Accepted Manuscript

Exploring the heterogeneity for node importance by von Neumann entropy

Xiangnan Feng, Wei Wei, Renquan Zhang, Jiannan Wang, Ying Shi, Zhiming Zheng

PII: S0378-4371(18)31427-4
DOI: https://doi.org/10.1016/j.physa.2018.11.019
Reference: PHYSA 20337

To appear in: *Physica A*

Received date : 31 July 2018
Revised date : 24 September 2018

*Highlights (for review)

Highlights:

- ✓ A new definition of node heterogeneity measurement based on von Neumann entropy
- ✓ Examples and experiments on real-world networks
- ✓ An approximation method to calculate the entropy
- ✓ Experiments on reducing the Estrada heterogeneity index
- ✓ Experiments on reducing the average clustering coefficient

Exploring the Heterogeneity for Node Importance by von Neumann Entropy

Xiangnan Feng[a,b], Wei Wei[a,b,c,*], Renquan Zhang[d], Jiannan Wang[a,b], Ying Shi[a,b], and Zhiming Zheng[a,b,c]

[a] *School of Mathematics and Systems Science, Beihang University, Beijing 100191, China*
[b] *Key Laboratory of Mathematics Informatics Behavioral Semantics, Ministry of Education 100191, China*
[c] *Beijing Advanced Innovation Center for Big Data and Brain Computing, Beihang University, Beijing 100191, China*
[d] *School of Mathematical Sciences, Dalian University of Technology, Dalian 116024, China*
[*] *weiw@buaa.edu.cn*

Abstract

When analyzing and describing the statistical and topological characteristics of complex networks, the heterogeneity can provide profound and systematical recognition to illustrate the difference of individuals, and many node significance indices have been investigated to describe heterogeneity in different perspectives. In this paper a new node heterogeneity index based on the von Neumann entropy is proposed, which allows us to investigate the differences of nodes features in the view of spectrum eigenvalues distribution, and examples in reality networks present its great performance in selecting crucial individuals. Then to lower down the computational complexity, an approximation calculation to this index is given which only depends on its first and second neighbors. Furthermore, in reducing the network heterogeneity index by Estrada, this entropy heterogeneity presents excellent efficiency in Erdös-Rényi and scale-free networks compared to other node significance measurements; in reducing the average clustering coefficient, this node entropy index could break down the cluster structures efficiently in random geometric graphs, even faster than clustering coefficient itself. This new methodology reveals the node heterogeneity and significance in the perspective of spectrum, which provides a new insight into networks research and performs great potentials to discover essential structural features in networks.

Keywords: Complex Network; Heterogeneity; Entropy

1. Introduction

Networks provide us a useful tool to analyze a wide range of complex systems, including WWW [1], the social structure [2], the economic behaviors [3], and the biochemical reactions [4]. Since the 1990's, a great number of interdisciplinary studies involving network both in theories and empirical work, have come up and developed new models and techniques to shed a light on the complex structure behind the particular subjects.

To extract the characteristics from networks, a number of indices have been created to illustrate the topological and statistical features of networks. Among these studies, to analyze the structural complexity, heterogeneity has attracted a lot of attention. In order to describe the heterogeneity of complex networks, it is necessary to find computationally efficient methods to measure it. Snijders [5] and Bell [6] used the variance of node degrees to measure the heterogeneity of networks, which was regarded as the first measurement of the network heterogeneity. Albertson [7] proposed that the sum of differences of degrees of nodes on the same edges could be applied to work as the heterogeneity measurement. The Gini coefficient [8] of degree distribution in networks serves as a great heterogeneity of networks, which has been widely used in the economics and sociology as the measurement of inequality. Jacob et al. [9] got a new heterogeneity index based on the distribution and creatively used it to compare and quantify the structural complexity of different chaotic attractors in the recurrence networks. Estrada [10] proposed a unique measurement of heterogeneity which is based on the differences of function of degrees and this index could be represented by the Laplacian matrix of the network, which means this heterogeneity index could be expressed by the spectrum. Later, Hancock et al. [11] compared the von Neumann entropy with Estrada's heterogeneity index and concluded that the entropy could work as a better classifier for networks and also performed the features of the eigenvalues distribution in different networks. These heterogeneity indices are widespread in the complex networks research.

Another crucial subject in complex network which has received considerable attention is measuring the significance or importance of nodes. A number of methods distinguishing different individuals on a large-scale system have been proposed to solve this problem and may of them could be viewed as descriptions of heterogeneity of nodes in their own perspectives. The degree of nodes [37] is a natural description of node difference concerning the number of its neighbors, and high-degree nodes provide great heterogeneity under some average degree. The clustering coefficient [12, 13] reflects the clustered patterns concerning some local structures, which produces another useful tool to measure the heterogeneity for different targets. Besides, there are many such indices to illustrate the variance in connection and function in the network, e.g. the PageRank by Larry Page et al. [17] would demonstrates the heterogeneity of node neighbors links and qualities, and the collective influence [18] could reflect the heterogeneity of nodes on the biggest eigenvalue of non-backtracking matrix.

In this paper, based on the performance of the von Neumann entropy in measuring the network heterogeneity, we propose to define and analyze the entropy heterogeneity in the view point of individuals, which we find that could be used as index of node importance or significance in the networks. In section 2 the von Neumann entropy and its ability in measuring the heterogeneity of networks are introduced. Commencing from this entropy, the heterogeneity of nodes is defined with some examples, and the specific calculation related to the entropy, including approximation, is presented. Next in section 3 some experiments are implemented to show the efficiency of out proposed index in reducing Estrada heterogeneity and average clustering coefficient. These researches build the heterogeneity for microscopic objects in networks with the spectrum and demonstrate their superior in describing the roles played by nodes in a new perspective.

2. Von Neumann Entropy and Heterogeneity

2.1. Spectral Distribution of Eigenvalues

Given an undirected network $G(V, E)$, V (or $V(G)$) is a finite set whose elements are nodes of the network G and E (or $E(G)$) is the edges set. E is composed of unordered pairs of nodes who belong to V, namely, when $(v_i, v_j) \in E$, we have $(v_j, v_i) \in E$ and $v_i, v_j \in V$. The edge in the form of (v_i, v_i) is called a self-loop. In this paper we only talk about the networks without self-loops. The *adjacency matrix* is an $N \times N$ matrix, where $N = |V|$. Using $A(G)$ to denote the adjacency matrix of G, the columns and rows of $A(G)$ are labeled by the vertices of G, and the (i, j) entry of $A(G)$ is 1 if and only if $(v_i, v_j) \in E(G)$, namely the adjacency matrix $A(G)$ could be defined as follows:

$$[A(G)]_{i,j} = \begin{cases} 1 & \text{if } (v_i, v_j) \in E, \\ 0 & \text{if } (v_i, v_j) \notin E. \end{cases} \tag{1}$$

Before the introduction of von Neumann entropy, firstly the normalized Laplacian matrix is introducted [19]. The degree of a vertex $v_i \in G$, denoted as $d_G(v_i)$ or d_i, is the total number of edges touching this vertex. In this way we could define the *degree matrix* which is an $N \times N$ diagonal matrix and denoted as $D(G)$. The entries in the degree matrix are defined as follows:

$$[D(G)]_{i,j} = \begin{cases} d_G(v_i) & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases} \tag{2}$$

The *combinatorial Laplacian matrix* $L(G)$ could be define as $L(G) = D(G) - A(G)$:

$$[L(G)]_{i,j} = \begin{cases} d_G(v_i) & \text{if } i = j, \\ -1 & \text{if } i \neq j,\ (v_i, v_j) \in E, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

It is worth noting that the Laplacian matrix will not change if the self-loop is added or deleted. As we can see, the Laplacian matrix is a diagonally dominant Hermite matrix, thus it is positive semi-defined [20].

