RWE: A Random Walk Based Graph Entropy for the Structural Complexity of Directed Networks

Chong Zhang, Cheng Deng, Luoyi Fu, *Member, IEEE*, Xinbing Wang, *Senior Member, IEEE*, Guihai Chen, *Fellow, IEEE*, Lei Zhou, and Chenghu Zhou

***Abstract*—This paper studies a graph entropy measure to characterize the structural complexity of directed networks. Since the von Neumann entropy (VNE) has found applications in many tasks by networked data, but suffers from a high computational complexity in computing the Laplacian spectrum, we aim to propose a simple yet effective alternative method. Considering that local nodal interactions collectively induce the information flow of the entire graph, we are motivated to use the probability flow of random walks as the proxy of information flow in real-world digraphs, and correspondingly propose a random walk based entropy (RWE) to measure the average information that the random walker reaches each node. Inspired by the close relation between Laplacian spectrum and the Perron vector, we prove that RWE serves as a good approximation for the VNE in digraphs with a guaranteed entropy gap. This approximation is further applied to a digraph similarity measure based on the Jensen-Shannon divergence. Therefore, RWE exhibits interpretability, scalability and capability of well capturing the structural complexity of digraphs as empirically verified. We further extend RWE to two dimensions by incorporating community structures, which characterizes information flow both between and within communities. By proving that the difference between the one-dimensional and two-dimensional RWE reflects the extent to which the community structure is preserved, we convert the community detection problem into the minimization of two-dimensional RWE, and design a greedy algorithm. Various experiments confirm the superiority of our approach over the baselines.**

***Index Terms*—Directed graph, graph entropy, random walk, structural complexity, von Neumann entropy.**

Manuscript received 6 April 2023; revised 4 December 2023; accepted 6 December 2023. Date of publication 5 January 2024; date of current version 23 February 2024. This work was supported in part by the National Key R&D Program of China under Grant 2023YFB3107401, in part by the National Natural Science Foundation of China under Grants 62161160337, 62132011, U21B2018, U20A20177, and 62206217, and in part by the Shaanxi Province Key Industry Innovation Program under Grants 2023-ZDLGY-38 and 2021ZD LGY01-02. Recommended for acceptance by Prof. Chao Gao. *(Corresponding author: Xinbing Wang.)*

Chong Zhang is with the School of Computer Science and Technology, Xi'an Jiaotong University, Xi'an Shaanxi 710049, China (e-mail: zhangchong@xjtu.edu.cn).

Cheng Deng, Luoyi Fu, Xinbing Wang, and Guihai Chen are with the School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong Univesity, Shanghai 200240, China (e-mail: davendw@sjtu.edu.cn; yiluofu@sjtu.edu.cn; xwang8@sjtu.edu.cn; gchen@cs.sjtu.edu.cn).

Lei Zhou is with the School of Oceanography, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: zhoulei1588@sjtu.edu.cn).

Chenghu Zhou is with the State Key Laboratory of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100045, China (e-mail: zhouch@ireis.ac.cn).

This article has supplementary downloadable material available at https://doi.org/10.1109/TNSE.2023.3342197, provided by the authors.

Digital Object Identifier 10.1109/TNSE.2023.3342197

I. Introduction

GRAPH is the most ubiquitous structure to illustrate relationships in massive real-world data. Social networks, communication networks, flight networks and citation networks are just a few examples that can be represented by graphs. One salient issue in graph analysis is to measure the complexity of graphs [1], [2], referring to the way in which nodes and links are arranged or the level of organization of the structural features such as degree distribution, graph spectrum, etc. In order to characterize the inherent structural complexity of graphs, many entropy based measures have been proposed [2], [3], [4], [5], [6], each of which is a specific form of Shannon entropy for different types of distribution extracted from graph structures.

