Structural Information and Dynamical Complexity of Networks

Angsheng Li and Yicheng Pan

***Abstract*— In 1953, Shannon proposed the question of quantification of structural information to analyze communication systems. The question has become one of the longest great challenges in information science and computer science. Here, we propose the first metric for structural information. Given a graph G, we define the K-dimensional structural information of G (or structure entropy of G), denoted by $\mathcal{H}^K(G)$, to be the minimum overall number of bits required to determine the K-dimensional code of the node that is accessible from random walk in G. The K-dimensional structural information provides the principle for completely detecting the natural or true structure, which consists of the rules, regulations, and orders of the graphs, for fully distinguishing the order from disorder in structured noisy data, and for analyzing communication systems, solving the Shannon's problem and opening up new directions. The K-dimensional structural information is also the first metric of dynamical complexity of networks, measuring the complexity of interactions, communications, operations, and even evolution of networks. The metric satisfies a number of fundamental properties, including additivity, locality, robustness, local and incremental computability, and so on. We establish the fundamental theorems of the one- and two-dimensional structural information of networks, including both lower and upper bounds of the metrics of classic data structures, general graphs, the networks of models, and the networks of natural evolution. We propose algorithms to approximate the K-dimensional structural information of graphs by finding the K-dimensional structure of the graphs that minimizes the K-dimensional structure entropy. We find that the K-dimensional structure entropy minimization is the principle for detecting the natural or true structures in real-world networks. Consequently, our structural information provides the foundation for knowledge discovering from noisy data. We establish a black hole principle by using the two-dimensional structure information of graphs. We propose the natural rank of locally listing algorithms by the structure entropy minimization principle, providing the basis for a next-generation search engine.**

***Index Terms*— Shannon entropy, structural information, dynamical complexity of networks, graph characterisation, networks.**

Manuscript received September 24, 2015; revised January 31, 2016; accepted April 10, 2016. Date of publication April 21, 2016; date of current version May 18, 2016. The work of A. Li was supported in part by the Hundred Talent Program of Chinese Academy of Sciences. The work of A. Li and Y. Pan supported in part by the China Basic Research Program (973) under Grant 2014CB340302, in part by the Institute of Software, Chinese Academy of Sciences through the Grand Project entitled Network Algorithms and Digital Information and in part by the National Natural Science Foundation of China under Grant 61161130530.

The authors are with the State Key Laboratory of Computer Science, and Institute of Software, Chinese Academy of Sciences, Beijing 100190, China (e-mail: angsheng@ios.ac.cn; yicheng@ios.ac.cn).

Communicated by C. E. Koksal, Associate Editor for Communication Networks.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2016.2555904

I. Introduction

SHANNON defined the information of a distribution $\mathbf{p} = (p_1, p_2, \cdots, p_n)$ to be the entropy as

$$H(p_1, \ldots, p_n) = -\sum_{i=1}^{n} p_i \log_2 p_i.$$

This metric has been the foundation for our understanding of "information" in the current state of the art in all areas of sciences.

Shannon [48] proposed the question to establish a structural theory of information to support the analysis of communication systems. Brooks, Jr., [13] listed the question of quantification of structural information, that is, the Shannon's 1953 question, as the first of the three great challenges for half-century-old computer science.

In the 21st century, data of new types, such as biological data, web data, topographical maps and medical data etc have appeared. Analysing the new data and discovering the orders and knowledge of the new data are new challenges in the current computer science. For this new mission, Shannon's definition of information is apparently insufficient, for which a new metric of structural information is urgently called for.

Equally important, a metric of structural information is central for us to understand the complexity of networks, including the networking computing systems, and to develop the network theory. The reason is that, structural entropy may well define the dynamical complexity of networks, that is, the measure of the complexity of interactions, communications, operations and even evolution of the networks. For this challenge, there is no even a nontrivial progress in the current state of the art.

Complex networks were assumed to be randomly evolved in the early history of network study, for which Erdös and Rényi [22], [23] proposed a classic model, referred to as the ER model. This model was used to explore many properties of random graphs such as the existence of giant connected components, the small diameter property etc. Much more recently, Watta and Strogatz [51] proposed a simple model in which random edges are added to a grid graph (or the like), and Kleinberg [25] introduced a model in which edges are added with endpoints chosen with probability inversely proportional to a power of the distances in the grid. The graphs generated by these models have small diameters and also witness clustering effects, whereby two nodes are more likely to be adjacent if they share common neighbors. Barabási and Albert [5] proposed the scale-free model by

introducing preferential attachment as an explicit mechanism, which generates graphs with a power law degree distribution. Following this, a number of new models were introduced by using randomness and some local rules, including the copying model [28], [29], the forest fire model [30], the random walk model and nearest neighbor model [50], the random-surfer model [8] and the hierarchical model [45]. These models provide theoretical approaches to studying notions of statistical robustness. The current state of the art, however, continues to indicate that real networks are complex in ways not captured by existing models. This brings the fundamental question of characterizing the complexity of complex networks into sharper focus.

Newman and Girvan [42] defined the notion of modularity to measure the quality of community structure in a network. Mowshowitz [40] defined the entropy of a graph G as the entropy of a probability distribution associated with the orbits of the automorphism group of G. This definition is concerned with the symmetries of graphs, which have been used for characterising molecules and chemical structures. Bonchev and Trinajstić [10] introduced a notion of entropy determined by distances, i.e. they considered the entropy of the distributions on the number of nodes accessible within certain distances. A Markov entropy centrality of a Markov chain associated with a system, was also proposed [41]. For degrees, there is a naive way to define an entropy of the degree distribution. Dehmer [19] proposed the notion of information functional, allowing some more theoretical analyses. These measures of graph entropy, together with more but similar measures are reviewed in a recent survey by Dehmer and Mowshowitz [21]. All these existing measures of graph entropy can be principally categorised as two approaches, one is the global approach and the other the local approach. The global approach is to measure the entropy of the distribution $p(G)$, where $p(G)$ is the probability that G occurs in a random construction of graphs. The local approach is to measure the entropy of a distribution $p(i)$ where $p(i)$ is the probability that node i of a graph G occurs, in which $p(i)$ is associated with a selected function f from the vertices to real numbers.

Rosvall and Bergstrom [47] proposed an algorithm to find communities of a network by minimising the description length of nodes of the network.

Bianconi [6] introduced a notion of entropy for network ensembles, which was called the "structural entropy of network ensembles" and is now known as the Gibbs entropy of network ensembles. The Gibbs entropy of a network ensemble is the number of bits required to determine the code of the network constructed by the ensemble. Anand and Bianconi [2] proposed the notion of Shannon entropy of network ensembles, which corresponds to the number of bits needed to determine the description of a classical system generated by the ensemble. Braustein *et al.* [11] proposed the notion of von Neumann entropy of a network. The von Neumann entropy of a network ensemble is interpreted as the number of bits needed to determine the description of the quantum system generated by the ensemble. The Gibbs entropy, Shannon entropy and von Neumann entropy are defined by a statistical mechanics approach, providing an approach to comparing the different models of networks.

In the present paper, we define the high-dimensional structure entropy of a graph to be the minimum overall number of bits required to determine the high-dimensional code of the node that is accessible from random walk in the graph. Intuitively, for a graph G and a natural number K, the K-dimensional structural information of G is the measure that controls the formation of the K-dimensional structure $\mathcal{T}$ of G satisfying: (i) $\mathcal{T}$ consists of the rules, regulations and order of G, and (ii) $\mathcal{T}$ excludes the perturbation of random variations occurred in G. Consequently, the metric of the high-dimensional structural information is a measure of structural information that perfectly supports: (1) analysis of structured and noisy data, (2) analysis of networking systems, (3) constructions of structures for unstructured big data, (4) extracting natural or true structures from corrupted structures, and (5) discovery of knowledge from noisy data, including both structured and unstructured. Our metric also measures the complexity of the interactions, communications, and operations in networks, and even the evolution of the networks, referred to as a dynamical complexity of networks. We find that our K-dimensional structural information of graphs satisfies a number of fundamental properties. We establish the theory of one- and two-dimensional structure entropy of arbitrarily given graphs, of classical structures in nature, and of networks generated by well-known models. We show that our K-dimensional structure entropy provides the principle to distinct the order from disorder in a noisy structure, and to detect the natural or true knowledge from noisy data. We propose the natural rank algorithms by the principle of two-dimensional structure entropy minimisation, providing the basis for a next-generation search engine.

We organize the paper as follows. In Section II, we review the static complexity of network ensembles. In Section III, we introduce the challenges. In Section IV, we introduce the overall ideas of our structural information. In Section V, we introduce our notion of K-dimensional structural information for graphs. In Section VI, we define the natural K-dimensional structures of networks by using our K-dimensional structural information. In Section VII, we establish some basic properties of the K-dimensional structural information of graphs. In Section VIII, we define the notion of K-dimensional structural complexity of networks. In Section IX, we establish the lower bound of positioning entropy of typical graphs i.e., the graphs of either simple or with balanced weights. In Section X, we establish the general principles of two-dimensional structural information of graphs. In Section XI, we establish the lower bounds of two-dimensional structural information of classical data structures and the networks of the small world model. In Section XII, we establish the upper bounds of two-dimensional structural information for classical data structures and for networks of the small world model. In Section XIII, we establish the lower bounds of two-dimensional structural information for expander graphs, and for the networks of classical models. In Section XIV, we propose the homophyly/kinship model and establish

the theory of two-dimensional structural information of the networks generated by the homophyly/kinship model. In Section XV, we establish a black hole principle of networks. In Section XVI, we introduce the algorithm for finding the natural K-dimensional structure of networks and the personalised listing rank algorithm for smart searching by using the K-dimensional structure entropy minimisation principle of networks. In Section XVII, we introduced the three-dimensional gene map developed by Li *et al.* [36] to show that structure entropy minimisation is indeed the principle for discovering true structures in Nature. In Section XVIII, we propose a new search engine, the natural rank, on the basis of structure entropy minimisation. Finally in Section XIX, we summarize the conclusions of our paper and discuss some open questions left by this research.

II. Existing Measures of Graph Entropy

For graphs and models of graphs, many measures of group entropy have been proposed. We review the representative such measures.

We will see that each of the measures is a specific form of the Shannon entropy for different types of distributions.

A. Global Measures

Rashevsky [44] proposed the first entropy measure of the following form: For a connected graph G with n nodes,

$$\mathcal{I}(G) = -\sum_{i=1}^{k} \frac{n_i}{n} \cdot \log_2 \frac{n_i}{n} \tag{1}$$

where n_i is the number of topologically equivalent vertices in the i-th vertex orbit of G, and k is the number of different orbits.

In this definition, vertices are considered as topologically equivalent if they belong to the same orbit of a graph.

Similar notion can be defined by using edges and edge equivalent classes.

Bonchev and Trinajstić [10] defined the entropy measure for a graph G based on the distribution of metric $d(i, j)$ where i, j are vertices of i and j.

The different choices of $d(i, j)$ lead to a number of entropy measures of graphs by using the same idea.

B. Local Entropy Measures

Raychaudhury *et al.* [46] proposed the first local measure of graph entropy. Such a measure is interpreted as a kind of vertex complexity [46].

A similar such notion is as follows. Given a connected graph G with n nodes, for each pair of nodes i and j, let $d(i, j)$ be the distance between i and j in G, and let $d(i)$ be the sum of $d(i, j)$ for all j.

For each i, we define the entropy of i in G by

$$\mathcal{I}^G(i) = -\sum_{j=1}^{n} \frac{d(i, j)}{d(i)} \cdot \log_2 \frac{d(i, j)}{d(i)}. \tag{2}$$

Local measures of the j-sphere cardinalities and the parametric information measures have also been defined [20], [26].

C. Parametric Graph Entropy

Dehmer [19] defined an interesting parametric measure as follows. Given a network $G = (V, E)$ and a function f from V to positive real numbers, define

$$p(i) = \frac{f(i)}{\sum_{j=1}^{n} f(j)}.$$

Then the parametric measure is the Shannon entropy of the distribution $(p(1), p(2), \cdots, p(n))$.

D. Gibbs Entropy

The Gibbs entropy per node in a network of n nodes, denoted Σ, was defined for network ensembles following a statistical mechanics approach. A microcanonical network ensemble is defined as the set of all the networks satisfying a set of constraints. Examples of typical constraints include having a fixed number of links per node, having a given degree distribution or community structure. The Gibbs entropy of a microcanonical ensemble is given by

$$\Sigma = \frac{1}{n} \cdot \log N, \tag{3}$$

where N is the number of the networks in the ensemble.

By definition, the Gibbs entropy of a network ensemble is the number of bits needed to determine the code of the network generated by the ensemble.

E. Shannon Entropy

For a network of n nodes, for each pair of nodes (i, j), create a link of weight α with probability $\pi_{i,j}(\alpha)$. Then, the probability Π of the canonical undirected network ensemble, defined by its adjacency matrix $\{a_{i,j}\}$, is defined by

$$\Pi = \Pi_{i<j} \pi_{i,j}(a_{i,j}), \tag{4}$$

for which the log-likelihood function is given by

$$\mathcal{L} = -\sum_{i<j} \log \pi_{i,j}(a_{i,j}). \tag{5}$$

The entropy of a canonical ensemble is the logarithm of the number of typical networks in the ensemble and is given by

$$S = \langle \mathcal{L} \rangle_\Pi = -\sum_{i<j} \sum_{\alpha} \pi_{i,j}(\alpha) \log \pi_{i,j}(\alpha), \tag{6}$$

which takes exactly the form of a Shannon entropy. This defines the notion of Shannon entropy of a network ensemble.

In particular, for the case of a simple undirected network, where $\alpha = 0$ or 1, we have

$$S = -\sum_{i<j} p_{i,j} \log p_{i,j} - \sum_{i<j} (1 - p_{i,j}) \log(1 - p_{i,j}), \tag{7}$$

where $p_{i,j} = \pi_{i,j}(1)$ is the probability of having a link between nodes i and j.

By definition, the Shannon entropy of a network ensemble is the number of bits needed to determine the expression of the network generated by the ensemble.

Bianconi *et al.* [7] introduced an indicator function, denoted Θ, on the basis of Shannon entropy to quantitatively measure the dependence of the structures of a network to the semantic characteristics of the nodes of the network.

F. Von Neumann Entropy

Braustein *et al.* [11] proposed the notion of von Neumann entropy of a network. The entropy is constructed from a density matrix ρ associated with the network. The density matrix must be a positive semidefinite matrix with unitary trance. The matrix ρ is defined as $\rho = L/\sum_{i,j} a_{i.j}$, where L is the Laplacian matrix of the network, with $L_{i,j} = \sum_r a_{ir}\delta_{i,j} - a_{i,j}$. Given ρ as above, we define the von Neumann entropy of an ensemble as

$$S_{VN} = -\langle \mathrm{Tr}\rho \log \rho \rangle_{\Pi}. \tag{8}$$

The von Neumann entropy is therefore defined by the spectra of the Laplacian of the network.

Anand and Bianconi [2] showed that the Gibbs entropy is equal to the Shannon entropy in the thermodynamic limit for random graphs, and that for scale-free networks, the Shannon entropy and the von Neumann entropy are linearly related. Anand *et al.* [3] showed that the Shannon entropy and the von Neumann entropy are correlated for heterogeneous networks.

In statistical mechanics, for configurations drawn from canonical ensembles, the Shannon entropy corresponds to the entropy of classical systems, while the von Neumann entropy provides the statistical description of quantum systems. Therefore, the Shannon entropy and the von Neumann entropy are the numbers of bits needed to determine the description of the classical system and the quantum system generated by the ensembles, respectively. According to this understanding, we can intuitively interpret the Shannon entropy and von Neumann entropy of a network ensemble as the *descriptive complexity* of the networks generated by the ensemble.

G. Structural Entropy of Models of Networks

Choi and Szpankowski [15] defined a structural entropy for a model of networks. Given a random graph model $\mathcal{M}$, let $\mathcal{S}$ be the set of all graphs of the same type generated by model $\mathcal{M}$. For graphs $G, H \in \mathcal{S}$, we have that G and H have the same structure, if they are isomorphic. For a graph $G \in \mathcal{S}$, the probability of G, denoted by $p(G)$, is the sum of $p(H)$ for all the graphs H that are in $\mathcal{S}$ and are isomorphic to G. Then the structural entropy $H_{\mathcal{S}}$ of $\mathcal{S}$ is the Shannon entropy of the distribution $p(G)$ for all $G \in \mathcal{S}$, that is,

$$H_{\mathcal{S}} = -\sum_{G\in\mathcal{S}} p(G)\log p(G)). \tag{9}$$

III. The Challenges

A. Quantification of Structural Information

Brooks, Jr., [13] proposed the question of quantification of structural information as one of the three great challenges for half-century-old computer science: *"We have no theory, however, that gives us a metric for the information embedded in structure, especially in physical structures $\cdots$ I consider this missing metric to be the most fundamental gap in the theoretical understandings of information science and of computer science."* The question was early observed by Rashevsky [44] and by Trucco [49]. In fact, the same problem was implicitly stated by Shannon himself in his 1953 paper [48].

To better understand the great challenge left by Shannon, we examine the following example.

Suppose that we are given a structured data $G = (V, E)$ which is a graph. We want to analyse G. According to Shannon's information, we can get the Shannon information of G only by the following approach:

(1) (De-structuring) To define a distribution $\mathbf{p}$ of G, such as the distribution of degrees, or distances of G.
(2) (Shannon's information) To compute the Shannon information of $\mathbf{p}$, i.e., $I = H(\mathbf{p})$.
(3) (Information of G) We obtain the information I of G.

Step (1) gives an unstructured vector $\mathbf{p}$ by removing the structure of G. This step may lose the most interesting properties of G. Step (3) gives us the Shannon information I of G, which is just a number. The question is: what properties of G can we find from the Shannon information I of G?

Therefore, Shannon information gives us only a number for every graph G. However, we can not analyse the properties of graph G from the Shannon number I.

The problem above is fundamental, because, graphs are perhaps the most general and most intuitive mathematical structures for computer science. However, Shannon's information does not help much for us to analyse the graphs, i.e., structured data.

By the same reason, none of the existing measures in Section II supports the analysis of structured data, because, all the measures are a specific form of the Shannon entropy.

The situation above becomes worse when we analyse the large-scaled networking data, and unstructured big data.

To solve the problem left by Shannon, we need a new metric of structural information that supports the analysis of graphs, networks, structured data and even unstructured big data. Here we provide such a metric.

Before introducing our metric, we describe the general problem we will solve as follows:

Given a graph G, suppose that

(i) G is a structured, but noisy data,
(ii) G is obtained from evolution,
(iii) there are rules controlling the evolution of G, and
(iv) there are random variations occurred in the evolution of G.

Our questions for analysis of G include:

(1) How to measure the amount of randomness in the evolution of G?
(2) How to extract exactly the part of G constructed by rules, excluding the random variations?
(3) Can we distinct the part of G generated by rules and the part of G perturbed by random variations?

It is conceivable that the general problem is fundamental to many areas of both information science and computer science.

As we will see, our metric of high-dimensional structural information provides the principle for us to perfectly solve the problem, in general.

B. Dynamical Complexity of Networks

Networks are complex. However, what are the complexity measures of networks?

Naturally evolving networks are hard even to describe, to define or to store. This concerns the *static complexity of networks*. We notice that all the existing notions of various entropies of graphs and models of graphs can be regarded as static measures of complexity of networks.

However, millions of interactions, communications and operations may occur in every second of time in a real world network, and the real world networks are dynamically evolving. The complexity of a network is largely the dynamical complexity of the network to quantitatively measure the complexity of the interactions, communications, operations and evolution of the networks. Therefore, the great challenge is to measure the complexity of interactions, communications, operations and evolution of the real world networks, referred to as *dynamical complexity of networks*.

Real world networks are either naturally evolving in nature and society, or else are engineered. In a given real world network, individuals interact and communicate frequently and simultaneously. The complexity of such a dynamical network is largely a quantitative measure of the non-determinism of the interactions and communications in the network.

Real world networks have roughly two categories. The first category consists of those networks that are evolved by natural rules in nature and society such as citation networks, social networks and protein-to-protein interaction networks. The second category consists of those networks which are created in engineering such as computer networks, communication networks, and service networks.

For a network naturally evolving in nature and society, self-organisation is the basic form of social organisation in the network, leading to natural social groups of the network. Networks constructed in industry involve both self-organisation behaviours and engineering requirements. In either case, a real world network must have a natural structure such as the natural community structure and the natural hierarchical structure of the network. This natural structure, either the natural community structure or the natural hierarchical structure of the network, must play the central role in the dynamics of the network. Therefore, identification of the natural structure of a network is central to network dynamics. Before resolving this issue, we have to solve some fundamental problems.

To establish our theory, we have to resolve the following challenges: What is the dynamical complexity of networks?

In the present paper, we will establish the measures of structural information and dynamical complexity of networks, resolving the challenges.

IV. Overall Ideas

As the availability of large-scaled, noisy, folded, corrupted and structured data, the challenge we are facing is: Given a noisy or corrupted graph G, to define the information H of G such that H determines the essential structure T and true knowledge K of G. We assume that the true knowledge K of G consists of the rules, regulations and laws of G, that the true knowledge K of G is placed in the essential structure T of G, that the essential structure T of G is obtained from G by excluding the maximum amount of nondeterminism, uncertainty, and noises that occurred in G, and that the essential structure T of G is determined by the structural information H of G. Such a metric, once defined, allows us to decode the truth from noisy data, and supports analysis for networking systems.

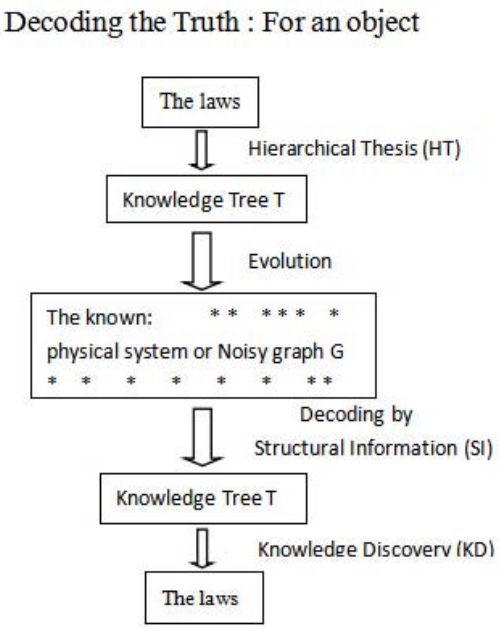

Fig. 1. Decoding the truth by structural information.

Here, we propose such metrics, and establish the fundamental theory of the metrics. First of all, we introduce the overall ideas of our concepts. The overall ideas of our theory is summarised in Figure 1. According to Figure 1, our theory provides the following approach to decoding the truth from physical systems or noisy data by the following steps: Given an object,

(1) Suppose that there are unknown laws of the object.
(2) (Hierarchical Thesis) Suppose that there is a natural structure of the object generated by the laws of the object, and that the natural structure is a hierarchical structure T.
The idea of the hierarchical thesis is to encode the graph by a high-dimensional encoding system and to measure the information required to position the vertices in the graph by using this high-dimensional encoding system. Given a network $G = (V, E)$, if G is evolved in nature, society or industry, then G has a natural hierarchical structure $\mathcal{T}$ such that for every vertex $v \in V$, v is encoded by a k-tuple $(i_1, i_2, \cdots, i_k)$ that determines the vertex v, for some k. For example, a person in a society is affiliated to "province", "city",

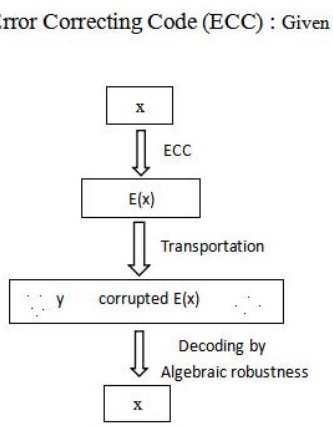

Fig. 2. Decoding error correcting code.

"institution" and "department", and so the *codes* of the "province", "city", "institution" and "department" uniquely determine the person. In this example, if the natural hierarchical structure $\mathcal{T}$ of G is given, then the knowledge of G can be easily extracted from $\mathcal{T}$, and furthermore, using $\mathcal{T}$, the dynamical complexity of G is the information required to determine the k-dimensional code $(i_1, i_2, \cdots, i_k)$ of the vertex v that is accessible from random walk in G.

The hierarchical thesis may hold in many systems in Nature and Society, although it is not easy to mathematically proven.

(3) (Physical or noisy graph) Suppose that the known data of the object form a physical system or noisy graph G generated by the laws of the object perturbed by noises.

(4) (Decoding by structural information) Our structural information determines and decodes the natural hierarchical structure T of G, referred to *the knowledge tree* of G.

(5) (Knowledge discovery) The laws of the object can be found from the knowledge tree T of G by knowledge discovering.

The framework of our theory is similar to decoding an error correcting code (ECC), depicted in Figure 2. According to Figure 2, in information transportation, we usually encode a string x by an ECC E to a string $E(x)$ such that even if the receiver gets a corrupted string y, from which we are able to decode the string x of the sender. This is the procedure of our framework in Figure 1.

Finally, we notice that the principle of decoding the error correcting code is the algebraic robustness of the codewords, and the principle of decoding the truth from data is to minimise the uncertainty by the structural information.

V. Graph Structural Information

A. One-Dimensional Structural Information

A real world network is a highly connected graph which is dynamically evolving, and in which individuals frequently interact with their neighbors. We use random walks to capture the interactions of nodes in a network, giving a notion of entropy which reflects the dynamical complexity of the network. This is the one-dimensional structural information of the network.

Before defining our notion, we recall the Shannon entropy for a distribution.

For a probability vector $\mathbf{p} = (p_1, \ldots, p_n)$, with $\sum_{i=1}^{n} p_i = 1$, the entropy function of $\mathbf{p}$ is defined by

$$H(p_1, \ldots, p_n) = -\sum_{i=1}^{n} p_i \log_2 p_i.$$

Intuitively speaking, in the definition of $H(p_1, p_2, \cdots, p_n)$, for every i, $l = -\log p_i$ is the length of binary representation of number $\frac{1}{p_i}$. This means that $\frac{1}{p_i}$ is one of the 2^l many numbers. For this reason, we interpret $-\log p_i$ as the "self-information of p_i". This also means that $-\log p_i$ is the amount of information needed to determine the code of i. Therefore $-\sum_{i=1}^{n} p_i \log_2 p_i$ is the average amount of information needed to determine the code of i chosen with probability distribution $\mathbf{p} = (p_1, p_2, \cdots, p_n)$.

1) Definitions: Given a graph $G = (V, E)$ with n nodes and m edges, we define the *one-dimensional structural information of* G or *positioning entropy of* G by using the stationary distribution of the degrees of G and the Shannon entropy function H.

Definition 1 (One-Dimensional Structural Information of Connected and Undirected Graphs): Let $G = (V, E)$ be an undirected and connected graph with n nodes and m edges. For each node $i \in \{1, 2, \cdots, n\}$, let d_i be the degree of i in G, and let $p_i = \frac{d_i}{2m}$. Then the stationary distribution of random walk in G is described by probability vector $\mathbf{p} = (p_1, p_2, \cdots, p_n)$. We define the one-dimensional structural information of G or the positioning entropy of G as follows:

$$\mathcal{H}^1(G) = H(\mathbf{p}) = H\left(\frac{d_1}{2m}, \ldots, \frac{d_n}{2m}\right) = -\sum_{i=1}^{n} \frac{d_i}{2m} \cdot \log_2 \frac{d_i}{2m}. \tag{10}$$

The one-dimensional structural information in Definition 1 is a special case of the parametric entropy in Subsection II-C. The differences are: 1) our one-dimensional structural information allows natural extensions to weighted and directed graphs, and 2) the parametric entropy is determined by arbitrarily given function f.

The one-dimensional structural information $\mathcal{H}^1(G)$ of a connected graph G measures the information required to determine the one-dimensional code of the node that is accessible from random walk in G with stationary distribution. By this understanding, it is also referred to as *positioning entropy* of G.

The definition of one-dimensional structural information or positioning entropy allows natural extension to weighted graphs (networks).

Definition 2 (Weighted Degree and Volume): Given a network $G = (V, E)$, suppose that the weights assigned to edges is defined by a weight function $w : E \to \mathbb{R}^+$. Define the weighted degree of node u to be $d_u = \sum_{v \in N(u)} w((u, v))$, where $N(u)$ is the set of neighbors of u. We say that a weighted

graph has k-bounded weight if for every edge e, $w(e) \leq k$. For a subset $U \subseteq V$, define the volume *of U to be $vol(U) = \sum_{v \in U} d_v$. Define $vol(G) = \sum_{v \in V} d_v$ to be the volume of G.*

For each node $u \in V$, let $p_u = \frac{d_u}{\mathrm{vol}(G)}$. We consider the random walk in which at each step, the probability of choosing a neighbor $v \in N(u)$ of the current node u is proportional to the weight of (u, v). It is easy to verify that the stationary distribution of this random walk in G is described by probability vector $\mathbf{p} = (p_1, p_2, \cdots, p_n)$, where p_i is the probability of visiting the i-th node.

The positioning entropy of weighted graph G is defined similarly to that in Equation (10).

Definition 3 (One-Dimensional Structural Information of Weighted and Connected Networks): Let $G = (V, E)$ be a connected and weighted graph with n nodes, m edges and weight function w. We define the one-dimensional structural information of G or positioning entropy of G as follows:

$$\mathcal{H}^1(G) = H(\mathbf{p}) = H\left(\frac{d_1}{vol(G)}, \ldots, \frac{d_n}{vol(G)}\right) = -\sum_{i=1}^{n} \frac{d_i}{vol(G)} \log_2 \frac{d_i}{vol(G)}. \tag{11}$$

Definition 4 (One-Dimensional Structural Information of Directed and Connected Graph): Given a connected and directed graph $G = (V, E)$, let $n = |V|$ be the number of nodes and m be the number of directed edges in G. For a node $v \in V$, we use d_v^{in} and d_v^{out} to denote the in- and out-degree of v in G respectively. We define the one-dimensional structural information of G or the positioning entropy of G as follows:

$$\mathcal{H}^1(G) := -\sum_{v \in V} \frac{d_v^{in}}{m} \log_2 \frac{d_v^{in}}{m}. \tag{12}$$

Finally, we define the one-dimensional structural information of a disconnected graph.

Definition 5 (One-Dimensional Structural Information of Disconnected Graphs): Given an undirected graph $G = (V, E)$,

1. *If $E = \emptyset$, then define $\mathcal{H}^1(G) = 0$.*
2. *Otherwise. Then Suppose that $G_1, G_2, \cdots, G_L$ are the induced subgraphs of all the connected components of G. Then we define the positioning entropy of G is the weighted average of the positioning entropies of G_i for all i's. That is,*

$$\mathcal{H}^1(G) = \frac{1}{\mathrm{Vol}(G)} \sum_{j=1}^{L} \mathrm{Vol}(G_j) \cdot \mathcal{H}^1(G_j), \tag{13}$$

where $\mathrm{Vol}(G)$ *is the volume of G,* $\mathrm{Vol}(G_j)$ *is the volume of G_j.*

In Definition 5, we have that: (1) is reasonable for which the reason is that there is no random walk which occurs in G, and (2) is reasonable since it simply follows the additivity of the Shannon entropy function.

2) Negative Weights: According to Definitions 1, 3 and 4, if there is a node i of a connected and weighted graph G such that the total weight of i is $w_i < 0$, then the definition of the one-dimensional structural information is no longer valid, since $\log \frac{w_i}{W}$ is undefined, where $W > 0$ is the total weight of the graph. In this case, we have to reduce the graph G to a new graph H such that for every node i, the total weight of i in H is greater than 0. Certainly, there are natural ways to reduce the graph G. For example, we keep all the nodes and edges of G for H, and for each edge e in G, we define the weight of e in H is e^w or 2^w, where $w = w(e)$ is the weight of e in G.

B. Two-Dimensional Structural Information

1) Structural Information of Graphs Given by a Partition: Let $G = (V, E)$ be a connected network. Suppose that $\mathcal{P} = \{X_1, X_2, \cdots, X_L\}$ is a partition of V, in which each X_j is called a module or a community.

By using the partition $\mathcal{P}$, we encode a node $v \in V$ by a pair of codes (i, j) such that j is the code of the community containing v, and i is the code of the node within its own community.

Considering the random walk with stationary distribution again, we need to define the *structural information of G given by $\mathcal{P}$.*

Definition 6 (Structural Information of a Network by a Partition): Given an undirected and connected graph $G = (V, E)$, suppose that $\mathcal{P} = \{X_1, X_2, \cdots, X_L\}$ is a partition of V. We define the structural information of G by $\mathcal{P}$ as follows:

$$\mathcal{H}^{\mathcal{P}}(G) = \sum_{j=1}^{L} \frac{V_j}{2m} \cdot H\left(\frac{d_1^{(j)}}{V_j}, \ldots, \frac{d_{n_j}^{(j)}}{V_j}\right) - \sum_{j=1}^{L} \frac{g_j}{2m} \log_2 \frac{V_j}{2m} = -\sum_{j=1}^{L} \frac{V_j}{2m} \sum_{i=1}^{n_j} \frac{d_i^{(j)}}{V_j} \log_2 \frac{d_i^{(j)}}{V_j} - \sum_{j=1}^{L} \frac{g_j}{2m} \log_2 \frac{V_j}{2m}, \tag{14}$$

where L is the number of modules in partition $\mathcal{P}$, n_j is the number of nodes in module X_j, $d_i^{(j)}$ is the degree of the i-th node of X_j, V_j is the volume of module X_j which is the sum of degrees of nodes in X_j, and g_j is the number of edges with exactly one endpoint in module X_j.

The intuition of $\mathcal{H}^{\mathcal{P}}(G)$ is as follows. Given a network G and a partition $\mathcal{P}$ of vertices of G, we encode every node v in G by a pair of codes (j, i) such that the code of the module X containing v is j, and the code of v in its own module X is i. The first term of $\mathcal{H}^{\mathcal{P}}(G)$ is the number of bits needed to determine the code of the node, v say, in its own module X, where v is the node accessible from a step of random walk in G, and the second term of $\mathcal{H}^{\mathcal{P}}(G)$ is the number of bits needed to determine the code of a module, X say, in the network, where X is the module accessible from a step of random walk from nodes outside of X. Notice that, for a random walk in G, the arrival node must be in a module of the partition $\mathcal{P}$, so the first term of $\mathcal{H}^{\mathcal{P}}(G)$ is always non-empty. However, the second term of $\mathcal{H}^{\mathcal{P}}(G)$ is nonzero only if the random walk enters a new module, otherwise,

we have already known the code of the module. This means that if a random walk is within the same module, then the code of the module is already known, so the corresponding second term is zero, in which case, we only need to determine the code of the accessible node in its known module. Therefore, $\mathcal{H}^{\mathcal{P}}(G)$ is the number of bits needed to determine the code (j, i) of the node accessible from a step of random walk in G, by using the partition $\mathcal{P}$.

Based on this understanding, we can interpret $\mathcal{H}^{\mathcal{P}}(G)$ as a 2-dimensional structural information, and $\mathcal{E}^1(G)$ as a 1-dimensional structural information of network.

The two-dimensional structural information of weighted graphs are defined similarly to that in Equations (14) - (19). Precisely, we have:

Definition 7 (Structural Information of Weighted Networks by a Partition): Let $G = (V, E)$ be a weighted graph with n nodes, m edges and weight function w. Suppose that $\mathcal{P} = \{X_1, X_2, \cdots, X_L\}$ is a partition of V. We define the structural information of G by $\mathcal{P}$ as

$$\begin{aligned}\mathcal{H}^{\mathcal{P}}(G) &= \sum_{j=1}^{L} \frac{V_j}{vol(G)} \cdot H\left(\frac{d_1^{(j)}}{V_j}, \ldots, \frac{d_{n_j}^{(j)}}{V_j}\right) \\ &\quad - \sum_{j=1}^{L} \frac{g_j}{vol(G)} \log_2 \frac{V_j}{vol(G)} \\ &= -\sum_{j=1}^{L} \frac{V_j}{vol(G)} \sum_{i=1}^{n_j} \frac{d_i^{(j)}}{V_j} \log_2 \frac{d_i^{(j)}}{V_j} \\ &\quad - \sum_{j=1}^{L} \frac{g_j}{vol(G)} \log_2 \frac{V_j}{vol(G)}, \end{aligned} \tag{15}$$

where L is the number of modules in partition $\mathcal{P}$, n_j is the number of nodes in module X_j, $d_i^{(j)}$ is the degree of the i-th node of X_j, V_j is the volume of module X_j, and g_j is the sum of the weights of the edges with exactly one endpoint in module X_j.

Our definitions of positioning entropy and structural information can be extended to directed graphs.

Given a directed graph $G = (V, E)$, let $n = |V|$ be the number of nodes and m be the number of directed edges in G. For a node $v \in V$, we use d_v^{in} and d_v^{out} to denote the in- and out-degree of v in G respectively.

For a subset of nodes $S \subseteq V$, we define the volume of S to be the total in-degrees of nodes in S, that is, $\text{vol}(S) = \sum_{v \in S} d_v^{\text{in}}$.

In this case, the volume of G is just the number of edges m.