The *normalized Laplacian matrix* is defined as $\mathcal{L} = D^{-1/2} L D^{-1/2}$ and the elements are:

$$[\mathcal{L}(G)]_{i,j} = \begin{cases} 1 & \text{if } i = j,\ d_G(v_i) \neq 0, \\ -\frac{1}{\sqrt{d_G(v_i) d_G(v_j)}} & \text{if } i \neq j,\ (v_i, v_j) \in E, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

The spectral decomposition of $\mathcal{L}(G)$ is $\mathcal{L}(G) = \Phi \Lambda \Phi$, where $\Lambda = diag(\lambda_i)_{i=1}^{N}$ is a diagonal matrix of eigenvalues with order $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N$ and Φ is a matrix whose columns are orthonormal eigenvectors corresponding to the ordered eigenvalues. Notice that the normalized Laplacian matrix is also semi-defined so all the eigenvalues are non-negative.

For the Laplacian matrix, one of the most important indices is the eigenvalues and they are directly related to the topological properties of the network: the number of eigenvalues equalling to zero is the number of connected components in this network and there is only one zero-eigenvalue for a connected network; λ_2 is referred as the algebraic connectivity and the corresponding eigenvector is known as Fiedler vector [21] [22], which is frequently used to network partition [23]. By the way, for the normalized Laplacian matrix, all the eigenvalues satisfy $0 \leq \lambda_i \leq 2$, $1 \leq i \leq N$ and the upper limit 2 is achieved only when the network is bipartite. Evaluating the accurate ranges of each eigenvalue is still an open problem.

Since the eigenvalues are crucial features of a matrix, we believe the distribution of the spectrum (distribution) is fatal to a network concerning the topological characteristics. Since there is a one-to-one mapping between the normalized Laplacian matrices and networks, this matrix contains all the topological characteristics of a network, thus the spectrum eigenvalues could be viewed as a natural data reduction of information in statistical viewpoint. If different nodes are removed from the network, various changes will be brought to its spectrum distribution. An example of Zachary's karate club network [24] is show in Figure 1. Considering the roles played by node 34 and node 1 in the club, they are more significant than node 3, thus removing node 34 or 1 will bring larger changes on the spectrum than removing node 3. Capturing the variations in spectrum distribution will lead to a significant understanding in structural changes of the network.

2.2. Entropy and Node Heterogeneity

As a crucial way of depicting distributions, entropy could be used to signify the features of spectrum distribution [19]. The von Neumann entropy, commencing from normalized Laplacian matrix, could be regarded as a well-designed and sophisticated representation of network. There are many researches related to von Neumann entropy and its function in describing the network structure, which receives quite a lot of attention in many applications [25] [4] [15]. This index integrates the complete values and properties of all the eigenvalues and thus could reflect global structural complexity and characteristics.

The von Neumann entropy of a network G associated with its normalized Laplacian matrix $\mathcal{L}(G)$, denoted as $S(G)$, is defined as [25] [19] [26]:

$$S(G) = -\sum_{i=1}^{N} \frac{\lambda_i}{2} \ln \frac{\lambda_i}{2}, \tag{5}$$

where $\lambda \log \lambda = 0$ when $\lambda = 0$.

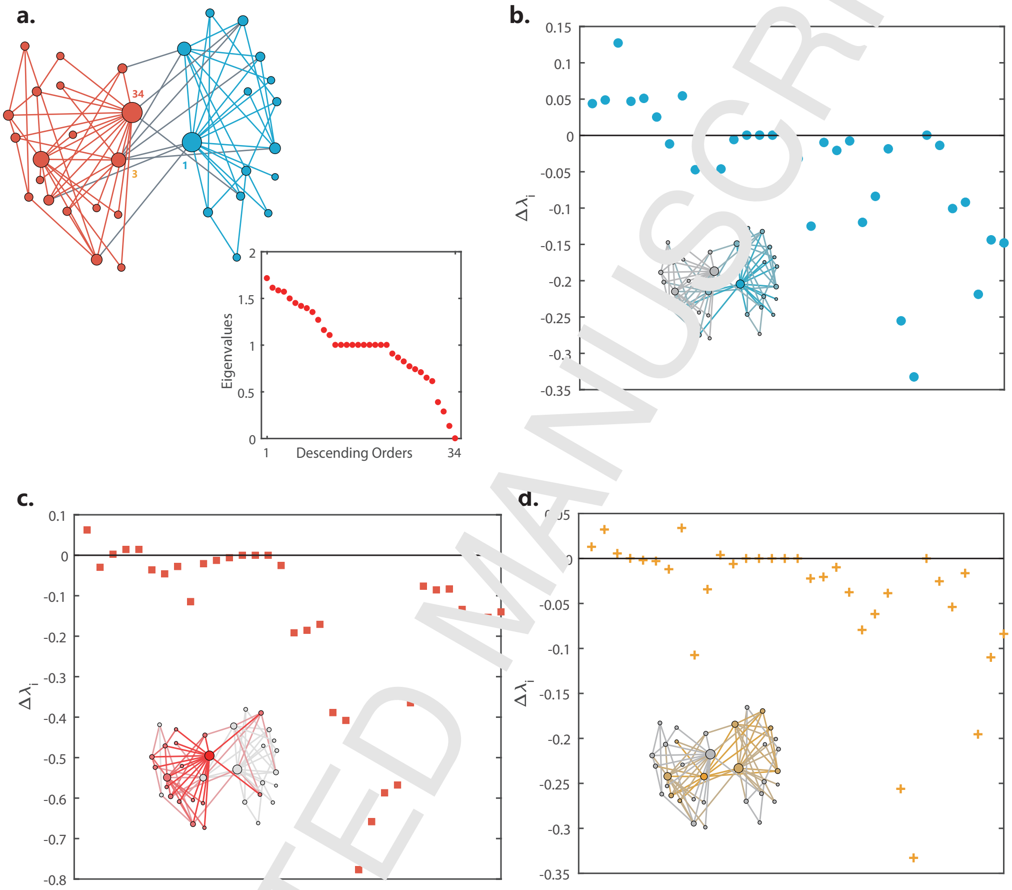

Figure 1: Zachary's karate club network. There are 34 members in the club and 78 links outside the club. Node 1 is the instructor and node 34 is the club administrator or president. A conflict has happened between the instructor and administrator during the study, which led to the split of the club. We could see that removing different nodes will bring different changes to the spectrum. Different nodes are marked by different colors. **a.** The connection relationship between the members. The red dots are the eigenvalues of the Laplacian matrix in descending order. **b.** The changes of eigenvalues after node 1 (blue) is removed. The eigenvalues are sort in decreasing order and the points in the graph stands for the variations of eigenvalues. **c.** The changes of eigenvalues after node 34 (red) is removed. **d.** The changes of eigenvalues after node 3 (yellow) is removed.