A. Motivations

However, the existing measures are primarily designed for simple undirected graphs, which cannot be applied to the measurement of more complex structures with directionality as well as communities. Recall the aforementioned scenarios like social networks, flight networks and etc., there is not a solitary exception. As an alternative, the von Neumann entropy (VNE), which is built on the Laplacian spectrum, is a natural way to capture the graph complexity since the Laplacian spectra is well-known to incorporate rich information about the multi-scale structure of graphs, thus spawning tremendous great interests to researchers [7], [8], [9], [10], [11], [12], [13]. In 2005, Chung extended the Laplacian matrix to directed graphs [14], rendering the possibility of VNE in digraphs. Despite the popularity received, VNE suffers from a high computational complexity of graph spectrum, which is cubic in network size, and is hard to interpret by intuitive structural patterns. Although there emerge a bunch of approximation approaches [10], [11], [12], [13] later on, most of them are mainly for undirected graphs, which cannot cope with directed case for the asymmetric directionality. Therefore, *there is a strong desire to design an efficient, effective and interpretable measure for the structural complexity of digraphs.*

To this end, this paper aims to circumvent the computational demanding graph spectra and seek for an alternative method that can capture rich information of graph structures. Note that in many realistic networks individuals frequently interact with their neighbors in various forms (e.g., hyperlinks between WWW, likes or comments on tweets in social networks). As a result, the underlying structural patterns of the graphs can be reflected through those interactions that collectively induce

the information flow of the entire graphs. This implies that if we can continuously sample these local interactions, the entire structural features can be gradually obtained. A very typical existing method based on this idea belongs to the structural information (SI) [2], where the approach of random *calls* are launched to collect the average information that each node is the receiver of calls. Based on this principle, SI is defined as the Shannon entropy for degree distribution in undirected graphs and in-degree distribution for directed counterpart. Although such geometric features works well in terms of higher efficiency than graph spectrum, it still does not suffice to well capture the structural features of digraphs by merely the in-degree distribution. Imagine that if each node specifies a fixed number of in-links, the value of SI will keep fixed no matter how these links are connected regularly or randomly.

B. Contributions

The efficiency of SI nevertheless drives us to remedy the mechanism to better suit directed cases. In doing so, we note that the random *calls* involved in SI only consider the receivers of information, ignoring the process of how the information transfers from the senders *all the way* to the receivers. As a remedy, we use a random *walker* instead to record all the directed paths along which the information flows from senders to the receivers. Thus the random walk serves as a proxy of information flow, and accordingly we propose a random walk based digraph complexity measure, called RW entropy (RWE) to represent the average information that the random walker reaches each node. Motivated by the close relation between Perron vector [14] and the graph spectrum, we prove that our RWE turns out to be a fairly good approximation for the VNE in digraphs with a difference range only between 0 and $\frac{\log_2 e}{\varphi} \cdot \mathrm{tr}(M^2)$, where φ is the minimum positive component of the Perron vector, and $\mathrm{tr}(\mathrm{M}^2)$ regards to out-degrees, Perron vector and the direction of links. As we will also provably demonstrate in later sections (Sections IV and VI), RWE is interpretable, scalable and sufficient simultaneously. Based on this approximation, RWE is further applied to an effective and efficient digraph similarity measure.

Next, we further enrich RWE in terms of its description of information flow by incorporating community structures in digraphs. Compared to the one-dimensional counterpart, the definition of two-dimensional RWE embodies both the intra-community and inter-community probability flow of random walks. As we will show in Section V, our designed two-dimensional RWE is able to retain unique codewords for each community while reusing the codewords of nodes within different communities, thus incurring fewer bits in the description of information flow than that in one-dimensional case. In this way, a larger difference between the one and two dimensional RWEs means a closer partition structure to the true underlying communities, allowing us to equivalently convert the community detection problem into a compressing problem, where the partition corresponding to the minimum two-dimensional RWE is exactly the communities we find.

In a nutshell, we make the following contributions.

- *Theory and interpretability:* We propose a new random walk based digraph complexity measure, called the RW entropy. As far as we know, we for the first time approximate the von Neumann entropy in directed graphs with provable accuracy, interpretability, scalability and sufficiency for capturing the structural complexity of digraphs.
- *Applications and efficient algorithm:* Using RWE as a proxy of VNE in digraphs, we develop a new digraph distance measure based on Jensen-Shannon divergence with both efficiency and effectiveness, and design an incremental algorithm to compute distance between adjacent graphs in a graph stream.
- *Extension by considering community structure:* We further extend RWE to two-dimensional structural complexity of digraphs, hence converting the community detection problem into a compressing problem, and propose an efficient global greedy merge algorithm to approximate the minimization of the two-dimensional RWE, whose results can be refined by a local movement policy.
- *Extensive experiments and evaluations:* Numerical results on real-world networks and synthetic networks confirm the superior performance of RWE in the following aspects: i) characterizing the structural complexity of digraphs, ii) approximating von Neumann entropy, and iii) identifying community structures in digraphs.

