. D-SCORE is applied to detect communities in the three periods respectively. Figure 7 shows their scree plots, suggesting that there are 2 and 3 communities in the first two networks, respectively. Table 12 lists the top 5 high in-degree papers in each community of the first two networks. From 2001 to 2006, there are two communities in the citation network. The first community in the period of 2001 to 2006 is a complex community, which mainly involves papers about variable selection, dimension reduction and nonparametric regression. The second community mainly involves papers in the field of false discovery rate. From 2001 to 2012, there are three communities detected by D-SCORE, namely "Variable Selection" community, "False Discovery Rate" community and "Dimension Reduction" community.

Table 10. Two sub-communities from "False Discovery Rate" community

Community	Title	Journal	Year	In-degree
False Discovery Rate	A direct approach to false discovery rates	JRSS-B	2002	74
	Empirical Bayes analysis of a microarray experiment	JASA	2001	70
	Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach	JRSS-B	2004	53
	A stochastic process approach to false discovery control	AoS	2004	46
	Operating characteristics and extensions of the false discovery rate procedure	JRSS-B	2002	42
Testing	Generalizations of the familywise error rate	AoS	2005	13
	Design sensitivity in observational studies	Biometrika	2004	12
	Optimality, variability, power: Evaluating response-adaptive randomization procedures for treatment comparisons	JASA	2003	10
	Estimation of a convex function: Characterizations and asymptotic theory	AoS	2001	10
	Exact and approximate stepdown methods for multiple hypothesis testing	JASA	2005	9

Table 11. Top 3 keywords of each community by three different methods

Comunity	SCORE	Modularity	FluidC
1	Sparsity(26)	Sparsity(54)	Sparsity(58)
	Variable selection(25)	Lasso(40)	Model selection(54)
	Nonparametric regression(22)	Model selection(35)	Lasso(45)
2	Model selection(26)	Bootstrap(37)	Bootstrap(39)
	Bootstrap(25)	Markov Chain Monte Carlo(26)	Markov Chain Monte Carlo(36)
	Markov Chain Monte Carlo(24)	Asymptotic normality(25)	Nonparametric regression(35)
3	Nonparametric regression(82)	Nonparametric regression(33)	Markov Chain Monte Carlo(37)
	Markov Chain Monte Carlo(80)	Variable selection(32)	Nonparametric regression(28)
	Model selection(79)	Markov Chain Monte Carlo(31)	Bootstrap(22)
4	Markov Chain Monte Carlo(25)	Markov Chain Monte Carlo(33)	Variable selection(44)
	Model selection(21)	Model selection(27)	Sparsity(43)
	Consistency(19)	Nonparametric regression(26)	Consistency(37)

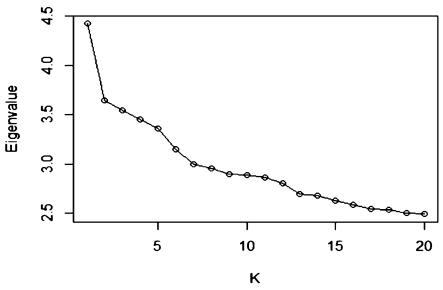
Figure 8 shows the community evolution from 2001 to 2008. For "False Discovery Rate" community, the size of this community rose slowly from 2001 to 2018, and the number of nodes (i.e. papers) increased from 225 to 703 in the period of 18 years. The in-degree of paper "A direct approach to false discovery rate" increases from 21 to 64 in the second period. However it only rises 26 in the third period. The slowdown trend in growth also appears in many other papers in this community, meaning that the popularity of this community is gradually slowing down.

We find that the paper "The control of the false discovery rate in multiple testing under dependency" is divided into "False Discovery Rate" community in the network from 2001-2012, and has the largest value of in-degree. However, in the network from 2001 to 2018, the paper is divided into "Variable Selection" community. This result may be due to the high closeness centrality of the paper (as shown in Table 4). It means that the paper is much closer to other papers in the network. In other words, the method proposed in this paper can be applied in many other statistical fields more directly, especially in the field of variable selection. As the rapid development of variable selection in recent years, more and more papers cites the paper, so it is possible to be divided into "Variable Selection" community.