First proposed in thermodynamics, entropy has been widely used to measure the orderliness of systems. The von Neumann entropy (or quantum entropy) has shown great success in qualifying the organization structure and levels in networks, and can be applied in networks as an index to quantify the network heterogeneous (homogeneous) characteristics. Although there are plenty indices which describe this characteristic, like the variance of degrees, they are confined to the degree distribution and neglect the specific connection patterns and structure in the networks. For networks with similar degree distribution yet different connection, these indices would not accurately reflect the intrinsic topological structure. Experiments by Han et al. [11] supports this idea. In their work, networks of 5 objects form the COIL dataset [27] are selected and the von Neumann entropy and Estrada Heterogeneity of each networks are calculated. For each object, the networks are even and close to regular networks and they own small heterogeneity values. For different objects, the heterogeneity values of each networks are quite similar, which makes it hard to distinguish the 5 objects by the Estrada heterogeneity values. Yet the von Neumann entropy performs good efficiency and is able to classify the objects much more accurately. It is believed that as the representation of the spectrum distribution of eigenvalues, the von Neumann entropy would work as an effective measurement of network heterogeneity.

Commencing from this spectrum heterogeneity, we propose the node heterogeneity for each node in networks. Accordingly, the heterogeneity of node v can be defined as the variation of von Neumann Entropy when removing this node and edges linked to it from the network. Using $H_E(v)$ to denote the *node entropy heterogeneity* of node v, we have

$$H_E(v) = |S(G) - S(G \setminus v)|. \tag{6}$$

Similarly, the heterogeneity could be defined on other structures in graph. Let s be a subnetwork of G, and denote $G \setminus s$ to be the network remained after deleting the nodes in s and edges linked with these nodes. The entropy heterogeneity of subnetwork s in network G could be defined as:

$$H_E(s) = |S(G) - S(G \setminus s)|. \tag{7}$$

For a problem with global targets, many of them are NP hard and for a lot of cases, it is hard to find an ultimate solution. To solve these problems, there are many methods focusing on the global or local targets. The strategies with global aims often work better, for example the collective influence algorithm and its local generalization in identifying influential spreaders. Similar to this, our target here is to lower the global heterogeneity of the network. It is a complicated target considering the huge number of possibilities. To achieve this goal, the von Neumann entropy is applied and localized to decompose the global target into local ones, namely, to decompose the global heterogeneity into the heterogeneity of nodes or sub-component. Based on this, the node heterogeneity is proposed to decouple the global heterogeneity represented by the von Neumann entropy into local structure. Usually the heterogeneity refers to the state of a global or sub-component network and it is impossible to define the heterogeneity for a single node or individual. Here, the H_E of one node is applied to present the roles play by it in the whole network and this definition is based on the interactions this node have involved. It is proposed to indicate the influence of this node to the irregularity of the network and the value depends on the global network and its local connection to the neighbors with various orders.

We believe that the importance and significance of a node originate from its heterogeneity in the network. For a regular graph with all nodes owning similar degrees and other characteristics like betweenness and closeness, they will undoubtedly have similar importance and it is hard to discriminate their roles played in the network. However, in an uneven network, when there exist nodes that perform high irregularity, like owing extremely high degrees or playing crucial “bridging” roles, the significance of these nodes stands out. A number of node indices focus on these features and different indices emphasis on different ones. These features could be regarded as the heterogeneity in different perspectives since they make some nodes different from others and nodes with high significance in networks would express high heterogeneity. The von Neumann entropy, which is a useful heterogeneity measurement, can reflect the regularity and complexity of the network and is effective to characterize the global structure of networks. When a node is removed, the network will change, which leads to the change of the Laplacian matrix and its eigenvalues, thus the von Neumann entropy of the network will finally change. If deleting node x brings larger change of $S(G)$ than deleting node y, namely node x owns higher heterogeneity than node y, it proves that deleting node x could cause more significant change on the spectrum distribution and network structure. Since the relationship between the network and its spectrum is elaborate and profound, when removing a node from the network, the change in von Neumann entropy will exactly present the impact of this node on the whole network structure, which makes the entropy heterogeneity of nodes a great index for node importance and significance.

2.3. Some Examples

To further analyze the node entropy heterogeneity, some specific networks are used to perform and compare different node centralities in this subsection. The results of betweenness centrality (BC) [16], closeness centrality (CC) [14, 15], degree centrality (DC) and the entropy heterogeneity on Padgett Florentine families network and the gift-giving network are shown in Figure 2.

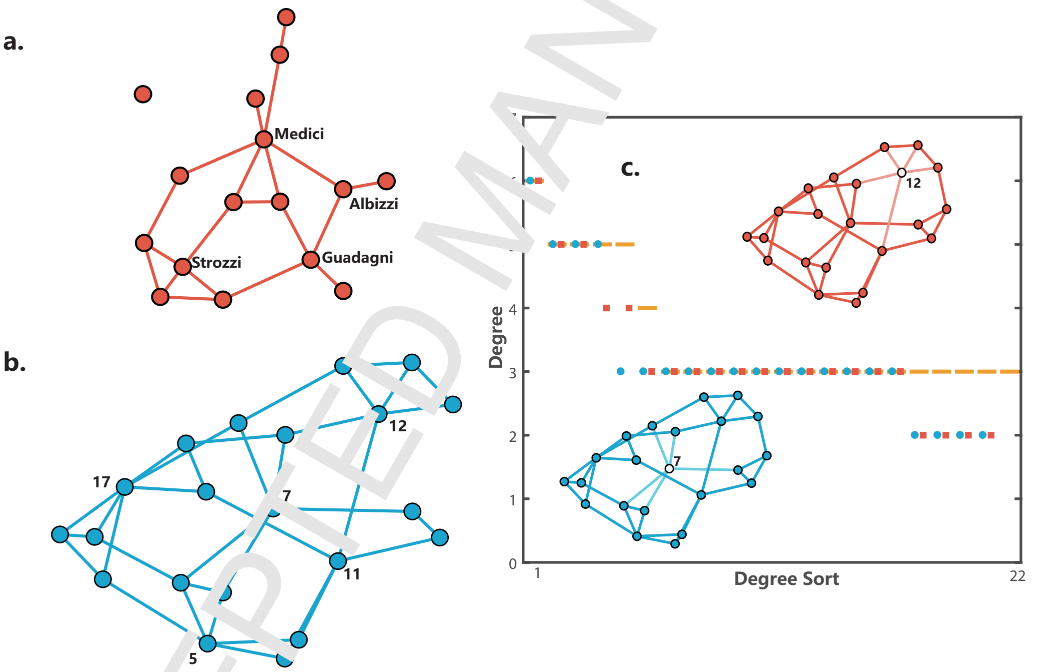

Figure 2: **a.** Padgett Florentine families marital ties network. 16 nodes and 20 edges are contained in this network. The Pucci family did not have martial tie with others, so the major part of the network is a component with 15 nodes. **b.** The gift-exchange network in a Papuan village. There are 22 nodes and 39 edges. Each node stands for a household and each edge stands for gift exchange. **c.** The degree distributions before and after the removal of node 7 or node 12 in the gift-exchange network. The x axis is the nodes sorted by degree in decreasing order. The oranges short lines stand for the degrees of nodes in original network. Blue circles are the degrees of nodes in the network with node 7 removed. Red squares are the degrees of nodes in the network with node 12 removed.