Organization: The rest of this paper is organized as follows. We review two related issues in Section II. In Section III, we introduce the random walk, von Neumann entropy and structural information of directed graphs. Based on the random walk, we propose one-dimensional RW entropy in Section IV. We not only prove its approximation accuracy with von Neumann entropy, but also present the application on graph distance. We extend the RW entropy to characterize two-dimensional structures in Section V. Section VI provides experimental results, and we conclude the paper and list the directions for future work in Section VII.

II. Related Works

A. Graph Entropy

To capture structural complexity of digraphs, Rashevsky [3] applied the automorphism based entropy measure to digraphs in the following form: For a connected graph G with n nodes,

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

where n_i is the number of topologically equivalent nodes in the i-th node orbit of G, and k is the number of different orbits, hence this definition is only concerned with the symmetries of graphs. Raychaudhury et al. [4] proposed a local measure of graph entropy which defines the distance complexity of node i as follows,

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

where $d(i,j)$ is the distance between i and j in G, and $d(i)$ is the sum of $d(i,j)$ for all j. Li and Pan [2] defined the structural information of digraphs as the Shannon entropy of in-degree distribution, but has limitations in well capturing the structural features of digraphs as we mentioned. Based on the Laplacian spectrum, the von Neumann entropy has recently found applications of complex network, but the main obstacle is the computational inefficiency, and also the lack of interpretability. Although Ye et al. [15] extended the von Neumann entropy in digraphs and tried to approximate it, the quadratic approximation to the Shannon entropy function and the approximation of Perron vector they adopt make their relative error of approximation high. Besides, they still do not address the issue of interpretability. Despite the effort to overcome the computational inefficiency of VNE [10], [11], [12], [13], as far as we know, most of them were designed for undirected graphs except [15]. Among them, Liu et al. [13] bridged the gap between structural information and VNE in undirected graphs with a provable accuracy. However, we will show that such a conclusion does not hold in digraphs, but their thoughts really inspired our work.

B. Community Detection

Complex networks have been used to model connections of individuals in various domains, which brought emerging issues on network science, such as link prediction [16], information dissemination [17], source detection [18] and so on. Among them, one salient problem is the recognition of community structures. In the last decades, the community detection problem has attracted huge attention [19], [20], [21], [22], [23], [24], which plays an indispensable role in a wide range of applications, such social ties analysis [25], functional studies [26] and healthcare [27]. However, only a limited number of the existing works are aimed for clustering in digraphs. For instance, Arenas et al. [28] proposed a topological approach based on *modularity* optimization, which is defined as the sum of the total weight of all links in each module minus the expected weight. Rosvall and Bergstrom [29] proposed classic Infomap to identify communities via a map of information flow. A more detailed introduction to Infomap can be referred to Section 3 in Supplementary Material. As we will show in Sections V and VI, although our method is also based on information flow, we actually capture a different information of random walks, and the performance is even better than Infomap in some real-world datasets.

III. Preliminaries

In this paper, we study the simple directed network $G(V,E,A)$ of order n with positive edge weights, where E is the set of edges with cardinality $|E|=m$, and $A \in \mathcal{R}_+^{n\times n}$ is the weight matrix with positive entry A_{ij}. The in-degree of node $i \in G$ (resp. out-degree), denoted by d_i^{in} (resp. d_i^{rmout}), is defined to be the sum of weights of edges that point to (resp. from) i, that is $d_i^{\text{in}} = \sum_{j=1}^{n} A_{ji}$ (resp. $d_i^{rmout} = \sum_{j=1}^{n} A_{ij}$). For convenience, we list the key parameters that will be used late in Table I.