Definition 8 (Two-Dimensional Structural Information of Directed Graphs): Given a directed and connected graph $G = (V, E)$ and a partition $\mathcal{P} = \{V_1, \ldots, V_L\}$ of G, we define the two-dimensional structural information of G by $\mathcal{P}$ as follows:

$$\begin{aligned}\mathcal{H}^{\mathcal{P}}(G) &:= -\sum_{j=1}^{L} \frac{vol(V_j)}{m} \sum_{v \in V_j} \frac{d_v^{\text{in}}}{vol(V_j)} \log_2 \frac{d_v^{\text{in}}}{vol(V_j)} \\ &\quad - \sum_{v \in V_j} \frac{g_j}{m} \log_2 \frac{vol(V_j)}{m}, \end{aligned} \tag{16}$$

where g_j is the number of edges going to nodes in V_j from nodes outside of V_j.

For weighted directed graphs, the definitions are similar. Let $G = (V, E)$ be a directed graph with weight function $w : E \to \mathbb{R}^+$. Define the weighted in-degree of a node v to be the sum of the weights of edges whose head endpoints are all v, that is, $d_v^{\text{in}} = \sum_{u:(u,v)\in E} w((u, v))$, where (u, v) is the directed edge from u to v. The volume of a subset S is defined to be the sum of weighted degrees of nodes in it. Then the positioning entropy and structural information of a weighted network can be defined similarly.

2) Two-Dimensional Structural Information of Graphs:

Definition 9 (Two-Dimensional Structural Information of Connected Networks): Let $G = (V, E)$ be connected graph. We define the two-dimensional structural information of G (also referred to such as the module entropy or as the local positioning entropy, of G) as follows:

$$\mathcal{H}^2(G) = \min_{\mathcal{P}} \{\mathcal{H}^{\mathcal{P}}(G)\}, \tag{17}$$

where $\mathcal{P}$ runs over all the partitions of G.

In the definition of two-dimensional structural information above, we require that the given graph G is connected. However, this condition is not necessary. In fact, there are many ways to extend the definitions to disconnected graphs. In our theory, we define the notion by additivity, following the law of Shannon entropy function.

Definition 10 (Two-Dimensional Structural Information of Disconnected Graphs): Given a graph G, suppose that $G_1, G_2, \cdots, G_L$ are the induced subgraphs of all the connected components of G. Then we define the two-dimensional structural information of G to be the weighted average of the two-dimensional structure entropies of all the subgraphs G_i's. That is,

$$\mathcal{H}^2(G) = \frac{1}{\text{Vol}(G)} \cdot \sum_{j=1}^{L} \text{Vol}(G_j) \cdot \mathcal{H}^2(G_j), \tag{18}$$

where $\text{Vol}(G)$, for each j, $\text{Vol}(G_j)$ is the volume of G_j.

To make sure that Definition 10 is well-defined, we notice that it is possible that there is a j such that G_j contains a single isolated node. In this case, the two-dimensional structural information of G_j is $\lim_{p \to 0} -p \log_2 p = 0$. So, for any graph G, if there is no edge among the nodes of G, then $\mathcal{E}(G) = 0$.

3) Normalised Two-Dimensional Structural Information of Networks: We may also define a normalized version of the two-dimensional structural information.

Definition 11 (Normalized Two-Dimensional Structural Information of Networks): For any graph $G = (V, E)$, we define the normalized two-dimensional structural information of G as follows:

$$\tau(G) = \frac{\mathcal{H}^2(G)}{\mathcal{H}^1(G)}. \tag{19}$$

C. High-Dimensional Structural Information

Real world networks generally have a hierarchical structure such that a module of a network may consist of a number

of submodules, which leads to a natural extension of the two-dimensional structural information to high-dimensional cases.

To define high-dimensional structural information, we introduce a partitioning tree of graphs. First, we consider the two-dimensional case. For a graph $G = (V, E)$, and a partition $\mathcal{P} = \{X_1, X_2, \cdots, X_L\}$ of V, we interpret the partition $\mathcal{P}$ by a partitioning tree $\mathcal{T}$ of hight 2 as follows: 1) first, we introduce the root node λ, and define a set of nodes $T_\lambda = V$, 2) we introduce L immediate successors for the root node denoted $\alpha_i = \lambda\hat{}\langle i\rangle$, where $i = 1, 2, \cdots, L$, and associate the set X_i with node α_i; thus, we define $T_{\alpha_i} = X_i$, and 3) for each α_i, we introduce $|X_i|$ immediate successors denoted $\alpha_i\hat{}\langle j\rangle$ for all $j \in \{1, 2, \cdots, |X_i|\}$, and each successor $\alpha_i\hat{}\langle j\rangle$ is associated with an element in X_i; thus, we define $T_{\alpha_i\hat{}\langle j\rangle}$ as the singleton of a node in $T_{\alpha_i} = X_i$.

Therefore, $\mathcal{T}$ is a tree of height 2, and all of the leaves of $\mathcal{T}$ are associated with singletons. For every node $\alpha \in \mathcal{T}$, T_α is the union of T_β for all of β values (of the immediate successors) of α, and the union of T_α for all of the nodes α values at the same level of the tree $\mathcal{T}$ is a partition of V.

Thus, the partitioning tree of a graph $G = (V, E)$ is a set of nodes such that each node is associated with a nonempty subset of vertices of graph G, and can be defined as follows:

Definition 12: (Partitioning Tree of Graphs): Let $G = (V, E)$ be an undirected and connected network. We define the partitioning tree $\mathcal{T}$ of G as a tree $\mathcal{T}$ with the following properties:

(1) *For the root node denoted λ, we define the set $T_\lambda = V$.*
(2) *For every node $\alpha \in \mathcal{T}$, the immediate successors of α are $\alpha\hat{}\langle j\rangle$ for j from 1 to a natural number N ordered from left to right as j increases.*
Therefore, $\alpha\hat{}\langle i\rangle$ is to the left of $\alpha\hat{}\langle j\rangle$ written as $\alpha\hat{}\langle i\rangle <_{\mathrm{L}} \alpha\hat{}\langle j\rangle$, if and only if $i < j$.
(3) *For every $\alpha \in \mathcal{T}$, there is a subset $T_\alpha \subset V$ that is associated with α.*
For α and β, we use $\alpha \subset \beta$ to denote that α is an initial segment of β. For every node $\alpha \neq \lambda$, we use α^- to denote the longest initial segment of α, or the longest β such that $\beta \subset \alpha$.
(4) *For every i, $\{T_\alpha \mid h(\alpha) = i\}$ is a partition of V, where $h(\alpha)$ is the height of α (note that the height of the root node λ is 0, and for every node $\alpha \neq \lambda$, $h(\alpha) = h(\alpha^-) + 1$).*
(5) *For every α, T_α is the union of T_β for all β's such that $\beta^- = \alpha$; thus, $T_\alpha = \cup_{\beta^- = \alpha} T_\beta$.*
(6) *For every leaf node α of $\mathcal{T}$, T_α is a singleton; thus, T_α contains a single node of V.*

We define the entropy of G by a partitioning tree $\mathcal{T}$ of G.

Definition 13: (Structural Information of a Graph by a Partitioning Tree): For an undirected and connected network $G = (V, E)$, suppose that $\mathcal{T}$ is a partitioning tree of G. We define the structural information of G by $\mathcal{T}$ as follows:

(1) *For every $\alpha \in \mathcal{T}$, if $\alpha \neq \lambda$, then define*

$$H^{\mathcal{T}}(G; \alpha) = -\frac{g_\alpha}{2m} \log_2 \frac{V_\alpha}{V_{\alpha^-}}, \tag{20}$$

where g_α is the number of edges from nodes in T_α to nodes outside T_α, V_β is the volume of set T_β, namely, the sum of the degrees of all the nodes in T_β.
(Remark: For an edge-weighted graph $G = (V, E)$, g_α is the sum of the weights of all the edges between T_α and nodes outside T_α, and the degree of a node $v \in V$ in G is the sum of the edge weights of all the edges incident to v. For a non-weighted graph, we regard the weight of an edge as 1.)
(2) *We define the structural information of G by the partitioning tree $\mathcal{T}$ as follows:*

$$\mathcal{H}^{\mathcal{T}}(G) = \sum_{\alpha \in \mathcal{T}, \alpha \neq \lambda} H^{\mathcal{T}}(G; \alpha). \tag{21}$$

For a weighted or directed graph $G = (V, E)$, if G is connected, and $\mathcal{T}$ is a partitioning tree of G, then the structural information of G given by $\mathcal{T}$ is defined similarly to Definition 13 by the same way as that for the one- and two-dimensional cases.

Definition 14 (K-Dimensional Structural Information): Let $G = (V, E)$ be a connected network. We define the K-dimensional structural information of G as follows:

$$\mathcal{H}^K(G) = \min_{\mathcal{T}} \{\mathcal{H}^{\mathcal{T}}(G)\}, \tag{22}$$

where $\mathcal{T}$ ranges over all of the partitioning trees of G of height K.

For a disconnected graph G, the K-dimensional structural information of G is defined similarly to Definition 10 as follows:

$$\mathcal{H}^K(G) = \frac{1}{\mathrm{Vol}(G)} \cdot \sum_{j=1}^{L} \mathrm{Vol}(G_j) \cdot \mathcal{H}^K(G_j), \tag{23}$$

where $\mathrm{Vol}(G)$, for each j, $\mathrm{Vol}(G_j)$ is the volume G_j, and $G_1, G_2, \cdots, G_L$ are all the connected components of G.

Our definition of the structural information defines the law of self-organisation of individuals in nature and society by implying the following *self-organisation principle*: minimising the non-determinism of a structure is the principle for the self-organisation of structures within naturally evolving networks.

In particular, the K-dimensional structural information of a graph G implies that minimisation of non-determinism is the principle of the self-organisation of a K-dimensional structure of a graph for $K > 1$. For $K = 1$, the K-dimensional structural information of a graph G is the positioning entropy of G.

D. Structural Information

Our K-dimensional structural information supports the analysis of graphs, networks, structured data and even unstructured data. To understand this, we introduce the following:

Definition 15 (Structural Information): Given a graph G, a natural number K, a height K partitioning tree $\mathcal{T}$ of G and a positive number $\delta \leq 1$:

(1) *We define the K-dimensional structure information of G by*

$$\mathcal{I}^K(G) = \mathcal{H}^K(G). \tag{24}$$

(2) *We say that $\mathcal{T}$ is a knowledge tree of G, if:*

$$\mathcal{H}^{\mathcal{T}}(G) = \mathcal{H}^{K}(G). \tag{25}$$

(3) *We say that $\mathcal{T}$ is a δ-knowledge tree of G, if:*

$$\mathcal{H}^{\mathcal{T}}(G) \leq \frac{1}{\delta} \cdot \mathcal{H}^{K}(G). \tag{26}$$

According to the Definition 15, for a given graph G, we compute the K-dimensional structural information $\mathcal{H}^{K}(G)$ of G, and simultaneously find a height K knowledge tree $\mathcal{T}$, satisfying the following principle:

1) The K-dimensional structural information $\mathcal{H}^{K}(G)$ of G is achieved or approximated by the K-dimensional structure $\mathcal{T}$ of G.
2) The K-dimensional structural information $\mathcal{H}^{K}(G)$ provides the controlling principle for the formation of the K-dimensional knowledge tree $\mathcal{T}$ by minimising the uncertainty or non-determinism of the K-dimensional structures of G.

   Therefore, $\mathcal{T}$, a knowledge tree of G, consisting of the rules, regulations, and orders of G, is found by minimising the random variations of the K-dimensional structures of the graphs, where the random variations are determined by our K-dimensional structural information.

Our K-dimensional structural information of graphs is completely different from the Shannon entropy.

According to the principles 1) and 2) above, our K-dimensional structural information of graphs is essentially the metric that allows us to fully or maximally detect the K-dimensional structure consisting of the rules, regulations, and orders of the graphs against the random variations occurring in the graphs. Our definition of the K-dimensional structural information perfects supports us to fully extract the order from un-ordered graphs, and to distinguish the order from disorder in a structured data. Our definition does not lose any property of the graphs. Our definition perfectly supports full analysis of networking data, and unstructured big data.

E. Dynamical Complexity of Networks

According to the definitions, the K-dimensional structural information of a graph G is exactly the measure of uncertainty of the K-dimensional structures of G. Hence we have:

Definition 16 (Dynamical Complexity of a Network): Given a network G and a natural number K, we define the K-dimensional dynamical complexity of G by

$$\mathcal{C}^{K}(G) = \mathcal{H}^{K}(G). \tag{27}$$

Definition 16 defines the dynamical complexity of G, that is, the K-dimensional complexity of interactions, communications, operations and evolution in G.

VI. Characterisation of Natural Structures of Networks

A. One-Dimensional Natural Structure

According to Definitions 1, 3, 4 and 5, for a graph G, it is easy to compute the one-dimensional structural information of G. However, what we can do with unstructured big data?

Networking is the current approach to processing the big data that are usually unstructured in the real world. The question is what is the principle to process the unstructured data in the real world.

Consider the following scenarios:

Case 1: Suppose that $v_1, v_2, \cdots, v_n$ are n objects such that for each pair (v_i, v_j), there is a weight $w(i, j)$ between the two objects v_i and v_j. Among the weights $w(i, j)$'s, there is only a small number of the weights that are non-trivial and meaningful, and many of the weights are probably just noise or trivial.

In this case, we are asked to construct a weighted graph G of nodes the objects $v_1, v_2, \cdots, v_n$ such that the weighted edges of G are exactly the edges between the objects with the non-trivial (i.e., large) weights. That is, in the construction of G, we must remove the noisy or trivial (or small) weights and maintain the nontrivial weights between the pairs of the objects. The desired graph G allows a systemical, unbiased, and connecting approach to analysing the natural structure of the objects $v_1, v_2, \cdots, v_n$.

Our one-dimensional structural information suggests a principle for constructing the desired graph G as follows:

(1) G is highly connected.
(2) The one-dimensional structural information of G is minimised.

(1) ensures that the connections of the unstructured objects $v_1, v_2, \cdots, v_n$ are well-defined, and (2) ensures that the non-determinism of networking of the unstructured objects $v_1, v_2, \cdots, v_n$ is minimised.

Case 2: Suppose that G is a graph, and v is a new node such that for each node u, there is some weight $w(v, u)$ between v and u, among which many of the weights $w(v, u)$ are trivial or simply noisy. However, we actually don't know which weights are non-trivial, and which are trivial or noisy.

In this case, we are asked to create the links from v to nodes in G such that all the important weights are maintained, and all the trivial or noisy weights are removed.

To solve this problem, the one-dimensional structural information provides a principle. It is actually the same principle as that in Case 1 above.

Therefore, the significance of the one-dimensional structural information is that it provides a principle for networking of unstructured big data. This points to a new direction of networking engineering.

B. Two-Dimensional Natural Structure

According to Darwin's evolution theory, animals form social groups for survival in the evolution of species. Similarly, in a network that is naturally evolving in nature and society, individuals form natural communities. The natural communities of a network are largely the results of self-organisation behaviours. The self-organisation of individuals in a network includes many factors, such as, maximisation of individual interests, competitions among individuals, maximisation of group interests, social organisation and randomness etc. To understand, let us consider the following examples.

(1) A graduate seeks for a job
When a graduate seeks for a job, he/she may know several candidate institutions, from which he/she eventually decides to join based on the negotiations and a few rounds of interviews with each of the candidate institutions. In this procedure, the graduate certainly wants to maximise his/her own interests, minimise his/her cost, the institutions want to maximise their own interests, and every stage of the procedure, there is always something by chance, i.e., the factor of randomness.

(2) A person joins a club
Someone joins a certain club, because, he/she may realise his/her own interests through the club. In this case, the motivation of the person is to realise his/her own interests, instead of the club interests.

(3) A person attends a meeting or a seminar
In this case, person knows something about the meeting and seminar, from which, the person would expect something for his/her own interests.

The examples above show that natural communities do form in real world networks, and that the natural communities have many factors. In our definition of the two-dimensional structural information of a graph, we use the non-determinism of random walks in the graph as a measure to unify the various factors of the self-organisation behaviours and social organisations. According to the definition of the two-dimensional structural information, we have the following principles:

(1) Minimisation of the two-dimensional structural information is equivalent to the minimisation of non-determinism of community structures of the network.

(2) Two-dimensional structural information minimisation of a network is the principle for both the self-organisation of individuals in the network and the formation of natural communities of the network.

We better understand both (1) and (2) above as follows. By definition, given a connected graph G and a vertex partition $\mathcal{P}$ of G, the structural information of G by $\mathcal{P}$ is the overall amount of information to position the two-dimensional code of the node that is accessible from random walk in G with stationary distribution. By the definition of the two-dimensional structural information, $\mathcal{H}^2(G)$ is the minimum of $\mathcal{H}^{\mathcal{P}}(G)$ over all partitions $\mathcal{P}$ of G. This means that $\mathcal{H}^2(G)$ captures the partition $\mathcal{P}$ of G that achieves the minimum information of the two-dimensional structure of G. By definition, structural information is the measure of the non-determinism of the structure of the graph. Therefore, the two-dimensional structural information may very well characterise the principle of the natural structure in the real world networks and the principle of the self-organisation of the individuals in the real world. This principle also implies that in the organisation of social relations, people fear uncertainty or non-determinism and usually follow the choices that minimising uncertainty or non-determinism.

For $K = 2$ and for a network G, we say that a partition $\mathcal{P}$ of G forms the *natural community structure of* G, if the two-dimensional structural information of G given by $\mathcal{P}$ is approximately the two-dimensional structural information of G, that is, $\mathcal{H}^2(G) \approx \mathcal{H}^{\mathcal{P}}(G)$.

Finally, we notice that, the two-dimensional structural information of networks provides a novel measure for the quality of community identification algorithms of networks, and for $K > 2$, the K-dimensional structural information provides the ideas for exploring the high-dimensional natural structures of networks. Li *et al.* [32] have shown that a simple greedy algorithm based on the two-dimensional structural information precisely identifies the natural communities of networks by models and precisely identifies the ground-truth communities in real world social networks.

For community detection, Newman and Girvan [42] had proposed the modularity maximisation principle. By the definition of modularity, this principle is to maximise the function of the modules of a network. It captures the role of social organisation in a network. Because, the principle characterises the partition $\mathcal{P}$ of a network G such that each module in $\mathcal{P}$ plays a maximum role as a functional module.

As mentioned above, social organisation is only one factor for the formation of natural communities. Therefore, the modularity maximisation principle fails to characterise the natural communities of networks. This is in deed the case, as shown in Li *et al.* [32].

C. High-Dimensional Natural Structure

A real world network G may have a natural structure in which a natural module contains several submodules. Our K-dimensional structural information of a network characterises the hierarchical structure through a partitioning tree $\mathcal{T}$ such that the non-determinism of the positioning by a K-dimensional coding system is minimised.

Given a network G, we say that a height K partitioning tree $\mathcal{T}$ is the K-dimensional natural structure of G, if the K-dimensional structural information of G given by $\mathcal{T}$ is approximately the K-dimensional structural information of G, that is, $\mathcal{H}^K(G) \approx \mathcal{H}^{\mathcal{T}}(G)$.

Therefore, the K-dimensional structural information of networks characterises the K-dimensional coding natural structure of the networks. This discovering has been verified by Li *et al.* [36].

The results established for the principle for high-dimensional natural structures of networks. We remark that for high-dimensional structures of networks, our K-dimensional structural information is the first such principle.

Remark: 1) For each $K \geq 2$, our K-dimensional structural information characterises the natural K-dimensional structure of networks. Our interest is to identify the natural structures of networks, instead of the algorithmic structures of the networks, which are simply the outputs of certain algorithms. 2) Our definition of the one-dimensional structural information of establishes the principle for us to construct the natural network G for an unstructured data, U say. 3) For $K \geq 2$, our K-dimensional structural information of a network G establishes the principle for us to identify the natural K-dimensional structure of the network G.

Li *et al.* [36] have proposed the method of three-dimensional gene map for cancer cell types and subtypes based on the K-dimensional structural information of networks for $k = 1,\ 2$ and 3.

The experiments in [31], [32], and [36] have shown that for both $K = 2$ and 3, K-dimensional structural information minimisation is in deed the currently best strategy to detect the natural structure, that is, the ground-truth structures of real world networks.

D. Natural Structures of Networks

This progress leads to the following:

Definition 17 (Natural Structure of Networks): Given a network G, a natural number $K > 1$, a positive constant $\delta \leq 1$, and a partitioning tree $\mathcal{T}$ of G with height K, we say that $\mathcal{T}$ is a K-dimensional δ-natural structure of G, if:

$$\mathcal{H}^{\mathcal{T}}(G) \leq \frac{1}{\delta} \cdot \mathcal{H}^{K}(G) \tag{28}$$

Definition 17 provides an approach to theoretically analysing the high-dimensional natural structures of networks (that is, the high-dimensional ground-truth structures of networks).

VII. Basic Theorems

In this section, we prove some basic properties of the K-dimensional structural information of graphs.

Theorem 18 (Locality Theorem): Given a connected graph G, let $\mathcal{P}$ be the partition of nodes of G such that each module X of $\mathcal{P}$ contains a single node of V, and let $\mathcal{Q}$ be the partition of G containing only one module of the whole set V. Then, we have

$$\mathcal{H}^{\mathcal{P}}(G) = \mathcal{H}^{\mathcal{Q}}(G). \tag{29}$$

Proof: By the definition of the structural information of G given by a partition. □

Theorem 18 ensures that if we start with the trivial partition $\mathcal{P}$ in which each module is a singleton, and merge the modules towards minimising the two-dimensional structural information of the graph, it is impossible to get the trivial partition with the single module of the whole set of nodes. Therefore, by minimising the structural information, none of the modules can grow up as the whole network. The same result can be established for the K-dimensional structural information of graphs for any $K > 2$.

Theorem 18 indicates that the simple idea of greedily merging the modules has already been a well-defined strategy for computing the two-dimensional structural information of a network.

Remark: Modularity maximisation fails to satisfy this property. It has been shown that it is possible (even always) to obtain the trivial partition consisting of the unique module of the whole set of nodes by merging modules towards maximising modularity from the trivial partition in which each module is a singleton. This is the reason why there is always an outside condition governing the algorithms of modularity maximisation. Our experiments in [31], [32], and [36] also showed that in most cases, the algorithms of modularity maximisation fail to detect natural communities (or ground-truth communities) of real world networks.

Theorem 19 (Separation Theorem): Let $G = (V, E)$ ne a connected graph. Suppose that $\mathcal{P}$ is a partition of V, and X and Y are two modules of $\mathcal{P}$. If there is no edge between the nodes in X and the nodes in Y. Let $Z = X \cup Y$. Let $\mathcal{Q}$ be the partition consisting Z and all the modules of $\mathcal{P}$ other than X and Y. Then, we have:

$$\mathcal{H}^{\mathcal{P}}(G) < \mathcal{H}^{\mathcal{Q}}(G). \tag{30}$$

Proof: By the definition of the two-dimensional structural information of graphs given by a partition, a proof is referred to Li *et al.* [32]. □

Theorem 19 ensures that the two-dimensional structural information minimisation forbids to merge two disconnected modules into one module.

Clearly, Theorem 19 can be extended to K-dimensional structural information of graphs for $K > 2$. Therefore, structural information minimisation ensures that disconnected parts of a network must be in distinct modules of the network.

VIII. Fundamental Properties of Structural Information and Dynamical Complexity of Networks

Network dynamics studies the laws of dynamical movement in a network. Among the various kinds of actions, interactions and communications in a network, positioning is the most fundamental operation. Given a network G, the K-dimensional structural information of G is the minimum overall number of bits needed to determine the K-dimensional code of the node that is accessible from a step of random walk by the stationary distribution in the network G.

By Definitions 15 and 16, for a network G and a natural number K,

- The K-dimensional structural information of G is

$$\mathcal{I}^{K}(G) = \mathcal{H}^{K}(G).$$

- The K-dimensional dynamical complexity of G is

$$\mathcal{C}^{K}(G) = \mathcal{H}^{K}(G).$$

Therefore, the K-dimensional structural information provides both the quantification of structural information and dynamical complexity of networks simultaneously.

The fundamental properties of the K-dimensional structural information explore simultaneously the principles of structural information and dynamical complexity of networks.

Before establishing our theory of structural information and dynamical complexity of networks, we outline some of the basic properties of the K-dimensional structural information of networks.

Let G be a graph and K be a natural number. By definition, $\mathcal{H}^{K}(G)$ satisfy the following properties:

(1) *Network dependency*
$\mathcal{H}^{K}(G)$ is a function of the sizes of G that is uniquely determined by the network G.

This property fails to hold for the existing measures of complexity of networks, because:
(i) For the modularity of a graph G, it is a number in the interval $[0, 1]$ which is independent of the sizes of the graph G.
(ii) For the existing entropy measures of graphs, including the Gibbs entropy, Shannon entropy and the von Neumann entropy, the the measures are determined by an ensemble which may generate the given graph, instead of the given graph itself.

(2) *Additivity*
$\mathcal{H}^K(G)$ follows an additivity law due to the additivity of the entropy function H.
This property fails to be satisfied by any of the existing entropy measures of graphs, including the Gibbs, Shannon and von Neumann entropies. However, additivity is essential to the theory of structural information and dynamical complexity of networks.

(3) *Locality*
This has been established in Theorem 18.

(4) *Dynamics*
As analysed before, $\mathcal{H}^K(G)$ is the quantitative measure of the K-dimensional structural information and the complexity of the dynamical actions in the network by using a K-dimensional coding system. Therefore, $\mathcal{H}^K(G)$ characterises both the structural information and the dynamics of network G.
This property fails to be satisfied by the existing entropy measures, including the Gibbs, Shannon and von Neumann entropies, because the later measures characterise the complexity for determining the code or the description of the graph from a network ensemble, which is completely irrelevant to the dynamics of the given network. As a matter of fact, each of the existing measures, including the Gibbs, Shannon and von Neumann entropies defines the measured quantity of a given network that is largely determined by an ensemble outside the given network.

(5) *Essentiality*
The notion of $\mathcal{H}^K(G)$ measures both the K-dimensional structural information and the complexity of K-dimensional positioning of G. Due to the fact that positioning is the primitive operation of various actions in a network, the notion $\mathcal{H}^K(G)$ should be fundamental to network dynamics.

(6) *Robustness*
According to the definition of $\mathcal{H}^K(G)$, if the graph G evolves to a graph G' such that the total number of changes in nodes and edges is small, then

$$\mathcal{H}^K(G) \approx \mathcal{H}^K(G'). \tag{31}$$

Therefore, the K-dimensional structural information and the dynamical complexity of networks are robust for any changes of small-scales.
In particular, we conjecture that for a classical model $\mathcal{M}$ and for a given type of the model, the K-dimensional structural information of the networks of the model with the given type is robust to the variations of the model.

(7) *Linking nature to science*
$\mathcal{H}^K$ may capture the laws of nature that can be applied to the new information sciences by the following principles:
(i) Structural information (K-dimensional) minimisation is the principle of nature evolving; and
(ii) Structural information minimisation is the principle for networking and for network information processing.

(8) *Local computability*
Theorem 19 ensures that structural information minimisation considers the actions of an individual or a module to the objects that link to the given individual or module. This property guarantees the local computability of the structural information of networks.

(9) *Incremental computability*
The additivity and separability in Theorem 19 ensure that in the evolution of networks, the corresponding structural information of the networks can be computed by an incremental approach in the sense that at any stage, we compute only the small number of modifications of the networks.

(10) *Applicability*
We have seen that $\mathcal{H}^K$ provides the principle for identifying the natural K-dimensional structure of networks, and defines the dynamical complexity of K-dimensional structures of networks. Due to the essentiality and primitivity of the measure, it is easy to think of other applications of the theory, including such as, in network controls, network security, security robustness, and game theory in networks etc. To the authors, it is conceivable that the applications of the theory maybe more than what we can image at this stage due to the fact that the theory maybe developed and extended to other disciplines including mathematics, information theory, computer science, machine learning, network, physics, biological and medical sciences.

IX. One-Dimensional Structural Information of Graphs

In this section, we establish both the lower and upper bounds of graphs. The graphs considered in this section are undirected, connected, and either simple or with balanced weights to be defined later. We prove the result by dividing into two cases, one for simple graphs, and the other for graphs with balanced weights.

Recall that we also call the one-dimensional structure entropy as positioning entropy. For convenience, we use both the two names.

A. Simple Graphs

A graph is called simple, if there is no self-loop or multi-edge in the graph. For simple graphs, we have:

Theorem 20 (Lower Bound of Positioning Entropy of Simple Graphs): Let $G = (V, E)$ be an undirected, connected, and simple graph with m edges, i.e., $|E| = m$. Then:

$$\mathcal{H}^1(G) \geq \frac{1}{2}\left(\log_2 m - 1\right).$$

Proof: Let $|V| = n$. For every node $v \in V$, let d_v be the degree of v in G. Consider the random walk in G with stationary distribution $\mathbf{p} = (p_1, \ldots, p_n)$, where $p_i = \frac{d_i}{2m}$. By definition, $\mathcal{H}^1(G)$ is the Shannon entropy of distribution $\mathbf{p}$. It is well-known that $\mathcal{H}^1(G)$ is a lower bound on the lengths of encoding schemes of the nodes of G by using alphabet $\{0, 1\}$.

Suppose that we have already had an encoding $\{C(u) \in \{0,1\}^* : u \in V\}$ for all the nodes in V. By assumption, $C(u)$ is a binary representation of length $|C(u)| = \lceil -\log_2 \frac{d_u}{2m} \rceil$. Moreover, suppose that $C(\cdot)$ is prefix-free. By Kraft-McMillan Theorem, there exists such an encode $C(\cdot)$. Furthermore, the Huffman code is the shortest such a prefix code. For this encoding, the average length of codewords is at most $\mathcal{H}^1(G) + 1$, which almost touches the lower bound.

For encoding for the edges in E, we consider an undirected edge (u, v) to be two directed ones $u \to v$ and $v \to u$. We encode every $u \to v$ with the codeword $C(u) \circ C(v)$, in which $\circ$ represents concatenation. The prefix-freeness of $C(\cdot)$ guarantees that the codeword of an edge defined by this way is unique.

In the random walk, denoted by P, under the stationary distribution $\mathbf{p}$, every directed edge has equal probability to be crossed over. So the entropy of the distribution on edges is $\log_2 2m$. We use the codewords of edges to represent the random walk. That is, at each step, we use the codeword of the edge being crossed over to represent this step. Thus a lower bound of the expected length of codeword for each step is the entropy of the uniform distribution on edges, i.e., $\log_2 2m$. Moreover, by the definition of the random walk, this expected length can be calculated as follows. Let $N(u)$ be the set of neighbors of node u, we have that

$$
\begin{aligned}
&E_{(u,v)\in P}[|C(u) \circ C(v)|] \\
&= \sum_{u\in V} \left[\frac{d_u}{2m} \left(\left\lceil -\log_2 \frac{d_u}{2m} \right\rceil + \frac{1}{d_u} \sum_{v\in N(u)} \left\lceil -\log_2 \frac{d_v}{2m} \right\rceil \right) \right] \\
&= \sum_{u\in V} \frac{d_u}{2m} \left\lceil -\log_2 \frac{d_u}{2m} \right\rceil + \sum_{u\in V} \sum_{v\in N(u)} \frac{1}{2m} \left\lceil -\log_2 \frac{d_v}{2m} \right\rceil \\
&= \sum_{u\in V} \frac{d_u}{2m} \left\lceil -\log_2 \frac{d_u}{2m} \right\rceil + \sum_{v\in V} \sum_{u\in N(v)} \frac{1}{2m} \left\lceil -\log_2 \frac{d_v}{2m} \right\rceil \\
&= \sum_{u\in V} \frac{d_u}{2m} \left\lceil -\log_2 \frac{d_u}{2m} \right\rceil + \sum_{v\in V} \frac{d_v}{2m} \left\lceil -\log_2 \frac{d_v}{2m} \right\rceil \\
&\leq 2(\mathcal{H}^1(G) + 1).
\end{aligned}
$$

Recall that $E_{(u,v)\in P}[|C(u) \circ C(v)|]$ is lower bounded by the entropy of the uniform distribution on directed edges, i.e., $\log_2 2m$. So we have $2(\mathcal{H}^1(G) + 1) \geq \log_2 2m$, and thus

$$\mathcal{H}^1(G) \geq \frac{1}{2} \log_2 2m - 1 \geq \frac{1}{2} \left(\log_2 m - 1 \right).$$

□

B. Graphs With Balanced Weights

In this subsection, we consider the weighted graphs where self-loops and multi-edges are allowed. Since the multi-edges between a single pair of nodes can be viewed as a single edge with the weight that is the sum of those of the multi-edges, we assume that there are no multi-edges in graphs. So the number of edges in a graph of size n is at most $n(n+1)/2$ but their weights vary. For the positioning entropy of graphs with balanced weights, we have the following lower bound.

Theorem 21 (Lower Bound of Positioning Entropy of Graphs of Balanced Weights): Let $G = (V, E)$ be a connected graph with weight function w. Let $m = |E|$ be the number of edges. If the ratio of maximum weight and minimum weight is at most m^ϵ, that is $\frac{\max_{e\in G}\{w(e)\}}{\min_{e\in G}\{w(e)\}} \leq m^\epsilon$, for some constant $\epsilon < 1$, then:

$$\mathcal{H}^1(G) \geq \frac{1}{2} \left[(1 - \epsilon) \log_2 m \ - 1 \right].$$

Proof: The proof is similar to that of Theorem 20. Here we consider the random walk in weighted graphs. We also assume that we have already had a prefix free code $\{C(u) \in \{0,1\}^* : u \in V\}$ for V, where $C(u)$ has length $|C(u)| = \lceil -\log_2 \frac{d_u}{\mathrm{vol}(G)} \rceil$. Then we consider every undirected edge (u, v) to be two directed ones and denote the directed graph by G', in which each edge $u \to v$ has weight $w((u, v))$. Then we encode every directed edge $u \to v$ by the codeword $C(u) \circ C(v)$. The prefix-freeness of $C(\cdot)$ guarantees the uniqueness of the codeword for each directed edge of G.

In the random walk, denoted by P, it is easy to verify that $\mathbf{p} = \left(\frac{d_1}{\mathrm{vol}(G)}, \ldots, \frac{d_n}{\mathrm{vol}(G)} \right)$ is a stationary distribution, and every directed edge $u \to v$ is visited with probability $w((u, v))/\mathrm{vol}(G)$. Then when we use the codewords of edges to represent the random walk, the expected length of codeword that represents one step random walk is hence

$$
\begin{aligned}
&E_{(u,v)\in P}[|C(u) \circ C(v)|] \\
&= \sum_{u\in V} \Big[\frac{d_u}{\mathrm{vol}(G)} \Big(\left\lceil -\log_2 \frac{d_u}{\mathrm{vol}(G)} \right\rceil \\
&\quad + \frac{w((u,v))}{d_u} \sum_{v\in N(u)} \left\lceil -\log_2 \frac{d_v}{\mathrm{vol}(G)} \right\rceil \Big) \Big] \\
&= \sum_{u\in V} \frac{d_u}{\mathrm{vol}(G)} \left\lceil -\log_2 \frac{d_u}{\mathrm{vol}(G)} \right\rceil \\
&\quad + \sum_{u\in V} \sum_{v\in N(u)} \frac{w((u,v))}{\mathrm{vol}(G)} \left\lceil -\log_2 \frac{d_v}{\mathrm{vol}(G)} \right\rceil \\
&= \sum_{u\in V} \frac{d_u}{\mathrm{vol}(G)} \left\lceil -\log_2 \frac{d_u}{\mathrm{vol}(G)} \right\rceil \\
&\quad + \sum_{v\in V} \sum_{u\in N(v)} \frac{w((u,v))}{\mathrm{vol}(G)} \left\lceil -\log_2 \frac{d_v}{\mathrm{vol}(G)} \right\rceil \\
&= \sum_{u\in V} \frac{d_u}{\mathrm{vol}(G)} \left\lceil -\log_2 \frac{d_u}{\mathrm{vol}(G)} \right\rceil \\
&\quad + \sum_{v\in V} \frac{d_v}{\mathrm{vol}(G)} \left\lceil -\log_2 \frac{d_v}{\mathrm{vol}(G)} \right\rceil \\
&\leq 2(\mathcal{H}^1(G) + 1).
\end{aligned}
$$

On the other hand, $E_{(u,v)\in P}[|C(u) \circ C(v)|]$ is lower bounded by the entropy of the probability distribution on edges for the

random walk in G'. Thus

$$E_{(u,v)\in P}[|C(u)\circ C(v)|] \geq -\sum_{e\in G'}\frac{w(e)}{\mathrm{vol}(G)}\log_2\frac{w(e)}{\mathrm{vol}(G)} \geq \min_{e\in G'}\left\{-\log_2\frac{w(e)}{\mathrm{vol}(G)}\right\}. \quad (32)$$

Since for every edge e, $\frac{\max_{e\in G}\{w(e)\}}{\min_{e\in G}\{w(e)\}} \leq m^{\epsilon}$, the above value is at least

$$\log_2\frac{2m\cdot\min_{e\in G}\{w(e)\}}{\max_{e\in G}\{w(e)\}} \geq \log_2\frac{2m}{m^{\epsilon}} = (1-\epsilon)\log_2 m \ +1.$$

So $2(\mathcal{H}^1(G)+1) \geq (1-\epsilon)\log_2 m + 1$. Theorem 21 follows. □

We say that a graph is *typical*, if it is undirected, connected, and either simple or with weights satisfying the conditions in Theorem 21.