The Padgett Florentine families network is a network of marital ties among Renaissance Florentine families [28, 29]. This network is built based on historical documents and an edge between two nodes means there existed marriage alliance between the two corresponding families. The network includes families who were involved in the struggle for the control of the city in politics around 1430s. 16 families are contained in the network and there is a major component consisted by 15 of them. The ranks of each index is shown in Table 1. As we could see, all the indices rank the Medici as the most influential one. This actually coincides with historical fact since the Medici family is one of the most famous families in history who reached peak in Italian upper classes during Renaissance. The node entropy heterogeneity is able to find out the most influential families correctly as others: all the first three families in the H_E sort appear in other three sorts and the Medici family has the highest rank. This reflects that the node heterogeneity based on the von Neumann entropy could work as a new reasonable and accurate index of node importance in the network and is able to exactly capture the nodes which work as the most influential ones and are crucial to the whole network.

Table 1: Ranks of nodes of Padgett Florentine families marital ties network in BC, CC, DC and H_E.

Ranks	BC	CC	DC	H_E
1st	Medici	Medici	Medici	Medici
2nd	Guadagni	Ridolfi	Guadagni/Strozzi	Guadagni
3rd	Albizzi	Albizzi/Tornabuon	Albizzi/Bischeri/...	Albizzi

The other example, the gift-giving network, shows the gift exchange relations among 12 households in a Papuan village [30, 31]. In this network, if two households exchange gifts, there will be an edge between them. In this village, the gift-exchanging is significant in life because it is regarded as a method to request political and economic assistance from others and works as the pristine market. Although there may exist deep contents and meanings behind the whole process in the network, yet it is natural to realize that the family who exchanges gifts with more persons and have higher degrees may have larger influence on the whole village. At the same time, since the exchange process could be long and complicated, like the family A may ask family B to ask family C to assist A, the betweenness and closeness will also point out influential households or persons in the network. Thus it is incomplete to evaluate the network with only one single index and multiple heterogeneity measurements are required to help understand the structure and information behind the network better. As shown in the Table 2, the entropy heterogeneity of nodes performs its potential to work as an all-round index on the node significance: the first five nodes in H_E sort have highest ranks in other sorts, like node 11 ranks first in BC sort and CC sort, node 12 ranks the second in DC sort. The all-round property of H_E allows us to find more meaningful information in the network.

Table 2: Ranks of nodes of gift-giving network in BC, CC, DC and H_E.

Ranks	BC	CC	DC	H_E
1st	11	11	17	17
2nd	7	7	5/7/11/12	11
3rd	17	1[illegible]/19	4	12
4th	12	12/16	1/2/3...	7
5th	5	4/18		5

There are more interesting things in the gift-giving network. A natural question is since node 7 ranks higher than 12 in both BC and CC and they rank the same in DC, why node 12 ranks higher than node 7?

We infer that this may result from the degree distribution changes of the graph when the nodes are removed. Node 12 is linked with the hub 11 and when node 12 or node 7 is removed, the degrees of the remained graphs are shown in table 3 and Figure 2. Here to measure the changes of degree distribution, we use the entropy $-\sum_i p_i \log p_i$ to describe the degree distributions. The entropy of the degree distribution in original graph is 0.8226. While node 12 or 7 is removed, the entropy of degree distribution is changed into 1.2235 and 1.0357. Since the entropy indicates the irregularity of corresponding distribution p_i, it is suggested that the removal of node 12 brings larger variations in the degree distributions and makes the network more even. This helps explain why node 12 ranks higher than node 7.

From networks above, the node entropy heterogeneity could be viewed as a comprehensive measure of node importance. The H_E takes the global network structure and heterogeneity of the whole networks into account and has excellent performance in selecting significant nodes in network.

Table 3: Degree distributions when node 7 or node 12 is removed from the gift-giving network.

Degrees	6	5	4	3	2
Original	1	4	1	16	0
Remove Node 12	1	2	2	12	4
Remove Node 7	1	3	0	13	4

2.4. Approximation to Node Heterogeneity

To calculate all or most of the eigenvalues of the matrix $\mathcal{L}$, a number of algorithms are studied. By a similarity transformation with orthogonal matrix Q, the matrix $\mathcal{L}$ could be transformed to an upper triangular matrix $T = Q^T \mathcal{L} Q$ where T and $\mathcal{L}$ own the same eigenvalues. Eigenvalues of T could be calculated by methods like QL algorithm [32] with complexity $O(N)$. The orthogonal matrix Q could be calculated by various methods. For a symmetric matrix, the Q could be researched by Householder algorithm [33] with complexity $O(N^3)$. For a sparse matrix, Lanczos algorithm [34] could find Q with complexity $O(MN)$ where M is the edge number of the network.

However, since in reality the scale of networks could be enormous, the complete algorithms above are not applicable considering the time consuming. To efficiently apply the entropy heterogeneity of nodes, the approximation of entropy will be discussed in this subsection.

Expanding at $x = 1$ we could easily get that

$$\ln(x) = x - 1 - \sum_{k=2}^{\infty} \frac{(1-x)^k}{k}. \tag{8}$$

This series could be applied to approximate the entropy by cutting off at some index k, and we use $\ln(x) = x - 1$ to approach the entropy. In this situation, the entropy is calculated as:

$$\begin{aligned} S &= -\sum_{i=1}^{N} \frac{\lambda_i}{2} \ln \frac{\lambda_i}{2} \backsimeq \sum_{i=1}^{N} \frac{\lambda_i}{2}(\frac{\lambda_i}{2} - 1) \\ &= \frac{1}{2}\sum_{i=1}^{N} \lambda_i - \frac{1}{4}\sum_{i=1}^{N} \lambda_i^2. \end{aligned} \tag{9}$$

Since $Tr(\mathcal{L}^n) = \sum_i(\lambda_i^n)$, the approximated entropy could be written as:

$$S_1 = \frac{1}{2}Tr(\mathcal{L}) - \frac{1}{4}Tr(\mathcal{L}^2). \tag{10}$$

According to the definition in equation (4), $Tr(\mathcal{L}) = |V|$.

To calculate $Tr(\mathcal{L}^2)$, with some linear algebra knowledge, it is concluded that

$$\begin{aligned} Tr(\mathcal{L}^2) &= Tr(\mathcal{L} \times \mathcal{L}) = \sum_i \sum_j \mathcal{L}_{ij}\mathcal{L}_{j,i} \\ &= \sum_i \sum_j \mathcal{L}_{ij}^2 = \sum_{i=j} \mathcal{L}_{ij}^2 + \sum_{i \neq j} \mathcal{L}_{ij}^2 \\ &= |V| + \sum_{i \sim j} \frac{1}{d_i d_j}, \end{aligned} \tag{11}$$

where $i \sim j$ means node v_i and node v_j are connected. In this way, the von Neumann entropy is approximated as:

$$S_1 = \frac{|V|}{2} - \frac{|V|}{4} - \sum_{i \sim j} \frac{1}{4d_i d_j} = \frac{|V|}{4} - \sum_{i \sim j} \frac{1}{4d_i d_j}. \tag{12}$$