TABLE I
Notation and Definition

Notation	Definition
$G(V,E,A)$	simple directed network with weight matrix A
n	number of nodes
m	number of edges
d_i^{in} (d_i^{out})	in-degree (out-degree) of node i
P	transition matrix of random walk
$\boldsymbol{\phi}$	Perron vector of P
L	combinatorial Laplacian matrix of G
$\text{vol}(G)$	volume of G (sum of edge weight)
$\mathcal{H}_{rw}^1(G)$	one-dimensional RWE of G
$\mathcal{H}_{vn}(G)$	von Neumann entropy of G
$\mathcal{H}_{SI}^1(G)$	structural information of G
$\Delta\mathcal{H}(G)$	digraph entropy gap ($\mathcal{H}_{rw}^1(G) - \mathcal{H}_{vn}(G)$)
$\mathcal{H}_{rw}^{\mathcal{P}}(G)$	the RWE of G by $\mathcal{P}$
$\mathcal{H}_{rw}^2(G)$	Two-dimensional RWE of G

A. Random Walk on Digraphs

The random walk on weighted digraphs is defined by a transition probability matrix P, where $P(u,v)$ denotes the probability that a walker moves from a node u to a neighbor v, such that

$$P(u,v) = \frac{A_{u,v}}{\sum_z A_{uz}}. \tag{1}$$

Obviously, for an unweighted graph G, $P(u,v)$ is reduced to $1/d_u^{rmout}$ if (u,v) is an edge. According to the Perron-Frobenius Theorem [30], the transition probability matrix P of a strongly connected and aperiodic directed graph has a unique left eigenvector $\boldsymbol{\phi}$ with $\boldsymbol{\phi}(v) > 0$ for all v, which satisfies $\boldsymbol{\phi}P = \boldsymbol{\phi}$. We can normalize and choose $\boldsymbol{\phi}$ to satisfy $\sum_v \boldsymbol{\phi}(v) = 1$, which is called the Perron vector [14] of P. Unlike the undirected graph, where $\boldsymbol{\phi}$ is consistent with the node degree distribution, there is no closed form for $\boldsymbol{\phi}$ for general digraphs. Nevertheless, there are polynomial-time algorithms to evaluate $\boldsymbol{\phi}$ efficiently.

If G is strongly connected and aperiodic, the random walk will converge to the unique stationary distribution ϕ, i.e. the Perron vector. However, when the digraph has one or more absorbing sets, where each of them is a vertex set that a random walk enters but can never go out, the stationary distribution is not unique. In this case, from the view of Markov chain, researchers often introduce teleportations in the random walk as a modification to make the Markov chain primitive. For example, in PageRank algorithm proposed by Page and Brin [31], a positive probability α of teleporting to any node uniformly in each step is introduced, which is 0.15 by experience. Initially, the state is a uniform distribution, then the random walk will converge to the unique stationary distribution.

B. Von Neumann Entropy and Structural Information

Originally, the von Neumann entropy was defined on the density matrix ρ of a quantum mechanical system [32]. Braunstein et al. [33] proposed to use the graph Laplacian to map graphs to quantum states, so they introduced the density matrix of a graph as the combinatorial Laplacian of the graph normalized to have unit trace. For a digraph G with transition

probability matrix P, and the Perron vector $\boldsymbol{\phi}$, if we let $\Phi = \text{diag}[\boldsymbol{\phi}(1), \boldsymbol{\phi}(2), \ldots, \boldsymbol{\phi}(n)]$, then the combinatorial Laplacian matrix [14] of G is defined as

$$L = \Phi - \frac{\Phi P + P^T \Phi}{2}, \tag{2}$$

and its density matrix is that $\rho = L/\text{tr}(L)$.

Let $\{\lambda_i\}_{i=1}^n$ be the sorted eigenvalues of L, which is the Laplacian spectrum of G. When G is strongly connected and aperiodic, the transition matrix P is exactly defined in (1), and we can easily derive that $\text{tr}(L) = \sum_{i=1}^n \lambda_i = \text{tr}(\Phi) = 1$. However, this does not hold for general digraphs due to the introduction of teleportations. For the convenience of delineation, we define the function $f(x) \triangleq x \log_2 x$ on the support $[0, \infty)$ where $f(0) \triangleq \lim_{x \to 0} f(x) = 0$. We present formal definitions of VNE and SI of digraphs as below.