By Shannon's theory, for typical graphs, there is a trivial upper bound of the positioning entropy of the graphs. This is, for a network G of size n, $\mathcal{H}^1(G) \leq \log_2 n$ holds. The result follows from the convexity of logarithmic function. This gives rise to the following

Theorem 22 (Exact Bounds of Positioning Entropy of Typical Graphs): For a typical graph G, the positioning entropy of G is

$$\mathcal{H}^1(G) = \Theta(\log n), \quad (33)$$

where n is the number of nodes in G.

Proof: By the arguments above. □

According to Theorems 20, 21 and 22, for an undirected and connected graph G of n nodes and m edges with balanced weights, the one-dimensional structural information or positioning entropy of G is $\mathcal{H}^1(G) = \alpha\cdot\log n$ for some constant $\alpha \leq \frac{1}{2}$, where the constant α is determined by the density of the graph, i.e., the number m of edges of the graph G. However, if the weights of G are highly unbalanced in the sense that $\frac{\max_{e\in G}\{w(e)\}}{\min_{e\in G}\{w(e)\}} > m^{\delta}$ for some constant δ, then the one-dimensional structure entropy of G could be arbitrarily small. Our results demonstrate that the one-dimensional structure entropy of G is $\beta\cdot\log n$ for some β determined by both the density of G, and the weights ratio $\frac{\max_{e\in G}\{w(e)\}}{\min_{e\in G}\{w(e)\}}$. Therefore, the one-dimensional structural information is a quantitative measure that simultaneously characterises both the density and the weights balance of graphs.

For any graph G, by definition, the one-dimensional structural information of G is very easy to compute. However, this property is no longer true for the two-dimensional structural information of a graph. Of course, the one-dimensional structural information gives rise a trivial upper bound for the two-dimensional structural information.

For any graph G, by definition, $\mathcal{H}^1(G) = \mathcal{H}^{\mathcal{P}}(G)$ for the partition $\mathcal{P}$ in which each module contains a single node. Therefore $\mathcal{H}^2(G) \leq \mathcal{H}^1(G)$.

By Equation (33), we have that a trivial upper bound for the two-dimensional structural information for both simple and weight balanced graphs.

Proposition 23: For a graph G, if G is either simple or weights balanced, then

$$\mathcal{H}^2(G) = O(\log n). \quad (34)$$

Proof: By definition of the one- and two-dimensional structural information of graphs. □

X. General Principles of Two-Dimensional Structural Information of Graphs

In this section, we establish some basic relationships among the well-known notions of modularity, conductance and the one- and two-dimensional structural information.

First we recall the notion of modularity and conductance.

Newman and Girvan [42] defined the notion of modularity to quantitatively measure the quality of community structure of a network. It is built based on the assumptions that random graphs are not expected to have community structures and that a network has a community structure if it is far from random graphs.

Let $G = (V, E)$ be a network with n nodes and m edges. Given a partition $\mathcal{P}$ of G, the modularity of G by the partition $\mathcal{P}$ is defined by

$$\sigma^{\mathcal{P}}(G) = \frac{1}{2m}\Sigma_{i,j}(A_{ij}-P_{ij})\delta(C_i,C_j), \quad (35)$$

where the sum runs over all pairs of vertices, A is the adjacency matrix, P_{ij} is the expected number of edges between vertices i and j in a null graph, i.e., a random copy of G. $\delta(C_j,C_j)=1$ if $C_i=C_j$, and 0 otherwise, C_k is a module of the partition $\mathcal{P}$.

By Equation (35), $\sigma^{\mathcal{P}}(G)$ is intuitively the distance between G by partition $\mathcal{P}$ and its own random copy P.

A standard null model assumes that the expected degree after averaging over all possible configurations matches the actual degree of the original graph [27]. Such a null model is essentially equivalent to the configuration model [37], [39], in which each node i is associated with d_i half-edges, where d_i is the degree of node i in G, and all the half-edges are joined randomly. It is easy to see that $P_{ij} = d_i d_j/2m$. Hence the modularity of G by $\mathcal{P}$ can then be rewritten as follows:

$$\sigma^{\mathcal{P}}(G) = \sum_{l=1}^{L}\left[\frac{k_l}{m}-\left(\frac{V_l}{2m}\right)^2\right], \quad (36)$$

where L is the number of modules in partition $\mathcal{P}$, k_l is the number of edges whose both endpoints are in module l, and V_l is the volume of module l. Note that the first term of each sum term represents the fraction of edges of G inside the module and the second term represents the expected fraction of edges that would be in the null model.

We define the *modularity of G* by

$$\sigma(G) = \max_{\mathcal{P}}\{\sigma^{\mathcal{P}}(G)\}. \quad (37)$$

Many people, especially physicists, take modularity as a measure of community structure of networks due to the intuition that if the modularity $\sigma(G)$ of G is large, then G is far from the random copy of G, and if $\sigma(G)$ is small, then G is close to a random copy of G. In communication theory,

conductance is used to measure the power of communications and information spreading. This principle implies that the power of communications and high modularity of a network cannot be achieved simultaneously.

Now we are ready to establish our results.

We first introduce the notion of conductance of a graph which will be frequently used in our results and proofs.

Given a graph $G = (V, E)$, and a subset S of V, the conductance of S in G is given by

$$\Phi(S) = \frac{|E(S, \bar{S})|}{\min\{\mathrm{vol}(S), \mathrm{vol}(\bar{S})\}}, \tag{38}$$

where $E(S, \bar{S})$ is the set of edges with one endpoint in S and the other in the complement of S, i.e. $\bar{S}$, and $\mathrm{vol}(X)$ is the sum of degrees d_x for all $x \in X$. The conductance of G is defined to be the minimum of $\Phi(S)$ over all subsets S's, that is:

$$\Phi(G) = \min_{S \subset V}\{\Phi(S)\}. \tag{39}$$

A. Modularity Principle

We first investigate the notion of modularity. Modularity is built based on the assumption that random graphs are not expected to have community structure and that a network has a community structure if it is far from random graphs. Intuitively, the modularity of a graph G is the distance between G and its own random copy. Our first result is the following upper bound of modularity for arbitrarily given graphs.

For Newman's modularity, we have the following general principle.

Theorem 24 (Modularity Principle): Given an arbitrary graph $G = (V, E)$, let $\Phi(G)$ be the conductance of G. and $\sigma(G)$ be the modularity of G. Then, the following inequality holds:

$$\sigma(G) \leq 1 - \Phi(G). \tag{40}$$

Proof: Suppose that $\mathcal{P} = \{X_1, X_2, \cdots, X_L\}$ is a partition of V with which $\sigma(G)$ is achieved, that is $\sigma(G) = \sigma^{\mathcal{P}}(G)$. Recall that for the partition $\mathcal{P}$,

$$\sigma^{\mathcal{P}}(G) = \sum_{l=1}^{L}\left[\frac{k_l}{m} - \left(\frac{V_l}{2m}\right)^2\right],$$

where L is the number of modules X_l's in partition $\mathcal{P}$, k_l is the number of edges whose both endpoints are in module X_l, V_l is the volume of module X_l, and g_l is the number of global edges from X_l to nodes outside of X_l.

Suppose without loss of the generality that for every l, $V_l \leq m$, where m is the number of edges in G.

By the definition of $\Phi(G)$, for each l,

$$\frac{g_l}{V_l} \geq \Phi(G).$$

Thus

$$\frac{\sum_{l=1}^{L} g_l}{\sum_{l=1}^{L} V_l} \geq \Phi(G).$$

Note that $\sum_{l=1}^{L} V_l = 2m$ and $\sum_{l=1}^{L} g_l$ is exactly twice of the number of edges that are between different modules X_j's of $\mathcal{P}$, which is also $2(m - \sum_{l=1}^{L} k_l)$.

Thus

$$\sigma^{\mathcal{P}}(G) = \sum_{l=1}^{L}\left[\frac{k_l}{m} - \left(\frac{V_l}{2m}\right)^2\right] \leq \frac{\sum_{l=1}^{L} k_l}{m} \leq 1 - \Phi(G).$$

Therefore

$$\sigma(G) = \sigma^{\mathcal{P}}(G) \leq 1 - \Phi(G).$$

□

By definition, the modularity of a graph G is intuitively the distance between G and its random copy, and the conductance $\Phi(G)$ is a measure of combinatorial optimization. The former is a probabilistic measure, and the latter is a combinatorial measure. By this reason, the modularity principle is highly nontrivial. Furthermore, by the modularity principle, we have that for every graph G, if the conductance of G is large, then G is close to its random copy, so that G has a random structure, and that if the modularity of G is large, then G is far from its random copy, implying that G has a well-defined structure.

B. Difference of the Positioning Entropy and the Structural Information

In this subsection, we establish a useful relationship between the one-dimensional and the two-dimensional structure entropies of an arbitrarily given graph.

Theorem 25 (Difference Principle of the Positioning Entropy and the Structural Information): Let $G = (V, E)$ be a connected graph. Suppose that $\mathcal{P}$ is a partition of V with the notations the same as that in the definitions of $\mathcal{H}^1(G)$ and $\mathcal{H}^{\mathcal{P}}(G)$. Then the positioning entropy of G, $\mathcal{H}^1(G)$, and the structural information of G by given $\mathcal{P}$, i.e., $\mathcal{H}^{\mathcal{P}}(G)$, satisfy the following properties:

(1) *The positioning entropy of G satisfies:*

$$\mathcal{H}^1(G) = -\sum_{j=1}^{L}\frac{V_j}{2m}\sum_{i=1}^{n_j}\frac{d_i^{(j)}}{V_j}\log_2\frac{d_i^{(j)}}{V_j} - \sum_{j=1}^{L}\frac{V_j}{2m}\log_2\frac{V_j}{2m}. \tag{41}$$

(2)

$$\mathcal{H}^1(G) - \mathcal{H}^{\mathcal{P}}(G) = -\sum_{j=1}^{L}\frac{V_j - g_j}{2m}\log_2\frac{V_j}{2m} \tag{42}$$

(3) *Assume that for each j, $V_j \leq m$, for $m = |E|$. Then*

$$\mathcal{H}^1(G) - \mathcal{H}^{\mathcal{P}}(G) = -\sum_{j=1}^{L}(1 - \Phi(X_j))\frac{V_j}{2m}\log_2\frac{V_j}{2m} \tag{43}$$

where $\Phi(X_j)$ is the conductance of X_j in G.

Proof: By Definitions 1 and 6, for the partition $\mathcal{P}$ of V,

$$\mathcal{H}^{\mathcal{P}}(G) = -\sum_{j=1}^{L} \frac{V_j}{2m} \sum_{i=1}^{n_j} \frac{d_i^{(j)}}{V_j} \log_2 \frac{d_i^{(j)}}{V_j} - \sum_{j=1}^{L} \frac{g_j}{2m} \log_2 \frac{V_j}{2m}, \tag{44}$$

and

$$\mathcal{H}^1(G) = H\left(\frac{d_1}{2m}, \ldots, \frac{d_n}{2m}\right) = -\sum_{i=1}^{n} \frac{d_i}{2m} \cdot \log_2 \frac{d_i}{2m}. \tag{45}$$

By the additivity of the entropy function, for the partition $\mathcal{P}$,

$$\mathcal{H}^1(G) = -\sum_{j=1}^{L} \frac{V_j}{2m} \sum_{i=1}^{n_j} \frac{d_i^{(j)}}{V_j} \log_2 \frac{d_i^{(j)}}{V_j} - \sum_{j=1}^{L} \frac{V_j}{2m} \log_2 \frac{V_j}{2m}.$$

(1) follows.

The information saved by $\mathcal{P}$ is

$$\mathcal{H}^1(G) - \mathcal{H}^{\mathcal{P}}(G) = -\sum_{j=1}^{L} \frac{V_j - g_j}{2m} \log_2 \frac{V_j}{2m}.$$

(2) follows.

For every $j \in \{1,\ 2, \cdots, L\}$, let Φ_j be the conductance of X_j in G, i.e., $\Phi_j = \Phi(X_j)$. Then for every $j \in \{1,\ 2, \cdots, L\}$, if $V_j \leq |E|$, then $\Phi_j = \frac{g_j}{V_j}$.

Assume that for each j, $V_j \leq m$, for $m = |E|$. By (2), we have

$$\mathcal{H}^1(G) - \mathcal{H}^{\mathcal{P}}(G) = -\sum_{j=1}^{L} (1 - \Phi_j) \frac{V_j}{2m} \log_2 \frac{V_j}{2m}.$$

(3) follows.

Equation (43) builds a bridge between the positioning entropy and structural information of a graph through conductances of the modules of the partition $\mathcal{P}$. By using this equation, we are able to establish a normalized structural information principle and a structural information principle. □

C. Normalized Structural Information Principle and Structural Information Principle

Recall Definition 11 for the notion of normalised structural information of a graph G, denoted $\tau(G) = \frac{\mathcal{H}^2(G)}{\mathcal{H}^1(G)}$.

By using Equations (41) and (43), we are able to establish a normalized structural information principle and a structural information principle for general graphs. Given a graph $G = (V, E)$, let $\Phi(G)$ be the conductance of G. Then we have:

For the normalized two-dimensional structural information, we have the following general principle.

Theorem 26 (Normalized Structural Information Principle): Given a graph $G = (V, E)$, let $\Phi(G)$ be the conductance of G. Then the normalized structural information of G satisfies the following:

$$\tau(G) \geq \Phi(G). \tag{46}$$

Proof: By Equation (41),

$$\mathcal{H}^1(G) = -\sum_{j=1}^{L} \frac{V_j}{2m} \sum_{i=1}^{n_j} \frac{d_i^{(j)}}{V_j} \log_2 \frac{d_i^{(j)}}{V_j} - \sum_{j=1}^{L} \frac{V_j}{2m} \log_2 \frac{V_j}{2m}. \tag{47}$$

Let $H_1 = -\sum_{j=1}^{L} \frac{V_j}{2m} \sum_{i=1}^{n_j} \frac{d_i^{(j)}}{V_j} \log_2 \frac{d_i^{(j)}}{V_j}$, and $H_2 = -\sum_{j=1}^{L} \frac{V_j}{2m} \log_2 \frac{V_j}{2m}$.

So $\mathcal{H}^1(G) = H_1 + H_2$.

By Equation (43), we have

$$\begin{aligned} \tau^{\mathcal{P}}(G) =: \frac{\mathcal{H}^{\mathcal{P}}(G)}{\mathcal{H}^1(G)} &= \frac{H_1 - \sum_{j=1}^{L} \frac{g_j}{2m} \log_2 \frac{V_j}{2m}}{H_1 + H_2} \\ &= \frac{H_1 - \sum_{j=1}^{L} \frac{\Phi_j V_j}{2m} \log_2 \frac{V_j}{2m}}{H_1 + H_2} \\ &\geq \frac{H_1 + \Phi(G) \cdot H_2}{H_1 + H_2} \\ &\geq \Phi(G). \end{aligned} \tag{48}$$

The result above holds for any nontrivial partition $\mathcal{P}$ of G. Therefore $\tau(G) \geq \Phi(G)$. □

By the result in (46), we have the following theorem.

Theorem 27 (Structural Information Principle): For any graph G, the structural information of G follows:

$$\mathcal{H}^2(G) \geq \Phi(G) \cdot \mathcal{H}^1(G), \tag{49}$$

where $\Phi(G)$ is the conductance of G, and $\mathcal{H}^1(G)$ is the positioning entropy of G.

Proof: By Theorem 26. □

The structural information principle is a result of the normalized structural information principle. The normalized structural information principle may have fundamental implications in both theory and new applications. For instance, it says that the information theoretical notion of normalized structural information can be lower bounded by the combinatorial measure of conductance. Notice that, both structural information and conductance are fundamental measures in complex systems, including networks. This implies that both the normalized structural information principle and the structural information principle may play an essential role in not only networks, but also other complex systems in nature and society.

The structural information principle poses a dilemma for communication networks. Let $G = (V, E)$ be a communication network. To make sure that information can be easily and quickly spread in the whole network G, we usually require that for any set $S \subset V$, the conductance of S in G is large, so that the conductance of G is large. However, by the normalized structural information principle, if $\Phi(G)$ is large, then local positioning system in G becomes hard. This means that the optimization of information spreading in networks and the best positioning system of the networks can not be achieved simultaneously. This is a *network dilemma* explored by our structural information principle.

The network dilemma can be interpreted as either the *incompressible principle of network information* or the *hardness principle of network positioning*. The reason is that in engineering networks, traditionally, it is required that the networks are expanders in the sense that the conductances of the networks are large. However, network information compression and positioning in networks require that the two-dimensional structural information of the networks are minimised. Theorem 27 indicates that if the conductance $\Phi(G)$ of network G is large then the two-dimensional structural information $\mathcal{H}^2(G)$ of G cannot be small. The challenge must be resolved by new theory of networks, consisting of both a local theory and a global theory such that the minimisation of two-dimensional structural information is realised by the local theory, and the expanding requirement is satisfied by a global core of the network that is an expander, details are referred to [31] and [34].

XI. Lower Bounds of Two-Dimensional Structural Information for Graphs

In this section, we establish the lower bounds of two-dimensional structural information for simple graphs and the graphs with balanced weights.

A. Simple Graphs

Theorem 28 (Lower Bounds of Two-Dimensional Structural Information of Simple Graphs): Let $G = (V, E)$ be an undirected, connected and simple graph with number of edges $|E| = m$. Then the two-dimensional structural information of G satisfies

$$\mathcal{H}^2(G) = \Omega(\log_2 \log_2 m). \tag{50}$$

Proof: At the beginning of this proof, we introduce a useful property called *additivity* of our positioning entropy, which is in general the *additivity* of the entropy function H and can be verified directly. We omit its proof here.

Lemma 29 (Additivity Law): For any partition $\mathcal{P} = \{V_1, \ldots, V_L\}$ of V, the positioning entropy of graph G defined as Equation (10) satisfies

$$\mathcal{H}^1(G) = \sum_{j=1}^{L} \frac{vol(V_j)}{2m} \cdot H\left(\frac{d_1^{(j)}}{vol(V_j)}, \ldots, \frac{d_{n_j}^{(j)}}{vol(V_j)}\right) - \sum_{j=1}^{L} \frac{vol(V_j)}{2m} \log_2 \frac{vol(V_j)}{2m}. \tag{51}$$

By comparing this with the definition of structural information Equation (14), we find that the only difference between these two is the coefficients in the second summation. Since we have known by Theorem 20 that $\mathcal{H}^1(G) \geq \frac{1}{2}(\log_2 m - 1)$, we only need to show that the value will not decrease too much after being converted to the form of $\mathcal{H}^{\mathcal{P}}$ for any partition $\mathcal{P}$.

Then we assume that each module V_j in partition $\mathcal{P}$ is connected. We can make this assumption for the following reason. Suppose without loss of generality that V_1 is disconnected, that is, $V_1 = U_1 \cup U_2$ and there is no edge between U_1 and U_2. Define a new partition $\mathcal{P}' = \{U_1, U_2, V_2, V_3, \ldots, V_L\}$. We can prove that $\mathcal{H}^{\mathcal{P}}(G) \geq \mathcal{H}^{\mathcal{P}'}(G)$, and thus a lower bound for $\mathcal{H}^{\mathcal{P}'}(G)$ is also a lower bound for $\mathcal{H}^{\mathcal{P}}(G)$.

Lemma 30:

$$\mathcal{H}^{\mathcal{P}}(G) \geq \mathcal{H}^{\mathcal{P}'}(G).$$

Proof: For node sets X and Y, let $e(X, Y)$ denote the number of edges between X and Y, and $e(X, X)$ denote that within X. Note that in the expression of $\mathcal{H}^{\mathcal{P}}(G) - \mathcal{H}^{\mathcal{P}'}(G)$, the terms unrelated to V_1 can be canceled. Since $e(V_1, \overline{V}_1) = e(U_1, \overline{U}_1) + e(U_2, \overline{U}_2)$, we have

$$\begin{aligned}
&\mathcal{H}^{\mathcal{P}}(G) - \mathcal{H}^{\mathcal{P}'}(G) \\
&= -\sum_{v \in V_1} \frac{d_v}{2m} \log_2 \frac{d_v}{\mathrm{vol}(V_1)} + \left(\sum_{v \in U_1} \frac{d_v}{2m} \log_2 \frac{d_v}{\mathrm{vol}(U_1)} + \sum_{v \in U_2} \frac{d_v}{2m} \log_2 \frac{d_v}{\mathrm{vol}(U_2)}\right) - \frac{e(V_1, \overline{V}_1)}{2m} \log_2 \frac{\mathrm{vol}(V_1)}{2m} \\
&\quad + \left(\frac{e(U_1, \overline{U}_1)}{2m} \log_2 \frac{\mathrm{vol}(U_1)}{2m} + \frac{e(U_2, \overline{U}_2)}{2m} \log_2 \frac{\mathrm{vol}(U_2)}{2m}\right) \\
&= \frac{1}{2m}\left(\sum_{v \in U_1} d_v \log_2 \frac{\mathrm{vol}(V_1)}{\mathrm{vol}(U_1)} + \sum_{v \in U_2} d_v \log_2 \frac{\mathrm{vol}(V_1)}{\mathrm{vol}(U_2)}\right) \\
&\quad + \frac{1}{2m}(e(U_1, \overline{U}_1) \log_2 \mathrm{vol}(U_1) \\
&\quad + e(U_2, \overline{U}_2) \log_2 \mathrm{vol}(U_2) - e(V_1, \overline{V}_1) \log_2 \mathrm{vol}(V_1)) \\
&= \frac{1}{2m}\left(e(U_1, U_1) \log_2 \frac{\mathrm{vol}(V_1)}{\mathrm{vol}(U_1)} + e(U_2, U_2) \log_2 \frac{\mathrm{vol}(V_1)}{\mathrm{vol}(U_2)}\right) \\
&\geq 0.
\end{aligned} \tag{52}$$

Lemma 30 follows. □

To simplify the notations, for each $j \in [L]$, we denote

$$\rho_j := \frac{\mathrm{vol}(V_j)}{2m},$$
$$H_j := -\sum_{v \in V_j} \frac{d_v}{\mathrm{vol}(V_j)} \log_2 \frac{d_v}{\mathrm{vol}(V_j)},$$

and

$$\Phi(V_j) := \frac{g_j}{\mathrm{vol}(V_j)}.$$

Here H_j is in fact the positioning entropy within module V_j and $\Phi(V_j)$ is its conductance. Then we show that for each module V_j, either its conductance is not too low or the positioning entropy within it is not too small. Let $0 < \gamma < 1$ be a constant. Then we have

Lemma 31: For each module $V_j \in \mathcal{P}$, we have that either $\Phi(V_j) \geq 1/\log_2^{\gamma} m$ or $H_j \geq \frac{\gamma}{3} \log_2 \log_2 m$ holds.

Proof: We only have to show that if $\Phi(V_j) < 1/\log_2^{\gamma} m$, then $H_j \geq \frac{\gamma}{3} \log_2 \log_2 m$.

Consider the following multi-graph G_j: the node set of G_j is just V_j and all edges of G within V_j are preserved. For the edges in G with exactly one endpoint in V_j, we

cut off them and match all the remaining half-edges in V_j freely (self-loops are permitted and each accounts for one to degree). We denote the resulting multi-graph by G_j. By this construction, the maximum weight (multiple number) of edges in G_j is at most g_j. Note that the positioning entropy of G_j is exactly H_j since the degrees of all nodes are preserved from G. Noting that G_j is connected, by Equation (32) in the proof of Theorem 21, we know that

$$2(H_j+1) \geq \log_2 \frac{\mathrm{vol}(G_j)}{\max_{e\in G_j}\{w_j(e)\}},$$

where w_j is the weight function in G_j. Since G is a simple graph, $\max_{e\in G_j}\{w_j(e)\} \leq g_j$. Since $\Phi(V_j) = g_j/\mathrm{vol}(G_j) < 1/\log_2^{\gamma} m$, we have that

$$H_j \geq \frac{1}{2}\log_2 \frac{\mathrm{vol}(G_j)}{g_j} - 1 \geq \frac{\gamma}{3}\log_2\log_2 m$$

holds for sufficiently large m. Lemma 31 follows. □

Recall that

$$\mathcal{H}^1(G) = \sum_{j\in[L]} \rho_j \cdot \left(H_j - \log_2 \rho_j\right) \geq \frac{1}{2}\log_2 m \ - 1,$$

and

$$\mathcal{H}^{\mathcal{P}}(G) = \sum_{j\in[L]} \rho_j \cdot \left(H_j - \Phi(V_j)\log_2 \rho_j\right).$$

Now, to complete the proof of Theorem 28, we divide the modules in $\mathcal{P}$ into two subsets according to Lemma 31. Define

$$A = \{j \in [L] : \Phi(V_j) \geq \frac{1}{\log_2^{\gamma} m}\},$$

and

$$B = [L] \setminus \ A.$$

By Lemma 31, we know that for any $j \in B$, $H_j \geq \frac{\gamma}{3}\log_2\log_2 m$. Let $0 < c < 1/2$ be a constant. We consider the following two cases:

Case 1: If $\sum_{j\in A}\rho_j\left(H_j - \log_2\rho_j\right) \geq c\log_2 m - 1$, then

$$\begin{aligned}\mathcal{H}^{\mathcal{P}}(G) &\geq \sum_{j\in A}\rho_j\cdot\left(H_j - \Phi(V_j)\log_2\rho_j\right)\\ &\geq c\cdot\Phi(V_j)\log_2 m - 1 \geq c\log_2^{1-\gamma} m - 1.\end{aligned}$$

Case 2: Otherwise, $\sum_{j\in A}\rho_j\left(H_j - \log_2\rho_j\right) < c\log_2 m - 1$, and thus $\sum_{j\in B}\rho_j\left(H_j - \log_2\rho_j\right) > \left(\frac{1}{2} - c\right)\log_2 m$. We consider the following two subcases:

Subcase 1: If $\sum_{j\in B}\rho_j H_j \geq \left(\frac{1}{4} - \frac{c}{2}\right)\log_2 m$, then certainly

$$\begin{aligned}\mathcal{H}^{\mathcal{P}}(G) &\geq \sum_{j\in B}\rho_j\cdot\left(H_j - \Phi(V_j)\log_2\rho_j\right)\\ &\geq \left(\frac{1}{4} - \frac{c}{2}\right)\log_2 m.\end{aligned}$$

Subcase 2: Otherwise, $\sum_{j\in B}\rho_j H_j < \left(\frac{1}{4} - \frac{c}{2}\right)\log_2 m$, and thus $-\sum_{j\in B}\rho_j\log_2\rho_j > \left(\frac{1}{4} - \frac{c}{2}\right)\log_2 m$. Since $-\log_2\rho_j \leq \log_2\frac{1}{2m} = \log_2 m + 1$, we know that

$$\sum_{j\in B}\rho_j \geq \frac{1}{4} - \frac{c}{2}$$

holds for sufficiently large m. Therefore,

$$\begin{aligned}\mathcal{H}^{\mathcal{P}}(G) &\geq \sum_{j\in B}\rho_j\cdot\left(H_j - \Phi(V_j)\log_2\rho_j\right)\\ &\geq \sum_{j\in B}\rho_j H_j \geq \frac{\gamma}{3}\left(\frac{1}{4} - \frac{c}{2}\right)\log_2\log_2 m.\end{aligned}$$

Putting all cases together, we know that $\mathcal{H}^{\mathcal{P}}(G) = \Omega(\log_2\log_2 m)$ for any partition $\mathcal{P}$. Theorem 28 follows. □

B. Graphs With Balanced Weights

Theorem 32 (Lower Bound of Two-dimensional Structural Information of Graphs with Balanced Weights): Let $G = (V, E)$ be a connected graph with weight function w. Let $m = |E|$ be the number of edges. If the ratio of maximum weight and minimum weight is at most $\log_2^{\epsilon} m$, that is $\frac{\max_{e\in G}\{w(e)\}}{\min_{e\in G}\{w(e)\}} \leq \log_2^{\epsilon} m$, for some constant $\epsilon < 1$, then the structural information of G satisfies

$$\mathcal{H}^2(G) = \Omega(\log_2\log_2 m). \tag{53}$$

Proof: The proof is similar to that of Theorem 28. It is easy to verify that for weighted graphs with corresponding definitions of degree and volume given in Definition 2, Lemma 29 and 30 still hold.

Suppose that the maximum weight of edges in G is W^+ and the minimum one is W_-. So $W^+/W_- \leq \log_2^{\epsilon} m$. For a node set $S \subseteq V$, let $\partial(S)$ denote the set of edges in G which have exactly one endpoint in S, and $w(\partial(S))$ denote the sum of the weights of edges in $\partial(S)$. Denote

$$\begin{aligned}\rho_j &:= \frac{\mathrm{vol}(V_j)}{\mathrm{vol}(G)},\\ H_j &:= -\sum_{v\in V_j}\frac{d_v}{\mathrm{vol}(V_j)}\log_2\frac{d_v}{\mathrm{vol}(V_j)},\end{aligned}$$

and

$$\Phi(V_j) := \frac{w(\partial(V_j))}{\mathrm{vol}(V_j)}.$$

As an analogue to Lemma 31, the following lemma can be proved. Let $\epsilon \leq \gamma < 1$ be a constant.

Lemma 33: For each module $V_j \in \mathcal{P}$, we have that either $\Phi(V_j) \geq 1/\log_2^{\gamma} m$ or $H_j \geq \frac{1}{3}(\gamma - \epsilon)\log_2\log_2 m$ holds.

Proof: Similarly to the proof of Lemma 31, we only have to show that if $\Phi(V_j) < 1/\log_2^{\gamma} m$, then $H_j \geq \frac{1}{3}(\gamma - \epsilon)\log_2\log_2 m$.

Construct weighted multi-graph G_j as follows: the node set of G_j is V_j and all edges of G within V_j are preserved. For the edges in G with exactly one endpoint in V_j, each of them forms a self-loop and its weight is maintained. If there are more than one self-loops associating to a single node, then merge them to be one self-loop and its weight is the sum of the weights of them. By the construction, the maximum weight

of edges in G_j is at most $w(\partial(V_j))+W^+$. Since the degree of each node in G_j has been preserved from G, the positioning entropy of G_j is exactly H_j. By Equation (32) again, we know that

$$2(H_j+1) \geq \log_2 \frac{\mathrm{vol}(G_j)}{\max_{e\in G_j}\{w_j(e)\}} \geq \log_2 \frac{\mathrm{vol}(G_j)}{w(\partial(V_j))+W^+},$$

where w_j is the weight function in G_j. Since G is connected, $\partial(V_j)$ is not empty, and so $w(\partial(V_j))$ must be at least W_-. Since $\Phi(V_j) = w(\partial(V_j))/\mathrm{vol}(V_j) < 1/\log_2^\gamma m$, we have that

$$w(\partial(V_j)) \geq W_- \geq \frac{W^+}{\log_2^\epsilon m},$$

and thus, $W^+ \leq w(\partial(V_j))\cdot \log_2^\epsilon m$. Therefore,

$$\begin{aligned} H_j &\geq \frac{1}{2}\log_2 \frac{\mathrm{vol}(G_j)}{w(\partial(V_j))+W^+} - 1 \\ &\geq \frac{1}{2}\log_2 \frac{\mathrm{vol}(G_j)}{(1+\log_2^\epsilon m)\cdot w(\partial(V_j))} - 1 \\ &\geq \frac{1}{2}\log_2 \frac{\log_2^\gamma m}{1+\log_2^\epsilon m} \\ &\geq \frac{1}{3}(\gamma-\epsilon)\log_2\log_2 m \end{aligned}$$

holds for sufficiently large m. Lemma 33 follows. □

Then we divide the modules in $\mathcal{P}$ into three subsets as follows. For each $j \in [L]$, let $m(G_j)$ denote the number of edges in G_j. So G_j is also the number of edges in G each of which has at least one endpoint in V_j. Define

$$A = \{j \in [L] : \Phi(V_j) \geq \frac{1}{\log_2^\gamma m}\},$$
$$B = \{j \in [L] : m(G_j) \leq \log_2^{\frac{1-\epsilon}{2}} m\},$$

and

$$C = [L] \setminus (A \cup B)$$

By Lemma 33, we know that for any $j \in C$, $\Phi(V_j) < 1/\log_2^\gamma m$, and so $H_j \geq \frac{1}{3}(\gamma-\epsilon)\log_2\log_2 m$. Recall that by Theorem 21,

$$\mathcal{H}^1(G) \geq \frac{1}{2}\big[(1-\epsilon)\log_2 m - 1\big].$$

Consider the following three cases:

Case 1: If $\sum_{j\in A}\rho_j\left(H_j - \log_2\rho_j\right) \geq \frac{1}{6}(1-\epsilon)\log_2 m - \frac{1}{2}$, then

$$\begin{aligned} \mathcal{H}^{\mathcal{P}}(G) &\geq \sum_{j\in A}\rho_j\cdot\left(H_j - \Phi(V_j)\log_2\rho_j\right) \\ &\geq \left[\frac{1}{6}(1-\epsilon)\log_2 m - \frac{1}{2}\right]\cdot\Phi(V_j) \\ &\geq \frac{1}{6}(1-\epsilon)\log_2^{1-\gamma} m - \frac{1}{2}. \end{aligned}$$

Case 2: Otherwise, if $\sum_{j\in B}\rho_j\left(H_j - \log_2\rho_j\right) \geq \frac{1}{6}(1-\epsilon)\log_2 m$, then

$$\begin{aligned} \Phi(V_j) &\geq \frac{W_-}{\mathrm{vol}(G_j)} \geq \frac{W_-}{2m(G_j)\cdot W^+} \\ &\geq \frac{1}{2\log_2^{\frac{1-\epsilon}{2}} m\cdot\log_2^\epsilon m} = \frac{1}{2\log_2^{\frac{1+\epsilon}{2}} m}. \end{aligned}$$

So

$$\begin{aligned} \mathcal{H}^{\mathcal{P}}(G) &\geq \sum_{j\in B}\rho_j\cdot\left(H_j - \Phi(V_j)\log_2\rho_j\right) \\ &\geq \left[\frac{1}{6}(1-\epsilon)\log_2 m\right]\cdot\Phi(V_j) \\ &\geq \frac{1}{12}(1-\epsilon)\log_2^{\frac{1-\gamma}{2}} m. \end{aligned}$$

Case 3: Otherwise, we know that $\sum_{j\in A\cup B}\rho_j\left(H_j - \log_2\rho_j\right) < \frac{1}{3}(1-\epsilon)\log_2 m - \frac{1}{2}$. Since $[L] = A\cup B\cup C$, we have that $\sum_{j\in C}\rho_j\left(H_j - \log_2\rho_j\right) \geq \frac{1}{6}(1-\epsilon)\log_2 m$. We consider the following two subcases:

Subcase 1: If $\sum_{j\in C}\rho_j H_j \geq \frac{1}{12}(1-\epsilon)\log_2 m$, then

$$\begin{aligned} \mathcal{H}^{\mathcal{P}}(G) &\geq \sum_{j\in C}\rho_j\cdot\left(H_j - \Phi(V_j)\log_2\rho_j\right) \\ &\geq \sum_{j\in C}\rho_j H_j \geq \frac{1}{12}(1-\epsilon)\log_2 m. \end{aligned}$$

Subcase 2: Otherwise, $\sum_{j\in C}\rho_j H_j < \frac{1}{12}(1-\epsilon)\log_2 m$, and thus $-\sum_{j\in C}\rho_j\log_2\rho_j > \frac{1}{12}(1-\epsilon)\log_2 m$. Since for any $j \in C$,

$$\begin{aligned} -\log_2\rho_j &= \log_2\frac{\mathrm{vol}(G)}{\mathrm{vol}(V_j)} \\ &\leq \log_2\frac{2m\cdot W^+}{2m(G_j)\cdot W_-} \\ &\leq \log_2\frac{m\log_2^\epsilon m}{\log_2^{\frac{1-\epsilon}{2}} m} \\ &\leq \log_2 m + \log_2\log_2 m \\ &\leq 2\log_2 m, \end{aligned}$$

we have that

$$\sum_{j\in C}\rho_j \geq \frac{1}{24}(1-\epsilon).$$

Therefore,

$$\begin{aligned} \mathcal{H}^{\mathcal{P}}(G) &\geq \sum_{j\in C}\rho_j\cdot\left(H_j - \Phi(V_j)\log_2\rho_j\right) \\ &\geq \sum_{j\in C}\rho_j H_j \\ &\geq \frac{1}{24}(1-\epsilon)\cdot\frac{1}{3}(\gamma-\epsilon)\log_2\log_2 m \\ &= \frac{1}{72}(1-\epsilon)(\gamma-\epsilon)\log_2\log_2 m. \end{aligned}$$

Putting all cases together, we know that $\mathcal{H}^{\mathcal{P}}(G) = \Omega(\log_2\log_2 m)$ for any partition $\mathcal{P}$. Theorem 32 follows. □

We notice that the proofs of Theorems 28 and 32 contain a general method for establishing lower bounds for the 2-dimensional structural information of graphs. The method consists of a few lemmas, that is, the additivity property in Lemma 29, the connected property in Lemma 30, the module property in Lemma 31, and the analysis of divided cases etc.

XII. Upper Bounds of Two-Dimensional Structural Information of Classic Data Structures

In this section, we will establish the upper bounds of two-dimensional structure entropies of the graphs corresponding to classical data structures. We will show that trees, grids and the networks generated from the small world model with appropriate choice of parameters satisfy a (two-dimensional) structural information minimization principle.

A. Trees

For simplicity, we suppose without loss of the generality that the trees are complete binary trees. A complete binary tree is a tree whose non-leaf nodes has exactly two children and every leaf node has the same depth (In this section, for notational simplicity, we define the depth of a node to be the number of nodes on the unique path from this node to the root). So the complete binary tree of depth H has exactly 2^H-1 nodes. Then we have the following theorem.