By this calculation, let the network with node v_i removed be G', the von Neumann entropy centrality of node v_i is

$$\begin{aligned} H_E(v_i) &\backsimeq |S_1(G) - S_1(G')| \\ &= |\frac{|V(G)|}{4} - |\frac{|V(G')|}{4}| - \sum_{j \sim k} \frac{1}{4d_j d_k} + \sum_{j \sim k} \frac{1}{4d'_j d'_k} \\ &= |\frac{|V(G)| - |V(G')|}{4}| - \sum_{j, i \sim j} \frac{1}{4d_i d_j} \\ &\quad - \sum_{j,k,i \sim j \sim k} \frac{1}{4d_j d_k} + \sum_{j,k,i \sim j \sim k} \frac{1}{4d'_j d'_k} \end{aligned} \tag{13}$$

If node v_j is linked with v_i, then $d'_j = d_j - 1$. Hence,

$$H_E(v_i) \backsimeq \frac{1}{4} - \sum_{j, i \sim j} \frac{1}{4d_i d_j} + \sum_{j,k,i \sim j \sim k} \frac{1}{4(d_j - 1)d_j d_k} \tag{14}$$

To cut off the series in equation (8) at a higher k could help improve the accuracy. In order to calculated $\sum_i \lambda_i (1 - \lambda_i)^k$, the sum of eigenvalues with higher power $\sum_i \lambda_i^t = Tr(\mathcal{L}^t)$ for $2 \leq t \leq k + 1$ need to be solved. They could be calculated in other perspective. Taking $Tr(\mathcal{L}^3)$ as example, since

$$\mathcal{L} = D^{-1/2} L D^{-1/2} = I - D^{-1/2} A D^{-1/2}, \tag{15}$$

it could be got easily that

$$\begin{aligned}
& Tr[(I-\mathcal{L})^3] \\
&= Tr(D^{-1/2}AD^{-1/2}D^{-1/2}AD^{-1/2}D^{-1/2}AD^{-1/2}) \\
&= Tr(D^{-1/2}AD^{-1}AD^{-1}AD^{-1/2}) \\
&= \sum_i \sum_j \sum_k \frac{1}{\sqrt{d_i}} A_{ij} \frac{1}{d_j} A_{jk} \frac{1}{d_k} A_{ki} \frac{1}{\sqrt{d_i}} \\
&= \sum_i \sum_j \sum_k \frac{1}{d_i d_j d_k} A_{ij} A_{jk} A_{ki}.
\end{aligned} \tag{16}$$

Since

$$\begin{aligned}
Tr[(I-\mathcal{L})^3] \quad &= Tr(I - 3\mathcal{L} + 3\mathcal{L}^2 - \mathcal{L}^3) \\
&= Tr(I - 3\mathcal{L}) + 3Tr(\mathcal{L}^2) - Tr(\mathcal{L}^3),
\end{aligned} \tag{17}$$

then we have

$$\begin{aligned}
Tr(\mathcal{L}^3) \quad &= -2|V| + 3|V| + \sum_{i\sim j} \frac{3}{d_i d_j} - \sum_i \sum_j \sum_k \frac{1}{d_i d_j d_k} A_{ij} A_{jk} A_{ki} \\
&= |V| + \sum_{i\sim j} \frac{3}{d_i d_j} - \sum_{i\sim j\sim k\sim i} \frac{1}{d_i d_j d_k}.
\end{aligned} \tag{18}$$

So the approximated entropy when cutting off at $k = 2$ in equation (8) is

$$S_2(G) = \frac{5}{16}|V| - \sum_{i\sim j} \frac{11}{16 d_i d_j} + \sum_{i\sim j\sim k\sim i} \frac{1}{16 d_i d_j d_k}. \tag{19}$$

Similar derivation could be applied to the situation when $k > 2$.

By a breadth-first search algorithm [35], the neighbor-relation of nodes in a network could quickly be achieved. Then to calculate the entropy heterogeneity of a node, we only need to consider all the paths starting from this node whose lengths are less than the order of cutting-off for approximation. If the neighbor-relations for each node are stored well, the calculation complexity will be reduced hugely and the global calculation is simplified and degenerated to local situations. This approximation will accelerate the calculation of node ranks by H_E and the time complexity will be reduced to $O(N)$.

To view the performance of this approximation method, examinations on random networks are conducted. The nodes sorts by complete calculation and approximation are compared. Since compared to the specific values of H_E, the nodes sort is what we finally get and our target, we examine the similarity of nodes in both sorts at certain percentages. This examine is conducted on Erdös-Rényi (ER) networks and the similarities of nodes at several top percentages are calculated. For example, if there are k nodes in both sorts at top y nodes ranked by the complete calculation and approximation, then we say the similarity at y is k/y. The results are presented in Table 4. As we could see, it is performed that the methods mentioned above could work as an efficient and reasonable approximation to the von Neumann entropy method in capturing the most significant nodes.

Table 4: Similarity of nodes sorts by complete calculation and approximation on ER networks at first 10%. Each random network contains 1,000 nodes and 2,000 edges.

Percentage	1%	2%	3%	4%	5%
Similarity	0.88 ± 0.09	0.92 ± 0.05	0.92 ± 0.04	0.91 ± 0.04	0.89 ± 0.03
Percentage	6%	7%	8%	9%	10%
Similarity	0.88 ± 0.03	0.89 ± 0.03	0.90 ± 0.04	0.92 ± 0.03	0.94 ± 0.02

3. Experiments

To further explore the properties and features of the entropy heterogeneity of nodes, first we discuss its performance in reducing the heterogeneity index of networks by Estrada. Then the variations in average clustering coefficient will be presented.

3.1. Heterogeneity Index by Estrada

There exist a number of heterogeneity indices and one of the most popular measurements is proposed by Estrada [10]. Firstly, the irregularity of link connecting node v_i and v_j is defined as:

$$I_{i,j} = [f(d_i) - f(d_j)]^2. \tag{20}$$

This irregularity will be zero if the pair of nodes connected by the link have the same degree, which usually appears in regular networks. Taking the

$f(d) = d^{-1/2}$ and summing the irregularity of all the links in the network, the heterogeneity index of network G is defined as (assuming $d \neq 0$):

$$\rho'(G) = \sum_{i\sim j}(d_i^{-1/2} - d_j^{-1/2})^2. \tag{21}$$

This quantity is zero for regular networks, and it will increase as the differences between the degrees of adjacent nodes increase. This index could be expressed by Laplacian matrix. Taking $|\mathbf{d}\rangle = (d_1^{-1/2}, d_2^{-1/2}, \ldots, d_N^{-1/2})$, this heterogeneity index could be calculated as:

$$\rho'(G) = \sum_{i\sim j}(d_i^{-1/2} - d_j^{-1/2})^2 = \frac{1}{2}\langle \mathbf{d}^{-1/2}|L|\mathbf{d}^{-1/2}\rangle = n - 2\sum_{i\sim j}(d_i d_j)^{-1/2}. \tag{22}$$

The lower bound of $\rho'(G)$ is attained for regular graphs, which is zero, and the upper bound is attained for star graphs, which is $|V| - 2\sqrt{|V| - 1}$. In this way, the normalized heterogeneity index is written as:

$$\rho(G) = \frac{\sum_{i\sim j}(d_i^{-1/2} - d_j^{-1/2})^2}{|V| - 2\sqrt{|V| - 1}}, \tag{23}$$

where $0 \leq \rho(G) \leq 1$.