Definition 1 (von Neumann entropy of digraph): The von Neumann entropy of a digraph G is defined by $\mathcal{H}_{vn}(G) = -\sum_{i=1}^n f(\lambda_i/\text{tr}(L))$, where $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_n = 0$ are the eigenvalues of the combinatorial Laplacian matrix $L = \Phi - \frac{\Phi P + P^T \Phi}{2}$ of the graph G.

Definition 2 (Structural information of digraph): The structural information of a digraph $G = (V, E, A)$ is defined as $\mathcal{H}_{SI}(G) = -\sum_{i=1}^n f(d_i^{\text{in}}/\text{vol}(G))$, where $\text{vol}(G) = \sum_{i=1}^n \sum_{j=1}^n A_{ij}$ denotes the volume of G. Particularly, $\text{vol}(G) = m$ for an unweighted graph G.

IV. One-Dimensional Structural Complexity

In this section, we introduce one-dimensional RWE, and show that RWE is a good proxy of VNE in digraphs with provable accuracy.

A. Introduction to One-Dimensional RWE

We commence by giving a brief overview of the intuition of SI. Imagine that messages can be delivered between nodes through edges. A *call* is a flow of message from a sender i to a receiver j, where $(i, j) \in E$, and an exogenous process is launched to continuously collect such calls uniformly at random. Hence, at any moment, the probability that a node v is the message receiver is $d_v^{\text{in}}/\text{vol}(G)$. The authors [2] encode the network based on this probability distribution, and take the average number of bits as a measure of the structural information of digraphs. However, this is different from the random walk where the receiver of a call is the sender of the next call, thus cannot fully capture the structural complexity of a digraph. To this end, we use random walk as a proxy of the entire information flow in digraphs and present the definition of random walk based structural complexity of digraphs, which we call the RW entropy (RWE) as follows.

Definition 3 (One-dimensional RWE of a Directed Network): Given a connected and weighted digraph $G = (V, E, A)$, suppose a random walker with transition matrix P on G, and $\mathbf{P} = [p_1, p_2, \ldots, p_n]$ is the ergodic node visit frequency vector within the random walk, the one-dimensional random walk based graph entropy of G is defined as:

$$\mathcal{H}_{rw}^1(G) = -\sum_{i=1}^n p_i \log_2 p_i. \tag{3}$$

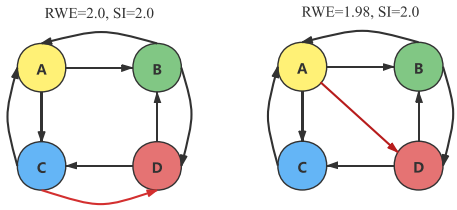


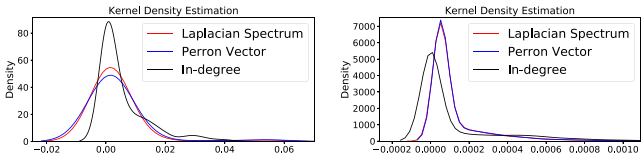
Fig. 1. Simple illustration on the sufficiency of RWE.

Although the definition is based on a connected graph, the RWE of disconnected digraph can be derived based on the additivity of Shannon entropy function. Note that for strongly connected and aperiodic digraphs, P is well defined in (1), which guarantees a unique stationary distribution. While for general digraphs, due to the existence of dangling nodes or sinks, we modify P with a teleportation parameter $\alpha = 0.15$ corresponding to the well known PageRank transition matrix by Page and Brin [31], in which case, $\mathbf{P}$ is also called the PageRank vector. We then illustrate that RWE exhibits interpretability, scalability and sufficiency for capturing the structural complexity of digraphs, respectively.

- *Interpretability:* RWE measures the average information that each node is traversed by the random walk. Compared to VNE, which is hard to interpret by simple structural patterns, RWE provides a more intuitive interpretation of the structural complexity by tracking the heterogeneity of the information flow induced by local nodal interactions.
- *Scalability:* The computational efficiency of $\mathbf{P}$ can benefit from various efficient algorithms of power iteration such as Arnoldi method [34] and Poor man's algorithm [35], whose complexity are $O(nk(k+l))$ and $O(tn^2)$ respectively, where the orders of the parameters k, l and t are much less than the network size n, making RWE more efficient than VNE ($O(n^3)$) and suitable for large-scale directed networks.
- *Sufficiency:* As a proxy of the information flow, the random walk can record all the directed paths along which the information flows from senders to the receivers. As a result, RWE is sufficient to capture the slight structural change of even a single edge direction, since the pattern of information flow will change eventually. As the simple illustration in Fig. 1, the only difference from the left graph to the right is the change of edge from $C \to D$ to $A \to D$, resulting in a regular structural pattern changing to a more heterogeneous one, which can be captured by RWE but not SI. More demonstration can be found in Section VI.