Theorem 34 (Upper Bounds of Structural Information of Trees): Let T be a complete binary tree of depth h and thus of size $n=2^h-1$. Then the structural information of T satisfies

$$\mathcal{H}^2(T)\le \log_2\log_2 n+4+o(1). \tag{54}$$

Proof: It suffices to define a partition $\mathcal{P}$ of the nodes in T such that $\mathcal{H}^{\mathcal{P}}(T)\le \log_2\log_2 n\ +4+o(1)$. We define $\mathcal{P}$ as follows. Let $1\le k\le h$ be an integer. We partition every subtree whose root is a node of depth $h-k+1$ as a module and the remaining part consisting of all the nodes of depth at most $h-k$ as a module. Now we have 2^{h-k} complete binary subtrees, each of which, denoted by $T_j,\ j=1,2,\ldots,2^{h-k}$, has a size 2^k-1 and another complete binary subtree, denoted by T', which has a size $2^{h-k}-1$. A simple calculation indicates that for each T_j, its volume $\mathrm{vol}(T_j)=2^{k+1}-3$, and the volume of T' is $\mathrm{vol}(T')=3\cdot\ 2^{h-k}-4$.

For each T_j, we have

$$\begin{aligned}
&-\sum_{v\in T_j}\frac{d_v}{2m}\log_2\frac{d_v}{\mathrm{vol}(T_j)}\\
&\quad=-(2^{k-1}-1)\cdot\frac{3}{2m}\log_2\frac{3}{2^{k+1}-3}\\
&\qquad-2^{k-1}\cdot\frac{1}{2m}\log_2\frac{1}{2^{k+1}-3}\\
&\quad\le\frac{1}{2m}\Big[\ (2^{k-1}-1)\cdot\ 3(k+1)+2^{k-1}(k+1)\Big]\\
&\quad\le\frac{2^{k+1}}{2m}(k+1).
\end{aligned}$$

So

$$\begin{aligned}
&-\sum_{j=1}^{2^{h-k}}\frac{\mathrm{vol}(T_j)}{2m}\sum_{v\in T_j}\frac{d_v}{\mathrm{vol}(T_j)}\log_2\frac{d_v}{\mathrm{vol}(T_j)}\\
&\quad=-\sum_{j=1}^{2^{h-k}}\sum_{v\in T_j}\frac{d_v}{\mathrm{vol}(T_j)}\log_2\frac{d_v}{\mathrm{vol}(T_j)}\\
&\quad\le\ 2^{h-k}\cdot\frac{2^{k+1}}{2m}(k+1).
\end{aligned}$$

Note that each T_j has exactly one global edge connecting to T'. So the number of global edges for each T_j is $g_j=1$. We have

$$\begin{aligned}
&-\sum_{j=1}^{2^{h-k}}\frac{g_j}{2m}\log_2\frac{\mathrm{vol}(T_j)}{2m}\\
&\quad=-2^{h-k}\cdot\frac{1}{2m}\log_2\frac{2^{k+1}-3}{2m}\\
&\quad=\frac{2^{h-k}}{2m}\cdot\left[\log_2\ 2m-(k+1)+O\left(\frac{1}{2^k}\right)\right].
\end{aligned}$$

Then consider the subtree T'. Note that all the nodes in T' except for the root of T which has degree 2, have degree 3. So

$$\begin{aligned}
&-\sum_{v\in T'}\frac{d_v}{2m}\log_2\frac{d_v}{\mathrm{vol}(T')}\\
&\quad=-(2^{h-k}-2)\cdot\frac{3}{2m}\log_2\frac{3}{3\cdot\ 2^{h-k}-4}\\
&\qquad-\frac{2}{2m}\log_2\frac{2}{3\cdot\ 2^{h-k}-4}\\
&\quad\le\frac{2^{h-k}}{2m}\cdot\ 3(h-k).
\end{aligned}$$

Note that T' has $g_{T'}=2^{h-k}$ global edges, each of which joins a subtree T_j. We have

$$\begin{aligned}
&-\frac{g_{T'}}{2m}\log_2\frac{\mathrm{vol}(T')}{2m}\\
&\quad=-\frac{2^{h-k}}{2m}\log_2\frac{3\cdot\ 2^{h-k}-4}{2m}\\
&\quad=\frac{2^{h-k}}{2m}\cdot\left[\log_2\ 2m-(h-k)+O\left(\frac{1}{2^{h-k}}\right)\right].
\end{aligned}$$

So in all, noting that $\log_2\ 2m=\log_2(2^{h+1}-4)\le h+1$, the structural information of T by partition $\mathcal{P}$ is

$$\begin{aligned}
\mathcal{H}^{\mathcal{P}}(T)&=-\sum_{j=1}^{2^{h-k}}\frac{\mathrm{vol}(T_j)}{2m}\sum_{v\in T_j}\frac{d_v}{\mathrm{vol}(T_j)}\log_2\frac{d_v}{\mathrm{vol}(T_j)}\\
&\quad-\sum_{j=1}^{2^{h-k}}\frac{g_j}{2m}\log_2\frac{\mathrm{vol}(T_j)}{2m}\\
&\quad-\sum_{v\in T'}\frac{d_v}{2m}\log_2\frac{d_v}{\mathrm{vol}(T')}-\frac{g_{T'}}{2m}\log_2\frac{\mathrm{vol}(T')}{2m}\\
&\le\frac{2^{h-k}}{2m}\cdot\ 2^{k+1}(k+1)\\
&\quad+\frac{2^{h-k}}{2m}\cdot\left[\log_2 2m-(k+1)+O\left(\frac{1}{2^k}\right)\right]\\
&\quad+\frac{2^{h-k}}{2m}\cdot\ 3(h-k)\\
&\quad+\frac{2^{h-k}}{2m}\cdot\left[\log_2\ 2m-(h-k)+O\left(\frac{1}{2^{h-k}}\right)\right]\\
&\le\frac{2^{h-k}}{2^{h+1}-4}\cdot\Big[(2^{k+1}+1)(k+1)+4(h-k)\\
&\quad+\ O\left(\frac{1}{2^k}\right)+O\left(\frac{1}{2^{h-k}}\right)\Big]
\end{aligned}$$

$$\leq (k+1) + \frac{4(h-k)}{2^{k+1}} + O\left(\frac{k+1}{2^k} + \frac{k+1}{2^{h-k}} + \frac{h-k}{2^h}\right).$$

When we choose $k+1 = \lceil \log_2 h \rceil$, the above value is at most $\lceil \log_2 h \rceil + 4 + o(1)$, which is $\log_2 \log_2 n + 4 + o(1)$. Theorem 34 follows. □

We emphasize that the upper bounds in Theorem 34 can be established for all trees of constant bounded degrees, for which the proofs are all similar to that for the complete binary trees.

B. Grid Graphs

An $n \times n$ grid $G = (V, E)$ is a graph defined on the node set $V = \{v_{i,j} : i, j \in \mathbb{Z}^+,\ 1 \leq i, j \leq n\}$ and the edge set $E = \{(v_{i,j}, v_{i,j'}) : |j - j'| = 1\} \bigcup \{(v_{i,j}, v_{i',j}) : |i - i'| = 1\}$. For the structural information of grid graphs, we have the following theorem.

Theorem 35 (Upper Bound of Two-Dimensional Structural Information of Grid Graphs): Let $G = (V, E)$ be an $n \times n$ grid graph. Then the two-dimensional structural information of G satisfies

$$\mathcal{H}^2(G) \leq 2 \log_2 \log_2 n + O(1). \tag{55}$$

Proof: Note that the size of G is $|V| = n^2$ and the number of edges is $m = |E| = 2n(n-1)$. Similarly to the proof of Theorem 34, we find a partition $\mathcal{P}$ for the nodes in G to witness the upper bound.

We divide G into sub-grids of size $k \times k$. For notational simplicity, assume that n can be divided by k. So we have exactly $\left(\frac{n}{k}\right)^2$ such sub-grids. For each sub-grid, denoted by G_j, let $d_i^{(j)}$ denote the degree of the i-th node, which is 4 for most nodes, 3 for border nodes and 2 for corner nodes of G. By the extremum property of the entropy function $H(\cdot)$, the positioning entropy within G_j satisfies

$$H\left(\frac{d_1^{(j)}}{\mathrm{vol}(G_j)}, \cdots, \frac{d_{k^2}^{(j)}}{\mathrm{vol}(G_j)}\right) \leq \log_2 k^2 = 2 \log_2 k.$$

So

$$\sum_j \frac{\mathrm{vol}(G_j)}{2m} \cdot H\left(\frac{d_1^{(j)}}{\mathrm{vol}(G_j)}, \cdots, \frac{d_{k^2}^{(j)}}{\mathrm{vol}(G_j)}\right) \leq 2 \log_2 k.$$

Since the total number of global edges is

$$\sum_j g_j = 2n\left(\frac{n}{k} - 1\right),$$

and noting that $m = 2n(n-1)$, we have

$$-\sum_j \frac{g_j}{2m} \log_2 \frac{\mathrm{vol}(G_j)}{2m} \leq \left(\sum_j g_j\right) \cdot \frac{1}{2m} \log_2 2m \leq \frac{n-k}{2k(n-1)} \cdot (2 \log_2 n + 2) \leq \frac{\log_2 n + 1}{k}.$$

So in all, we have that the structural information of G by partition $\mathcal{P}$ is

$$\mathcal{H}^{\mathcal{P}}(G) = \sum_j \frac{\mathrm{vol}(G_j)}{2m} \cdot H\left(\frac{d_1^{(j)}}{\mathrm{vol}(G_j)}, \cdots, \frac{d_{k^2}^{(j)}}{\mathrm{vol}(G_j)}\right) - \sum_j \frac{g_j}{2m} \log_2 \frac{\mathrm{vol}(G_j)}{2m} \leq 2 \log_2 k + \frac{\log_2 n + 1}{k}.$$

Let $k = O(\log_2 n)$, then $\mathcal{H}^{\mathcal{P}}(G) \leq 2 \log_2 \log_2 n + O(1)$. Theorem 35 follows. □

Although we have only focused on the two-dimensional grid, our analysis can be generalized to k-dimensional grids for any constant k (natural number). In the k-dimensional case, for a grid of size n^k, the structural information can be shown to be at most $k \log_2 \log_2 n + O(1)$.

Finally, we have the following.

Theorem 36 (Structural Information of Classical Data Structures): If G is a tree or a grid, then, almost sure, we have

$$\mathcal{H}^2(G) = \Theta(\log_2 \log_2 n).$$

Proof: By combining Theorems 34, 35, 28 and 32. □

XIII. Two-Dimensional Structural Information of Networks

In this section, we establish the upper bounds of two-dimensional structural information of the expander graphs, and the networks of the PA model.

A. Expander Graphs

For expander graphs, we have

Theorem 37 (Expanders): Let $\{G_n\}$ be a family of expanders, each of which is either a simple graph or a graph with balanced weights on edges. Then for each $G = G_n$, we have that

$$\mathcal{H}^2(G) = \Omega(\log n). \tag{56}$$

Proof: By definition of expanders, there is a constant $0 < \alpha < 1$ such that for any n, and $G = G_n$, the conductance of G satisfies $\Phi(G) \geq \alpha$. By the normalized structural information principle (Theorem 26), the two-dimensional structural information of G, satisfies $\mathcal{H}^2(G) \geq \Phi(G) \cdot \mathcal{H}^1(G)$. By Theorem 22, $\mathcal{H}^1(G) = \Omega(\log n)$. This gives the desired result $\mathcal{H}^2(G) = \Omega(\log n)$. □

B. Preferential Attachment Model

For nontrivial networks of the PA model, we have

Theorem 38 (Scale-Free Networks): For $d \geq 2$, let G be a network of n nodes generated by the PA model with edge parameter d. Then with probability $1 - o(1)$, we have that

$$\mathcal{H}^2(G) = \Omega(\log n). \tag{57}$$

Proof: For a network, G say, generated by the PA model, it has been proved that, when $d \geq 2$, there is a large

constant $\alpha > 0$ such that almost surely, i.e., with probability $1-o(1)$, the conductance of G is larger than or equal to α, i.e., $\Phi(G) \geq \alpha$ [38]. Note that in the PA model, the number of multiple edges between any two nodes is at most d. Combing Theorems 20 and 21, the normalized structural information principle gives the desired result of the theorem. □

Notice that the theorem holds only for $d \geq 2$. When $d = 1$, the graph constructed by the PA model is a tree, for which the two-dimensional structural information is $\Theta(\log\log n)$, to be established in the next two sections.

C. Uniform Attachment Model

Li *et al.* [33] proposed the security model of network. The authors [34] have shown that for appropriately large affinity exponent, the networks generated by the security model are provably secure resisting global cascading failure for any attacks of $\log^{O(1)} n$ scales. In the same paper, the authors showed that the networks generated by the security model have an expander global core, satisfying the engineering requirement of efficient transportation and communication of information in the whole network. The global core is realised by the uniform attachment given below.

Definition 39 (Uniform Attachment Model [34]): Given a natural number d, we construct a network by stages as follows.

Stage 0*: Let G_0 be an arbitrarily given initial graph.*

Stage $t+1$*: Let G_t be the graph constructed by the end of stage t.*

(1) *Create a new node v.*
(2) *Crate d edges from v to nodes in G_t each of which is chosen randomly and uniformly among the nodes in G_t.*

For the networks of the uniform attachment model, we have

Theorem 40 (Expander Theorem of the Uniform Attachment Model): Given a natural number d, let G be a network generated by the attachment model. Then, with probability $1-o(1)$, the following properties hold:

(1) *If $d = 1$, then G is a tree with height bounded by $O(\log n)$.*
(2) *If $d > 1$, then there is a constant α (related to d) such that the conductance $\Phi(G)$ of G is at least α.*

Proof: By the *Expander Core Theorem* of networks of the security model, that is, [34, Th. 6.7]. □

Theorem 41 (Two-Dimensional Structural Information of the Networks of Uniform Attachment Model): Given d, let G be a network of the uniform attachment model. Then, with probability $1-o(1)$, the following properties hold:

(1) *If $d = 1$, then $\mathcal{H}^2(G) = \Theta(\log\log n)$.*
(2) *If $d > 1$, then $\mathcal{H}^2(G) = \Omega(\log n)$.*

Proof: By Theorems 40, 27 and 36. □

Remark: The uniform attachment model is in fact a dynamical version of random model of graphs. Theorem 41 indicates that a connected network generated randomly and uniformly cannot have low two-dimensional structural information.

D. ER Model

For the ER model, the random graph of the model [22], are disconnected, in which case, the two-dimensional structural information of the graph is the weighted summation of the two-dimensional structural information of all the connected components. However, for p is large, Li and Peng [35] have shown

Theorem 42 (Li and Peng [35]): Let $p = \frac{\ln(n\cdot\omega(n))}{n}$, where $\omega(n)$ is a function goes to infinity arbitrarily slow. Let G be the network of the ER model. Then, probability $1-o(1)$, the following properties hold:

(1) *G is connected.*
(2) *There is a constant α such that the conductance of G is at least α.*

We thus have

Theorem 43 (Two-Dimensional Structural Information of Networks of the ER Model): Let $p = \frac{\ln(n\cdot\omega(n))}{n}$, where $\omega(n)$ is a function goes to infinity arbitrarily slow. Let G be the network of the ER model. Then, probability $1-o(1)$, the following property holds:

$$\mathcal{H}^2(G) = \Omega(\log n).$$

Proof: By Theorems 21, 27 and 42. □

E. Small World Model

The upper bound result for the grids can be easily extended to perturbed grids such as the grid-like graphs generated by Kleinberg's well-known small world model [25], in which an underlying structure of the grid is endowed with mild randomness which reduces the diameter exponentially.

A network G of size $n \times n$ is constructed in the following way: Let $\widehat{G}$ be a grid of size $n \times n$, $r \geq 0$ be a constant and $d(u,v)$ denote the distance of nodes u and v on $\widehat{G}$. Each node u chooses a long-contact v independently with probability proportional to $d(u,v)^{-r}$ and edge (u,v) is added. We will show that the structural information of the small world model with parameter $r \geq 2$ is $O(\log\log n)$ with high probability (over the construction of networks).

Theorem 44 (Upper Bound of Two-Dimensional Structural Information of Networks Generated by Kleinberg's Small World Model With $r \geq 2$): Let $G = (V,E)$ be a network of size $n \times n$ constructed from the small world model with parameter $r \geq 2$. Then with probability $1 - \exp\left\{-\Omega\left(\frac{n^2}{\log n}\right)\right\}$, the two-dimensional structural information of G is $O(\log\log n)$.

Proof: Note that in the small world model, each node u chooses a long-contact v independently. We can partition V first and then consider the edges (including the long-contact edges and grid edges in $\widehat{G}$) among modules.

For a node u, all the nodes whose distance to u is at most k in $\widehat{G}$ for some integer $k \geq 1$ form a diamond of size $2k^2 + 2k + 1$ whose center is u (when u is far from the border of $\widehat{G}$). Then V can be partitioned into $\Theta(n^2/k^2)$ such diamonds with incomplete ones on the border.

By the extremum property of the entropy function $H(\cdot)$ again, the positioning entropy within each diamond, denoted by G_j, is at most $\log_2 |G_j| = O(\log k)$, and so

$$\sum_j \frac{\mathrm{vol}(G_j)}{2m} \cdot H\left(\frac{d_1^{(j)}}{\mathrm{vol}(G_j)}, \cdots, \frac{d_{|G_j|}^{(j)}}{\mathrm{vol}(G_j)}\right) = O(\log k).$$

Then we consider the number of edges among the diamonds. For a fixed G_j, let $\widehat{g}_j$ denote the number of edges in $\widehat{G}$ that have exactly one end in G_j, and let $\overline{g}_j$ denote the number of long-contact edges have exactly one end in G_j. Then the total number of edges with exactly one endpoint in G_j is $g_j = \widehat{g}_j + \overline{g}_j$. A simple calculation gives

$$\widehat{g}_j = 4(2k+1).$$

Then we turn to calculate $\overline{g}_j$.

Note that $\overline{g}_j$ is a random variable depending on the long-contacts. For each node $u \in G_j$ and $v \in \neg G_j$, the complement of G_j, let $X_{u,v} \in \{0,1\}$ be the indicator random variable to represent whether or not there is a long-contact edge between u and v. Let $X_{u,\neg G_j} = \sum_{v \in G_j} X_{u,v}$.

Note that on grid $\widehat{G}$, the number of nodes whose distance is exactly d from u is $4d$. Let

$$A = \sum_{d=1}^{n} 4d \cdot d^{-r} = 4\sum_{d=1}^{n} d^{1-r}.$$

Note that for a node $u \in G_j$ whose distance in $\widehat{G}$ to the center of G_j is i, its distance to $\neg G_j$ is $k-i$. So the expectation of $\overline{g}_j$ is

$$\begin{aligned} E(\overline{g}_j) &= \sum_{u \in G_j} E[X_{u,\neg G_j}] \\ &= \sum_{u \in G_j} \sum_{v \in \neg G_j} E[X_{u,v}] \\ &= \sum_{u \in G_j} \sum_{v \in \neg G_j} \frac{d_{u,v}^{-r}}{A} \\ &= \Theta\left(\frac{1}{A} \cdot \sum_{i=1}^{k} \sum_{d=k-i}^{n} d^{1-r}\right). \end{aligned} \tag{58}$$

For $r = 2$, $A = \Theta(\log n)$, and so

$$\begin{aligned} E(\overline{g}_j) &= \Theta\left(\frac{1}{A} \cdot \sum_{i=1}^{k} (\log n - \log(k-i))\right) \\ &= \Theta\left(\frac{1}{A} \cdot (k \log n - k \log k)\right) \\ &= \Theta\left(\frac{k}{\log n} \log \frac{n}{k}\right). \end{aligned}$$

So the total number of long-contact edges is expected to be

$$\begin{aligned} E\left(\sum_j \overline{g}_j\right) &= \Theta\left(\left(\frac{k}{\log n} \log \frac{n}{k}\right) \cdot \left(\frac{n}{k}\right)^2\right) \\ &= \Theta\left(\frac{n^2}{k}\left(1 - \frac{\log k}{\log n}\right)\right). \end{aligned}$$

We will use the following form of Chernoff bound.

Lemma 45 (Chernoff Bound [14]): Let $X_1, \ldots, X_n$ be independent random variables with $\Pr[X_i = 1] = p_i$ and $\Pr[X_i = 0] = 1 - p_i$. Denote the sum by $X = \sum_{i=1}^{n} X_i$ with expectation $E(X) = \sum_{i=1}^{n} p_i$. Then we have

$$\Pr[X \le E(X) - \lambda] \le \exp\left(-\frac{\lambda^2}{2E(X)}\right),$$

$$\Pr[X \ge E(X) + \lambda] \le \exp\left(-\frac{\lambda^2}{2(E(X) + \lambda/3)}\right).$$

Let $k = \Theta(\log n)$. Since the long-contact edges are created independently, by the Chernoff bound, when we choose $\lambda = E(\sum_j \overline{g}_j)/2$, we know that with probability $1 - \exp\{-\Omega(n^2/\log n)\}$, the total number of long-contact edges is at most $O(n^2/\log n)$.

Thus, with probability $1 - \exp\{-\Omega(n^2/\log n)\}$, the total number of edges among the diamonds is

$$\begin{aligned} \sum_j g_j &= \sum_j (\widehat{g}_j + \overline{g}_j) \\ &= 4(2k+1) \cdot O\left(\frac{n^2}{k^2}\right) + O\left(\frac{n^2}{k}\right) \\ &= O\left(\frac{n^2}{\log n}\right). \end{aligned}$$

Therefore, with probability $1 - \exp\left\{-\Omega\left(\frac{n^2}{\log n}\right)\right\}$, the total number of edges

$$m = 2n(n-1) + \frac{1}{2}\sum_j g_j = \Theta(n^2),$$

and

$$-\sum_j \frac{g_j}{2m} \log_2 \frac{\mathrm{vol}(G_j)}{2m} \le \left(\sum_j g_j\right) \cdot \frac{1}{2m} \log_2 2m = O(1).$$

The structural information of G by partition $\mathcal{P}$ is

$$\begin{aligned} \mathcal{H}^{\mathcal{P}}(G) &= \sum_j \frac{\mathrm{vol}(G_j)}{2m} \cdot H\left(\frac{d_1^{(j)}}{\mathrm{vol}(G_j)}, \cdots, \frac{d_{k^2}^{(j)}}{\mathrm{vol}(G_j)}\right) \\ &\quad - \sum_j \frac{g_j}{2m} \log_2 \frac{\mathrm{vol}(G_j)}{2m} \\ &= O(\log k) + O(1) \\ &= O(\log \log n). \end{aligned}$$

For $r > 2$, $A = O(1)$, and by Equation (58),

$$E(\overline{g}_j) = \begin{cases} O(k^{3-r}) & \text{if } 2 < r < 3, \\ O(\log k) & \text{if } r = 3, \\ O(1) & \text{if } r > 3. \end{cases}$$

When $k = \Theta(\log n)$, we can always show that with probability $1 - \exp\left\{-\Omega\left(\frac{n^2}{\log n}\right)\right\}$, the total number of edges among the diamonds is $O(n^2/\log n)$. The same result can be obtained by a similar proof.

This completes the proof of Theorem 52. □

Theorems 34, 35 and 52 demonstrate that the 2-dimensional structural information of the graphs for classical data structures is $O(\log \log n)$. We will show that this upper bound is actually the lower bound.

F. Phase Transition Theorem of Structural Information of Networks of Kleinberg's Small World Model

In this section, we establish the full characterizations of the two dimensional structural information of networks of the small world model introduced by Kleinberg.

Theorem 46 (Phase Transition Theorem of Two-Dimensional Structural Information of Networks of the Small World Model): Let G be a network generated from the small world model with parameter $r \geq 0$. Then the two-dimensional structural information has a sharp phase transition at the point $r = 2$. That is,

(1) *if $r \geq 2$, then with probability $1 - o(1)$, $\mathcal{H}^2(G) = O(\log\log n)$;*
(2) *if $r < 2$, then with probability $1 - o(1)$, $\mathcal{H}^2(G) = \Omega(\log n)$.*

Proof: The first item is just Theorem 52. We only have to prove (2).

When $r < 2$, Flaxman has proved that the edge expansion, which is the minimum ratio between the outgoing edges of a subset of nodes and its volume, is a constant with probability $1-o(1)$ [24]. That is, for every subset S of nodes, the number of edges between S and its complement $\neg S$, denoted by $e(S, \neg S)$, is at least $\delta|S|$ for some constant $\delta > 0$. Note that the edges whose both ends are in S consist of two parts: those from the underlying grid and those contributed by long-contacts. The former ones is at most $2|S|$ and the latter ones is at most $|S|$. So $\mathrm{vol}(S) \leq 6|S| + e(S, \neg S)$. Thus the conductance of S satisfies

$$\Phi(S) = \frac{e(S,\neg S)}{\mathrm{vol}(S)} \geq \frac{e(S,\neg S)}{6|S| + e(S,\neg S)} \geq \frac{\delta}{6+\delta},$$

which is a constant. By Theorem 40, $\mathcal{H}^2(G) = \Omega(\log n)$.

This completes the proof of Theorem 46. □

By Theorems 36, 46, 38, 41 and 43, we observe an interesting phenomenon that for networks of either classical data structure, or generated by classical models, the two-dimensional structure entropies of the graphs are either $O(\log\log n)$ or $\Omega(\log n)$, where n is the number of nodes of the graphs. This poses an interesting question: Does the nature really have only two types of networks with two-dimensional structural information either $O(\log\log n)$, meaning at most $c \cdot \log\log n$ for some constant c, or $\Omega(\log n)$, meaning at least $d \cdot \log n$ for some constant d? We will answer this question by introducing a new model of networks, the homophyly/kinship model.

XIV. Two-Dimensional Structural Information of Nature Evolving

A. Homophyly/Kinship Model

Real world networks consist of natural communities which play an essential role in the networks. To understand the intrinsic mechanism of the natural communities of the networks that are naturally evolving, the authors and colleagues [31] proposed the homophyly/kinship model of networks to realise the networks naturally evolving in the real world. The model is established by paralleling the Darwin's evolution theory [18]. According to Darwin's theory, animals from ants to people from social groups in which most individuals work for the common good, which is a type of fitness of species in the evolution. For the social groups of animals in the evolution, kinship plays the key role. By paralleling Darwin's theory, we introduce the notion of *affinity exponent* to capture the idea of kinship in Darwin's evolution theory, and propose the *homophyly/kinship model* by following the ideas of Darwin's individual fitness proposal and natural selection.

The model proceeds as follows.

Definition 47 (Homophyly/Kinship Model [31]): Given affinity exponent $a \geq 0$ and natural number d:

1) *Let G_d be an initial d-regular graph in which each node is associated with a distinct colour and is called a seed. (The initial graph could be an arbitrarily given graph, which does not change the results of the model.)*
   For $i > d$, let G_{i-1} be the graph constructed at the end of step $i-1$, and let $p_i = \frac{1}{(\log i)^a}$.
2) *At step i, we create a new node v.*
3) *(Preferential attachment) With probability p_i, v chooses a new colour, in which case*
   a) *we call v a seed, and*
   b) *create d edges from v to nodes in G_{i-1}, chosen with probability proportional to the degrees in G_{i-1}.*
4) *(Homophyly/kinship) Otherwise, then v chooses an old colour, in which case*
   a) *v chooses randomly and uniformly an old colour as its own colour. and*
   b) *creates d edges from v to nodes of the same colour in G_{i-1}, chosen with a probability proportional to the degrees in G_{i-1}.*

It has been shown [31] that the networks generated by the homophyly/kinship model truthfully reflects many real networks, including both biological networks and social networks. We have also demonstrated that the homophyly/kinship model provides a methodology for community analyses of networks [31].

B. Fundamental Properties

We call the networks generated by our model *homophyly networks*. In this section, we establish some basic properties of the homophyly networks, which will be frequently used throughout the paper.

The probabilistic tools and inequalities useful in the proofs are given in Appendix A.

We use $\mathcal{H}(n, a, d)$ to denote the set of networks of n nodes, constructed from the homophyly model with homophyly exponent a and average number of edges d. Let $G = (V, E) \in \mathcal{H}(n, a, d)$ be a network of n nodes generated from our homophyly model. We use G_t to denote the graph obtained at the end of time step t of the construction of G, and C_t to denote the set of seed nodes of G_t. Recall that every node $v \in V$ is associated with a color. The vertices in V is partitioned naturally as the homochromatic sets, each of which is a connected set of nodes of the same color. Every homochromatic set contains a seed node, which is the first

node of the community. For an edge $e = (u, v)$, we call e a *local edge*, if the two endpoints u, v share the same color, and a *global edge*, otherwise.

In this subsection, we will prove some basic properties about the number of seed nodes, the sizes, volumes and the numbers of global edges of homochromatic sets.

Theorem 48 (Basic properties): Given $a \geq 0$, and $d \geq 2$, let $G = (V, E)$ be a graph of n nodes generated from our homophyly model. Let $T_1 = \log^{a+1} n$ and $T_2 = \frac{n}{\log^b n}$ for some positive constant b. Then the following properties hold:

(1) *With probability $1-o(1)$, for all $t \geq T_1$, $\frac{t}{2\log^a t} \leq |C_t| \leq \frac{2t}{\log^a t}$.*
(2) *For each homochromatic set S, if $t > t_S \geq T_1$, then the expectation of its size at time step t is $\Theta(\log^{a+1} t - \log^{a+1} t_S)$, where t_S is the time step at which the seed node of S is created.*
(3) *With probability $1-o(1)$, every homochromatic set in G has a size upper bounded by $O(\log^{a+1} n)$.*
(4) *For each homochromatic set S, if $t_S \geq T_2$, then the number of global edges in G associated to S, denoted by g_S, satisfies that, for sufficiently large n,*
   (i) *if $a > 1$, then $E(g_S) \leq \frac{5}{2}(a+1)db^2(\log\log n)^2$;*
   (ii) *if $a = 1$, then $E(g_S) \leq 8db^2(\log\log n)^2$;*
   (iii) *if $0 < a < 1$, then $E(g_S) \leq 5db^2(\log\log n)^2$.*

Proof: The proof of Theorem 48 is given in Appendix B. □

C. Power Law and Holographic Law

In this section, we prove that the degrees of the networks generated by the homophyly model follows a *power law*, and a *holographic law*. Let G be such a network. We say that a homochromatic set of G is a *natural community* or *community* of G. We will show that there exists a constant β such that almost surely, the degrees of a natural community, the degrees of the induced subgraph of a natural community, and the degrees of the whole network G, all follow a power law with power exponent the same constant β.

The networks of the homophyly/kinship model follow the following new phenomenon:

Definition 49 (Holographic Law): Given a network $G = (V, E)$, we say that G follows a holographic law, if there exists a constant β, and a partition $\mathcal{P}$ of the vertices of G satisfying both (i) and (ii) below:

(i) *G follows the power law with power exponent β, and*
(ii) *almost all communities of $\mathcal{P}$ follow the power law with power exponent β.*

By Definition 49, the power exponent of the power law of G is the same as the power exponent of the power law of an community of G.

Theorem 50 (Power Law and Holographic Law): For $a \geq 0$ and $d \geq 2$, let $G = (V, E)$ is a network constructed from $\mathcal{H}(n, a, d)$. Then we have:

(1) *For $a = 0$, the degree of nodes in G follows a power law distribution.*
(2) *For $a > 0$, there exists a constant β satisfying:*
   (i) *The degrees of the induced subgraph of almost all homochromatic sets follow a power law with power exponent β.*
   (ii) *The degrees of nodes of almost all homochromatic sets follow a power law with power exponent β.*
   (iii) *(Power law) Degrees of nodes in G follow a power law with power exponent β.*

By Theorem 48 (3), a natural community is small. By Theorem 50, a natural community is interpretable by common attributes of the nodes, the common color here, and almost all the communities have degree patterns similar to that of the whole graph, showing a type of *self-similarity* of networks. (2) (i), (ii) and (2) (iii) of Theorem 50 explore a *holographic law* of the homophyly networks, that the power exponent of the whole network is contained in a natural community of the network.

Proof: For (1). If $a = 0$, then the homophyly model degenerates to the classic PA model, and the power law degree distribution with $\beta = 3$ can be obtained by the canonical proof of the PA model [5].

For (2). We prove the following two items together, which proves (2) (i) and (2) (ii), respectively.

(A) For almost every homochromatic set X, the degree distribution of the induced subgraph G_X follows a power law, and
(B) For almost every homochromatic set X, the degrees of nodes in X follow a power law.

(2) (iii) will follow immediately from (B) by observing that the union of several power law distributions with the same power exponent is also a power law distribution.

The proofs of (A) and (B) are similar to the canonical proof of the PA model. Additionally, we will verify that the contribution of degrees of a community from global edges is negligible compared with those from its local edges. So the construction of a homochromatic set basically follows the classic preferential attachment scheme, and the number of global edges created by other seed nodes is negligible. We will realize this idea gradually in the proofs below.

Let $T_3 = (1 - \delta_1)n$, where $\delta_1 = \frac{1}{\log^{a/2} n}$. Let X denote a homochromatic set of a fixed color and t_X be the time step at which X is created. Suppose that $T_2 \leq t_X \leq T_3$. By Theorem 48, since there are $\Theta(n/\log^a n)$ homochromatic sets in G_n and each of them has size $O(\log^{a+1} n)$ with extremely high probability, it is easy to show that if $b > 2a+1$ for $T_2 = n/\log^b n$, then almost all homochromatic sets are created in the time interval (T_2, T_3) with probability $1 - o(1)$. So to prove (A) and (B), we only analyze the homochromatic sets created in this time interval.

For positive integers s and k, define $A_{s,k}$ to be the number of nodes of degree k in X when $|X|$ reaches s, $B_{s,k}$ the number of nodes of degree k in the induced subgraph of X when $|X|$ reaches s, and $g_{s,k}$ the number of global edges associated with the nodes in X of degree k in the induced subgraph of X when $|X|$ reaches s. Obviously, we have $A_{1,d} = 1$, $A_{1,k} = 0$ for all $k > d$, and $B_{1,k} = 0$ for all k. Then we establish the recurrence formula for the expectation of both $A_{s,k}$ and $B_{s,k}$.

Define $T(s)$ to be the time step at which the size of X becomes to be s, and s' to be the number of global edges

connecting X in the case that $|X| = s$ (note that probably at several consecutive time steps, $|X|$ keeps s). We consider the time interval $(T(s-1), T(s))$.

By Theorem 48, since $T_1 \le T_2 \le t_X \le T_3$, the size of X at time step n is expected to be $E(|X|) = \Theta(\log^{a+1} n - \log^{a+1} t_S) = \Omega\left(\log^a n \cdot \log \frac{1}{1-\delta_1}\right) = \Omega(\log^{a/2} n)$, and the number of global edges is expected to be $O(\log\log n)^2 = o(E(|X|))$. Thus, in the time interval $(T(s-1), T(s))$, the number of global edges that link to a node in X is $o(s^{-1})$.

For $s > 1$ and $k > d$, we have

$$E(A_{s,k}) = A_{s-1,k} \cdot \left[1 - \frac{kd}{2d(s-1)+s'} - o\left(\frac{1}{s}\right)\right] + A_{s-1,k-1} \cdot \left[\frac{(k-1)d}{2d(s-1)+s'} + o\left(\frac{1}{s}\right)\right] + O\left(\frac{1}{s^2}\right),$$

where the error terms caused by the case that more than one edge joins to a single node are absorbed in the term $O(1/s^2)$. Taking expectations on both sides, we have

$$E(A_{s,k}) = E(A_{s-1,k}) \cdot \left[1 - \frac{k}{2(s-1)+s'/d} - o\left(\frac{1}{s}\right)\right] + E(A_{s-1,k-1}) \cdot \left[\frac{(k-1)}{2(s-1)+s'/d} + o\left(\frac{1}{s}\right)\right] + O\left(\frac{1}{s^2}\right). \quad (59)$$

When $k = d$,

$$E(A_{s,d}) = E(A_{s-1,d}) \cdot \left[1 - \frac{d}{2(s-1)+s'/d} + o\left(\frac{1}{s}\right)\right] + 1 + O\left(\frac{1}{s^2}\right). \quad (60)$$

Similarly, for $s > 1$ and $k > d$,

$$E(B_{s,k}) = B_{s-1,k} - \frac{d \cdot (kB_{s-1,k} + g_{s-1,k})}{2d(s-1)+s'} + \frac{d \cdot [(k-1)B_{s-1,k-1} + g_{s-1,k-1}]}{2d(s-1)+s'} + O\left(\frac{1}{s^2}\right).$$

Taking expectations on both sides, we have

$$E(B_{s,k}) = E(B_{s-1,k}) \cdot \left[1 - \frac{kd}{2d(s-1)+s'}\right] + E(B_{s-1,k-1}) \cdot \frac{(k-1)d}{2d(s-1)+s'} + \frac{E(g_{s-1,k-1} - g_{s-1,k})}{2d(s-1)+s'} + O\left(\frac{1}{s^2}\right). \quad (61)$$

When $k = d$,

$$E(B_{s,d}) = B_{s-1,d} - \frac{d \cdot (dB_{s-1,d} + g_{s-1,d})}{2d(s-1)+s'} + 1 + O\left(\frac{1}{s^2}\right) = B_{s-1,d} \cdot \left[1 - \frac{d}{2(s-1)+s'/d}\right] + \left[1 - \frac{g_{s-1,d}}{2(s-1)+s'/d}\right] + O\left(\frac{1}{s^2}\right),$$

and

$$E(B_{s,d}) = E(B_{s-1,d}) \cdot \left[1 - \frac{d}{2(s-1)+s'/d}\right] + \left[1 - \frac{E(g_{s-1,d})}{2(s-1)+s'/d}\right] + O\left(\frac{1}{s^2}\right). \quad (62)$$

To solve these recurrences, we introduce the following lemma that is used in the canonical proof of the preferential attachment model.