An interesting problem is how to reduce the network heterogeneity as fast as possible by removing nodes. It is regarded that the star network which has only one center node and $N-1$ leaves has the highest heterogeneity and the $\rho(G)$ equals to 1. Removing the center is the fastest method to reduce the heterogeneity of the star network and $\rho(G)$ will decrease to zero, which makes sense since obviously the center node is the most heterogeneous one. Yet in real world, the networks are much more complex than star networks and it is hard to find the centers or hubs. Also, although nodes in the regular networks have the same degrees, the structure of the network could be various, which makes the statuses of nodes in networks vary a lot. The hubs are not determined by degree anymore. More information is required to accurately signify the node importance and significance.

We proposed to apply node entropy heterogeneity to this problem. Firstly the H_E of all nodes are calculated and node with the highest H_E value is removed from the network, and then the same process is repeated until the desired heterogeneity is achieve.

Figure 3**a** shows the variations of $\rho(G)$ as nodes are removed in ER networks. The H_E is compared to several other indices on the same networks: DC, BC, CC, eigenvector centrality (EC) [36], Page Rank (PR) [17], high degree adaptive (HDA), and collective influence (CI) [18]. It is illustrated that the node entropy heterogeneity outperforms all the other ones. Similar results are shown in scale-free (SF) [37] networks in Figure 3**b**.

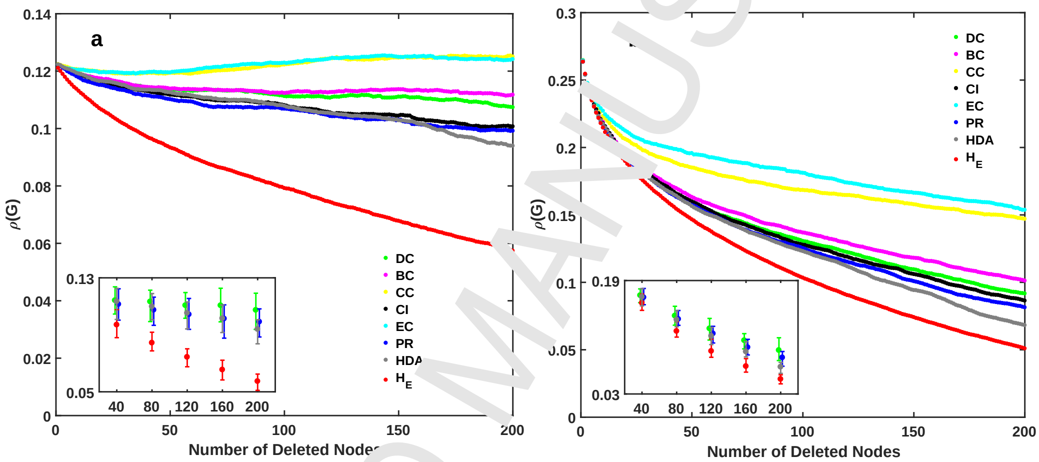

Figure 3: **a.** $\rho(G)$ in ER network. The results are the average values of 20 ER networks. Each network contains 1,000 nodes and 2,000 edges. The performances of degree centrality, betweenness centrality, closeness centrality, collective influence, eigenvector centrality, Page Rank, high degree adaptive and node entropy heterogeneity are represented in different colors. Inside figure presents several error bars of the data. **b.** $\rho(G)$ in SF network. The points are the average of 20 network with 1,000 nodes and $\gamma = 3$. Inside figure presents several error bars of the data.

According to definition, an ER network is composed by nodes and links between each pair of nodes with the same probability, thus the node degrees are similar to each other and the whole network looks homogeneous in their connection pattern. That's why compared to SF networks, the ER networks own lower $\rho(G)$ values and the centralities work less effective in reducing the heterogeneity of networks. Yet it is observed that the node entropy heterogeneity works well in both ER and SF networks and it could efficiently capture the nodes with high heterogeneity and increase the network homogeneity. In the SF networks, the similar effects are observed in several measurements including DC, BC, HDA and PR. That's because there exist extremely high-degree nodes, and the heterogeneity of the whole network concentrates on these hubs, which leads to similar findings by these centrality.

3.2. Average Clustering Coefficient

Another fascinating phenomenon related to the node entropy heterogeneity is the variation in average clustering coefficient. The global average clustering coefficient of a network is defined as $\bar{C} = \frac{1}{n}\sum_{i=1}^{n} C_i$, where C_i is the clustering coefficient (CLC) of node v_i

$$C_i = \frac{2|\{(v_j, v_k) \in E : (v_i, v_j) \in E, (v_i, v_k) \in E\}|}{d_i(d_i - 1)}, \tag{24}$$

which indicates how well the neighbors of node v_i are connected. This index is also a measurement of network heterogeneity concerning the connection of node neighbors. If the neighbors of a node are highly connected, then this node owns high CLC and it is safe to say that the local region where this node belongs to is dense, which means connective heterogeneity in this region compared to low CLC ndoes.

In the random geometric graphs (RGGs) [38], every time a nodes with the highest H_E is removed, the global average clustering coefficient is calculated. We find that in the comparison with others including PR, DC, CC, CLC, EC, the H_E brings a much more rapid decrease of $\bar{C}$ (Figure 4**a**). It is worth noting that the reduction caused by H_E is even more significant than by CLC itself, which suggests that the removal by the node entropy heterogeneity brings more structural damages than others to RGGs. This phenomenon does not appear in SF networks and ER networks.

Actually, this phenomenon is deeply related to the special topological features of RGGs. The RGGs are the networks whose nodes are scattered randomly in d dimension space. If the distance between two nodes is less than a specific threshold r, then these two nodes are linked. One of the most important property of RGGs is that the cluster or modularity structure is striking and there are a lot of large or small clusters in each RGG. Nodes inside each cluster are densely connected and less connected to outliers. This point is also supported by the significance of BC in reducing the size of giant component (Figure 4**c**) since the BC breaks down the giant components fast, which means there are a few nodes working as bridges between clusters and own highest BC values.

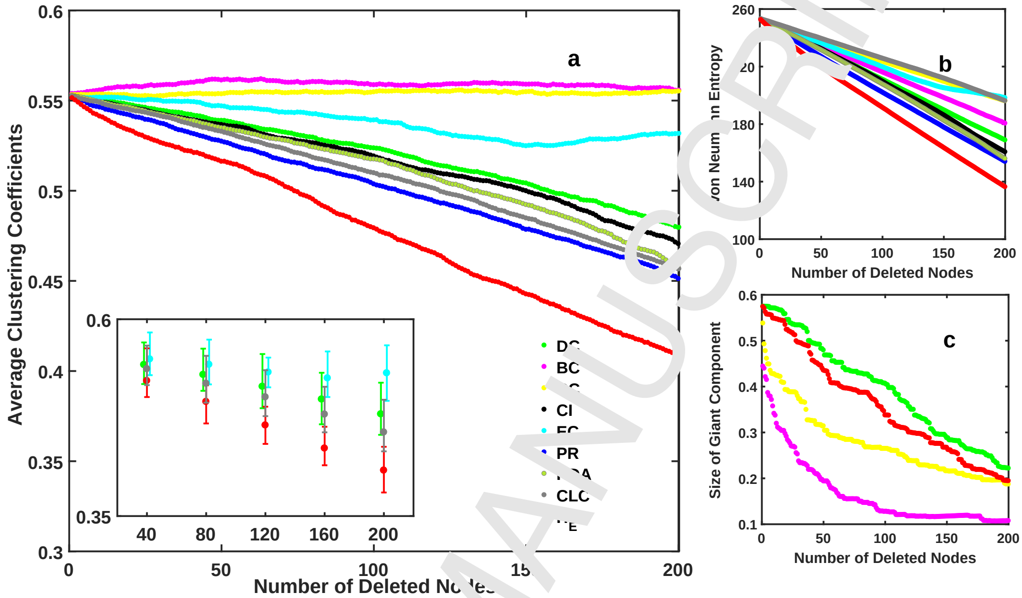

Figure 4: **a.** Average clustering coefficient in RGGs. The results are the average values of 20 RGGs. Each network contains 1,000 nodes scattered in a 3-dimension space and the average degree around 4.3. The performances of degree centrality, betweenness centrality,closeness centrality, collective influence, eigenvector centrality, PageRank, HDA, clustering coefficient, and node entropy heterogeneity are represented in different colors. Inside figure presents several error bars of the data. **b.** The variations of the von Neumann entropy in RGGs. **c.** Size of giant connective component in RGGs.