B. Approximation for Von Neumann Entropy in Digraphs

To prove that RWE is a good approximation for the VNE in digraphs, we first define the difference between RWE and VNE as *digraph entropy gap* that $\Delta\mathcal{H}(G) = \mathcal{H}_{rw}^1(G) - \mathcal{H}_{vn}(G)$.



(a) Lake (b) Wiki-Vote

Fig. 2. Close distribution between Perron vector and Laplacian spectrum in two real-world directed networks.

To strictly bound $\Delta\mathcal{H}(G)$, we are inspired by the fact that there is a close relation between Perron vector and Laplacian spectrum. As illustrated in Fig. 2, the kernel density estimation curve of Perron vector almost coincides with that of Laplacian spectrum in two representative real-world digraphs, while the in-degrees separate from them, which indicates that our RWE may be a good approximation of VNE. Based on that, we provide the additive approximation errors in Theorem 1.

Theorem 1 (Approximation Accuracy): For any strongly connected and aperiodic directed graph $G = (V, E, A)$ with combinatorial Laplacian matrix L defined in (2), we denote $M = (\Phi P + P^T\Phi)/2$ and $E_{\rightleftarrows} = \{(u,v)|(u,v) \in E \text{ and } (v,u) \in E\}$ as the bidirectional edge set such that $E_{\rightleftarrows} \subseteq E$, then the inequality

$$0 \leq \Delta\mathcal{H}(G) \leq \frac{\log_2 e}{\varphi} \cdot \mathrm{tr}(M^2), \tag{4}$$

holds, where

$$\mathrm{tr}(M^2) = \frac{1}{2}\left[\sum_{(u,v)\in E} \frac{\boldsymbol{\phi}^2(u)}{{d_u^{rmout}}^2} + \sum_{(u,v)\in E_{\rightleftarrows}} \frac{\boldsymbol{\phi}(u)\cdot\boldsymbol{\phi}(v)}{d_u^{rmout}\cdot d_v^{rmout}}\right], \tag{5}$$

and $\varphi = \min\{\boldsymbol{\phi}(i) \,|\, \boldsymbol{\phi}(i) > 0\}$.

Before proving Theorem 1, we first introduce an important conception that will be used.

Lemma 1 (Majorization [36]): Given two vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$, we denote by $\boldsymbol{x}^{\downarrow}, \boldsymbol{y}^{\downarrow} \in \mathbb{R}^d$ with the same components, but sorted in descending order. We say that $\boldsymbol{x}$ majorizes $\boldsymbol{y}$, written as $\boldsymbol{x} \succ \boldsymbol{y}$, if and only if $\sum_{i=1}^{k} x_i^{\downarrow} \geq \sum_{i=1}^{k} y_i^{\downarrow}$ for $k = 1, \ldots, d$, and $\sum_{i=1}^{d} x_i = \sum_{i=1}^{d} y_i$. Moreover, the following conditions are equivalent: (i) $\boldsymbol{x} \succ \boldsymbol{y}$; (ii) $Q\boldsymbol{x} = \boldsymbol{y}$ for some doubly stochastic matrix Q.

The intuition of majorization is that if $\boldsymbol{x}$ majorizes $\boldsymbol{y}$, then the entries of $\boldsymbol{y}$ are more mixed than those of $\boldsymbol{x}$. Based on the Schur-Horn theorem [36], the diagonal elements of the Hermitian matrix are majorized by its eigenvalues. Recall that for strongly connected and aperiodic graph G, $\mathrm{tr}(L) = 1$, so we have $\mathcal{H}_{vn}(G) = -\sum_{i=1}^{n} f(\lambda_i)$. The combinatorial Laplacian matrix L defined in (2) is positive semidefinite symmetric whose diagonal elements form the Perron vector $\boldsymbol{\phi}$ and eigenvalues form the spectrum $\boldsymbol{\lambda}$. Therefore, we obtain $\boldsymbol{\lambda} \succ \boldsymbol{\phi}$, indicating that there exists some doubly stochastic matrix $Q = (q_{ij}) \in [0,1]^{n\times n}$ such that $Q\boldsymbol{\lambda} = \boldsymbol{\phi}$. Based on that, we now proceed to prove Theorem 1.