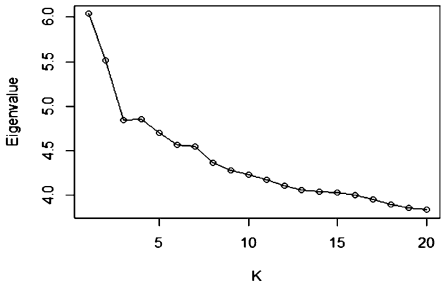
Nowadays, variable selection is a very "hot" area and develops rapidly. The number of nodes (i.e., papers) increased from 748 to 2,074 in the period of 18 years. Paper "Variable selection via nonconcave penalized likelihood and its oracle properties" is one of the critical papers in the area of variable selection, and its in-degree increases from 11 to 335. As Figure 8 shows, lasso gradually becomes a research hotspot and more and more popular in the area of variable selection from 2006. The paper "The adaptive lasso and its oracle properties" published in 2006 is a crucial work of the

Table 12. The result of community detection by D-SCORE for two periods

Period	Community	Title	Journal	Year	In-degree
2001-2006	Variable Selection & Dimension reduction & Nonparametric regression	Variable selection via nonconcave penalized likelihood and its oracle properties	JASA	2001	17
		Bayesian measures of model complexity and fit	JRSS-B	2002	17
		Least angle regression	AoS	2004	14
		Gibbs sampling methods for stick-breaking priors	JASA	2001	14
		Dimension reduction for conditional mean in regression	AoS	2002	13
		Semiparametric and nonparametric regression analysis of longitudinal data	JASA	2001	13
		Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics	JRSS-B	2001	13
	False Discovery Rate	The control of the false discovery rate in multiple testing under dependency	AoS	2001	26
		Empirical Bayes analysis of a microarray experiment	JASA	2001	26
		A direct approach to false discovery rates	JRSS-B	2002	21
		Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach	JRSS-B	2004	16
		Operating characteristics and extensions of the false discovery rate procedure	JRSS-B	2002	15
2001-2012	Variable Selection	Variable selection via nonconcave penalized likelihood and its oracle properties	JASA	2001	152
		Least angle regression	AoS	2004	108
		The adaptive lasso and its oracle properties	JASA	2006	102
		High-dimensional graphs and variable selection with the Lasso	AoS	2006	88
		The Dantzig selector: Statistical estimation when p is much larger than n	AoS	2007	67
	False Discovery Rate	The control of the false discovery rate in multiple testing under dependency	AoS	2001	70
		Empirical Bayes analysis of a microarray experiment	JASA	2001	65
		A direct approach to false discovery rate	JRSS-B	2002	64
		A stochastic process approach to false discovery control	AoS	2004	45
		Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach	JRSS-B	2004	44
	Dimension reduction	An adaptive estimation of dimension reduction space	JRSS-B	2002	39
		Dimension reduction for conditional mean in regression	AoS	2002	34
		Dimension reduction for the conditional kth moment in regression	JRSS-B	2002	24
		Sufficient dimension reduction via inverse regression: A minimum discrepancy approach	JASA	2005	23
		Contour regression: A general approach to dimension reduction	AoS	2005	21



(a) The First Period (2001 to 2006)



(b) The Second Period (2001 to 2012)

Figure 7. Scree Plots of two period of networks.

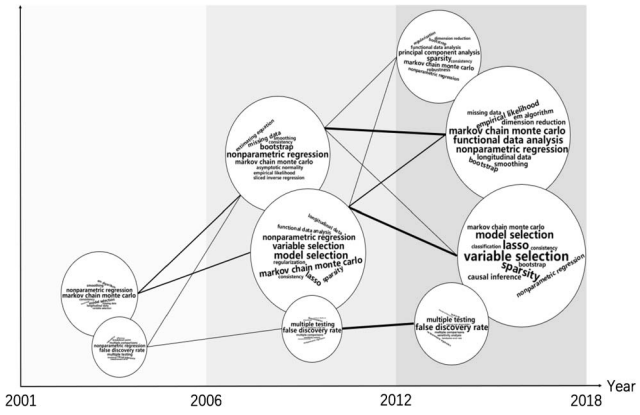


Figure 8. Community evolution from 2001 to 2008.

lasso. Its in-degree has increased to 208 in the period of 12 years. Besides, Markov Chain Monte Carlo, dimension reduction and functional data analysis are also current research hotspots.