Lemma 51 ([16], Lemma 3.1): Suppose that a sequence $\{a_s\}$ satisfies the recurrence relation

$$a_{s+1} = \left(1 - \frac{b_s}{s+s_1}\right) a_s + c_s \text{ for } s \ge s_0,$$

where the sequences $\{b_s\}, \{c_s\}$ satisfy $\lim_{s\to\infty} b_s = b > 0$ and $\lim_{s\to\infty} c_s = c$ respectively. Then the limit of $\frac{a_s}{s}$ exists and

$$\lim_{s\to\infty} \frac{a_s}{s} = \frac{c}{1+b}.$$

Note that since $|X|$ goes to infinity as $n \to \infty$. By Theorem 48, the terms s'/d in equalities (59) and (60) are comparatively negligible. Since $g_{s,k}$, for each k, is expected to be $o(E(|X|))$, the terms $\frac{E(g_{s-1,k-1}-g_{s-1,k})}{2d(s-1)+s'}$ and $\frac{E(g_{s-1,d})}{2d(s-1)+s'}$ in equalities (61) and (62) are also comparatively negligible. By Lemma 51, $\frac{E(A_{s,k})}{s}$ and $\frac{E(B_{s,k})}{s}$ must have the same limit as t goes to infinity. Thus we will only give the proof of the power law distribution for $E(A_{s,k})$, which also holds for $E(B_{s,k})$.

Denote by $S_k = \lim_{s\to\infty} \frac{E(A_{s,k})}{s}$ for $k \ge d$. In the case of $k = d$, we apply Lemma 51 with $b_s = d/2 + o(1)$, $c_s = 1 + O(1/s^2)$, $s_1 = -1$, and get

$$S_d = \lim_{s\to\infty} \frac{E(A_{s,d})}{s} = \frac{1}{1+\frac{d}{2}} = \frac{2}{2+d}.$$

For $k > d$, assume that we already have $S_{k-1} = \lim_{s\to\infty} \frac{E(A_{s,k-1})}{s}$. Applying Lemma 51 again with $b_s = k/2 + o(1)$, $c_s = \frac{E(A_{s-1,k-1})}{s-1} \cdot \frac{k-1}{2} + O(1/s^2)$, $s_1 = -1$, we get

$$S_k = \lim_{s\to\infty} \frac{E(A_{s,k})}{s} = \frac{S_{k-1} \cdot \frac{k-1}{2}}{1+\frac{k}{2}} = S_{k-1} \cdot \frac{k-1}{k+2}.$$

Thus recurrently, we have

$$S_k = S_d \cdot \frac{(d+2)!(k-1)!}{(d-1)!(k+2)!} = \frac{2d(d+1)}{k(k+1)(k+2)}. \quad (63)$$

This implies that for sufficiently large s, $E(A_{s,k}) = f(d)(1 + o(1)) \cdot k^{-3}s$ for some function f only depending on d. Since $s = \omega(1)$ goes to infinity as $n \to \infty$, we have $E(A_{s,k}) \propto k^{-3}$. For the same reason, $E(B_{s,k}) \propto k^{-3}$. This proves (A) and (B), and also completes the proof of (2) (i) and (2) (ii).

For (2) (iii), a key observation is that the union of several power law distributions is also a power law distribution if the power exponents are equal. We will give the same explicit expression of the expectation of the number of degree k nodes by combining those for the homochromatic sets, leading to a similar power law distribution.

To prove the power law degree distribution of the whole graph, we take the union of distributions of all homochromatic sets. Suppose that G has m homochromatic sets that are created in time interval (T_2, T_3). For $i = 1, \ldots, m$, let M_i

be the size of the i-th homochromatic set and $N_{s,k}^{(i)}$ denote the number of nodes of degree k when the i-th set has size s. For each i, we have

$$\lim_{n\to\infty} \frac{E(N_{M_i,k}^{(i)})}{M_i} = S_k.$$

Hence,

$$\lim_{n\to\infty} \frac{E(\sum_{i=1}^{m} N_{M_i,k}^{(i)})}{\sum_{i=1}^{m} M_i} = S_k.$$

Let M_0 denote the size of the union of all other homochromatic sets and $N_{s,k}^{(0)}$ denote the number of nodes of degree k in this union when it has size s. By Theorem 48, we know that $E(M_0) = o(n)$. Since the color of each node is chosen independently, by the Chernoff bound, $M_0 = o(n)$ holds with probability $1 - o(1)$. Define $N_{t,k}$ to be the number of nodes of degree k in G_t. Then we have

$$\lim_{n\to\infty} \frac{E(N_{n,k})}{n} = \lim_{n\to\infty} \frac{E(\sum_{i=0}^{m} N_{M_i,k}^{(i)})}{\sum_{i=0}^{m} M_i}.$$

For M_0, we have that

$$\lim_{n\to\infty} \frac{M_0}{\sum_{i=1}^{m} M_i} = \lim_{n\to\infty} \frac{M_0}{n - M_0} = 0$$

and

$$\lim_{n\to\infty} \frac{E(N_{M_0,k}^{(0)})}{n} \le \lim_{n\to\infty} \frac{M_0}{n} = 0$$

hold with probability $1 - o(1)$. So

$$\lim_{n\to\infty} \frac{E(N_{n,k})}{n} = \lim_{n\to\infty} \frac{E(\sum_{i=1}^{m} N_{M_i,k}^{(i)})}{\sum_{i=1}^{m} M_i} = S_k.$$

Recall that $S_k = \frac{2d(d+1)}{k(k+1)(k+2)}$. This implies that for sufficiently large n, $E(N_{n,k}) = f(d)(1+o(1)) \cdot k^{-3} n$ for some function f only depending on d, and so $E(N_{n,k}) \propto k^{-3}$. (2) (iii) follows.

This completes the proof of Theorem 50. □

D. Small Diameter Property

In this section, we show that the diameters of the networks generated from the homophyly model are small.

Theorem 52 (Small Diameter Property): For $a \ge 0$, let $G = (V, E)$ be a graph generated from $\mathcal{H}(n, a, d)$. Then the following properties hold:

1. *For $a = 0$, with probability $1 - o(1)$, the diameter of G is $O(\log n)$.*
2. *For $a > 0$, with probability $1 - o(1)$, the diameter of G is $O(\log^{a+2} n)$.*

We remark that the small world phenomenon consists of two properties, the first is the small diameter property, and the second is the clustering effect. Our theorem shows the first. For the second property, it is easy to see that if we add some local rules in the creation of local edges, i.e., the edges between nodes of the same color, then the clustering effect will be significantly amplified. It is important that this kind of modifications will never change all the other properties of the graphs. The idea of introducing local rules in creation of local edges here is interesting. It has been a hard open problem to combine the two properties of the small world phenomenon in modeling, as commented by Chung and Lu in [16]. To our knowledge, the only known theoretical results are on grid-like graphs by Kleinberg [25], leading to local algorithms for finding short paths. It would be an interesting problem to better fit our model by introducing some local rules to generate graphs with both small diameter property and the clustering effect and to allow local algorithms for communications in the graphs.

Now we turn to prove Theorem 52.

Proof: For (1). When $a = 0$, the homophyly model is the classic PA model, and the diameter has been proved to be $O(\log n)$ with probability $1 - o(1)$ by Bollobás and Riordan [9].

For (2). Let $a > 0$ and G be a graph in $\mathcal{H}(n, a, d)$. We consider the hierarchical structure of G. Let G' be the graph obtained from G by merging all non-seed nodes of each homochromatic set to the corresponding seed node. Hence the nodes of G' are the seed nodes of G, and the edges of G' are the global edges of G. Let G'' be the graph obtained from G by deleting all global edges. In so doing, G'' consists of the isolated homochromatic sets.

Given two nodes u and v in G, suppose that u_0 and v_0 are the seed nodes of colors as the same as that of u and v respectively. By definition, u_0 and v_0 are nodes in G'. We choose a path $P_{u,v}$ connecting two nodes u, v in G as follows. We consider two cases.

Case 1: u and v share the same color.

Let X be the homochromatic set of u. In this case, we define $P_{u,v}$ to be the shortest path between u and v in the induced subgraph of the natural community X.

Case 2: Otherwise. Then

Suppose that $u_0, u_1, \cdots, u_l = v_0$ is a shortest path in G', where each u_i is a seed node of G. For each i, let X_i be the set of nodes sharing the same color as that of u_i.

Let P_{u,u_0} and $P_{v_0,v}$ be the shortest paths between u and u_0 in X_0, and between v_0 and v in X_l, respectively. For every i with $0 \le i < l$, let $P_{u_i,u_{i+1}}$ be the path consisting of a shortest path in X_i, a shortest path in X_{i+1} and an edge between X_i and X_{i+1}. We define $P_{u,v}$ to be the path consisting the paths P_{u,u_0}, P_{u_0,u_1}, $\cdots$, P_{u_{l-1},u_l}, $p_{u_l,v}$. By definition, there are l edges in G' are used in the path $P_{u,v}$ and $P_{u,v}$ consists alternately of paths composed by local and global edges..

To estimate the number l, we recall a known result on random recursive trees. A random recursive tree is constructed by stages. At each stage, a new vertex is created which links to an earlier node randomly. If each node is picked uniformly, then we call it a uniform recursive tree. We use a result of Pittel in [43], saying that the height of a uniform recursive tree of size n is $O(\log n)$ with high probability.

Lemma 53 (Recursive Tree Lemma [43]): With probability $1 - o(1)$, the height of a uniform recursive tree of size n is asymptotic to $e \log n$, where e is the natural logarithm.

Consider G' as a union of d recursive trees. Notice that the earlier created homochromatic sets in G have larger expected volumes than that of the communities created later. So with higher probability than the uniform recursive tree, the height

of a recursive tree in G' is at most $e \log |C_n|$, where $|C_n|$ is the number of colors in G and is also the number of nodes in G'. This means that with probability $1 - o(1)$, the diameter of G' is an upper bounded by $2e \log |C_n| = O(\log n)$. Therefore the number l in the definition of $P_{u,v}$ is at most $O(\log n)$.

On the other hand, by Theorem 48, with probability $1 - o(1)$, every homochromatic set has a size upper bounded by $O(\log^{a+1} n)$. Hence, for every i, the length of $P_{u_i,u_{i+1}}$ is at most $O(\log^{a+1} n)$.

Therefore, with probability $1 - o(1)$, the diameter of G is upper bounded by $O(\log^{a+1} n) \cdot O(\log n) = O(\log^{a+2} n)$. (2) follows.

This completes the proof of Theorem 52. □

Since the diameter of a PA network is expected to be logarithmic of its size, the diameter of each homochromatic set is expected to be $O(\log \log n)$. This implies that the expected diameter of a homophyly network is $O(\log n \ \log \log n)$.

However, for concentration probability, our result in Theorem 52 is only a rough upper bound $O(\log^{a+2} n)$ for the diameter. It is an interesting open question to prove or disprove the concentration result of diameter $O(\log n \log \log n)$ for the homophyly networks.

Generally speaking, there are interesting problems left open by our model. The first is to modify the model to generate networks with both small diameter property and clustering effect simultaneously. The second is to develop a theoretical approach to algorithmic small world phenomenon, for which it would be very interesting to design algorithms to find short paths on networks of small diameters in time complexity polynomial of the diameters.

E. Structure Conductance, and Conductance Minimization Principle

1) Structure Conductance of Networks: Given a graph $G = (V, E)$, and a subset S of V, the conductance of S in G is given by

$$\Phi(S) = \frac{|E(S, \bar{S})|}{\min\{\text{vol}(S), \text{vol}(\bar{S})\}}, \tag{64}$$

where $E(S, \bar{S})$ is the set of edges with one endpoint in S and the other in the complement of S, i.e. $\bar{S}$, and $\text{vol}(X)$ is the sum of degrees d_x for all $x \in X$. The conductance of G is defined to be the minimum of $\Phi(S)$ over all subsets S's, that is,

$$\Phi(G) = \min_{S \subset V} \{\Phi(S)\}. \tag{65}$$

Then we extend the definition of conductance of a set to a class of sets.

Definition 54 (Structure Conductance by a Partition): Let G be a graph, and $\mathcal{P} = \{X_1, X_2, \cdots, X_L\}$ be a partition of vertices of G such that $vol(X_i) \leq vol(G)/2 = m$ for each i. Then we define the structure conductance of G by $\mathcal{P}$ as follows.

$$\theta^{\mathcal{P}}(G) = \frac{1}{n} \sum_{j=1}^{L} |X_j| \cdot \Phi(X_j). \tag{66}$$

$\theta^{\mathcal{P}}(G)$ is in fact the weighted average of conductance for each module of partition $\mathcal{P}$.

Definition 55 (Structure Conductance): Given a graph G, we define the structure conductance of G by

$$\theta(G) = \min_{\mathcal{P}} \{\theta^{\mathcal{P}}(G)\}. \tag{67}$$

The motivations of the structure conductance of a network are:

- Intuitively speaking, if a set $X \subset V$ has a small conductance, then the internal links of X are strong, and the external links of X are weak. This exactly captures the common convention of quality communities.
- By definition, if $\theta(G)$ is small, then there is a well-defined community structure $\mathcal{P}$ in G. It is clear that a good community structure indicated by low structure conductance requires a mass of nodes belonging to high-quality communities.

2) Structure Conductance Principle of Nature Evolving: In this subsection, we investigate the structure conductance of the networks generated by the homophyly model with different values of the homophyly exponent a.

We will establish a phase transition theorem of structure conductance and the structure conductance minimization principle of the existence of well-defined structures of power law networks. Precisely, we have:

Theorem 56 (Phase Transition of Structure Conductance, and Conductance Minimization Principle): For $a \geq 0$, $d \geq 2$, let $G = (V, E)$ be a graph generated from $\mathcal{H}(n, a, d)$. We have

1. *If $a = 0$, then there exists a positive constant α depending only on d such that, with probability $1 - o(1)$, $\theta(G) \geq \alpha$.*
2. *If $0 < a \leq 1$, then for almost every homochromatic set S, the expected conductance of S is $O\left(\frac{(\log \log n)^2}{\log^{a/2} n}\right)$, and $E(\theta(G)) = o(1)$.*
3. *If $a > 1$, then for $\beta = \frac{a-1}{4(a+1)}$, with probability $1 - o(1)$, for almost every homochromatic set S, the conductance of S, $\Phi(S) = O\left(\frac{1}{|S|^{\beta}}\right)$. Moreover, with probability $1 - o(1)$, $\theta(G) = o(1)$.*

Theorem 56 implies that the structure conductance of networks generated by the homophyly model decreases as a increases. Furthermore, the structure conductance of the networks decreases in three interesting ways. The results for $a = 0$ and $a > 1$ hold almost surely, and go to the opposite extremes. However for $0 < a \leq 1$, we are just able to prove the expectation result that, the expected structure conductance of network G is $o(1)$. Although the result also implies that even if $0 < a \leq 1$, with high probability, the structure conductance of the networks generated by the homophyly model is as small as $o(1)$, it could be possible that, with a non-negligible probability, the structure conductances of some networks of the model are larger than some constant $\alpha > 0$. It is interesting to prove or disprove the concentration result of the structure conductance for the case of $0 < a \leq 1$. Here we leave it as an open question.

Nevertheless the results in Theorem 56 demonstrate that structure conductance minimization is a principle of the existence of community structures of power law networks.

Proof of Theorem 56: For (1). If $a = 0$, then the homophyly model degenerates to the PA model. It has been shown that, with probability $1-o(1)$, networks generated by the PA model are expanders, whose structure conductances has been proved to be at least a constant α depending on d ([38, Th. 1]). Since for any partition $\mathcal{P}$ of V, $\theta^{\mathcal{P}}(G)$ is the weighted average over all modules of $\mathcal{P}$ and $\theta^{\mathcal{P}}(G) \geq \alpha$. This implies that $\theta(G) \geq \alpha$.

For (2). Recall that in Section XIV-C, we have shown that if $T_2 \leq t_S \leq T_3$, where $T_2 = \frac{n}{\log^b n}$, $b > 2a+1$, $T_3 = (1-\delta_1)n$, $\delta_1 = \frac{1}{\log^{a/2} n}$, then with probability $1 - o(1)$, almost all the homochromatic sets are created in the time interval (T_2, T_3). So we only have to show that every such homochromatic set, denoted by S, has expected conductance $O\left(\frac{(\log\log n)^2}{\log^{a/2} n}\right)$.

Let $0 < a \leq 1$. By Theorem 48, if $t_S \geq T_2$, then the expected number of global edges $E(g_S) = O(\log\log n)^2$. On the other hand, if $t_S \leq T_3$, then the size of S is expected to be $E(|S|) = \Theta(\log^{a+1} n - \log^{a+1} t_S) = \Omega\left(\log^a n \cdot \log\frac{1}{1-\delta_1}\right) = \Omega(\log^{a/2} n)$, and the expected volume of S is also $\Omega(\log^{a/2} n)$. By the Chernoff bound, the results above hold with probability $1 - o(1)$. Hence, $E(\Phi(S)) = O\left(\frac{(\log\log n)^2}{\log^{a/2} n}\right)$.

To show that $E(\theta(G)) = o(1)$, we only have to show that in the case of a specific partition $\mathcal{P}_1$ in G, $\theta^{\mathcal{P}_1}(G) = o(1)$. We define $\mathcal{P}_1$ by colors such that each homochromatic set S satisfying $T_2 \leq t_S \leq T_3$ is a module of $\mathcal{P}_1$ and the rest nodes form the last module of $\mathcal{P}_1$. Since $b > 2a+1$ and each homochromatic set has an expected size $O(\log^{a+1} n)$, the size of the last module of $\mathcal{P}_1$ is expected to be at most $O(\log^{a+1} n) \cdot T_2 + \delta_1 n = o(n)$. Thus

$$\begin{aligned} E(\theta(G)) &\leq E(\theta^{\mathcal{P}_1}(G)) \\ &= O\left(\frac{(\log\log n)^2}{\log^{a/2} n}\right) \cdot (1 - o(1)) + o(1) = o(1). \end{aligned}$$

For (3). Assume $a > 1$. Let $T_4 = (1-\delta_2)n$, where $\delta_2 = \frac{1}{\log^{(a-1)/2}} n$. By a similar argument to that for (2) above, we have that for $b = a+2$, almost all the homochromatic sets are created in the time interval (T_2, T_4). We will show that, with probability $1 - o(1)$, every homochromatic set S with $T_2 \leq t_S \leq T_4$ has conductance $O\left(\frac{1}{|S|^\beta}\right)$.

By the proof of Theorem 48, we can establish an upper bound for the volume of S. We will show a concentration result for the event that the volume of S is $O(\log^{a+1} n)$.

Recall that $D(S)[t]$ is the volume of S at time step t. We have:

Lemma 57 (Degree of Communities Lemma): With probability $1 - o(1)$, for any homochromatic set S created at time $t_S \geq T_2$, $D(S)[n] = O(\log^{a+1} n)$ holds.

Proof: We only have to show that for a fixed S with $t_S \geq T_2$, $D(S)[n] = O(\log^{a+1} n)$ holds with probability $1 - o(n^{-1})$. Then the lemma follows from the union bound. We assume the worst case that S is created at time step $t_S = T_2$. In the proof of Theorem 48, we know that the recurrence on $D(S)[t]$ can be written as

$$\begin{aligned} &E(D(S)[t] \mid D(S)[t-1], \mathcal{E}) \\ &\quad = D(S)[t-1] + \frac{1}{\log^a t} \cdot \frac{D(S)[t-1]}{2d(t-1)} \cdot d \\ &\qquad + \left(1 - \frac{1}{\log^a t}\right) \cdot \frac{2d}{|C_{t-1}|}, \end{aligned}$$

and thus

$$\begin{aligned} &E(D(S)[t] \mid D(S)[t-1], \mathcal{E}) \\ &\quad \leq D(S)[t-1]\left[1 + \frac{1}{2(t-1)\log^a t}\right] + \frac{4d\log^a t}{t}. \end{aligned}$$

These are the Inequalities (77) and (78), respectively. Applying Inequality (80) to (78), we have

$$\begin{aligned} &E(D(S)[t] \mid D(S)[t-1], \mathcal{E}) - 9d\log^{a+1}(t+1) \\ &\quad \leq \left[1 + \frac{1}{2(t-1)\log^a t}\right] \cdot (D(S)[t-1] - 9d\log^{a+1} t). \end{aligned}$$

Recall that

$$\theta_t = \prod_{i=t_S+1}^{t} \left[1 + \frac{1}{2(i-1)\log^a i}\right].$$

Define $X[t] = \frac{D(S)[t] - 9d\log^{a+1}(t+1)}{\theta_t}$. Then

$$E[X[t] \mid X[t-1], \mathcal{E}] \leq X[t-1].$$

Note that

$$\begin{aligned} &X[t] - E[X[t] \mid X[t-1], \mathcal{E}] \\ &\quad = \frac{D(S)[t] - E[D(S)[t] \mid D(S)[t-1], \mathcal{E}]}{\theta_t} \leq 2d. \end{aligned}$$

Since

$$D(S)[t] - D(S)[t-1] \leq 2d,$$

we have, for sufficiently large n (and so sufficiently large t),

$$\begin{aligned} &\mathrm{Var}[X[t] \mid X[t-1], \mathcal{E}] \\ &\quad = E[(X[t] - E(X[t]|X[t-1], \mathcal{E}))^2] \\ &\quad = \frac{1}{\theta_t^2} E[(D(S)[t] - E(D(S)[t] \mid D(S)[t-1], \mathcal{E}))^2] \\ &\quad \leq \frac{1}{\theta_t^2} E[(D(S)[t] - D(S)[t-1])^2 | D(S)[t-1], \mathcal{E}] \\ &\quad \leq \frac{2d}{\theta_t^2} E[D(S)[t] - D(S)[t-1] \mid D(S)[t-1], \mathcal{E}] \\ &\quad \leq \frac{2d}{\theta_t^2}\left[\frac{4d\log^a t}{t} + \frac{D(S)[t-1]}{2(t-1)\log^a t}\right] \\ &\quad = \frac{8d^2\log^a t}{t\theta_t^2} + \frac{d}{(t-1)\theta_t\log^a t} \cdot \frac{D(S)[t-1]}{\theta_t} \\ &\quad \leq \frac{8d^2\log^a t}{t\theta_t^2} + \frac{9d^2\log^{a+1} t}{(t-1)\theta_t^2\log^a t} + \frac{dX[t-1]}{(t-1)\theta_t\log^a t} \\ &\quad \leq \frac{10d^2\log^a t}{t\theta_t^2} + \frac{dX[t-1]}{(t-1)\theta_t\log^a t}. \end{aligned}$$

Since θ_t can be bounded as

$$\theta_t \sim \exp\left\{\sum_{i=T_2+1}^{t}\frac{1}{2(i-1)\log^a i}\right\} \in \left[\left(\frac{t}{T_2}\right)^{\frac{1}{2\log^a n}}, \left(\frac{t}{T_2}\right)^{\frac{1}{2\log^a T_2}}\right],$$

we have

$$\begin{aligned}&\sum_{i=T_2+1}^{t}\frac{10d^2\log^a i}{i\theta_i^2}\\ &\le\ 10d^2\log^a n\int_{T_2}^{t}\frac{1}{x}\cdot\left(\frac{T_2}{x}\right)^{\frac{1}{\log^a n}}dx\\ &\le\ 10d^2\log^a n\ \cdot\log n = 10d^2\log^{a+1} n,\end{aligned}$$

and

$$\begin{aligned}&\sum_{i=T_2+1}^{t}\frac{1}{(i-1)\theta_i\log^a i}\\ &\le\frac{2}{\log^a T_2}\int_{T_2}^{t}\frac{T_2^{\frac{1}{2\log n}}}{x\cdot x^{\frac{1}{2\log n}}}dx\\ &\le\frac{2\log n}{\log^a T_2}.\end{aligned}$$

Here we can assume that $X[t]$ is non-negative, which means that $D(S)[t]\ge\ 10\log^{a+1}(t+1)$, because otherwise, $D(S)[n]$ will be smaller and the event in the lemma holds with higher probability. Let $\lambda = 20\log^{a+1} n$. By Lemma 71,

$$\begin{aligned}&\Pr[X[t]=\omega(\log^{a+1} n)]\le\Pr[X[t]\ge X[T_2]+\lambda]\\ &\le\exp\left\{-\frac{\lambda^2}{2(10d^2\log^{a+1} n\ +\ (2\log n/\log^a T_2)\lambda+d\lambda/3)}\right\}\\ &\quad+O(n^{-2})=O(n^{-2}).\end{aligned}$$

This implies that $D(S)[n] = O(\log^{a+1} n)$ holds with probability $1-O(n^{-2})$. □

Then we consider the number of global edges associated with S with $T_2\le t_S\le T_4$. Suppose the event, denoted by $\mathcal{F}$, that for any $t\ge t_S$, $D(S)[t] = O(\log^{a+1} n)$, which holds with probability $1-o(n^{-1})$ by the proof Lemma 57. For each $t\ge t_S$, we define a random variable X_t to be the number of global edges that connect S at time t. We have

$$E(X_t|\mathcal{F}) = d\cdot\frac{1}{\log^a t}\cdot\frac{D(S)[t-1]}{2d(t-1)}\le\frac{\log^{1+\epsilon} n}{2(t-1)},$$

for arbitrarily small positive ϵ. Then

$$\begin{aligned}&E\left(\sum_{t=t_S}^{n}X_t|\mathcal{F}\right)\\ &\le\left(\log^{1+\epsilon} n\right)\cdot\sum_{t=t_S}^{n}\frac{1}{2(t-1)}\\ &\le a(\log^{1+\epsilon} n)(\log\log n).\end{aligned}$$

By the Chernoff bound (conditioned on the event $\mathcal{F}$), for sufficiently large n,

$$\Pr\left[\sum_{t=t_S}^{n}X_t\ge\ 2a(\log^{1+\epsilon} n)(\log\log n)|\mathcal{F}\right]\le n^{-2}.$$

That is, conditioned on the event $\mathcal{F}$, with probability at least $1-n^{-2}$, the total number of global edges joining S is upper bounded by $2a(\log^{1+\epsilon} n)(\log\log n)$. Since $\mathcal{F}$ happens with probability $1-o(n^{-1})$, we have

$$\Pr\left[\sum_{t=t_S}^{n}X_t\ge\ 2a(\log^{1+\epsilon} n)(\log\log n)\right]=o(n^{-1}).$$

On the other hand, it is easy to show that when $a>1$, with probability $1-o(1)$, every such S (satisfying $t_S\in[T_2,T_4]$) has a size $\Omega(\log^{\frac{a+1}{2}} n)$, and so a volume $\Omega(\log^{\frac{a+1}{2}} n)$. Let $0<\epsilon<\frac{a-1}{4}$. Then, with probability $1-o(1)$, for each such S,

$$\begin{aligned}\Phi(S)\ &=\ O\left(\frac{2a(\log^{1+\epsilon} n)(\log\log n)}{\log^{(a+1)/2} n}\right)\\ &\le\ O\left(\log^{-\frac{a-1}{4}} n\right)\\ &\le\ O\left(|S|^{-\frac{a-1}{4(a+1)}}\right).\end{aligned}$$

To show that $\theta(G)=o(1)$ holds with probability $1-o(1)$, as we have done in the proof of (2), we define a partition $\mathcal{P}_2$ and show that $\theta^{\mathcal{P}_2}(G)=o(1)$ with high probability. We define $\mathcal{P}_2$ such that each homochromatic set S satisfying $T_2\le t_S\le T_4$ is a module of $\mathcal{P}_2$ and the rest nodes form the last module of $\mathcal{P}_2$. In fact, by Lemma 57, with probability $1-o(1)$, the total number of nodes belonging to the homochromatic sets which appear before time T_2 or after T_4, that is the size of the last module of $\mathcal{P}_2$, is at most $O(\log^{a+1} n)\cdot\frac{n}{\log^{a+2} n}+\frac{n}{\log^{(a-1)/2} n}=o(n)$ for constant $a>1$. Therefore, $1-o(1)$ fraction of nodes of G belongs to a homochromatic set S of conductance bounded by $O\left(|S|^{-\frac{a-1}{4(a+1)}}\right)=o(1)$. Thus with probability $1-o(1)$,

$$\theta(G)\le\theta^{\mathcal{P}_2}(G)=o(1)(1-o(1))+o(1)=o(1).$$

This completes the proof of Theorem 56. □

The results and proofs of Theorem 56 imply that if $a=0$, then almost surely the networks of the homophyly model simply fail to have conductance-based community structure, that if $0<a\le 1$, then the networks of the homophyly model may have a conductance-based community structure, but the conductance-based community structures are non-robust, and that if $a>1$, then almost surely, the networks of the homophyly model have a well-defined conductance-based community structure.

It is interesting to notice that Theorem 56 (1) and (3) hold with probability $1-o(1)$, but Theorem 56 (2) is only an expectation result. This leads to some interesting open questions: Whether or not the result in Theorem 56 (2) can be strengthened to a concentration result? Is structure conductance a robust measure for characterizing the structures or complexity of networks?

According to Theorem 56, structure conductance of the networks of the homophyly/kinship model has three types corresponding to $a = 0$, $0 < a \leq 1$ and $a > 1$, respectively.

F. Modularity Principle of Nature Evolving

In this subsection, we will investigate modularity of networks. Firstly, we establish an upper bound of modularity for general graphs. Secondly, we establish a phase transition phenomenon of modularity of power law networks, and the modularity maximization principle of the existence of well-defined structures of power law networks.

In this subsection, we investigate the power law networks generated by our homophyly model. Precisely, for the homophyly networks, we have the following theorem.

Theorem 58 (Modularity Maximization Principle): Let $G = (V, E)$ be a graph of n nodes generated from the homophyly model with homophyly exponent $a \geq 0$, and average number of edges d. Then we have

(1) *If $a = 0$, then there exists a large constant α such that*

$$\Pr[\sigma(G) \leq 1 - \alpha] = 1 - o(1).$$

(2) *If $a > 0$, then*

$$\Pr[\sigma(G) = 1 - o(1)] = 1 - o(1).$$

By the definition of modularity, we can intuitively interpret the modularity of a graph G, $\sigma(G)$, as the distance between G and a random copy of G. By this reason, for a graph G, if $\sigma(G)$ is large, then G is far from its own random copy, and if $\sigma(G)$ is small, then G is close to its own random copy. In addition, we always assume that a random copy of a graph fails to have a well-defined structure.

Therefore, the principle explored in Theorem 58 implies that if $a = 0$, then the modularity of G, i.e., $\sigma(G)$, is almost surely small, so that G is close to its own random copy, or simply, we say that G is random, and that if $a > 0$, then the modularity of the networks of the homophyly model are almost surely far from their own random copies, meaning that the networks have well-defined structures.

It is remarkable that the results in both cases hold with certainty, i.e., almost surely, or with probability $1 - o(1)$. This feature is different from that of structure conductance in Theorem 56, showing some differences between modularity and structure conductance. This suggests that modularity maximization is a principle behind the existence of community structures of networks, and that modularity is a robust measure for characterizing the community structures of networks.

Now we turn to prove the theorem.

Proof of Theorem 58: For (1). In this case, the homophyly model is equivalent to the PA model, therefore, with probability $1-o(1)$, the conductance of G, $\Phi(G)$ is a constant $\alpha > 0$. Then the result follows from Theorem 24.

For (2). By definition, $\sigma(G)$ is the maximum $\sigma^{\mathcal{P}}(G)$ over all partitions $\mathcal{P}$'s. So it suffices to give a specific partition $\mathcal{P}$ such that $E(\sigma^{\mathcal{P}}(G)) = 1 - o(1)$.

Let $\mathcal{P}$ be the partition of G defined by colors such that each homochromatic set is a module of $\mathcal{P}$. Suppose that $\mathcal{P} = \{X_1, X_2, \cdots, X_L\}$. For each $l \in \{1, 2, \cdots, L\}$, let k_l be the number of edges whose endpoints are both in X_l, V_l be the volume of X_l, and g_l be the number of edges from X_l to nodes outside of X_l.

By Theorem 48, we have that with probability $1 - o(1)$, there are $L = |C_n| \leq \frac{2n}{\log^a n}$ homochromatic sets in G, and that with probability $1 - o(1)$, the number of global edges, the edges between different modules, is $d \cdot |C_n| \leq \frac{2dn}{\log^a n} = o(n)$. Thus

$$\Pr\left[m - \sum_{l=1}^{L} k_l = d \cdot |C_n| = o(n)\right] = 1 - o(1).$$

By the construction of G, $m = dn$, we thus have

$$\Pr\left[\sum_{l=1}^{L} k_l = (1 - o(1))m\right] = 1 - o(1) \tag{68}$$

By Theorem 48 again, we have that with probability $1 - o(1)$, each homochromatic set has a size upper bounded by $O(\log^{a+1} n)$.

Let $V_{\max} = \max\{V_1, \ldots, V_L\}$. Since the total number of global edges is $o(n)$, we have that with probability $1 - o(1)$, the following holds

$$\Pr[V_{\max} = o(m)] = 1 - o(1).$$

Thus

$$\Pr\left[\sum_{l=1}^{L} V_l^2 \leq V_{\max} \cdot \sum_{l=1}^{L} V_l = o(m^2)\right] = 1 - o(1) \tag{69}$$

Combining Inequality (68) and (69), we have that

$$\sigma(G) \geq \sigma^{\mathcal{P}}(G) = \frac{1}{m}\sum_{l=1}^{L} k_l - \frac{1}{4m^2}\sum_{l=1}^{L} V_l^2 = 1 - o(1)$$

holds with probability $1 - o(1)$.

This completes the proof of Theorem 58. □

Theorem 58 demonstrates that 0 is a threshold of a for modularity of the homophyly networks, which is either almost surely low in the case of $a = 0$, or almost surely as high as $1 - o(1)$ for the case of $a > 0$, and that the modularity results for both $a = 0$ and $a > 0$ are concentration results, which hold with probability $1 - o(1)$.

According to Theorem 58, modularity fails to distinct the networks of the homophyly/kinship model for different affinity exponents $a > 0$.

G. Structural Information Principle of Nature Evolving

In this section, we will investigate the positioning entropy, structural information and normalized structural information of networks. Our contributions are as follows:

- To establish a lower bound of the positioning entropy for general power law networks
- To establish an equation between positioning entropy and structural information of a general graph through conductances of the partition
- To establish a lower bound of normalized structural information for general graphs

- To establish a lower bound of structural information for general graphs
- To establish a structural information phase transition phenomenon and structural information minimization principle for the power law networks generated by our homophyly model.

At first, we propose an additive law of entropy function, and establish a difference principle of positioning entropy and structural information, a normalized structural information principle, and a structural information principle for general graphs.

1) Phase Transition of Structural Information, and Structural Information Minimization Principle of Homophyly Networks: For the homophyly networks, we have the following theorems.

Theorem 59 (Phase Transition, and Structural Information Minimization Principle): Given $a \geq 0$, and $d \geq 2$, let $G = (V, E)$ be a graph generated from $\mathcal{H}(n, a, d)$ and $\mathcal{N}$ be the partition of G given by the natural communities of G. Then with probability $1 - o(1)$, the two-dimensional structural information of G, satisfies the following:

(1) *If $a = 0$, then $\mathcal{H}^2(G) = \Theta(\log n)$.*
(2) *If $0 < a < 1$, then $\mathcal{H}^{\mathcal{N}}(G) = \Theta(\log^{1-a} n)$*[1]
(3) *If $a \geq 1$, then $\mathcal{H}^2(G) = \Theta(\log \log n)$.*

Theorem 59 implies that the structural information of the networks of the homophyly model with $a = 0$ are at least $c \cdot \log n$ for some constant $c > 0$, that if $0 < a < 1$, then structural information of the networks of the homophyly model is at most $c \cdot \log^{1-a} n$ for some constant $c > 0$, and that if $a \geq 1$, then the structural information of the networks of the homophyly model is at most $c \cdot \log \log n$ for some constant c. The results demonstrate that structural information of the networks of the homophyly model dramatically decreases as a increases, and that each of the results holds almost certainly. The theorem thus implies that structural information is a robust and well-defined measure to characterize the structures and complexity of complex networks.

Theorem 59 also indicates two thresholds 0 and 1 for the homophyly exponent a. (1) and (2) of the theorem show that if $a = 0$, then the structural information of the homophyly graphs is extremely high, and that if $a > 0$, then the structural information of the homophyly networks is close to 0. (2) and (3) show that if $a \geq 1$, then the structural information of the networks of the homophyly model is exponentially smaller than that of the networks with $a < 1$.

Theorem 59 demonstrates that nature evolving may have structure entropies in three classes: (i) at least $c \log n$ for some constant c, (ii) equal to $c \log^{1-a} n$ for some constant c and some affinity exponent $a < 1$, and (iii) equal to $c \log \log n$ for some constant c.

Theorem 60 (Phase Transition and Normalized Structural Information Minimization Principle): Given $a \geq 0$, and $d \geq 2$, let $G = (V, E)$ be a graph generated from $\mathcal{H}(n, a, d)$. We have the following

(1) *If $a = 0$, then*

$$\Pr[\tau(G) \geq \alpha] = 1 - o(1),$$

where α is the constant defined in Theorem 56.