Figure 5 shows a cluster composed of seven nodes. By definition, clustering coefficients of nodes v_1 to v_6 are all $\frac{2}{C_3^2} = \frac{2}{3}$ and node v_7 is $\frac{6}{C_6^2} = \frac{2}{5}$. Yet when node v_7 which owns the highest H_E is removed, the $\bar{C}$ of this whole cluster is brought down to zero. That's the reason why the node entropy heterogeneity causes larger reduction in average clustering coefficient than CLC to RGGs in the experiments. Also, this phenomenon suggests that H_E does obtain the centre nodes or hubs in the networks efficiently and is able to break down the cluster structures rapidly.

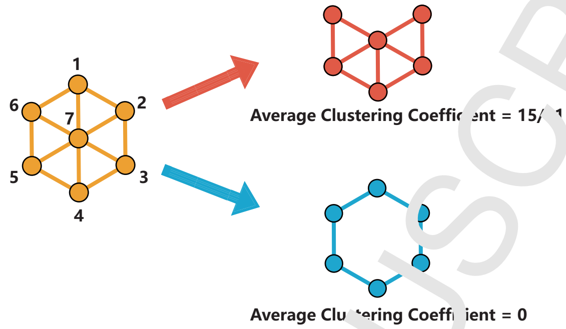

Figure 5: An example of cluster. Node v_7 owns the lowest clustering coefficient, yet removing node v_7 will decrease the $\bar{C}$ to 0.

3.3. General Relativity and Quantum Cosmology collaboration network

To apply our discoveries above, experiments of Estrada heterogeneity and average clustering coefficients are conducted on the General Relativity and Quantum Cosmology (GR-QC) collaboration network [46]. This network is a paper co-authorship network and captures the papers of the GR-QC category on arXiv from January 1993 to April 2003. The nodes in the network stand for researchers and if two researcher co-author one paper, then there will be an edge between the two nodes. Results on this network are presented on Figure 6.

As we could see, in accordance with the results on random networks, the H_E is still the fastest method to reducing the Estrada heterogeneity and average clustering coefficient. The H_E method could capture the most influential nodes in the networks accurately. Since the results on GR-QC network are more similar to the SF networks than ER networks, we infer that this network performs SF property to some extent. Also, removing the most significant nodes found by H_E could reduce the average clustering coefficient fast. This means that this network performs similar topological features to RGGs and there exist a number of small groups inside which the connection is denser. These nodes play crucial roles in many collaboration groups.

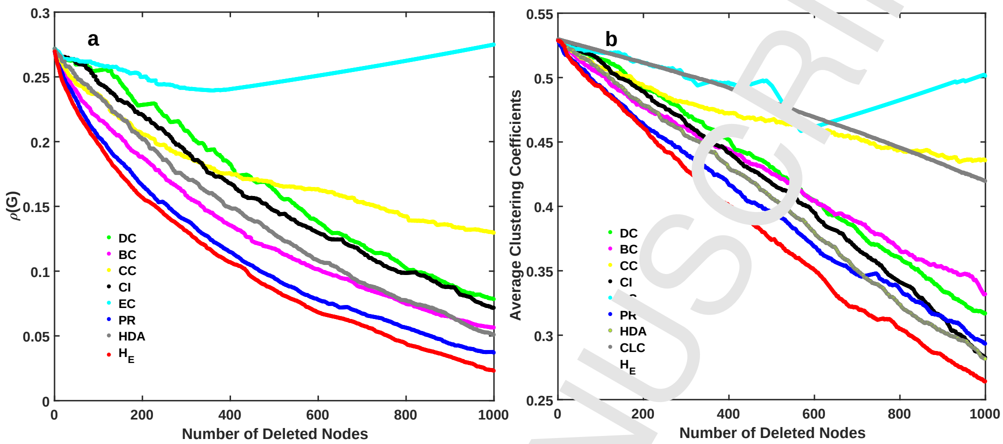

Figure 6: **a.** $\rho(G)$ in GR-QC network. This network contains 5,242 nodes and 14,496 edges. The performances of degree centrality, betweenness centrality, closeness centrality, collective influence, eigenvector centrality, Page Rank, high degree adaptive and node entropy heterogeneity are represented in different colors. **b.** Average clustering coefficient in GR-QC network.

4. Conclusion and Discussion

In this paper the node entropy heterogeneity based on von Neumann entropy is discussed, which makes it possible to study the significance of nodes in the perspective of structural complexity and heterogeneity. By comparing the heterogeneity of nodes with classical node centrality, it is shown that the H_E is an all-round measurement of node importance. By comparing the changes of Estrada heterogeneity of networks and average clustering coefficient with other heterogeneity indices when deleting high H_E nodes, it is concluded that this node entropy heterogeneity has an excellent performance in breaking down network structure and can capture the significant features.

This index could be applied to find the most influential nodes in real-world networks. It could be used to find the uneven parts in many kinds of networks and help managers identify crucial nodes. For example, this method could be applied to the markets networks to analyze the different roles played by various participants. Also, this measurement could help identify some topological features in the networks. More experiments on various networks with different structural characteristics are expected to uncover more features of this index.

Another advantage of the node heterogeneity is that this definition could be expanded to mesoscopic subjects, like motifs. In 2002, Alon et al. [39] introduced the idea of motif when they were studying the gene network, which is defined as the recurring, significant sub-networks and patterns in a network, and it is discovered that the frequencies of some specific motifs in realistic networks are much more significant by comparing with random networks [40]. Since motifs emphasize on the structure and connection patterns which could not be found by only observing single nodes, node centralities could not capture the structural characterizations completely. Also, for many node centralities, like eigenvector centrality and closeness centrality, they are hard to be generalized to motifs directly. The index provides an access to evaluate and measure the significance of specific structure on the global network and a new perspective to study network structural features.

Since a great number of real-world data is directed, it is worth defining and researching the von Neumann on directed networks. Chung provided a definition of Laplacian matrix on directed networks [41] using Perron-Frobenius Theorem [20] and based on this work, Ye et al. [42] proposed a method to approximate the von Neumann entropy of directed networks, which allows us to compute the von Neumann entropy in terms of in-degree and out-degree of nodes simply. However, these results only work on strongly-connected directed networks. Another definition involving incidence matrix [43], loses the direction information when calculating the Laplacian. It is still an open problem to define the von Neumann entropy on directed networks generally.