5. CONCLUSION

In this study, we collect papers from 2001 to 2018 published in the top 4 statistical journals and analyze the citation network for these papers. To explore the characteristics of the network, we investigate the development of productivity, in-degree distribution and centrality. The results suggest that authors are becoming more and more collaborative. Highly cited papers that lead the statistical development are relatively rare. Papers published in AoS tend to have high betweenness centrality and closeness centrality. Then, we apply the D-SCORE method to community detection of the citation network and detect 4 communities and 11 sub-communities. 11 sub-communities include but are not limited to "High-dimensional Model", "Variable Selection", "Covariance Matrix Analysis", "Markov Chain Monte Carlo", "Dirichlet Process" and "False Discovery Rate". We also compare the community detection results of different methods, including SCORE, FluidC, greedy modularity algorithm. The result suggests that these three methods are not as suitable as D-SCORE for the citation network. Finally, we focus on the dynamic nature of the communities. We find that the lasso becomes a research hotspot and more and more popular in "Variable selection" community nowadays and "False Discovery Rate" community is on the rise.

Several directions for future research are possible. First, constrained by computer resources, the data we collect is limited to the papers published in the top 4 statistical journals: AoS, Biometrika, JASA, and JRSS-B from 2001 to 2018. We recognize that many papers are not only published in these top 4 statistical journals, but also published in other journals such as Nature, Science, Statistical Science, and Physical Review. Besides, papers published in the top 4 statistical journals cite papers from other journals in addition to papers from the top 4 statistical journals. For these reasons, some of the results presented in this paper should be interpreted with caution. Second, the importance of a paper is measured by in-degree, betweenness centrality and closeness centrality in this paper. However, the rankings produced by these indicators are biased by age. It is of interest to explore better indicators to measure the importance of papers. Third, although the D-SCORE method removes the effect of degree heterogeneity, several other characteristics should be considered in natural network research. One is mixed-memberships [35]. Some papers focus not only on a single topic, but on multiple topics, so it is more reasonable to assign these paper into different communities. The other is sparsity. The sparsity levels may range significantly from one network to another, and may also range significantly from one node to another [35]. We are going to do further research on those features in our future work.

Another problem of great interest is global testing, which tests whether the network has only one community or there are more than one community. Recently, the global testing problem becomes more and more popular and many approaches have been proposed. [24] developed the Erdős-Zuckerberg (EZ) test for the global community structure. [34] proposed a class of test statistics called the graphlet counting (GC), which includes the EZ test as a special case. However, both tests are not optimally adaptive. To solve that problem, [35] proposed a class of new tests called Signed Polygon, including the Signed Triangle (SgnT) and the Signed Quadrilateral (SgnQ), which outperform EZ and GC, especially in the less sparse case. Global testing can

be combined with methods in community detection to estimate the number of communities. Therefore, being able to conduct global testing can offer fruitful insights about the network structure. It will be a topic to consider in our future research.

Received 25 February 2020

REFERENCES

[1] Amasino, D. R., Sullivan, N. J., Kranton, R. E. and Huettel, S. A. (2019). Amount and time exert independent influences on intertemporal choice. *Nature Human Behaviour* **3** 383–392.

[2] An, W. and Ding, Y. (2018). The landscape of causal inference: perspective from citation network analysis. *The American Statistician* **72** 265–277. MR3836449

[3] Bang-Jensen, J. and Guting, G. (2009). *Digraphs: Theory, Algorithms and Applications*. Springer, London. MR2472389

[4] Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *Science* **286** 509–512. MR2091634

[5] Bickel, P. J. and Doksum, K. A. (2015). *Mathematical Statistics: Basic Ideas and Selected Topics, Volumes I-II Package*. Chapman and Hall/CRC. MR3287337