(2) *If $0 < a < 1$, then*

$$\Pr\left[\tau(G) = O\left(\frac{1}{\log^a n}\right)\right] = 1 - o(1).$$

(3) *If $a \geq 1$, then*

$$\Pr\left[\tau(G) = O\left(\frac{\log \log n}{\log n}\right)\right] = 1 - o(1).$$

Proof of Theorem 60: By combining Theorems 27 and 59. □

Theorem 60 exactly reflects the results in Theorem 59 in a normalized form.

By Theorems 27, 59 and 60, we have that the structural information and normalized structural information of networks of the homophyly model decrease as a increases, for which 0 and 1 are the thresholds for dramatic decrease of the entropies, and that structural information minimization and normalized structural information minimization are the principles of the existence of well-defined structures of networks, and the principles of power law networks from high positioning entropy to low positioning entropy.

Now we are ready to prove Theorem 59.

Proof of Theorem 59: Let G be a homophyly network. Let $\mathcal{N}$ be the partition of G given by the natural communities of G.

For (1). By Theorem 27, for any partition $\mathcal{P}$ of G, $\mathcal{H}^{\mathcal{P}}(G) \geq \Phi(G) \cdot \mathcal{H}^1(G)$. Since $a = 0$, G is actually generated by the PA model, so that with probability $1 - o(1)$, the conductance $\Phi(G)$ of G is a constant $\alpha > 0$. By Theorem IX-B, with probability $1 - o(1)$, the positioning entropy $\mathcal{H}^1(G)$ of G is at least $c \log n$ for some constant $c > 0$. (1) follows.

For (2) and (3). For upper bounds proofs, it suffices to give a partition of the graph G with the desired properties. Let $\mathcal{N}$ be the natural partition of G given by the homochromatic sets of G. Then we build an upper bound for $\mathcal{H}^{\mathcal{N}}(G)$.

By Theorem 25, we have that for a partition $\mathcal{P}$ of V,

$$\mathcal{H}^{\mathcal{N}}(G) = -\sum_{j=1}^{L} \frac{V_j}{2m} \sum_{i=1}^{n_j} \frac{d_i^{(j)}}{V_j} \log_2 \frac{d_i^{(j)}}{V_j} - \sum_{j=1}^{L} \frac{g_j}{2m} \log_2 \frac{V_j}{2m}.$$

Set the first term of $\mathcal{H}^{\mathcal{N}}(G)$ by $H_1 = -\sum_{j=1}^{L} \frac{V_j}{2m} \sum_{i=1}^{n_j} \frac{d_i^{(j)}}{V_j} \log_2 \frac{d_i^{(j)}}{V_j}$, and for each homochromatic set X_j, set $L_j = -\sum_{i=1}^{n_j} \frac{d_i^{(j)}}{V_j} \cdot \log_2 \frac{d_i^{(j)}}{V_j}$. By Theorem 48, with probability $1 - o(1)$, for each j, $n_j = O(\log^{a+1} n)$. Since uniform distribution gives rise to the maximum entropy, we have that with probability $1 - o(1)$,

$$L_j \leq \log_2 n_j = O(\log \log n),$$

and by averaging, we have

$$H_1 = \sum_{j=1}^{L} \frac{V_j}{2m} \cdot L_j = O(\log \log n).$$

[1]It is an interesting open question to characterise exactly for $0 < a < 1$, the two-dimensional structural information of the networks of the homophyly/kinshio model.

Moreover, recall that by Lemma 57, with probability $1 - o(1)$, almost all nodes belongs to some homochromatic set of volume $O(\log^{a+1} n)$ (those created after T_2). We have

$$-\sum_{j=1}^{L} \frac{g_j}{2m} \log_2 \frac{V_j}{2m} = \Theta\left(\sum_{j=1}^{L} \frac{g_j}{m} \log_2 m\right) = \Theta\left(\frac{\log_2 m}{m} \cdot \sum_{j=1}^{L} g_j\right).$$

Let m_g be the number of global edges in G. Then $\sum_{j=1}^{L} g_j = 2m_g$.

By the construction of G, $m_g = d|C_n|$, where $|C_n|$ is the number of colors in G (and also the number of homochromatic sets in G and the number of modules in $\mathcal{P}$). By Theorem 48, with probability $1 - o(1)$, the size $|C_n|$ of C_n is at most $2n/\log^a n$ and at least $n/2\log^a n$. Therefore, noting that $m = dn$, the second term of $\mathcal{H}^{\mathcal{N}}(G)$ satisfies

$$-\sum_{j=1}^{L} \frac{g_j}{2m} \log_2 \frac{V_j}{2m} = \Theta\left(\frac{\log_2 m}{m} \cdot \frac{n}{\log^a n}\right) = \Theta(\log^{1-a} n).$$

Putting all together, we have that, with probability $1 - o(1)$,

$$\mathcal{H}^{\mathcal{N}}(G) = H_1 - \sum_{j=1}^{L} \frac{g_j}{2m} \log_2 \frac{V_j}{2m} = O(\log\log n) + \Theta(\log^{1-a} n).$$

If $0 < a < 1$, then with probability $1 - o(1)$,

$$\mathcal{H}^{\mathcal{N}}(G) = \Theta(\log^{1-a} n).$$

If $a \geq 1$, then with probability $1 - o(1)$,

$$\mathcal{H}^{\mathcal{N}}(G) = O(\log\log n).$$

By Theorem 32, this gives rise to

$$\mathcal{H}^2(G) = \Theta(\log\log n).$$

Both (2) and (3) follow.

This completes the proof of Theorem 59.

According to Theorem 59, the two-dimensional structural information function of the networks of the networks of the homophyly/kinship model are classified as the following types:

1) $\Theta(\log n)$;
2) $\Theta(\log^{1-a} n)$ for a with $0 < a < 1$; and
3) $\Theta(\log\log n)$.

Therefore, the two-dimensional structural information characterises the networks naturally evolving in nature and society by three types with varying affinity exponent a and hidden constant in the Θ of the complexity functions above.

So far, we considered the structural information for balanced graphs. However, what happens for highly unbalanced graphs? We will see that there will be a black hole in such networks.

XV. Black Hole Principle

A well-known feature of the Shannon entropy function $H(\mathbf{p})$ is that, loosely speaking, the more the distribution $\mathbf{p}$ looks like a uniform one, the larger $H(\mathbf{p})$ is. A skewed distribution often exhibits a low entropy. In the proofs of our results on lower bounds, the conditions of "simple graphs" and "graphs with balanced weights" guarantee that the degrees of nodes and volumes of any small subset do not diverge too much, and so the lower bounds can be proved. The positioning entropy and structural information do not depend on the absolute values of degrees, but on the relative ones, i.e., $d_v/\mathrm{vol}(G)$ and $\rho_j = \mathrm{vol}(V_j)/\mathrm{vol}(G)$. Conversely, if the degrees of nodes in a complex graph with a large amount of multi-edges or unbalanced weights skew the degree distribution severely, then both positioning entropy and structural information may diminish dramatically.

Theorem 61 (Black Hole Theorem - Necessity): Let $G = (V, E)$ be a connected weighted graph of size $n = |V|$ and weight function $w : E \to \mathbb{R}^+$.

(1) *If there is a subset $S \subseteq V$ of size s and volume $vol(S) = \rho \cdot vol(G)$ for some $0 < \rho \leq 1$, then both positioning entropy $\mathcal{H}^1(G)$ and structural information $\mathcal{H}^2(G)$ of G are at most*

$$H(1-\rho, \rho) + (1-\rho)\log_2(n-s) + \rho \log_2 s.$$

(2) *If $s = \log^{o(1)} n$ and $\rho \geq 1 - \frac{1}{\log n}$, then*

$$y\mathcal{H}^2(G) \leq \mathcal{H}^1(G) = o(\log\log n).$$

Proof: For (1). The proof is quite straightforward from the additivity of positioning entropy given by Lemma 29. Partition V into two subsets: $V = S \cup \overline{S}$. Noting that the positioning entropy in each subset is at most the logarithm base 2 of its size, we have

$$\mathcal{H}^1(G) \leq \frac{\mathrm{vol}(S)}{\mathrm{vol}(G)} \log_2 |S| + \frac{\mathrm{vol}(\overline{S})}{\mathrm{vol}(G)} \log_2 |\overline{S}| + H(1-\rho, \rho) = H(1-\rho, \rho) + (1-\rho)\log_2(n-s) + \rho \log_2 s.$$

For (2). This follows from (1) and Proposition 23.

Then Theorem 61 follows. □

The result follows from the fact that the small subset S absorbs almost all probability for a random walk on G. Once a random walk goes into S, it is hard to escape. This phenomenon is quite similar to the black holes in astronomy. We also call *a black hole of a network*, which means that a highly dense subset makes the positioning entropy and structural information break the lower bounds $\Omega(\log n)$ and $\Omega(\log\log n)$, respectively, for simple and balanced weight graphs.

Therefore, a black hole in a network makes the entropies decreasing dramatically.

Theorem 62 (Black Hole Theorem - Sufficiency): Let $G = (V, E)$ be a connected graph of size $n = |V|$ and volume $vol(G)$. If $\mathcal{H}^2(G) = o(\log\log n)$, then we have the following conclusions.

(1) *If $\mathcal{H}^1(G) = o(\log n)$, then there is a subset $S \subseteq V$ in G whose size is $n^{o(1)}$ and whose volume is $(1 - o(1)) \cdot vol(G)$.*

(2) *Otherwise, there is a subset $S \subseteq V$ in G whose volume is $vol(S) \geq \rho \cdot vol(G)$ for some constant $0 < \rho < 1$, and each node in S belongs to a subset of size $\log^{o(1)} n$ and conductance $o(1)$ (understood as a black hole, that is, S is composed by black holes). For the complement $\overline{S}$ of S, either its volume is $o(vol(G))$, in which case, the complement of S consists of only "tiny dusts" and it is trivial, or there is a subset $U \subseteq \overline{S}$ with size $|U| = n^{o(1)}$, volume $vol(U) = (1 - o(1)) \cdot vol(\overline{S})$ and conductance $\Phi(U) = o(1)$, in which case, U corresponds to a black hole.*

Proof: First, we consider the case that $\mathcal{H}^1(G) = o(\log n)$. That is, for any constant $\varepsilon > 0$ and sufficiently large n, $\mathcal{H}^1(G) \leq \varepsilon^2 \log_2 n$. Let

$$S = \{v \in V : \frac{d_v}{\mathrm{vol}(G)} \geq \frac{1}{n^{\varepsilon}}\},$$

and $\overline{S} = V \setminus I$. So we have that

$$\sum_{v \in \overline{S}} \frac{d_v}{\mathrm{vol}(G)} \cdot (\varepsilon \log_2 n) \leq \mathcal{H}^1(G) \leq \varepsilon^2 \log_2 \ n.$$

Thus,

$$\sum_{v \in \overline{S}} \frac{d_v}{\mathrm{vol}(G)} \leq \varepsilon,$$

and

$$\sum_{v \in S} \frac{d_v}{\mathrm{vol}(G)} \geq \ 1 - \varepsilon.$$

On the other hand, the size of S satisfies

$$|S| \leq \frac{\mathrm{vol}(G)}{\min\limits_{v \in S}\{d_v\}} \leq n^{\varepsilon}.$$

Because of the arbitrariness of ε, (1) has been proved.

Then we consider the case that $\mathcal{H}^2(G) = o(\log \log n)$ but $\mathcal{H}^1(G) = \Omega(\log n)$. That is, for any constant $\varepsilon > 0$ and sufficiently large n, $\mathcal{H}^2(G) \leq \varepsilon^3 \log_2 \log_2 n$, and meanwhile, there exists a constant $\varepsilon_0 > 0$ such that for sufficiently large n, $\mathcal{H}^1(G) \geq \varepsilon_0 \log_2 n$.

Let $\mathcal{P}$ be a partition of nodes in G such that $\mathcal{H}^{\mathcal{P}}(G) \leq \varepsilon^3 \log_2 \log_2 n$. Define

$$J = \{j \in \mathcal{P} : H_j \leq \varepsilon^2 \log_2 \log_2 n\},$$

and $\overline{J} = \mathcal{P} \setminus J$, where $H_j = -\sum\limits_{v \in V_j} \frac{d_v}{\mathrm{vol}(V_j)} \log_2 \frac{d_v}{\mathrm{vol}(V_j)}$ and V_j is the j-th module of $\mathcal{P}$. Since

$$\sum_{j \in \mathcal{P}} \frac{\mathrm{vol}(V_j)}{\mathrm{vol}(G)} \cdot H_j \leq \mathcal{H}^{\mathcal{P}}(G) \leq \varepsilon^3 \log_2 \log_2 n,$$

we have

$$\begin{aligned}\sum_{j \in \overline{J}} \frac{\mathrm{vol}(V_j)}{\mathrm{vol}(G)} \cdot \varepsilon^2 \log_2 \log_2 n \ &\leq \sum_{j \in \overline{J}} \frac{\mathrm{vol}(V_j)}{\mathrm{vol}(G)} \cdot H_j \\ &\leq \varepsilon^3 \log_2 \log_2 \ n.\end{aligned}$$

So

$$\sum_{j \in \overline{J}} \frac{\mathrm{vol}(V_j)}{\mathrm{vol}(G)} \leq \varepsilon,$$

and

$$\sum_{j \in J} \frac{\mathrm{vol}(V_j)}{\mathrm{vol}(G)} \geq 1 - \varepsilon,$$

which means that almost all volume of G is contributed by the modules in J.

Define $\mathrm{vol}(J) = \sum_{j \in J} \mathrm{vol}(V_j)$. By the additivity of entropy function $H(\cdot)$,

$$\begin{aligned}\mathcal{H}^1(G) = H\left(\frac{\mathrm{vol}(J)}{\mathrm{vol}(G)}, \frac{\mathrm{vol}(\overline{J})}{\mathrm{vol}(G)}\right) &+ \frac{\mathrm{vol}(J)}{\mathrm{vol}(G)} \cdot \mathcal{H}^1(J) \\ &+ \frac{\mathrm{vol}(\overline{J})}{\mathrm{vol}(G)} \cdot \mathcal{H}^1(\overline{J}) \\ &\geq \varepsilon_0 \log_2 n,\end{aligned}$$

where $\mathcal{H}^1(J)$ and $\mathcal{H}^1(\overline{J})$ are the entropy of nodes in $V_j (j \in J)$ and that in $V_j (j \in \overline{J})$, respectively. Since $\mathrm{vol}(\overline{J}) \leq \varepsilon \cdot \mathrm{vol}(G)$ and $\mathcal{H}^1(\overline{J}) \leq \log_2 n$ which is a trivial upper bound, there must be a constant $\varepsilon_1 > 0$ such that, for sufficiently large n,

$$\begin{aligned}\mathcal{H}^1(J) &= -\sum_{j \in J} \frac{\mathrm{vol}(V_j)}{\mathrm{vol}(J)} \cdot H_j - \sum_{j \in J} \frac{\mathrm{vol}(V_j)}{\mathrm{vol}(J)} \log_2 \frac{\mathrm{vol}(V_j)}{\mathrm{vol}(J)} \\ &\geq \varepsilon_1 \log_2 n.\end{aligned}$$

Note that $(1-\varepsilon) \cdot \mathrm{vol}(G) \leq \mathrm{vol}(J) \leq \mathrm{vol}(G)$. Thus there must be a constant $\varepsilon_2 > 0$ such that, for sufficiently large n,

$$-\sum_{j \in J} \frac{\mathrm{vol}(V_j)}{\mathrm{vol}(G)} \cdot H_j - \sum_{j \in J} \frac{\mathrm{vol}(V_j)}{\mathrm{vol}(G)} \log_2 \frac{\mathrm{vol}(V_j)}{\mathrm{vol}(G)} \geq \varepsilon_2 \log_2 n.$$

Let $0 < \delta < 1$ be an arbitrary constant, and $\phi = 1/\log^{\delta} n$. Define

$$A = \{j \in J : \Phi_j \leq \phi\},$$

and

$$B = J \setminus A = \{j \in J : \Phi_j > \phi\}.$$

Let $\mathrm{vol}(A) = \sum_{j \in A} \mathrm{vol}(V_j)$ and $\mathrm{vol}(B) = \sum_{j \in B} \mathrm{vol}(V_j)$. Assume that almost all volumes of modules in J are contributed by those in B. Then for the same reason as above, there must be a constant $\varepsilon_3 > 0$ such that

$$-\sum_{j \in B} \frac{\mathrm{vol}(V_j)}{\mathrm{vol}(G)} \cdot H_j - \sum_{j \in B} \frac{\mathrm{vol}(V_j)}{\mathrm{vol}(G)} \log_2 \frac{\mathrm{vol}(V_j)}{\mathrm{vol}(G)} \geq \varepsilon_3 \log_2 n. \tag{70}$$

Recall that

$$\begin{aligned}-\sum_{j \in B} \frac{\mathrm{vol}(V_j)}{\mathrm{vol}(G)} \cdot H_j - \sum_{j \in B} \Phi_j \cdot \frac{\mathrm{vol}(V_j)}{\mathrm{vol}(G)} \log_2 \frac{\mathrm{vol}(V_j)}{\mathrm{vol}(G)} \\ \leq \mathcal{H}^2(G) \leq \varepsilon^3 \log_2 \log_2 n.\end{aligned} \tag{71}$$

Comparing Equation (71) with (70) and noting that $\Phi_j > \phi = 1/\log^{\delta} n$ for $j \in B$ and $0 < \delta < 1$, we know that these two equations conflict for sufficiently large n. So there must be a constant $\rho_0 > 0$ such that $\mathrm{vol}(A) \geq \rho_0 \cdot \mathrm{vol}(J) \geq \rho_0(1 - \varepsilon) \cdot \mathrm{vol}(G)$.

Recall that for each $j \in J$, $H_j \leq \varepsilon^2 \log_2 \log_2 n$. Since

$$H_j = -\sum_{v \in V_j} \frac{d_v}{\mathrm{vol}(V_j)} \log_2 \frac{d_v}{\mathrm{vol}(V_j)},$$

there must be a subset of nodes $I_j \subseteq V_j$, such that for each $v \in I_j$,

$$\frac{d_v}{\mathrm{vol}(V_j)} \geq \frac{1}{\log_2^{\varepsilon} n},$$

and

$$\sum_{v \in I_j} \frac{d_v}{\mathrm{vol}(V_j)} \geq \ 1 - \varepsilon,$$

because otherwise, for each $v \in V_j \setminus I_j$, $\frac{d_v}{\mathrm{vol}(V_j)} < \frac{1}{\log_2^{\varepsilon} n}$, and so

$$\begin{aligned} H_j &\geq -\sum_{v \in V_j \setminus I_j} \frac{d_v}{\mathrm{vol}(V_j)} \log_2 \frac{d_v}{\mathrm{vol}(V_j)} \\ &> -\varepsilon \log_2 \left(\frac{1}{\log_2^{\varepsilon} n} \right) \\ &= \varepsilon^2 \log_2 \log_2 n, \end{aligned}$$

which is a contradiction. The size of I_j satisfies

$$|I_j| \leq \frac{\mathrm{vol}(V_j)}{\min_{v \in I_j}\{d_v\}} \leq \log_2^{\varepsilon} \ n.$$

This means that in each V_j for $j \in J = A \cup B$, there is a subset I_j of size $\log_2^{\varepsilon} n$ and volume $(1-\varepsilon) \cdot \mathrm{vol}(V_j)$.

In particular, consider the V_j's for $j \in A$. Let g_j be the total weights of edges each of which has exactly one end-point in V_j. So $g_j = \Phi_j \cdot V_j$. Since $\mathrm{vol}(I_j) \geq (1-\varepsilon) \cdot \mathrm{vol}(V_j)$, the conductance of I_j satisfies

$$\begin{aligned} \Phi(I_j) &\leq \frac{g_j + (\mathrm{vol}(V_j) - \mathrm{vol}(I_j))}{\mathrm{vol}(I_j)} \\ &\leq \frac{(1+\Phi_j) \cdot \frac{\mathrm{vol}(I_j)}{1-\varepsilon} - \mathrm{vol}(I_j)}{\mathrm{vol}(I_j)} \\ &= \frac{\varepsilon^{-1} \cdot \Phi_j + 1}{\varepsilon^{-1} - 1} \leq \frac{\varepsilon^{-1} \cdot \phi + 1}{\varepsilon^{-1} - 1} \leq \frac{2\varepsilon}{1-\varepsilon} \end{aligned}$$

for sufficiently large n. Set $S = \cup_{j \in A} I_j$ and $\rho = \rho_0(1-\varepsilon)$. Noting the arbitrariness of ε, the subset S has a volume $\mathrm{vol}(S) \geq \rho \cdot \mathrm{vol}(G)$, and each node in it belongs to a subset (some I_j) of size $\log^{o(1)} n$ and conductance $o(1)$.

Then we consider the complement $\overline{S}$ of S. It is exactly the union of V_j for j in $A \setminus S$, B and $\overline{J}$. Since for $j \in B$, $\Phi_j > \phi$, we know that if $\mathrm{vol}(B)$ is not negligible, that is, there is a constant $\eta > 0$ such that $\mathrm{vol}(B) \geq \eta \cdot \mathrm{vol}(G)$, then

$$-\sum_{j \in B} \frac{\mathrm{vol}(V_j)}{\mathrm{vol}(J)} \cdot H_j - \sum_{j \in B} \frac{\mathrm{vol}(V_j)}{\mathrm{vol}(J)} \log_2 \frac{\mathrm{vol}(V_j)}{\mathrm{vol}(J)} = o(\log_2 n),$$

because otherwise,

$$\begin{aligned} \mathcal{H}^2(G) &\geq -\sum_{j \in B} \frac{\mathrm{vol}(V_j)}{\mathrm{vol}(G)} \cdot H_j - \sum_{j \in B} \Phi_j \cdot \frac{\mathrm{vol}(V_j)}{\mathrm{vol}(G)} \\ &\quad \cdot \log_2 \frac{\mathrm{vol}(V_j)}{\mathrm{vol}(G)} \\ &= \Omega\left(-\sum_{j \in B} \frac{\mathrm{vol}(V_j)}{\mathrm{vol}(B)} \cdot H_j - \sum_{j \in B} \Phi_j \cdot \frac{\mathrm{vol}(V_j)}{\mathrm{vol}(B)} \right. \\ &\quad \left. \cdot \log_2 \frac{\mathrm{vol}(V_j)}{\mathrm{vol}(B)} \right) \\ &= \Omega\left(\log_2^{1-\delta} n \right), \end{aligned}$$

which contradicts to the fact that $\mathcal{H}^2(G) = o(\log \log n)$. Thus,

$$\begin{aligned} \mathcal{H}^1(B) &= -\sum_{j \in B} \frac{\mathrm{vol}(V_j)}{\mathrm{vol}(B)} \cdot H_j - \sum_{j \in B} \frac{\mathrm{vol}(V_j)}{\mathrm{vol}(B)} \log_2 \frac{\mathrm{vol}(V_j)}{\mathrm{vol}(B)} \\ &= o(\log_2 n). \end{aligned}$$

By a similar discussion to the prove of (1), we know that there is a subset $U \subseteq \cup_{j \in B} V_j$ with size $|U| = n^{o(1)}$, volume $\mathrm{vol}(U) = (1 - o(1)) \cdot \mathrm{vol}(B)$.

Since almost all volume of B is contributed by U and the volumes of both $A \setminus S$ and $\overline{J}$ are negligible compared to $\mathrm{vol}(G)$, we know that U's volume is approximately B's volume, and is also approximately $\overline{S}$'s volume. That is, $\mathrm{vol}(U) = (1 - o(1)) \cdot \mathrm{vol}(\overline{S})$, which is also $\Omega(\mathrm{vol}(G))$ because $\mathrm{vol}(\overline{S}) \geq \mathrm{vol}(B)$ is not negligible. Consider the edges that has exactly one end-point in U. The total weight of those whose the other end-point is in $\overline{S} \setminus U$ is at most $\mathrm{vol}(\overline{S} \setminus U) = o(\mathrm{vol}(G))$. The total weight of those whose the other end-point is in A is at most $\Phi(A) \cdot \mathrm{vol}(A)$. Note that for each $j \in A$, $\Phi_j \leq \phi$. Thus, $\Phi(A) \leq \phi$, and $\Phi(A) \cdot \mathrm{vol}(A) \leq \phi \cdot \mathrm{vol}(G) = o(\mathrm{vol}(G))$. So $\Phi(U) = o(1)$, and (2) has been proved.

This completes the proof of Theorem 62. □

On the other hand, for a connected network G, if $\mathcal{H}^2(G) = o(\log \log n)$, then there must be a set of nodes S such that the volume of S is huge while each node in S belongs to a tiny set of small conductance, so that a random walk in G is highly like to arrive at S, and once arrives at S, it is hard to escape from S.

We have seen that structural information of a network measures the dynamical complexity of the network as the least overall number of bits to determine the code of the node that is accessible from a step of random walk in the network. As we have mentioned early, structural information provides a measure to characterise the natural structures of a network. We will verify that this is indeed the case.

XVI. Structural Information Minimisation Principle

Given a network G, our K-dimensional structural information $\mathcal{H}^K(G)$ determines a K-dimensional structure $\mathcal{T}$ such that the K-dimensional structural information of the network given by $\mathcal{T}$ is minimised. Structural information minimisation is the problem to find such a K-dimensional structure $\mathcal{T}$ which minimises the K-dimensional structural information of the network.

In this section, we will investigate the roles of the structure $\mathcal{T}$ that minimises the structural information in the network.

A. Structural Information Minimization Principle for Discovering Natural Communities in Networks

Given a network $G = (V, E)$, our algorithm $\mathcal{E}$ is to find a partition $\mathcal{P}$ such that $L^{\mathcal{P}}(G)$ is minimized. The problem is clearly hard. Here we describe a simple greedy algorithm to find a partition which minimizes the structural information of the network G. Before describing the algorithm, we define the following notation.

Suppose that $\mathcal{P} = \{X_1, X_2, \cdots, X_L\}$ be a partition of V. For i, j with $1 \le i, j \le L$. We define $\Delta^{\mathcal{P}}_{i,j}(G)$ as follows:

$$\begin{aligned}\Delta^{\mathcal{P}}_{i,j}(G) = &-\frac{V_i}{2m}\sum_{k=1}^{n_i}\frac{d_k^{(i)}}{V_i}\log\frac{d_k^{(i)}}{V_i}\\ &-\frac{V_j}{2m}\sum_{k=1}^{n_j}\frac{d_k^{(j)}}{V_j}\log\frac{d_k^{(j)}}{V_j}\\ &+\frac{V_X}{2m}\sum_{k=1}^{n_i+n_j}\frac{d_k^{(i,j)}}{V_X}\log\frac{d^{(i,j)}}{V_X}\\ &-\frac{g_i}{2m}\log\frac{V_i}{2m}-\frac{g_j}{2m}\log\frac{V_j}{2m}+\frac{g_X}{2m}\log\frac{V_X}{2m}\\ =&\frac{1}{2m}[(V_i-g_i)\log V_i+(V_j-g_j)\log V_j\\ &-(V_X-g_X)\log V_X+(g_i+g_j-g_X)\log 2m],\end{aligned} \tag{72}$$

where $X = X_i \cup X_j$, V_X is the volume of X, g_X is the number of edges from X to nodes outside of X, $d_k^{(i,j)}$ is the degree of the k-th node in X.

By definition, we have (i) for every pair (i, j), $\Delta^{\mathcal{P}}_{i,j}(G)$ is locally computable, and (ii) for any i, j, if there is no edge between X_i and X_j, then $\Delta^{\mathcal{P}}_{i,j}(G) < 0$.

(i) is obvious.

For (ii). If there is no edge between X_i and X_j, then $g_X = g_i + g_j$, and $V_X = V_i + V_j$. By using these, we have that

$$\begin{aligned}\Delta^{\mathcal{P}}_{i,j}(G) =&\frac{1}{2m}[(V_i-g_i)\log V_i+(V_j-g_j)\log V_j\\ &-(V_X-g_X)\log V_X]\\ =&\frac{1}{2m}[(V_i-g_i)\log\frac{V_i}{V_i+V_j}\\ &+(V_j-g_j)\log\frac{V_j}{V_i+V_j}]\le 0.\end{aligned} \tag{73}$$

Definition 63 (Structural Information Minimisation Algorithm $\mathcal{E}$): We now describe our algorithm $\mathcal{E}$, which proceeds as follows.

1. *Suppose that $v_1, v_2, \cdots, v_n$ are all the nodes in V with ordering as they are listed. Set X_i to be the singleton $\{v_i\}$ for all i, which form the initial partition of V. Suppose that $\mathcal{P} = \{X_1, X_2, \cdots, X_L\}$ is a partition with ordering as they are listed.*
2. *If there is no $i < j$ such that $\Delta^{\mathcal{P}}_{i,j}(G) > 0$, then terminate with output $\mathcal{P}$, where $e(X_i, X_j)$ is the number of edges between X_i and X_j.*
3. *Otherwise, then let i_0, j_0 be such that $\Delta^{\mathcal{P}}_{i_0,j_0}(G)$ is maximized among $\Delta^{\mathcal{P}}_{i,j}(G)$ for all i, j's, set $X = X_{i_0} \cup X_{j_0}$, set $\mathcal{P} = \{X_1, \cdots, X_{i_0-1}, X_{i_0+1}, \cdots, X_{j_0-1}, X_{j_0+1}, \cdots, X_{L-1}, X\}$, and go back to step (2).*

In [32], we have shown that the algorithm $\mathcal{E}$ by structural information minimization exactly identifies or precisely approximates natural or true communities in large-scale networks.

B. Algorithm for Minimising the K-Dimensional Structural Information - Identifying Natural Structures in Networks

In this section, we describe an algorithm for finding a partitioning tree $\mathcal{T}$ of a graph G of height K to minimise the K-dimensional structural information of G by $\mathcal{T}$.

Two operators, the merging operator and the combining operator, are introduced, and a partitioning tree is developed by using the two operators.

First, we define the *merging operator*. Let $\mathcal{T}$ be a partitioning tree and let α and β be nodes of $\mathcal{T}$ with $\alpha <_{\mathrm{L}} \beta$ (meaning that α is to the left of β) and $\alpha^- = \beta^- = \gamma$ for some γ. In addition, let $\alpha = \gamma\,\hat{}\,\langle i\rangle$, and $\beta = \gamma\,\hat{}\,\langle j\rangle$ for $i < j$.

Definition 64 (Merging Operator): We define a partitioning tree below, which is obtained from $\mathcal{T}$ via the following merging operator: $\mathcal{M}(\mathcal{T}; \alpha, \beta)$:

Let $T_\alpha = \{x_1, x_2, \cdots, x_M\}$ and $T_\beta = \{y_1, y_2, \cdots, y_N\}$, which are ordered as listed in the sets. Then,

1. *Define $T_\alpha = \{x_1, x_2, \cdots, x_M, y_1, y_2, \cdots, y_N\}$, which are ordered as they are listed.*
2. *Set $h(\alpha) \leftarrow h(\alpha)$.*
3. *For each $s \in \{1, 2, \cdots, M\}$, define $T_{\alpha\,\hat{}\,\langle s\rangle} = \{x_s\}$ with $h(\alpha\,\hat{}\,\langle s\rangle) \leftarrow h(\alpha) + 1$.*
4. *For every t with $M + 1 \le t \le M + N$, define $T_{\alpha\,\hat{}\,\langle t\rangle} = \{y_{t-M}\}$ with $h(\alpha\,\hat{}\,\langle t\rangle) = h(\alpha) + 1$.*
5. *Delete β.*
6. *For every $j' > j$, if $\gamma\,\hat{}\,\langle j'\rangle$ is defined, then set*

$$T_{\gamma\,\hat{}\,\langle j'-1\rangle} \leftarrow T_{\gamma\,\hat{}\,\langle j'\rangle}.$$

Here, we use $\mathcal{T}^{\mathcal{T}}_G(\alpha, \beta)$ to denote the partitioning tree defined by $\mathcal{T}$ via the merging operator $\mathcal{M}(\mathcal{T}; \alpha, \beta)$ above.

We define the difference between the structure entropies of G obtained from a partitioning tree $\mathcal{T}$ and a partitioning tree obtained from $\mathcal{T}$ through a merging operator.

For a graph $G = (V, E)$ and a partitioning tree $\mathcal{T}$ of G, let $\alpha, \beta \in \mathcal{T}$ such that $\alpha^- = \beta^-$ and $h(\alpha) < K$. Then, define $\Delta^G(\mathcal{T}; \alpha, \beta) = L^{\mathcal{T}}(G) - L^{\mathcal{T}'}(G)$, where $\mathcal{T}' = \mathcal{T}^{\mathcal{T}}_G(\alpha, \beta)$. By definition, we have

$$\begin{aligned}\Delta^{\mathcal{M}}_G(\mathcal{T}; \alpha, \beta) = &-\sum_{\gamma\in\mathcal{T}:\ \alpha\subseteq\gamma\ \text{or}\ \beta\subseteq\gamma}\frac{g_\gamma}{2m}\log_2\frac{V_\gamma}{V_{\gamma^-}}\\ &+\sum_{\delta\in\mathcal{T}':\ \alpha\subseteq\delta}\frac{g_\delta}{2m}\log_2\frac{V_\delta}{V_{\delta^-}}.\end{aligned} \tag{74}$$

In this case, if α and β occur such that $\alpha <_{\mathrm{L}} \beta$, $\alpha^- = \beta^-$, $h(\alpha) < K$, and if $\Delta^{\mathcal{M}}_G(\mathcal{T}; \alpha, \beta) > 0$, then $\mathcal{M}(\mathcal{T}; \alpha, \beta)$ is defined and written as $\mathcal{M}(\mathcal{T}; \alpha, \beta)\downarrow$.

According to equation (74), $\Delta^{\mathcal{M}}_G(\mathcal{T}; \alpha, \beta)$ is locally computable.

Second, we define the *combining operator*.

Definition 65 (Combining Operator): Let $G = (V, E)$ be a graph, and let $\mathcal{T}$ be a partitioning tree of G.

For any $\alpha, \beta \in \mathcal{T}$, if:

(i) $\alpha^- = \beta^- = \gamma$ for some γ, and

(ii) for any $\delta \in \mathcal{T}$, if either $\alpha \subseteq \delta$ or $\beta \subseteq \delta$, then $h(\delta) < K$, then define the combining operator $\mathcal{C}(\mathcal{T}; \alpha, \beta)$ as follows:

– create a new node ξ with $T_\xi = T_\alpha \cup T_\beta$ and $\xi^- = \gamma$,

– let the two branches with root α and β in $\mathcal{T}$ be two branches of ξ, while maintaining the same order as in $\mathcal{T}$.

By definition, the Δ-function with the combining operator $\mathcal{C}(\mathcal{T}; \alpha, \beta)$ is as follows:

$$\Delta_G^{\mathcal{C}}(\mathcal{T}; \alpha, \beta) = H^{\mathcal{T}}(G; \alpha) + H^{\mathcal{T}}(G; \beta) - (H^{\mathcal{T}'}(G; \xi) + H^{\mathcal{T}'}(G; \alpha) + H^{\mathcal{T}'}(G; \beta)), \quad (75)$$

where $\mathcal{T}'$ is the tree obtained from $\mathcal{T}$ by the combing operator $\mathcal{C}(\mathcal{T}; \alpha, \beta)$.

If $\alpha <_{\mathrm{L}} \beta$ such that $\alpha^- = \beta^-$, and if $\alpha \subseteq \delta$ or $\beta \subseteq \delta$ implies $h(\delta) < K$ for every δ and $\Delta_G^{\mathcal{C}}(\mathcal{T}; \alpha, \beta) > 0$. Thus, $\mathcal{C}(\mathcal{T}; \alpha, \beta)$ is defined and written as $\mathcal{C}(\mathcal{T}; \alpha, \beta) \downarrow$.

In this case, it is clear that $\Delta_G^{\mathcal{C}}(\mathcal{T}; \alpha, \beta)$ is locally computable.

Finally, we introduce our algorithm denoted $\mathcal{E}^K$ by using both the merging and combining operators.

Definition 66 (Algorithm $\mathcal{E}^K$ for Detecting K-Dimensional Natural Structure of Networks): Let $G = (V, E)$ be a graph. Suppose that $\{v_1, v_2, \cdots, v_n\}$ is the set of all vertices in V ordered as they are listed in the set. The K-dimensional structural information algorithm on G proceeds as follows:

(1) *Define the initial partitioning tree $\mathcal{T}$ as follows:*
   – Set $T_\lambda = V$ with $h(\lambda) = 0$, and for every $i \in \{1, 2, \cdots, n\}$, define $T_{\lambda\hat{}\langle i\rangle} = \{v_i\}$ with $h(\lambda\hat{}\langle i\rangle) = h(\lambda) + 1$.

(2) *If there are $\alpha, \beta \in \mathcal{T}$ such that $\mathcal{M}(\mathcal{T}; \alpha, \beta) \downarrow$, then*
   a) *choose α and β such that $\Delta_G^{\mathcal{M}}(\mathcal{T}; \alpha, \beta)$ is maximised;*
   b) *let $\mathcal{T}'$ be the partitioning tree obtained from $\mathcal{T}$ by the merging operation of $\mathcal{T}$ with α and β;*
   c) *set $\mathcal{T} \leftarrow \mathcal{T}'$; and*
   d) *go back to step (2).*

(3) *If there are $\alpha, \beta \in \mathcal{T}$ such that $\mathcal{C}(\mathcal{T}; \alpha, \beta) \downarrow$, then*
   a) *choose α and β such that $\Delta_G^{\mathcal{C}}(\mathcal{T}; \alpha, \beta)$ is maximised;*
   b) *let $\mathcal{T}'$ be the partitioning tree obtained from $\mathcal{T}$ by the combining operation of $\mathcal{T}$ with α and β;*
   c) *set $\mathcal{T} \leftarrow \mathcal{T}'$; and*
   d) *go back to step (2).*

(4) *Otherwise, output the partitioning tree $\mathcal{T}$, and terminate the program.*

The algorithm $\mathcal{E}^K$ outputs a partitioning tree $\mathcal{T}$ of G. Clearly algorithm $\mathcal{E}^K$ works naturally on weighted networks.