Acknowledgements

This work is supported by the Fundamental Research Funds for the Central Universities, the National Natural Science Foundation of China (No.11201019), the International Cooperation Project No.2010DFR00700, Fundamental Research of Civil Aircraft No.MJ-F-2012-04, and the Fundamental Research Funds for the Central Universities (DUT18RC(4)066).

References

[1] C. Orsini, E. Gregori, L. Lenzini and D. Krioukov, "Evolution of the Internet k-Dense Structure," in IEEE/ACM Transactions on Networking,

vol.22, no.6, pp.1769-1780, Dec. 2014. doi: 10.1109/TNET.2013.2282756

[2] S. Wasserman and K. Faust, "Social network analysis: Methods and applications (Vol. 8)", Cambridge university press. (1994)

[3] P. Cano, O. Celma, M. Koppenberger and J. M. Buldu, "Topology of music recommendation networks". *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 16(1), 013107. (2006)

[4] M. Zitnik, M. Agrawal and J. Leskovec, "Modeling polypharmacy side effects with graph convolutional networks". *arXiv preprint*, arXiv:1802.00543.(2018)

[5] T. A. Snijders (1981) The degree variance: an index of graph heterogeneity. Social networks, 3(3), 163-174.

[6] F. K. Bell (1992) A note on the irregularity of graphs. Linear Algebra and its Applications, 161, 45-54.

[7] M. O. Albertson (1997) The irregularity of a graph. Ars Combinatoria, 46, 219-225.

[8] D. C. Colander (2001) Microeconomics. Boston, MA: McGraw-Hill

[9] R. Jacob, K. P. Harikrishnan, R. Misra and G. Ambika (2017) Measure for degree heterogeneity in complex networks and its application to recurrence network analysis. Royal Society open science, 4(1), 160757.

[10] E. Estrada (2010) Quantifying network heterogeneity. Physical Review E, 82(6), 066102.

[11] L. Han, E. R. Hancock and R. C. Wilson (2011, May) Entropy versus heterogeneity for graphs. In International Workshop on Graph-Based Representations in Pattern Recognition (pp. 32-41). Springer, Berlin, Heidelberg.

[12] P. W. Holland and S. Leinhardt (1971). Transitivity in structural models of small groups. Comparative group studies, 2(2), 107-124.

[13] D. J. Watts and S. H. Strogatz (1998). Collective dynamics of small-worldnetworks. nature, 393(6684), 440.

[14] A. Bavelas (1950) “Communication patterns in task-oriented groups”, *The J. Acoust. Soc. Am.* 22 725–730

[15] G. Sabidussi (1966) “The centrality index of a graph”, *Psychom.* 31 588–603

[16] L. C. Freeman (1977) “A set of measures of centrality based on betweenness”, *Sociom.* 35–41

[17] L, Page, S. Brin, R. Motwani, T. Winograd (1999) “The PageRank citation ranking: Bringing order to the web”. *Stanford InfoLab*

[18] F. Morone and H. A Makse (2015) “Influence maximization in complex networks through optimal percolation”, *Nature*, 524(7563), 65.

[19] F. R. Chung (1997) *Spectral graph theory*, (No. 92) American Mathematical Soc.

[20] R. A. Horn and C. R. Johnson (2012) *Matrix Analysis*, Cambridge University Press New York, NY, USA

[21] M. Fiedler, “Algebraic connectivity of graphs”, *Czechoslovak mathematical journal*, 1973, 23(2): 298-305

[22] M. Fiedler (1989) “Laplacian of graphs and algebraic connectivity”, *Banach Center Publications*, 25(1), 57-70.

[23] D. L. Powers (1988) “Graph partitioning by eigenvectors”, *Linear Algebra and its Applications*, 101, 121-133.

[24] W. W. Zachary (1977) “An information flow model for conflict and fission in small groups”, *Journal of anthropological research*, 33(4), 452-473

[25] F. Passerini and S. Severini (2008) “The von Neumann entropy of networks”, *ArXiv e-prints* 0812.2597

[26] L. Han, E. R. Hancock and R. C. Wilson (2011) “Characterizing graphs using approximate von Neumann entropy”, *In Iberian Conference on Pattern Recognition and Image Analysis* (pp. 484-491), Springer, Berlin, Heidelberg.

[27] S. A. Nene, S. K. Nayar and H. Murase (1996). Columbia Object Image Library (COIL-100).

[28] R. L. Breiger and P. E. Pattison (1986) "Cumulated social roles: The duality of persons and their algebras" *Soc. networks* 8 215–256

[29] R. C. Mueller (1981) "The Rise of the Medici Faction in Florence", *The ANNALS Am. Acad. Polit. Soc. Sci.* 455 191–192

[30] P. Hage and F. Harary (1983) *Structural models in anthropology*, Cambridge University Press Cambridge [Cambridgeshire] New York

[31] E. G. Schwimmer (1970) "Exchange in the social structure of the Orokaiva", *PhD Thesis* University of British Columbia Canada

[32] W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery (1992) "Numerical Recipes in C (2nd Ed.) The Art of Scientific Computing", Cambridge University Press, New York, NY, USA.

[33] A. S. Householder, "Unitary Triangularization of a Non-symmetric Matrix", *J. ACM* 5, 4 (October 1958), 339-342, DOI=http://dx.doi.org/10.1145/320941.320947

[34] C. Lanczos (1950) "An iteration method for the solution of the eigenvalue problem of linear differential and integral operators", Los Angeles, CA: United States Governm. Press Office.

[35] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, "Introduction to Algorithms, Second Edition", MIT Press and McGraw-Hill, 2001. ISBN 0-262-03293-7. Section 22.2: Breadth-first search, pp. 531C539.

[36] L. C. Freeman (1978) "Centrality in social networks: conceptual clarification", *Soc. Networks* 1, 215C239.

[37] M. E. J. Newman, "Networks: An Introduction", (Oxford Univ. Press, 2000)

[38] M. Penrose (2013) *Random geometric graphs*, Oxford University Press

[39] S. S. Shen-Orrand, R. Milo, S. Mangan and U. Alon (2002) "Network motifs in the transcriptional regulation network of Escherichia coli", *Nat. genetics* 31 64–68

[40] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii and U. Alon (2002) "Network motifs: simple building blocks of complex networks", *Science* 298 824–827

[41] F. Chung (2009) "Laplacians and the Cheeger inequality for directed graphs", *Annals Comb.* 9 1–19

[42] C. Ye, R. C. Wilson, Cé. H. Comin, L. F. Costa and E. R. Hancock (2014) "Approximate von Neumann entropy for directed graphs", *Phys. Rev. E* 89 052804

[43] C. Godsil and G. F. Royle (2013) *Algebraic graph theory*, Springer Science & Business Media 207

[44] N. de Beaudrap, V. Giovannetti, S. Severini and R. Wilson, "Interpreting the von Neumann entropy of graph Laplacians, and coentropic graphs", *ArXiv e-prints* 1304.7946 (2013)

[45] L. Han, F. Escolano, E. R. Hancock and R. C. Wilson, "Graph characterizations from von Neumann entropy", *Pattern Recognit. Lett.* 33 1958–1967 (2012)

[46] J. Leskovec, J. Kleinberg and C. Faloutsos. Graph Evolution: Densification and Shrinking Diameters. ACM Transactions on Knowledge Discovery from Data (ACM TKDD), 1(1), 2007.