[6] Bickel, P. J. and Sarkar, P. (2015). Hypothesis testing for automated community detection in networks. *Journal of the Royal Statistical Society: Series B,(Statistical Methodology)* **1** 253–273. MR3453655

[7] Chen, K. and Lei, J. (2018). Network cross-validation for determining the number of communities in network data. *Journal of the American Statistical Association* **113** 241–251. MR3803461

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

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

[10] Clauset, A., Shalizi, C. R. and Newman, M. E. (2009). Power-law distributions in empirical data. *SIAM Review* **51** 661–703. MR2563829

[11] Danon, L., Diaz-Guilera, A., Duch, J. and Arenas, A. (2005). Comparing community structure identification. *Journal of Statistical Mechanics: Theory and Experiment* **2005** 09008.

[12] Daudin, J.-J., Picard, F. and Robin, S. (2008). A mixture model for random graphs. *Statistics and Computing* **18** 173–183. MR2390817

[13] Deelen, J., Kettunen, J., Fischer, K., van der Spek, A., Trompet, S., Kastenmüller, G., Boyd, A., Zierer, J., van den Akker, E. B., Ala-Korpela, M. et al. (2019). A metabolic profile of all-cause mortality risk identified in an observational study of 44,168 individuals. *Nature Communications* **10** 1–8.

[14] Dorie, V., Hill, J., Shalit, U., Scott, M. and Cervone, D. (2019). Automated versus Do-It-Yourself Methods for Causal Inference: Lessons Learned from a Data Analysis Competition. *Statistical Science* **34** 43–68. MR3938963

[15] Duan, Y., Ke, T. and Wang, M. (2019). State aggregation learning from markov transition data. In *Advances in Neural Information Processing Systems* 4488–4497.

[16] Efron, B. (2019). Bayes, Oracle Bayes and Empirical Bayes. *Statistical Science* **34** 177–201. MR3983318

[17] Fahimnia, B., Sarkis, J. and Davarzani, H. (2015). Green supply chain management: A review and bibliometric analysis. *International Journal of Production Economics* **162** 101–114.

[18] Fan, J. and Li, R. (2001). Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties. *Journal of the American Statistical Association* **96** 1348–1360. MR1946581

[19] Fan, J. and Li, R. (2006). Statistical Challenges with High Dimensionality: Feature Selection in Knowledge Discovery. *In Proc. Int Congr. Mathematicians (eds M. Sanz-Sole, J. Soria, J. L. Varona and J. Verdera)* **3** 595–622. MR2275698

[20] Ferguson, T. S. (1973). A Bayesian Analysis of Some Nonparametric Problems. *Annals of Statistics* **1** 209–230. MR0350949

[21] Fortunato, S. (2009). Community detection in graphs. *Physics Reports* **486** 75–174. MR2580414

[22] Fowler, M. D., Kooperman, G. J., Randerson, J. T. and Pritchard, M. S. (2019). The effect of plant physiological responses to rising CO 2 on global streamflow. *Nature Climate Change* **9** 873–879.

[23] Freeman, L. C., Borgatti, S. P. and White, D. R. (1991). Centrality in valued graphs: A measure of betweenness based on network flow. *Social Networks* **13** 141–154. MR1135768

[24] Gao, C. and Lafferty, J. (2017). Testing for global network structure using small subgraph statistics. *arXiv preprint arXiv:1710.00862*.

[25] Gini, C. (1936). On the measure of concentration with special reference to income and statistics. *Colorado College Publication, General Series* **208** 73–79.

[26] Girvan, M. and Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America* **99** 7821–7826. MR1908073

[27] Holland, P. W. (1986). Statistics and Causal Inference. *Journal of the American Statistical Association* **81** 945–960. MR0867618

[28] 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.

[29] Hu, J., Qin, H., Yan, T. and Zhao, Y. (2019). Corrected Bayesian information criterion for stochastic block models. *Journal of the American Statistical Association* 1–13.

[30] Ji, P. and Jin, J. (2016). Coauthorship and Citation Networks for Statisticians. *The Annals of Applied Statistics* **10** 1779–1812. MR3592033

[31] Jin, J. (2015). Fast community detection by SCORE. *Annals of Statistics* **43** 57–89. MR3285600

[32] Jin, J., Ke, Z. T. and Luo, S. (2017). Estimating network memberships by simplex vertex hunting. *arXiv preprint arXiv:1708.07852*.