Time complexity of algorithm $\mathcal{E}^K$. For $K = 2$, the time complexity of $\mathcal{E}^2$ is $O(n^2)$ for all graphs, and is $O(n \cdot \log^2 n)$ for sparse networks [32], where n is the number of nodes in the graph. This algorithm is a nearly linear time algorithm for networks, which easily functions for networks that include millions of nodes. For $K = 3$, for every first level node α in the partitioning tree, the size of T_α does not decrease during the implementation of the algorithm. Therefore, $|T_\alpha| = M$ for M with $1 \le M \le n$. For a fixed M and a fixed T_α of size M, the number of operations associated with the children of α is the time complexity of an $\mathcal{E}^2$ with M graphs of M nodes; thus, $O(M^2)$ for general graphs, and $O(M \cdot \log^2 M)$ for networks. This analysis gives the time complexity with one first level node α of the partitioning tree $O(n^3)$ for all graphs and $O(n^2 \log^2 n)$ for sparse graphs. Because there are at most n first level nodes in the partitioning tree, the time complexity of $\mathcal{E}^3$ is bounded by $O(n^4)$ for all graphs, and $O(n^3 \cdot \log^2 n)$ for sparse networks, which is significantly larger than that of $\mathcal{E}^2$.

The time complexity analysis above clearly indicates that our algorithm $\mathcal{E}^3$ is not a hierarchical clustering algorithm with 3 levels. Because of the time complexity of $O(n^3 \log^2 n)$ for sparse networks or $O(n^4)$ for all graphs, although $\mathcal{E}^3$ is a polynomial time algorithm, it can, in practice, only manage graphs that contain thousands of nodes. Therefore, it is difficult to detect the natural K-dimensional structure of a network of large sizes for $K > 3$ (the time complexity generally increases by a factor of n^2 whenever the dimension increases by 1, for dimensions $K \ge 2$), which poses a new issue regarding the design of better algorithms for minimising the K-dimensional structural information of networks for each $K > 1$, including for the case of $K = 2$. We remark that the time complexity $O(n \log^2 n)$ of $\mathcal{E}^2$ for networks is in fact impractical for n as large as hundreds of millions. For this reason, there is a need to design better algorithms to find the minimal two-dimensional structural information of networks.

Finally, we note that $\mathcal{E}^K$ is a heuristic algorithm to compute the K-dimensional structural information of graphs and indicate that precisely computing the K-dimensional structural information of graphs is an extremely difficult problem that should be resolved in future computer science studies.

Remark: (i) the merging operator combines two sets X and Y into a set $Z = X \cup Y$ such that all of the nodes in Z are not distinguished and are allowed to re-group within Z in the future; (ii) the combining operator combines two sets X and Y into $Z = X \cup Y$ such that that the subtypes X and Y are kept within Z; (iii) the two operators are natural rules in real world clustering, which incorporates the idea of a mixture both bottom-up and top-down methods; (iv) our algorithm $\mathcal{E}^K$ is a basic greedy strategy for minimising the K-dimensional structural information, and new rules are required to design new algorithms to minimise the K-dimensional structural information of graphs; (v) we determined the K-dimensional structural information for small values of K because, in real world networks, hierarchical structures occur; however, the number of levels of a community within a community is indeed small.

Clearly, our algorithm $\mathcal{E}^K$ not only seeks to follow the principle of self-organisation of networks, but also the natural rules in the real world as the operators of the algorithm. The algorithms are designed to explore the natural two- and three-dimensional structures of networks rather than optimise an artificially defined object function. We use this strategy

because natural objects can be identified by following natural rules, and algorithm $\mathcal{E}^2$ has been shown to successfully detect natural communities in social networks [32]. The algorithm $\mathcal{E}^3$ can be regarded as a deep detecting algorithm that seeks to explore the natural hierarchical structure of networks.

In [36], we have shown that the cell types and subtypes of the cell sample networks constructed from the gene expression profiles identified by our algorithms $\mathcal{E}^2$ and $\mathcal{E}^3$ are defined by a unique gene expression pattern, leading to a high-definition, one-to-one map of tumour types and subtypes and the gene expression patterns. We have also shown that most cell samples within the same type or subtype identified by our algorithms $\mathcal{E}^2$ and $\mathcal{E}^3$ share similar survival times, survival indicators and International Prognostic (IPI) scores, and that cell samples from distinct types or subtypes identified by our algorithms $\mathcal{E}^2$ and $\mathcal{E}^3$ have distinct overall survival times, survival ratios and IPI scores.

C. Security Engineering of Networks

Li *et al.* [33] proposed a security model of networks. The authors [34] have shown that for appropriately large affinity exponent a, for sufficiently large networks generated by the security model, with probability almost 1, for any set S of the nodes of the network, if the size of S is bounded by $\log^{O(1)} n$, then the infection set of S in the network G has size at most $o(n)$, where n is the number of nodes of network G. The authors have shown that Theorem 59 holds for the networks of the security model. This result, together with the security theory in [34] demonstrate that as the affinity exponent a varies, the lower the two-dimensional structural information of the networks is, the more secure the networks of the security module are. This indicates that two-dimensional structural information minimisation would be a principle of network security. In particular, the normalised two-dimensional structural information could provide an index for the security of networks. Nevertheless, the two-dimensional structural information minimisation provides a principle for network security engineering.

D. Emergence of Cooperation in Evolutionary Games in Networks

According to Darwin's evolution theory [18], animals from ants to people form social groups in which most individuals work for the common good. The cooperation of social groups is certainly key to the survival and the evolution of species. Darwin suggested that kinship and reciprocity are the means for fitness for survival. Our homophyly/kinship model realises some of Darwin's idea by networks [31]. In [31], it was shown that for appropriately large affinity exponent a, and the networks of the homophyly/kinship model, cooperation is guaranteed to emerge in the evolutionary prisoner's dilemma games in the networks. According to Theorem 59, we now know that as the affinity exponent a varies, the two-dimensional entropy of the networks vary, and that the lower the two-dimensional structural information of the networks is, the higher the emergence of cooperation of evolutionary games in the networks. The results demonstrate the following principle:

- Structural information minimisation is the principle for the natural structures of a naturally evolving network.
- Structural information minimisation is the principle for the emergence of cooperation in evolutionary games in the network.

The principle above provides a significant first step towards a mathematical theory of evolutionary games in networks, nature and society.

XVII. Natural Structure Discovery

In Definition 17, we define the natural structure of a network to be the structure of the network that minimises the structural information of the network. To see that the natural structure defined here is in fact the true structure of the network, we noticed that Li *et al.* [36] have developed a three-dimensional gene maps of cancer cell types and subtypes for a number of cancers. The method consists of two steps, the first step is to construct a network of the unstructured gene expression profiles on the basis of one-dimensional structural information minimisation, and the second step is to find the cell types and subtypes by our algorithms $\mathcal{E}^2$ and $\mathcal{E}^3$ on the basis of the two- and three-dimensional structural information minimisation above.

For completeness of the paper, we introduce three figures of the gene maps for Lymphoma developed in [36]. (Details of the data and experiments can be found in [36].)

The Lymphoma data from Alizadeh *et al.* [1], which contain the expression of 4, 026 genes for 96 samples, which constitute 9 cell types. The 9 types consist of three different types of tumours, diffuse large B cell lymphoma (DLBCL), chronic lymphocytic leukaemia (CLL), and follicular lymphoma (FL), as well as normal B and T cells at different stages of cell differentiation, including germinal centre B, NL.lymph node/tonsil, activated blood B, resting/activated T, transformed cell lines, and resting blood B. On the basis of gene expression profiling, Alizadeh *et al.* [1] have suggested dividing the DLBCL type into two subtypes, GC B-like DLBCL and activated B-like DLBCL.

We use our algorithms $\mathcal{E}^K$ for $K = 2, 3$ to identify the modules and submodules of cancers and to compare our algorithms with the most frequently used algorithm, namely, the modularity maximisation algorithm $\mathcal{M}$ [17].

A. Gene Map of True Types

The true types of lymphoma consist of 9 types: DLBCL, germinal centre B, NL. lymph node/tonsil, activated blood B, resting/activated T, transformed cell lines, FL, resting blood B, and CLL.

Figure 3 shows the gene map of the true types of lymphoma listed as above.

Figure 3 reveals the following results:

1) All of the types are distinguishable because they are defined by different blocks of genes.
2) All of the types except DLBCL are expressed by different sets of genes.

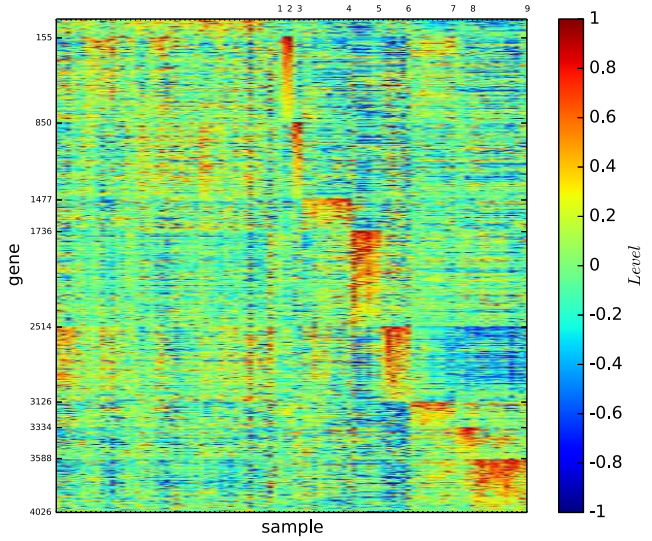

Fig. 3. Gene map of true types of the lymphomas.

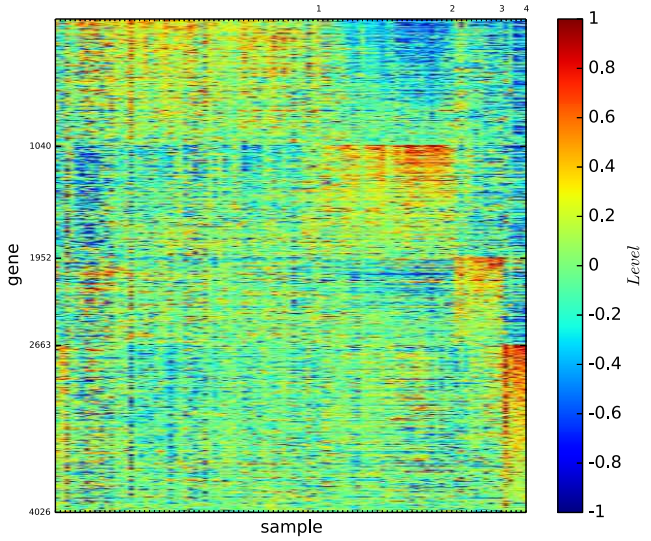

Fig. 4. Gene map of types of the lymphomas found by $\mathcal{M}$.

3) The type DLBCL is a large set; however, it is not well-expressed.
4) Four types (germinal centre B, NL. lymph node/tonsil, resting/activated T and transformed cell lines) are highly expressed by a set of many genes; thus, the blocks of genes expressing the types are large.
5) Except for DLBCL, the 8 remaining types are highly expressed by their corresponding blocks of genes.

These results imply that DLBCL is not a well-defined type, which will be shown in the gene map of the modules of lymphoma identified by the algorithm $\mathcal{E}^2$.

B. Gene Map of the Modules by Modularity Maximisation

The algorithm $\mathcal{M}$ found 4 communities of the lymphoma. Figure 4 depicts the gene map of the four modules of lymphoma identified by the modularity maximisation algorithm $\mathcal{M}$.

Figure 4 reveals that the four modules identified by $\mathcal{M}$ are distinguishable by different sets of genes. However, the modules identified by $\mathcal{M}$ are far from the true types. This means that gene expression patterns alone is insufficient for evaluating the modules identified by a community detection algorithm.

C. Gene Map of the Modules by Structural Information Minimisation Algorithm $\mathcal{E}^2$

Our algorithm $\mathcal{E}^2$ found 11 types. They are:

Module 1: OCI Ly3, OCI Ly1, WSU1, Jurkat, U937, OCI Ly12, OCI Ly13.2, SUDHL5, DLCL-0041. This module consists of DLBCL or Transformed cell lines.

Module 2: OCI Ly10, DLCL-0042, DLCL-0007, DLCL-0031, DLCL-0036;OCT, DLCL-0025, DLCL-0040, DLCL-0017, DLCL-0028, DLCL-0021, DLCL-0012. This module is the Activated B-like DLBCL, except for DLCL-0012.

Module 3: DLCL-0030, DLCL-0011, DLCL-0020, DLCL-0032, DLCL-0033, DLCL-0003, DLCL-0034, DLCL-0051, DLCL-0001, DLCL-0018, DLCL-0037, DLCL-0010. This module is the GC B-like DLBCL except for DLCL-0011.

Module 4: DLCL-0004, DLCL-0029, DLCL-0008, Tonsil GC B, Tonsil GC Centroblasts, SUDHL6, DLCL-0052. This module consists of the GC B-like DLBCL and the Germinal centre B.

Module 5: DLCL-0006, DLCL-0049, Tonsil, DLCL-0039, Lymph Node, DLCL-0002. This module is the Activated B-like DLBCL, together with Nl. lymph node/tonsil.

Module 6: DLCL-0015, DLCL-0026, DLCL-0023, DLCL-0027, DLCL-0024, DLCL-0005, DLCL-0013, DLCL-0016, DLCL-0014, DLCL-0048. This module is a DLBCL subtype.

Module 7: This module is exactly the Activated blood B.

Module 8: This module is exactly the Resting/activated T.

Module 9: This module is the FL with one error DLCL-0011.

Module 10: This module is exactly the Resting blood B.

Module 11: This module is exactly the CLL.

Figure 5 depicts the gene map of the modules of lymphoma classified by our algorithm $\mathcal{E}^2$ with ordering listed as above.

Figure 5 reveals the following properties: (1) Modules 1, 2, 3, 7, 8, 9, 10, and 11 essentially correspond to transformed cell lines, activated B-like DLBCL, GC B-like DLBCL, activated blood B, resting/activated T, FL, resting blood B, and CLL, respectively. (2) The DLBCL type is essentially divided into modules 2, 3, and 6. (3) Except for module 3, which contains the subtype GC B-like DLBCL, every module is highly expressed by a significantly large set of genes. (4) Module 2 is the subtype activated B-like DLBCL and is highly expressed by a set of more than 300 genes. (5) Module 3 contains the subtype GC B-like DLBCL (except for DLCL-0011) and is large. However, the module is well-expressed only by a set of less than 100 genes. This finding could be caused by i) the expression of the subtype GC B-like by only a small set of genes or ii) an

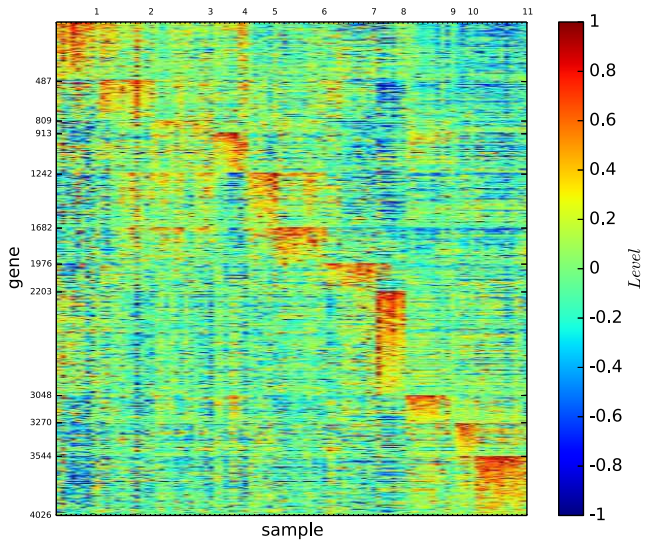

Fig. 5. Gene map of types of the lymphomas found by $\mathcal{E}^2$.

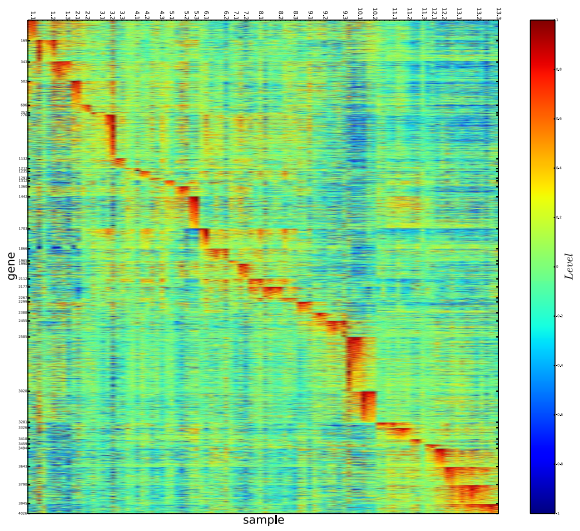

Fig. 6. Gene map of types of the lymphomas found by $\mathcal{E}^3$.

incomplete current gene expression array. (6) Module 6 contains a subset of DLBCL, and its biological and medical classification is unknown. However, our two-dimensional gene map shows that the module is highly expressed by a set of more than 290 genes. (7) Module 4 is a combination of GC B-like DLBCL, and germinal centre B. Our gene map shows that the module is highly expressed by a set of more than 300 genes. (8) Module 5 is a combination of activated B-like DLBCL and NL. lymph node/tonsil. Our gene map shows that the module is highly expressed by a set of more than 400 genes, implying that, it is a biologically meaningful type. (9) Module 8 is the resting/activated T, and it is highly expressed by a set of more than $1,800$ genes. (10) Module 11 is the CLL, and it is highly expressed by a set of more than 450 genes.

These results imply that modules 2, 3 and 6 could be new subtypes of DLBCL and modules 4 and 5 could be new subtypes of lymphoma.

D. Three-Dimensional Gene Map

The subtypes found by the algorithm $\mathcal{E}^3$ are principally a refined classification of the types by $\mathcal{E}^2$. For the lymphoma, algorithm $\mathcal{E}^3$ found 13 types, each of which consists of two or three subtypes.

Figure 6 depicts the gene map of the refined classification of lymphoma derived by our algorithm $\mathcal{E}^3$. Figure 6 establishes the three-dimensional gene map from the types and subtypes of the gene expression patterns, which shows the types and subtypes of lymphoma, and demonstrate that almost all of the subtypes may have a biological meaning related to the classification of tumour types and subtypes of lymphoma. In particular, it predicts some remarkable subtypes for DLBCL and lymphoma, which will be verified by clinical data analyses.

The gene maps show that each of the found type or subtype by our algorithms $\mathcal{E}^2$ and $\mathcal{E}^3$ is uniquely defined by a gene patten. More importantly, we have shown that most cell samples found by our algorithm share similar survival times, survival indicators and IPI scores, that the samples in different modules have significantly different overall survival times, survival ratios, and IPI scores. These results indicate that the classification of the cell samples by our algorithms are interpretable and distinguishable in clinical practice.

The results demonstrate the following principles:

- Two-dimensional structural information minimisation is the principle for defining tumour types.
- Three-dimensional structural information minimisation is the principle for defining cancer cell subtypes.

Generally, our results demonstrate that, for the networks naturally evolving in nature and society, the following results hold:

- High-dimensional structural information minimisation is the principle for the self-organisation of individuals in the networks.
- Two-dimensional structural information minimisation is the principle for natural communities in the networks.
- High-dimensional structural information minimisation is the principle for the natural high-dimensional structures of the networks.

XVIII. Natural Rank: Locally Listing Algorithms

We design a personalized web ranking algorithm based on the two dimensional structural information of networks. It is a local algorithm for searching and ranking by the notion of structural information.

Let $G = (V, E)$ be a network. Suppose that $X \subset V$ is a subset of vertices, and that $\{y_1, y_2, \cdots, y_N\}$ is the set of all nodes $y \in V \setminus X$. Let Y_j be the set of single node y_j. Then X and all Y_j's form a partition of G.

For j with $1 \leq j \leq L$, if we put y_j into X, giving a new partition of G, then by definition, the difference of structure entropies of the two partitions is given by

$$\Delta^G(X; j) = \frac{1}{2m}[(V_X - g_X)\log V_X - (V_{Z_j} - g_{Z_j})\log V_{Z_j} + (g_X + d_j - g_{Z_j})\log 2m], \quad (76)$$

where $Z_j = X \cup \{y_j\}$, V_X is the volume of X, g_X is the number of edges from X to nodes outside of X, d_j is the degree of y_j in G, and m is the number of edges of the graph G.

Our personalized web ranking by structural information, written by $\mathcal{E}^L$, is a greedy algorithm and proceeds as follows.

Definition 67 (Locally Listing Rank, $\mathcal{E}^L$): Given network $G = (V, E)$, and node $v \in V$:

(1) *enumerate v into X and set $l = 1$,*
(2) *if there is no j such that $\Delta^G(X; j) > 0$, then terminate with output X, and*
(3) *otherwise, then,*
   a) *enumerate the y_j with which $\Delta^G(X; j)$ is maximized, and*
   b) *go back to step (2).*

The algorithm $\mathcal{E}^L$ with input v outputs a list of nodes $v = u_1, u_2, \cdots, u_l$ for some l. In this case, the output is the set $X = \{u_1, u_2, \cdots, u_l\}$ with the priority as they are enumerated by the algorithm, which is the ranking from v by the algorithm.

Algorithm $\mathcal{E}^L$ generates a personalised rank by simply listing the nodes by the priority of minimising the structural information of the partition with the desired module and singleton modules for all the other nodes.

Let Δ be the largest degree in G. Then the complexity of $\mathcal{E}^L$ is $O(\Delta \cdot |X|)$, where X is the output set. Notice that the output set X cannot be too large by the arguments of the definition of structural information. This means that $\mathcal{E}^L$ simply lists its output, noting that the largest degree Δ is usually small for real world networks.

By Theorem 18, for every input node v, the output of $\mathcal{E}^L$ must be a proper subset of V. In fact, the sizes of the outputs are very small. In fact, we may set a boundary for the size of the outputs such as $\log^b n$ for some small constant $b \in [1, 2]$. In so doing, the time complexity of the algorithm is expected to be $O(\log^{O(1)} n)$.

We verify that algorithm $\mathcal{E}^L$ precisely identifies the natural communities of networks. We verify this fact by using the networks of the homophyly/kinship model.

Definition 68: Let $G = (V, E)$ be a network and X, Y be sets of nodes in V. We define the similarity between X and Y as follows:

$$s(X, Y) = \frac{|X \cap Y|}{\sqrt{|X| \cdot |Y|}}.$$

Given a network $G = (V, E)$ generated by the homophyly/kinship model, for every node v, we use X_v to node the natural community of v, i.e., the maximal set of nodes sharing the same color with v. Let Z_v be the output of $\mathcal{E}^L$ with input v such that Z_v is listed as the ordering as they are listed by the algorithm $\mathcal{E}^L$. Let $l = |X_v|$, and Y_v be the set of the first l elements of Z_v. We define $s(v)$ to be the similarity between X_v and Y_v. We investigate the distribution of $s(v)$ for all the nodes $v \in V$.

Table I describes the soundness of the locally listing rank algorithm $\mathcal{E}^L$ on the networks of the homophyly/kinship model.

TABLE I

SOUNDNESS OF THE LOCALLY LISTING RANK ALGORITHM $\mathcal{E}^L$ ON THE NETWORKS OF THE HOMOPHYLY/KINSHIP MODEL. T AND S DENOTE THE TYPES AND SOUNDNESS, RESPECTIVELY

% S / T	(0.0,0.4]	(0.4,0.6]	(0.6,0.8]	(0.8,1.0)	1.0
a=0.2,d=2	0.14%	0.03%	0.15%	0.02%	99.66%
a=0.2,d=4	0.13%	0.00%	0.00%	0.00%	99.87%
a=0.2,d=8	0.00%	0.00%	0.00%	0.00%	100.00%
a=0.3,d=2	0.29%	0.18%	0.10%	0.10%	99.33%
a=0.3,d=4	0.05%	0.04%	0.00%	0.00%	99.91%
a=0.3,d=8	0.18%	0.00%	0.00%	0.00%	99.81%
a=0.4,d=2	0.37%	0.02%	0.07%	0.27%	99.27%
a=0.4,d=4	0.19%	0.00%	0.00%	0.00%	99.81%
a=0.4,d=8	0.04%	0.00%	0.00%	0.00%	99.96%
a=0.5,d=2	0.65%	0.04%	0.11%	0.25%	98.95%
a=0.5,d=4	0.31%	0.00%	0.00%	0.00%	99.69%
a=0.5,d=8	0.02%	0.00%	0.00%	0.00%	99.98%
a=0.8,d=2	0.86%	0.00%	0.09%	0.52%	98.53%
a=0.8,d=4	0.26%	0.00%	0.00%	0.08%	99.66%
a=0.8,d=8	0.38%	0.00%	0.00%	0.00%	99.62%
a=1.0,d=2	0.18%	0.00%	0.04%	0.45%	99.33%
a=1.0,d=4	1.04%	0.00%	0.00%	0.02%	98.94%
a=1.0,d=8	0.52%	0.00%	0.00%	0.00%	99.48%
a=1.2,d=2	1.10%	0.07%	0.12%	0.34%	98.37%
a=1.2,d=4	1.74%	0.00%	0.04%	0.09%	98.13%
a=1.2,d=8	0.00%	0.00%	0.00%	0.00%	100.00%

TABLE II

SOUNDNESS OF PAGERANK $\mathcal{G}$ WITH $\alpha = 0.15$ AND $\epsilon = 0.00001$ ON THE NETWORKS OF THE HOMOPHYLY/KINSHIP MODEL. T AND S DENOTE THE TYPES AND SOUNDNESS, RESPECTIVELY

% S / T	(0.0,0.4]	(0.4,0.6]	(0.6,0.8]	(0.8,1.0)	1.0
a=0.2,d=2	1.22%	2.87%	5.17%	1.32%	89.42%
a=0.2,d=4	0.75%	1.51%	4.53%	1.36%	91.85%
a=0.2,d=8	0.18%	0.68%	2.02%	0.80%	96.32%
a=0.3,d=2	1.63%	3.01%	6.88%	3.40%	85.08%
a=0.3,d=4	0.59%	1.74%	5.81%	2.44%	89.42%
a=0.3,d=8	0.05%	0.61%	2.20%	2.06%	95.08%
a=0.4,d=2	1.10%	3.06%	7.59%	6.28%	81.97%
a=0.4,d=4	0.54%	1.47%	5.88%	5.30%	86.81%
a=0.4,d=8	0.06%	0.33%	1.91%	3.76%	93.94%
a=0.5,d=2	1.11%	2.99%	7.89%	8.71%	79.30%
a=0.5,d=4	0.15%	0.99%	5.03%	7.75%	86.08%
a=0.5,d=8	0.04%	0.20%	1.58%	5.83%	92.35%
a=0.8,d=2	0.37%	1.62%	6.20%	15.57%	76.24%
a=0.8,d=4	0.01%	0.21%	2.08%	10.62%	87.08%
a=0.8,d=8	0.00%	0.01%	0.16%	6.94%	92.89%
a=1.0,d=2	0.13%	0.81%	5.17%	20.62%	73.27%
a=1.0,d=4	0.00%	0.05%	0.98%	10.81%	88.16%
a=1.0,d=8	0.00%	0.00%	0.01%	6.58%	93.41%

According to Table I, we have that our locally listing rank algorithm exactly identifies almost all the natural communities of the networks generated by the homophyly/kinship model.

For Google's Personalized pagerank, we will use the local algorithm in [4] with the teleportation constant $\alpha = 0.15$ as recommended in [12] to rank the nodes of the networks of our model from every node. We use $\mathcal{P}$ to denote this algorithm.

Table II describes the soundness of the PageRank algorithm $\mathcal{G}$ on the networks of the homophyly/kinship model.

According to Table II, we observe the following results:

(1) For three types, the fractions of queries with precision 1 are between 73% and 79.3%.

(2) For most types, the fractions of queries with precision 1 are from 80% to 89%.
(3) For 7 types, the fractions of queries with precision 1 are greater than 90%.

The results demonstrate that the PageRank algorithm $\mathcal{G}$ finds the precise answers for most queries in the networks of the homophyly/kinship model.

The results demonstrate that our algorithm is a personalised, locally listing and smart rank algorithm that precisely identifies the natural communities for almost all the queries. Our algorithm is remarkably better than the current-generation search engines on the basis of PageRank in both the time complexity and accuracy of answers for queries.

More importantly, Li, Pan and Yin (Natural rank: Locally listing detection of natural communities in networks, submitted) have shown that for a model that generates networks with natural community structure, and for a network G generated by the model,

(1) For most communities X, our locally listing rank algorithm exactly identifies X from every query input in X,
(2) (1) above holds for every network of the model, and
(3) Both (1) and (2) hold for all the existing models that generate networks with natural communities.
(4) Our locally listing algorithm exactly identifies the cancer cell types for a number of cancers.

We also showed that (1) - (4) are no longer true for the PageRank algorithms.

We thus verified that our personalized web ranking algorithm $\mathcal{E}^L$ identifies or precisely approximates most or even almost all natural or ground-truth communities of large-scale networks. Our results demonstrated that precise searching and ranking in large-scale networks can be efficiently realized, for which structural information minimization is the principle, and that the current-generation search engine based on PageRank vectors fails to precisely identify most of the natural or ground-truth communities of large-scale networks.

XIX. Conclusions and New Challenges

We proposed the notions of positioning entropy (or one-dimensional structural information), structural information (or two-dimensional structural information), and K-dimensional structural information of graphs for $K \geq 3$. The K-dimensional structural information of a graph is defined to be the minimum overall number of bits needed to determine the K-dimensional code of the node that is accessible from a step of random walk in the graph, for each $K \geq 1$. For a network G and a dimensionality K, we use $\mathcal{H}^K(G)$ to denote the K-dimensional structural information of G.

The K-dimensional structural information of a graph is both the measure of structural information and the dynamical complexity of networks.

We demonstrated that the K-dimensional structural information functions of networks satisfy a number of important properties, including: (i) network dependency, (ii) additivity, (iii) locality, (iv) dynamics, (v) essentiality, (vi) robustness, (vii) linking nature to science, (viii) local computability, (ix) incremental computability and (x) applicability. These properties ensure that K-dimensional structural information is a well-defined measure for the dynamical complexity of networks, for each K.

Our K-dimensional structural information of graphs is the first metric of structural information, solving the great challenge suggested by Shannon in 1953 [13], [48].

The K-dimensional structural information of networks provides an approach to investigate the dynamical complexity of networks, and classify the networks, solving a fundamental challenge of network theory.

A. Conclusions

Given a network G, the K-dimensional structure, $\mathcal{T}$ say, of G that minimises the K-dimensional structural information of G naturally characterises the natural K-dimensional structure of the network. Our theory demonstrates that (K-dimensional) structural information minimisation is the principle for detecting the natural K-dimensional structure of networks.

We established the theory of one- and two-dimensional structural information of graphs and networks. The theory consists of the following results:

(1) *Positioning entropy principle*
If G is either simple or with balanced weights, then
$$\mathcal{H}^1(G) = \Theta(\log n).$$
Remark: The differences between distinct graphs are thus only in the hidden constant in Θ.
(2) *Modularity principle*
For an arbitrarily given connected graph G, the modularity of G is bounded by $1 - \Phi(G)$, where $\Phi(G)$ is the conductance of G.
(3) *Basic principle of two-dimensional structural information*
For an arbitrarily given graph G,
$$\mathcal{H}^2(G) \geq \Phi(G) \cdot \mathcal{H}^1(G).$$
This establishes a useful relationship between the one- and two-dimensional structural information of all the graphs through the combinatorial object of the conductance.
(4) *Lower bound of two-dimensional structural information of graphs*
If G is either simple or with balanced weights, then
$$\mathcal{H}^2(G) = \Omega(\log \log n).$$
This gives a lower bound of the two-dimensional structural information of all the usual networks.
(5) *Structural information of classical data structures*
a) If G is a tree, then
$$\mathcal{H}^2(G) = \Theta(\log \log n).$$
b) If G is a grid graph, then
$$\mathcal{H}^2(G) = \Theta(\log \log n).$$

This principle indicates that the most fundamental data structures such as trees or grids in computer science, and in nature and society may achieve the minimum two-dimensional structural information of the data structures. The results indicate that the most important data structures for the big data in the information age should achieve the minimum structural information of new data structures. The question is hence: what are the possible candidates for the new data structures?

(6) *Structural information of the networks of the preferential attachment model*
If G is a network generated by the preferential attachment model, then
a) If $d = 1$, then with probability $1 - o(1)$,
$$\mathcal{H}^2(G) = \Theta(\log\log n).$$
b) If $d \geq 2$, then with probability $1 - o(1)$,
$$\mathcal{H}^2(G) = \Theta(\log n).$$
This shows that nontrivial networks of the preferential attachment achieve the maximum two-dimensional structural information.

(7) *Structural information of the networks of dynamical random model*
If G is a network generated by the uniform attachment model, then
a) If $d = 1$, then with probability $1 - o(1)$,
$$\mathcal{H}^2(G) = \Theta(\log\log n).$$
b) If $d \geq 2$, then with probability $1 - o(1)$,
$$\mathcal{H}^2(G) = \Theta(\log n).$$
This result indicates that nontrivial networks generated by a dynamical random procedure have the maximum two-dimensional structural information.

(8) *Structural information of networks of the small world model*
If G is a network generated by the small world model introduced by Kleinberg, then
a) If $r \geq 2$, then with probability $1 - o(1)$,
$$\mathcal{H}^2(G) = \Theta(\log\log n).$$
b) If $r < 2$, then with probability $1 - o(1)$,
$$\mathcal{H}^2(G) = \Theta(\log n).$$
This result indicates that if $r \geq 2$, then the networks of the small world model achieve the minimum structural information, in which case, the networks behave like grid-like graphs. Interestingly, the result explores a phase transition phenomenon of the structural information of the networks of the small world model from $\Omega(\log n)$ to $O(\log\log n)$.

(9) *Structural information of nature evolving*
If G is a network generated by the homophyly/kinship model with affinity exponent a and average number of edges d, then
a) If $a = 0$ and $d = 1$, then
$$\mathcal{H}^2(G) = \Theta(\log\log n).$$
b) If $a = 0$ and $d > 1$, then
$$\mathcal{H}^2(G) = \Theta(\log n).$$
c) If $0 < a < 1$, then
$$\mathcal{H}^{\mathcal{N}}(G) = \Omega(\log^{1-a} n),$$
where $\mathcal{N}$ is the natural community structure of G.
d) If $a \geq 1$, then
$$\mathcal{H}^2(G) = \Theta(\log\log n).$$
This result explores a new principle of the networks naturally evolved in nature and society. The result demonstrates that naturally evolving networks may provide new data structures for computer science in the information age.

(10) *Black hole principle*
Given a connected network $G = (V, E)$ with weight function W, there is a black hole in G if and only if $\mathcal{H}^2(G) = o(\log\log n)$.
This principle indicates that our theory of structural information may have implications in physics.

B. New Challenges

Our theory here also partially solves or points to a solution to the following challenges:

1) What are the principles of self-organisation for individuals in networks?
This challenge is high level, but very interesting. Suppose that self-organisation is a natural selection and that natural selection is the controlling principle of the evolution from random variations. Then structural information minimisation is the principle of self-organisation. This implies that people fear uncertainty, and that minimisation of the uncertainty of an individual in a group is the principle for the individual to survive in a group and in the society. This result emphases that uncertainty or non-determinism of high-dimensional structures is a serious problem for a network or a society. The investigation of this challenge would solve some long standing challenges in social sciences.

2) What are the principles for networking of unstructured big data?
Networking of unstructured data is the first step of big data processing. Structural information minimisation provides the principle for this mission. A successful application of this principle has been developed in [36]. However, further investigation of the challenge is needed.

3) What are the principles determining the natural structures of real world networks?
If structural information minimisation is the principle of self-organisation of individuals, then the algorithms for minimising the structural information

would find the natural or ground truth social groups.

4) What are the principles for networking engineering?
This question is explicitly stated by Shannon [48]. Partial solution for this challenge has already been given by the current authors in [34]. The best communication systems may be designed by combining both the engineering requirements and the structural information entropy minimisation principle, for which the complete solution requires the combination of our theory and networking engineering.

Of course, each of the four challenges can only be fully resolved by separate papers following the theory in the present paper.

In addition, there are a number of interesting, but less challenging, open questions left by this research. The most interesting such questions include:

(1) Computational complexity and approximation algorithms for the problem of minimising the K-dimensional structural information of graphs.
(2) Locally testable algorithms for the two-dimensional structural information of graphs. That is to design sub-linear algorithms to test whether or not the structural information of a graph is either $\Omega(\log n)$ or $O(\log\log n)$ or far from the classes of graphs.
(3) Better applied algorithms for computing the K-dimensional structural information of graphs.
(4) Are there polynomial time algorithms on trees and grid that can be extended to the networks G with $\mathcal{H}^2(G) = O(\log\log n)$? Are the networks G with $\mathcal{H}^2(G) = O(\log\log n)$ new data structures for networking information?
The measure of our structural information is a notion between information science and computer science. The algorithmic theory of the structural information would bring the theory to practice. Questions (1) - (4) are the new algorithmic problems of the information theoretic notion.
(5) Is structural information minimisation really the principle for network security?
(6) Is structural information minimisation really the principle for the emergence and convergence of cooperation in evolutionary games in networks?
This problem is closely related to question (10) below, understanding the cooperation behaviors in the evolution of selfish individuals.
(7) What are the principles for networking of unstructured data?
Answering this question may provide new ideas for algorithms for big data. Structural information minimisation could be such a principle.
(8) Is there a new learning and deep learning theory that can be built based on the high-dimensional structural information?
(9) Is structural information minimisation a principle for networking computing systems?
This question is closely related to the principle of networking engineering above.
(10) What roles does structural information minimisation play in biology and in evolution of species? What is the relationship between structural information minimisation principle and natural selection?
Suppose that natural selection is the controlling principle of the evolution from random variations. Then by our theory, structural information minimisation is the principle of the natural selection in the evolution.
(11) Robustness of the K-dimensional structural information of the networks generated by the classical models?
This question is to prove that the K-dimensional structural information of all the networks generated by a model, the preferential attachment model say, with the same type, are almost the same, that is, the differences are within a constant bound that is independent of the size of the networks. This question allows us to develop a rich mathematical theory of the structural information of networks.
(12) To fully develop a structural information theory, a dynamical complexity theory of networks and a dynamical complexity theory of computation.
This points to new directions in information theory, network theory and computational theory based on the metric of the K-dimensional structural information of graphs. In particular, the current research suggests that it would be the new mission of information theory to distinct the certainty and the contingency of noisy data, for which our theory provides the foundation.

APPENDIX A
PROBABILISTIC TOOLS

At first, we introduce some useful probabilistic tools and inequalities.

We will use the following form of Chernoff bound.

Lemma 69 (Chernoff Bound [14]) Let $X_1, \ldots, X_n$ be independent random variables with $\Pr[X_i = 1] = p_i$ and $\Pr[X_i = 0] = 1 - p_i$. Denote the sum by $X = \sum_{i=1}^{n} X_i$ with expectation $E(X) = \sum_{i=1}^{n} p_i$. Then we have

$$\Pr[X \le E(X) - \lambda] \le \exp\left(-\frac{\lambda^2}{2E(X)}\right),$$
$$\Pr[X \ge E(X) + \lambda] \le \exp\left(-\frac{\lambda^2}{2(E(X) + \lambda/3)}\right).$$

We will use the following form of Azuma's inequality for martingales.

Lemma 70 (Azuma's inequality) Let $\boldsymbol{c} = (c_1, \ldots, c_n)$ be a vector of positive entries. Let a sequence of random variables $X_0, X_1, \ldots, X_n$ be a martingale. If it is $\boldsymbol{c}$-Lipschitz, that is, $|X_i - X_{i-1}| \le c_i$ for $i = 1, \ldots, n$, then for any $\lambda > 0$,

$$\Pr[X_n \le X_0 - \lambda] \le \exp\left(-\frac{\lambda^2}{2\sum_{i=1}^{n} c_i^2}\right),$$
$$\Pr[X_n \ge X_0 + \lambda] \le \exp\left(-\frac{\lambda^2}{2\sum_{i=1}^{n} c_i^2}\right).$$

We will use the following form of supermartingale inequality.

Lemma 71 (Supermartingale Inequality, [16, Th. 2.40]): For a filter $\{0,\Omega\} = \mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_n = \mathcal{F}$, suppose that a non-negative random variable X_i is $\mathcal{F}_i$-measurable for $0 \le i \le n$. Let B be the bad set associated with the following admissible conditions: (that is, the set of events that the conditions fail to hold.)

$$\begin{aligned} E(X_i|\mathcal{F}_{i-1}) &\le X_{i-1}, \\ Var(X_i|\mathcal{F}_{i-1}) &\le \sigma_i^2 + \phi_i X_{i-1}, \\ X_i - E(X_i|\mathcal{F}_{i-1}) &\le a_i + M, \end{aligned}$$

where σ_i, ϕ_i, a_i and M are non-negative constants. Then we have

$$\begin{aligned} &\Pr(X_n \ge X_0 + \lambda) \\ &\le \exp\left(-\frac{\lambda^2}{2(\sum_{i=1}^n(\sigma_i^2+a_i^2)+(X_0+\lambda)(\sum_{i=1}^n\phi_i)+M\lambda/3)}\right) \\ &\quad + \Pr(B). \end{aligned}$$

The following fact will also be very useful in our proofs.

Fact 72: For any real x,

$$\frac{1}{x+1} \le \log\left(1+\frac{1}{x}\right) \le \frac{1}{x}.$$

Proof: Note that $1+y \le e^y$ holds for all real y. The fact is obtained by replacing y with $-\frac{1}{x+1}$ and $\frac{1}{x}$, respectively. □

The following expansion of power series is folklore.

Fact 73: For any $u > 0$ and $|x| \le 1$,

$$\begin{aligned} (1\pm x)^u = \; & 1 \pm ux + \frac{u(u-1)}{2!}x^2 \pm \frac{u(u-1)(u-2)}{3!}x^3 \\ & + \cdots + (-1)^m\frac{u(u-1)\cdots(u-m+1)}{m!}x^m + \cdots. \end{aligned}$$

Appendix B
Proof of Theorem 48

Proof of Theorem 48: For (1). By the construction of G, the expectation of $|C_t|$ is

$$E[|C_t|] = \; 2 + \sum_{i=3}^{t}\frac{1}{\log^a i}.$$

By indefinite integral

$$\int\left(\frac{1}{\log^a x} - \frac{a}{\log^{a+1} x}\right)dx = \frac{x}{\log^a x} + C,$$

we know that if $t \ge T_1$ is large enough, then

$$\begin{aligned} \sum_{i=3}^{t}\frac{1}{\log^a i} &\le 1 + \int_2^t \frac{1}{\log^a x}dx \\ &\le \int_2^t \frac{6}{5}\left(\frac{1}{\log^a x} - \frac{a}{\log^{a+1} x}\right)dx \\ &\le \frac{4t}{3\log^a t}, \end{aligned}$$

where $\frac{6}{5}$ and $\frac{4}{3}$ are chosen arbitrarily among the numbers larger than 1. Similarly,

$$\begin{aligned} \sum_{i=3}^{t}\frac{1}{\log^a i} &\ge \int_2^t \frac{1}{\log^a x}dx \\ &\ge \int_2^t \left(\frac{1}{\log^a x} - \frac{a}{\log^{a+1} x}\right)dx \\ &\ge \frac{3t}{4\log^a t}. \end{aligned}$$

By the Chernoff bound, since $t \ge T_1$, with probability $1 - exp(-\Omega(\frac{t}{\log^a t})) = 1 - o(n^{-1})$, we have $\frac{t}{2\log^a t} \le |C_t| \le \frac{2t}{\log^a t}$. By the union bound, such an inequality holds for all $t \ge T_1$ with probability $1 - o(1)$.

For (2). By the construction of G, the expectation of $|S|$ at time step t is

$$E(|S|) = 1 + \sum_{i=t_S+1}^{t}\left(1 - \frac{1}{\log^a i}\right)\cdot\frac{1}{|C_i|}.$$

Before going on the proofs, we define the following:

Definition 74: Let $\mathcal{E}$ be the event that: for all $i \ge T_1$, $\frac{i}{2\log^a i} \le |C_i| \le \frac{2i}{\log^a i}$.

We will use this event several times throughout the proofs.

By (1), we know that $\mathcal{E}$ holds with probability $1 - o(1)$. Thus, at time step t,

$$\begin{aligned} E(|S|) &= \Theta\left(\sum_{i=t_S}^{t}\left(1 - \frac{1}{\log^a t}\right)\cdot\frac{\log^a t}{t}\right) \\ &= \Theta\left(\int_{t_S}^{t}\frac{\log^a x}{x}dx\right) \\ &= \Theta(\log^{a+1} t - \log^{a+1} t_S). \end{aligned}$$

For (3). It suffices to prove that with probability $1 - o(n^{-1})$, the homochromatic set of the first color κ has size $O(\log^{a+1} n)$. Then the result follows from the union bound.

Let S_κ be the set of nodes share color κ.

Conditioned on the event $\mathcal{E}$, in Definition 74, for large enough n,

$$\begin{aligned} E(|S_\kappa|) = \; & 1 + \sum_{i=3}^{n}\left(1 - \frac{1}{\log^a i}\right)\cdot\frac{1}{|C_i|} \\ &\le T_1 + \sum_{i=T_1+1}^{n}\left(1 - \frac{1}{\log^a i}\right)\cdot\frac{2\log^a i}{i} \\ &\le \; 3\log^{a+1} n. \end{aligned}$$

By the Chernoff bound,

$$\Pr[|S_\kappa| > 4\log^{a+1} n] = o(n^{-1}).$$

Therefore, with probability $1 - o(n^{-1})$, the size of S_κ is at most $4\log^{a+1} n$.

For (4). We need to bound the number of global edges with one endpoint in S.

For $t \ge t_S$, define $S[t]$ to be the snapshot of S at time step t, and $\partial(S)[t]$ to be the set of edges from $S[t]$ to $\overline{S[t]}$, the complement of $S[t]$. So $\partial(S)[t]$ is in fact the set of global edges of S at time step t and $g_S = |\partial(S)[n]|$. Denote by

$D(S)[t]$ the total degree of nodes in (volume of) $S[t]$. In our proof, we first give a recurrence for the expected value of $D(S)[t]$ at any time step $t > t_S$, and then show that $\partial(S)[n]$ is not expectedly too many.

By the construction of G, the recurrence on $D(S)[t]$ can be written as

$$\begin{aligned} &E[D(S)[t] \mid D(S)[t-1]] \\ &\quad = D(S)[t-1] + \frac{1}{\log^a t} \cdot \frac{D(S)[t-1]}{2d(t-1)} \cdot d \\ &\qquad + \left(1 - \frac{1}{\log^a t}\right) \cdot \frac{2d}{|C_{t-1}|}. \end{aligned} \tag{77}$$

We suppose the event $\mathcal{E}$ that for all $t \geq T_1 = \log^{a+1} n$, $\frac{t}{2\log^a t} \leq |C_t| \leq \frac{2t}{\log^a t}$, which almost surely holds by (1). It also holds for $t \geq T_2$ for sufficiently large n. On this condition,

$$\begin{aligned} &E(D(S)[t] \mid D(S)[t-1], \mathcal{E}) \\ &\quad \leq D(S)[t-1]\left[1 + \frac{1}{2(t-1)\log^a t}\right] \\ &\qquad + \frac{4d\log^a t}{t}. \end{aligned} \tag{78}$$

Taking expectation on both sides, we have

$$\begin{aligned} E(D(S)[t]) \leq E(D(S)[t-1])\left[1 + \frac{1}{2(t-1)\log^a t}\right] \\ + \frac{4d\log^a t}{t}. \end{aligned} \tag{79}$$

Then we analyze this recurrence for the cases of $a \geq 1$ and $a < 1$, respectively.

When $a \geq 1$, since for sufficiently large n and thus for sufficiently large t with $t \geq t_S \geq T_2$, we have

$$\begin{aligned} &9d\log^{a+1}(t+1) - \left[1 + \frac{1}{2(t-1)\log^a t}\right] \cdot 9d\log^{a+1} t \\ &\quad \geq 9d\log^a t \log\frac{t+1}{t} - \frac{9d\log t}{2(t-1)} \\ &\quad \geq \frac{9d\log^a t}{t+1} - \frac{9d\log^a t}{2(t-1)} \\ &\quad \geq \frac{4d\log^a t}{t}, \end{aligned} \tag{80}$$

where the second inequality follows from Fact 72. Applying it to Inequality (79), we have

$$\begin{aligned} &E(D(S)[t]) - 9d\log^{a+1}(t+1) \\ &\leq \left[1 + \frac{1}{2(t-1)\log^a t}\right] \cdot (E(D(S)[t-1]) - 9d\log^{a+1} t). \end{aligned}$$

Recursively, we have that

$$\begin{aligned} E(D(S)[t]) \leq \theta_t \cdot [E(D(S)[t_S]) \\ - 9d\log^{a+1}(t_S+1)] + 9d\log^{a+1}(t+1) \end{aligned}$$

holds for all $t_S < t \leq n$, where

$$\theta_t = \prod_{i=t_S+1}^{t}\left[1 + \frac{1}{2(i-1)\log^a i}\right].$$

Note that $E(D(S)[t_S]) = d$. So

$$\begin{aligned} E(D(S)[t]) \leq 9d\log^{a+1}(t+1) - \theta_t \\ \cdot [9d\log^{a+1}(t_S+1) - d]. \end{aligned} \tag{81}$$

When $0 < a < 1$, since for sufficiently large n and thus for sufficiently large t,

$$\begin{aligned} &\left[1 + \frac{1}{2(t-1)\log^a t}\right] \cdot 9d\log^{2a} t - 9d\log^{2a}(t+1) \\ &\quad = \frac{9d\log^a t}{2(t-1)} - 9d \cdot [\log^{2a}(t+1) - \log^{2a} t] \\ &\quad \geq \frac{9d\log^a t}{2(t-1)} - \frac{d\log^a t}{2t} \\ &\quad \geq \frac{4d\log^a t}{t}, \end{aligned} \tag{82}$$

where the first inequality follows from the fact that $\log(t+1) - \log t = \log\left(1 + \frac{1}{t}\right) \leq \frac{1}{t}$ and $a < 1$,

$$\begin{aligned} &\lim_{t\to\infty} \frac{\log^{2a}(t+1) - \log^{2a} t}{\frac{\log^a t}{t}} \\ &\quad = \lim_{t\to\infty} t \cdot \left[\frac{\log^a(t+1)}{\log^a t} - 1\right] \cdot (\log^a(t+1) + \log^a t) \\ &\quad \leq \lim_{t\to\infty} t \cdot \left[\frac{\log(t+1)}{\log t} - 1\right] \cdot (\log^a(t+1) + \log^a t) \\ &\quad \leq \lim_{t\to\infty} t \cdot \frac{\log(t+1) - \log t}{\log t} \cdot 2\log^a(t+1) \\ &\quad \leq \lim_{t\to\infty} \frac{2\log^a(t+1)}{\log t} = 0. \end{aligned}$$

Applying Inequality (82) to (79), we have

$$\begin{aligned} &E(D(S)[t]) + 9d\log^{2a}(t+1) \\ &\quad \leq \left[1 + \frac{1}{2(t-1)\log^a t}\right] \cdot (E(D(S)[t-1]) + 9d\log^{2a} t). \end{aligned}$$

Recursively, we have that

$$\begin{aligned} &E(D(S)[t]) \\ &\quad \leq \theta_t \cdot [E(D(S)[t_S]) + 9d\log^{2a}(t_S+1)] - 9d\log^{2a}(t+1) \end{aligned}$$

holds for all $t_S < t \leq n$, and so

$$\begin{aligned} &E(D(S)[t]) \\ &\quad \leq \theta_t \cdot (9d\log^{2a}(t_S+1) + d) - 9d\log^{2a}(t+1). \end{aligned} \tag{83}$$

Note that by the construction of G,

$$\begin{aligned} E(g_S) &= \sum_{t=t_S}^{n} \frac{1}{\log^a t} \cdot \frac{E(D(S)[t])}{2d(t-1)} \cdot d \\ &= \sum_{t=t_S}^{n} \frac{1}{\log^a t} \cdot \frac{E(D(S)[t])}{2(t-1)}. \end{aligned} \tag{84}$$

Next, we will bound $E(g_S)$ by using Inequalities (81) and (83) for different values of a.

When $a \geq 1$, we have

$$E(g_S) \leq \sum_{t=t_S}^{n} \frac{9d\log^{a+1}(t+1) - \theta_t \cdot [9d\log^{a+1}(t_S+1) - d]}{2(t-1)\log^a t}.$$

Since $\theta_t > 1$, for sufficient large n, we have

$$\begin{aligned}
&E(g_S)\\
&\quad\le \sum_{t=t_S}^{n} \frac{9d\log^{a+1}(t+1)-[9d\log^{a+1}(t_S+1)-d]}{2(t-1)\log^a t}\\
&\quad= \frac{9d}{2}\left[\sum_{t=t_S}^{n}\frac{\log t}{t-1}-\left[\log^{a+1}(t_S+1)-\frac{1}{9}\right]\right.\\
&\qquad\left.\cdot\sum_{t=t_S}^{n}\frac{1}{(t-1)\log^a t}\right]\\
&\quad\le\ 5d\cdot\left(\int_{t_S}^{n}\frac{\log x}{x}dx-\log^{a+1}t_S\int_{t_S}^{n}\frac{1}{x\log^a x}dx\right).
\end{aligned}$$

If $a > 1$, then

$$\begin{aligned}
&E(g_S)\\
&\quad\le\ 5d\cdot\left[\frac{1}{2}(\log^2 n-\log^2 t_S)-\frac{\log^{a+1}t_S}{1-a}\right.\\
&\qquad\left.\cdot(\log^{1-a}n-\log^{1-a}t_S)\right]\\
&\quad=\ 5d\log^2 n\ \cdot\left[\frac{1}{2}-\left(\frac{1}{2}+\frac{1}{a-1}\right)\left(\frac{\log t_S}{\log n}\right)^2\right.\\
&\qquad\left.+\frac{1}{a-1}\left(\frac{\log t_S}{\log n}\right)^{a+1}\right]\\
&\quad=\ 5d\log^2 n\ \cdot\left[\frac{1}{2}-\frac{a+1}{2(a-1)}\left(1-\frac{b\log\log n}{\log n}\right)^2\right.\\
&\qquad\left.+\frac{1}{a-1}\left(1-\frac{b\log\log n}{\log n}\right)^{a+1}\right].
\end{aligned}$$

By Fact 73,

$$\begin{aligned}
&\left(1-\frac{b\log\log n}{\log n}\right)^{a+1}\\
&\quad\le\ 1-\frac{(a+1)b\log\log n}{\log n}+\frac{(a+1)ab^2(\log\log n)^2}{2\log^2 n}.
\end{aligned}$$

Thus,

$$\begin{aligned}
&E(g_S)\\
&\le\ 5d\log^2 n\ \cdot\left[\frac{1}{2}-\frac{a+1}{2(a-1)}\left(1-\frac{2b\log\log n}{\log n}+\frac{b^2(\log\log n)^2}{\log^2 n}\right)\right.\\
&\quad\left.+\frac{1}{a-1}\left(1-\frac{(a+1)b\log\log n}{\log n}+\frac{(a+1)ab^2(\log\log n)^2}{2\log^2 n}\right)\right]\\
&=\frac{5}{2}db^2(a+1)(\log\log n)^2.
\end{aligned}$$

(4) (i) follows.

If $a = 1$, then

$$\begin{aligned}
&E(g_S)\\
&\le\ 5d\cdot\left(\int_{t_S}^{n}\frac{\log x}{x}dx-\log^2 t_S\int_{t_S}^{n}\frac{1}{x\log x}dx\right)\\
&=\ 5d\left[\frac{1}{2}(\log^2 n-\log^2 t_S)-\log^2 t_S\cdot(\log\log n-\log\log t_S)\right]\\
&=\ 5d\left[\frac{1}{2}(\log^2 n-\log^2 t_S)-\log^2 t_S\right.\\
&\qquad\left.\cdot\log\left(1+\frac{b\log\log n}{\log n-b\log\log n}\right)\right]\\
&\le\ 5d\left[\frac{1}{2}(\log^2 n-\log^2 t_S)-\log^2 t_S\cdot\frac{b\log\log n}{\log n}\right]\\
&=\ 5d\left[\frac{1}{2}\log^2 n-\frac{1}{2}(\log n-b\log\log n)^2\right.\\
&\qquad\left.-(\log n-b\log\log n)^2\cdot\frac{b\log\log n}{\log n}\right]\\
&=\ 5d\left[\frac{3}{2}b^2(\log\log n)^2-\frac{(b\log\log n)^3}{\log n}\right]\\
&\le\ 8db^2(\log\log n)^2.
\end{aligned}$$

(4) (ii) follows.

When $a < 1$, applying Inequality (83) to (84), we have

$$\begin{aligned}
&E(g_S)\\
&\le\sum_{t=t_S}^{n}\frac{\theta_t\cdot(9d\log^{2a}(t_S+1)+d)-\ 9d\log^{2a}(t+1)}{2(t-1)\log^a t}\\
&\le\frac{9d}{2}\cdot\left[\sum_{t=t_S}^{n}\frac{\theta_n\log^{2a}t_S}{(t-1)\log^a t}-\sum_{t=t_S}^{n}\frac{log^{2a}(t+1)}{(t-1)\log^a t}\right]\\
&=\frac{9d}{2}\cdot\left(\theta_n\log^{2a}t_S\cdot\int_{t_S}^{n}\frac{1}{x\log^a x}dx\right.\\
&\quad\left.-\int_{t_S}^{n}\frac{log^a x}{x}dx\right)+O\left(\frac{1}{n}\right)\\
&=\frac{9d}{2}\cdot\left(\theta_n\log^{2a}t_S\cdot\frac{\log^{1-a}n-\log^{1-a}t_S}{1-a}\right.\\
&\quad\left.-\frac{\log^{1+a}n-\log^{1+a}t_S}{1+a}\right)+O\left(\frac{1}{n}\right)\\
&=\frac{9d\theta_n}{2(1-a)}\log^{1-a}n\log^{2a}t_S-\frac{9d}{2}\left(\frac{\theta_n}{1-a}-\frac{1}{1+a}\right)\\
&\quad\cdot\log^{1+a}t_S-\frac{9d}{2(1+a)}\log^{1+a}n+O\left(\frac{1}{n}\right)\\
&=\frac{9d\theta_n}{2(1-a)}\log^{1+a}n\left(1-\frac{b\log\log n}{\log n}\right)^{2a}\\
&\quad-\frac{9d}{2}\left(\frac{\theta_n}{1-a}-\frac{1}{1+a}\right)\log^{1+a}n\cdot\left(1-\frac{b\log\log n}{\log n}\right)^{1+a}\\
&\quad-\frac{9d}{2(1+a)}\log^{1+a}n+O\left(\frac{1}{n}\right)\\
&=\frac{9d\theta_n}{2(1-a)}\log^{1+a}n\cdot\left[1-\frac{2ab\log\log n}{\log n}+\frac{2a(2a-1)}{2}\right.\\
&\quad\left.\left(\frac{b\log\log n}{\log n}\right)^2+O\left(\frac{\log\log n}{\log n}\right)^3\right]-\frac{9d}{2}\cdot\\
&\quad\left(\frac{\theta_n}{1-a}-\frac{1}{1+a}\right)\log^{1+a}n\cdot\left[\ 1-\frac{(1+a)b\log\log n}{\log n}\right.\\
&\quad\left.+\frac{(a+1)a}{2}\left(\frac{b\log\log n}{\log n}\right)^2+O\left(\frac{\log\log n}{\log n}\right)^3\right]\\
&\quad-\frac{9d}{2(1+a)}\log^{a+1}n+O\left(\frac{1}{n}\right)
\end{aligned}$$

$$= \left[\frac{9d\theta_n}{2(1-a)} - \frac{9d}{2}\left(\frac{\theta_n}{1-a} - \frac{1}{1+a}\right) - \frac{9d}{2(1+a)}\right] \cdot \log^{1+a} n + \left[-\frac{9d}{2(1-a)} \cdot \ 2ab + \frac{9d}{2}\left(\frac{\theta_n}{1-a} - \frac{1}{1+a}\right) \cdot (1+a)b\right] \cdot \log^a n \ \log\log n \ + O\left[\frac{(\log\log n)^2}{\log^{1-a} n}\right]$$

$$= \frac{9}{2}db(\theta_n - 1)\log^a n \log\log n + O\left[\frac{(\log\log n)^2}{\log^{1-a} n}\right].$$

To deal with the factor $(\theta_n - 1)$, we need the following lemma.

Lemma 75: For sufficiently large n,

$$\theta_n - 1 \le \frac{b\log\log n}{\log^a n}.$$

Note that by the above lemma, for sufficiently large n

$$E(g_S) \le \frac{9}{2}db \cdot \frac{b\log\log n}{\log^a n}\log^a n \log\log n + O\left[\frac{(\log\log n)^2}{\log^{1-a} n}\right] \le \ 5db^2(\log\log n)^2,$$

and hence (4) (iii) follows.

To complete the proof, we prove the lemma, i.e., Lemma 75.

Proof: Recall that

$$\theta_n = \prod_{i=t_S+1}^{n}\left[1 + \frac{1}{2(i-1)\log^a i}\right].$$

Then:

$$\log\theta_n = \sum_{i=t_S+1}^{n}\log\left[1 + \frac{1}{2(i-1)\log^a i}\right] \le \sum_{i=t_S+1}^{n}\frac{1}{2(i-1)\log^a i} \le \frac{1}{2}\int_{t_S}^{n}\frac{1}{x\log^a x} = \frac{1}{2(1-a)}\cdot(\log^{1-a} n - \log^{1-a} t_S) = \frac{\log^{1-a} n}{2(1-a)}\cdot\left[1 - \left(1 - \frac{b\log\log n}{\log n}\right)^{1-a}\right] = \frac{\log^{1-a} n}{2(1-a)}\cdot\left[(1-a)\cdot\frac{b\log\log n}{\log n} - \frac{(1-a)(-a)}{2}\cdot\left(\frac{b\log\log n}{\log n}\right)^2 + O\left(\frac{\log\log n}{\log n}\right)^3\right] = \frac{b\log\log n}{2\log^a n} + O\left[\frac{(\log\log n)^2}{\log^{1+a} n}\right].$$

Thus, for sufficiently large n, $\log\theta_n \le \frac{3b\log\log n}{4\log^a n}$, which implies that

$$\theta_n \le (\log n)^{\frac{3b}{4\log^a n}}.$$

A key observation is that, for any constant c, by l'Hôpital's rule,

$$\lim_{n\to\infty}\frac{(\log n)^{\frac{c}{\log^a n}} - 1}{\frac{\log\log n}{\log^a n}} = \lim_{y\to\infty}\frac{y^{\frac{c}{y^a}} - 1}{\frac{\log y}{y^a}} = \lim_{y\to\infty}\frac{\left(y^{\frac{c}{y^a}} - 1\right)'}{\left(\frac{\log y}{y^a}\right)'} = \lim_{y\to\infty}\frac{c(1 - a\log y)}{y^{1+a-\frac{c}{y^a}}}\cdot\frac{y^{1+a}}{1 - a\log y} = \lim_{y\to\infty} c\cdot y^{\frac{c}{y^a}} = \lim_{y\to\infty} c\cdot e^{\frac{c\log y}{y^a}} = \ c.$$

Thus, for any $\epsilon > 0$, if n is large enough, then

$$\theta_n - 1 \le \frac{3b}{4}(1+\epsilon)\cdot\frac{\log\log n}{\log^a n}.$$

Let $\epsilon = \frac{1}{3}$, then the lemma follows. □

This completes the proof of Theorem 48. □

ACKNOWLEDGMENT

The authors are grateful to Professor Andrew Lewis-Pye from London School of Economics for reading the paper, and for helpful suggestions when he was visiting the first author in Chinese Academy of Sciences. Both authors are grateful to Professor Mingshu Li for his long term support when he was the Director of the Institute. The authors would like to thank Professor Yucheng Li for his support for the execution of the project in the institute.

REFERENCES

[1] A. A. Alizadeh, "Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling," *Nature*, vol. 403, no. 6769, pp. 503–511, 2000.

[2] K. Anand and G. Bianconi, "Entropy measures for networks: Toward an information theory of complex topologies," *Phys. Rev. E*, vol. 80, no. 4, p. 045102, 2009.

[3] K. Anand, G. Bianconi, and S. Severini, "Shannon and von Neumann entropy of random networks with heterogeneous expected degree," *Phys. Rev. E*, vol. 83, no. 3, p. 036109, 2011.

[4] R. Andersen, F. Chung, and K. Lang, "Local graph partitioning using PageRank vectors," in *Proc. FOCS*, 2006, pp. 475–486.

[5] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, pp. 509–512, Oct. 1999.

[6] G. Bianconi, "Entropy of network ensembles," *Phys. Rev. E*, vol. 79, no. 3, p. 036114, 2009.

[7] G. Bianconi, P. Pin, and M. Marsili, "Assessing the relevance of node features for network structure," *Proc. Nat. Acad. Sci. USA*, vol. 106, no. 28, pp. 11433–11438, 2009.

[8] A. Blum, T.-H. H. Chan, and M. R. Rwebangira, "A random-surfer Web-graph model," in *Proc. ANALCO*, 2006, pp. 238–246.

[9] B. Bollobás and O. Riordan, "The diameter of a scale-free random graph," *Combinatorica* vol. 24, no. 1, pp. 5–34, 2004.

[10] D. Bonchev and N. Trinajstić, "Information theory, distance matrix, and molecular branching," *J. Chem. Phys.*, vol. 38, no. 10, pp. 4517–4533, 1977.

[11] S. L. Braunstein, S. Ghosh, and S. Severini, "The Laplacian of a graph as a density matrix: A basic combinatorial approach to separability of mixed states," *Ann. Combinat.*, vol. 10, no. 3, pp. 291–317, 2006.

[12] S. Brin and L. Page, "The anatomy of a large-scale hypertextual Web search engine," *Comput. Netw. ISDN Syst.*, vol. 30, nos. 1–7, pp. 107–117, 1998.

[13] F. P. Brooks, Jr., "Three great challenges for half-century-old computer science," *J. ACM*, vol. 50, no. 1, pp. 25–26, 2003.

[14] H. Chernoff, "A note on an inequality involving the normal distribution," *Ann. Probab.*, vol. 9, no. 3, pp. 533–535, 1981.

[15] Y. Choi and W. Szpankowski, "Compression of graphical structures," in *Proc. IEEE Int. Symp. Inf. Theory*, Jun./Jul. 2009, pp. 364–368.

[16] F. Chung and L. Lu, *Complex Graphs and Networks*. Providence, RI, USA: AMS, 2006.

[17] A. Clauset, M. E. J. Newman, and C. Moore, "Finding community structure in very large networks," *Phys. Rev. E*, vol. 70, no. 6, p. 066111, 2004.

[18] C. Darwin, *On the Origin of Species by Means of Natural Selection, Or, The Preservation of Favoured Races in the Struggle for Life*. London, U.K.: John Murray, 1859.

[19] M. Dehmer, "Information processing in complex networks: Graph entropy and information functionals," *Appl. Math. Comput.*, vol. 201, pp. 82–94, Jul. 2008.
[20] M. Dehmer and F. Emmert-Streib, "Towards network complexity," in *Complex Sciences* (Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering), vol. 4, J. Zhou, Ed. Berlin, Germany: Springer, 2009, pp. 707–714.
[21] M. Dehmer and A. Mowshowitz, "A history of graph entropy measures," *Inf. Sci.*, vol. 181, no. 1, pp. 57–78, 2011.
[22] P. Erdõs and A. Rényi, "On random graphs I," *Publ. Math. Inst. Hungar. Acad. Sci.*, vol. 6, pp. 290–297, 1959.
[23] P. Erdõs and A. Rényi, "On the evolution of random graphs," *Publ. Math. Inst. Hungar. Acad. Sci.*, vol. 5, pp. 17–61, 1960.
[24] A. D. Flaxman, "Expansion and lack thereof in randomly perturbed graphs," *Internet Math.*, vol. 4, nos. 2–3, pp. 131–147, 2007.
[25] J. Kleinberg, "The small-world phenomenon: An algorithmic perspective," in *Proc. 32nd ACM Symp. Theor. Comput.*, 2000, pp. 163–170.
[26] E. Konstantinova and A. A. Paleev, "Sensitivity of topological indices of polycyclic graphs," *Vychisl. Sistemy*, vol. 136, pp. 38–48, Jan. 1990.
[27] S. Fortunato, "Community detection in graphs," *Phys. Rep.*, vol. 486, pp. 75–174, Feb. 2010.
[28] R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins, "Trawling the Web for emerging cyber-communities," in *Proc. 8th Int. Conf. WWW*, 1999, pp. 403–416.
[29] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal, "Stochastic models for the Web graph," in *Proc. FOCS*, 2000, pp. 57–65.
[30] J. Leskovec, J. Kleinberg, and C. Faloutsos, "Graphs over time: Densification laws, shrinking diameters and possible explanations," in *Proc. ACM KDD*, 2005, pp. 177–187.
[31] A. Li, "Homophyly/kinship model: Naturally evolving networks," *Sci. Rep.*, vol. 5, p. 15140, Oct. 2015.
[32] A. Li, J. Li, and Y. Pan, "Discovering natural communities in networks," *Phys. A, Statist. Mech. Appl.*, vol. 436, pp. 878–896, Oct. 2015.
[33] A. Li, X. Li, Y. Pan, and W. Zhang, "Strategies for security of networks," *Sci. China Inf. Sci.*, vol. 58, no. 1, pp. 012107-1–012107-14, 2015.
[34] A. Li and Y. Pan, "A theory of network security: Principles of natural selection and combinatorics," *Internet Math.*, accepted.
[35] A. Li and P. Peng, "Community structures in classical network models," *Internet Math.*, vol. 7, no. 2, pp. 81–106, 2011.
[36] A. Li, X. Yin, and Y. Pan, "Three-dimensional gene map of cancer cell types: Structural entropy minimisation principle for defining tumour subtypes," *Sci. Rep.*, vol. 6, p. 20412, Feb. 2016.
[37] T. Łuczak, "Sparse random graphs with a given degree sequence," in *Proc. Symp. Random Graphs*, Poznan, Poland, 1989, pp. 165–182.
[38] M. Mihail, C. Papadimitriou, and A. Saberi, "On certain connectivity properties of the Internet topology," in *Proc. 44th Annu. IEEE Symp. Found. Comput. Sci.*, Oct. 2003, pp. 28–35.
[39] M. Molloy and B. Reed, "A critical point for random graphs with a given degree sequence," *Random Struct. Algorithms*, vol. 6, nos. 2–3, pp. 161–180, 1995.
[40] A. Mowshowitz, "Entropy and the complexity of graphs: I. An index of the relative complexity of a graph," *Bull. Math. Biophys.*, vol. 30, no. 1, pp. 175–204, 1968.
[41] C. R. Munteanu, "Markov entropy centrality: Chemical, biological, crime, and legislative networks," in *Towards an Information Theory of Complex Networks*, M. Dehmer, F. Emmert-Streib, and A. Mehler, Eds. New York, NY, USA: Springer, 2011, pp. 199–258.
[42] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Phys. Rev. E*, vol. 69, no. 2, p. 026113, 2003.
[43] B. Pittel, "Note on the heights of random recursive trees and random m-ary search trees," *Random Struct. Algorithms*, vol. 5, no. 2, pp. 337–347, 1994.
[44] N. Rashevsky, "Life, information theory, and topology," *Bull. Math. Biophys.*, vol. 17, no. 3, pp. 229–235, 1955.
[45] E. Ravasz and A.-L. Barabási, "Hierarchical organization in complex networks," *Phys. Rev. E*, vol. 67, p. 056104, Feb. 2003.
[46] C. Raychaudhury, S. K. Ray, J. J. Ghosh, and A. B. Basak, "Discrimination of isomeric structures using information theoretic topological indices," *J. Comput. Chem.*, vol. 5, no. 6, pp. 581–588, 1984.
[47] M. Rosvall and C. T. Bergstrom, "Maps of random walks on complex networks reveal community structure," *Proc. Nat. Acad. Sci. USA*, vol. 105, no. 4, pp. 1118–1123, 2008.
[48] C. Shannon, "The lattice theory of information," *IEEE Trans. Inf. Theory*, vol. 1, no. 1, pp. 105–107, Feb. 1953.
[49] E. Trucco, "A note on the information content of graphs," *Bull. Math. Biophys.*, vol. 18, no. 2, pp. 129–135, 1956.
[50] A. Vázquez, "Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations," *Phys. Rev. E*, vol. 67, p. 026112, May 2005.
[51] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, pp. 440–442, Jun. 1998.

Angsheng Li is a research professor of Institute of Software, Chinese Academy of Sciences. He was born in 1964. He got bachelor in Mathematics in Yunnan Normal University in 1984, and ph D in 1993 in Institute of Software, Chinese Academy of Sciences. He has been working for the Institute of Software, Chinese Academy of Sciences since 1993 after he finished his ph D. From 1998 to 2002, he was a visitor and research fellow in the University of Leeds, UK, working with Professor Barry Cooper (an academic descendant of Alan Turing) in Computability Theory. In 2003, he was awarded the Distinguished Young Investigator award of the National Natural Science Foundation of China. In 2008, he was selected by the Hundred Talent Program of Chinese Academy of Sciences. From 2008 to 2009, he was a visiting scientist in Computer Science Department, Cornell University, US, working with Professor Juris Hartmanis (the founder of Computational Complexity Theory). His research areas include Computability Theory, Computational Theory, Network Theory and Information Science. His current interests focus on network algorithms, and network information theory.

Yicheng Pan is now an assistant professor of Institute of Software, Chinese Academy of Sciences. He was born in 1983. He got ph D under the supervision of Prof. Angsheng Li in Institute of Software, Chinese Academy of Science in 2012. He finished post-doctoral study in Institute of Information Engineering, Chinese Academy of Sciences in 2015. His research interests include Computation Complexity, Local Algorithms and Network Theory.