[33] Jin, J., Ke, Z. T. and Luo, S. (2018). SCORE+ for network community detection. *arXiv preprint arXiv:1811.05927*.

[34] Jin, J., Ke, Z. T. and Luo, S. (2018). Network global testing by counting graphlets. *arXiv preprint arXiv:1807.08440*.

[35] Jin, J., Ke, Z. T. and Luo, S. (2019). Optimal adaptivity of signed-polygon statistics for network testing. *arXiv preprint arXiv:1904.09532*.

[36] Ke, Z. T., Shi, F. and Xia, D. (2019). Community Detection for Hypergraph Networks via Regularized Tensor Power Iteration. *arXiv preprint arXiv:1909.06503*.

[37] Ke, Z. T. and Wang, M. (2017). A new SVD approach to optimal topic estimation. *arXiv preprint arXiv:1704.07016*.

[38] Kim, Y., Son, S.-W. and Jeong, H. (2010). Finding communities in directed networks. *Physical Review E* **81** 016103.

[39] Latouche, P., Birmele, E. and Ambroise, C. (2012). Variational Bayesian inference and complexity control for stochastic block models. *Statistical Modelling* **12** 93–115. MR2953099

[40] Le, C. M. and Levina, E. (2015). Estimating the number of communities in networks by spectral methods. *arXiv preprint arXiv:1507.00827*.

[41] Lei, J. et al. (2016). A goodness-of-fit test for stochastic block models. *The Annals of Statistics* **44** 401–424. MR3449773

[42] Leicht, E. A. and Newman, M. E. (2008). Community structure in directed networks. *Physical Review Letters* **100** 118703.

[43] Lim, K. W. and Buntine, W. (2016). Bibliographic Analysis with the Citation Network Topic Model. *arXiv: Digital Libraries*.

[44] Liu, S., Wang, S. and Krishnan, R. (2014). Persistent community detection in dynamic social networks. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining* 78–89. Springer.

[45] Ma, S., Su, L. and Zhang, Y. (2018). Determining the number

of communities in degree-corrected stochastic block models. *arXiv preprint arXiv:1809.01028*.
[46] Matias, C. and Miele, V. (2016). Statistical clustering of temporal networks through a dynamic stochastic block model. *Journal of the Royal Statal Society. Series B (Statal Methodology)* **60** 12–31. MR3689311
[47] McDaid, A. F., Murphy, T. B., Friel, N. and Hurley, N. J. (2013). Improved Bayesian inference for the stochastic block model with application to large networks. *Computational Statistics & Data Analysis* **60** 12–31. MR3007016
[48] Meinshausen, N. and Buhlmann, P. (2006). High-dimensional graphs and variable selection with the Lasso. *Annals of Statistics* **34** 1436–1462. MR2278363
[49] Moon, J., Ma, W., Shin, J. H., Cai, F., Du, C., Lee, S. H. and Lu, W. D. (2019). Temporal data classification and forecasting using a memristor-based reservoir computing system. *Nature Electronics* **2** 480–487.
[50] Nakazawa, R., Itoh, T. and Saito, T. (2015). A visualization of research papers based on the topics and citation network. In *2015 19th International Conference on Information Visualisation* 283–289.
[51] Newman, M. E. (2001). Scientific collaboration networks. I. Network construction and fundamental results. *Physical Review E* **64** 016131. MR1975193
[52] Newman, M. E. J. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences of the United States of America* **103** 8577–8582.
[53] Newman, M. E. J. and Leicht, E. A. (2007). Mixture models and exploratory analysis in networks. *Proceedings of the National Academy of Sciences of the United States of America* **104** 9564–9569.
[54] Palla, G., Barabási, A.-L. and Vicsek, T. (2007). Quantifying social group evolution. *Nature* **446** 664–667.
[55] Parés, F., Gasulla, D. G., Vilalta, A., Moreno, J., Ayguadé, E., Labarta, J., Cortés, U. and Suzumura, T. (2017). Fluid communities: a competitive, scalable and diverse community detection algorithm. In *International Conference on Complex Networks and their Applications* 229–240. Springer.
[56] Peixoto, T. P. (2013). Parsimonious module inference in large networks. *Physical Review Letters* **110** 148701.
[57] Ramasco, J. J. and Mungan, M. (2008). Inversion method for content-based networks. *Physical Review E* **77** 036122. MR2495435
[58] Ritov, Y. and Bickel, P. J. (1990). Achieving Information Bounds in Non and Semiparametric Models. *Annals of Statistics* **18** 925–938. MR1056344
[59] Rosvall, M. and Bergstrom, C. T. (2007). An information-theoretic framework for resolving community structure in complex networks. *Proceedings of the National Academy of Sciences* **104** 7327–7331.
[60] Rosvall, M. and Bergstrom, C. T. (2008). Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences of the United States of America* **105** 1118–1123.
[61] Sabidussi, G. (1966). The centrality index of a graph. *Psychometrika* **31** 581–603. MR0205879
[62] Saldana, D. F., Yu, Y. and Feng, Y. (2017). How many communities are there? *Journal of Computational and Graphical Statistics* **26** 171–181. MR3610418
[63] Vandermeulen, R. A. and Scott, C. (2019). An operator theoretic approach to nonparametric mixture models. *Annals of Statistics* **47** 2704–2733. MR3988770
[64] Varin, C., Cattelan, M. and Firth, D. (2016). Statistical modelling of citation exchange between statistics journals. *Journal of the Royal Statistical Society. Series A,(Statistics in Society)* **179** 1. MR3461568
[65] Velden, T., Yan, S. and Lagoze, C. (2017). Mapping the cognitive structure of astrophysics by infomap clustering of the citation network and topic affinity analysis. *Scientometrics* **111** 1033–1051.
[66] Volpp, L. (2006). Complex networks: Structure and dynamics. *Physics Reports* **424** 175–308. MR2193621
[67] Wang, Y. R., Bickel, P. J. et al. (2017). Likelihood-based model selection for stochastic block models. *The Annals of Statistics* **45** 500–528. MR3650391
[68] Xiang, S., Yao, W. and Yang, G. (2019). An Overview of Semiparametric Extensions of Finite Mixture Models. *Statistical Science* **34** 391–404. MR4017520
[69] Yan, T., Leng, C., Zhu, J. et al. (2016). Asymptotics in directed exponential random graph models with an increasing bi-degree sequence. *The Annals of Statistics* **44** 31–57. MR3449761
[70] Yang, T., Chi, Y., Zhu, S., Gong, Y. and Jin, R. (2010). Directed network community detection: A popularity and productivity link model. In *Proceedings of the 2010 SIAM International Conference on Data Mining* 742–753.
[71] Yang, T., Chi, Y., Zhu, S., Gong, Y. and Jin, R. (2011). Detecting communities and their evolutions in dynamic social networks – a Bayesian approach. *Machine Learning* **82** 157–189. MR3108191
[72] Zhang, J. and Cao, J. (2017). Finding common modules in a time-varying network with application to the Drosophila melanogaster gene regulation network. *Journal of the American Statistical Association* **112** 994–1008. MR3735355
[73] Zhao, Y., Levina, E. and Zhu, J. (2011). Community extraction for social networks. *Proceedings of the National Academy of Sciences* **108** 7321–7326.
[74] Zou, H. (2006). The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association* **101** 1418–1429. MR2279469

Tianchen Gao
School of Statistics and Mathematics
Central University of Finance and Economics
Beijing
P. R. China
E-mail address: gaotianchen_cufe@qq.com

Yan Zhang
School of Statistics and Mathematics
Central University of Finance and Economics
Beijing
P. R. China
E-mail address: 2019210863@email.cufe.edu.cn

Siyu Wang
School of Statistics and Mathematics
Central University of Finance and Economics
Beijing
P. R. China
E-mail address: 240668046@qq.com

Yuehan Yang
School of Statistics and Mathematics
Central University of Finance and Economics
Beijing
P. R. China
E-mail address: yyh@cufe.edu.cn

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