Proof of Theorem 1: To begin with, we define a discrete random variable X_i with probability mass function $\sum_{j=1}^{n} q_{ij}\delta_{\lambda_j}(x)$ for each node i, where $\delta_a(x)$ is the Kronecker delta function. Then the expectation of X_i is $\mathbb{E}[X_i] = \sum_{j=1}^{n} q_{ij}\lambda_j = \boldsymbol{\phi}(i)$, and the variance $\mathrm{var}(X_i) = \mathbb{E}[{X_i}^2] - \mathbb{E}^2[X_i] = \sum_{j=1}^{n} q_{ij}{\lambda_j}^2 - \boldsymbol{\phi}^2(i)$. In the following, we first show that $\Delta\mathcal{H}(G)$ is non-negative, and then prove $\Delta\mathcal{H}(G) \leq \frac{\log_2 e}{\varphi} \cdot \mathrm{tr}(M^2)$, as well as deriving the expression $\mathrm{tr}(M^2)$.

By expressing the digraph entropy gap in terms of Perron vector sequence and Laplacian spectrum, we have

$$\Delta\mathcal{H}(G) = \mathcal{H}_{rw}^{1}(G) - \mathcal{H}_{vn}(G) = \sum_{i=1}^{n} f(\lambda_i) - \sum_{i=1}^{n} f(\boldsymbol{\phi}(i)).$$

Since $f''(x) = (\log_e)/x > 0$, $f(x) = x\log_2(x)$ is convex. Then, based on Jensen's equality, we have that $f(\boldsymbol{\phi}(i)) = f(\mathbb{E}[X_i]) \leq \mathbb{E}[f(X_i)]$ for any $i \in \{1, \ldots, n\}$. By summing over i, we have

$$\sum_{i=1}^{n} f(\boldsymbol{\phi}(i)) \leq \sum_{i=1}^{n} \mathbb{E}[f(X_i)] = \sum_{i=1}^{n}\sum_{j=1}^{n} q_{ij} f(\lambda_j) = \sum_{i=1}^{n} f(\lambda_i).$$

Hence, $\Delta\mathcal{H}(G) \geq 0$ is proved for any directed graphs, which means RWE is always no less than VNE.

Before we proceed to prove the upper bound for $\Delta\mathcal{H}(G)$, we present an important technique in the following lemma.

Lemma 2 (Jensen's gap [37]): Let X be a one-dimensional random variable with mean μ and support (a, b), where $-\infty \leq a < b \leq \infty$. Let $\psi(x)$ be a twice differentiable function on (a, b), and define the function $l(x) = \frac{\psi(x)-\psi(\mu)}{(x-\mu)^2} - \frac{\psi'(\mu)}{x-\mu}$, then $E(\psi(X)) - \psi(E(X)) \leq \sup_{x\in(a,b)}\{l(x)\} \cdot \mathrm{var}(X)$. Additionally, if $\psi'(x)$ is convex, then $l(x)$ is monotonically increasing in x, and if $\psi'(x)$ is concave, then $l(x)$ is monotonically decreasing in x.

Next, we use Jensen's gap to prove $\Delta\mathcal{H}(G) \leq \frac{\log_2 e}{\varphi} \cdot \mathrm{tr}(M^2)$. Applying the Jensen's gap to X_i and $f(x)$,

$$\mathbb{E}[f(X_i)] - f(\mathbb{E}[X_i]) \leq \sup_{x\in[0,1]}\{l_i(x)\} \cdot \mathrm{var}(X_i), \tag{6}$$

where,

$$l_i(x) = \frac{f(x) - f(\mathbb{E}[X_i])}{(x - \mathbb{E}[X_i])^2} - \frac{f'(\mathbb{E}[X_i])}{x - \mathbb{E}[X_i]}.$$

Due to the concavity of $f'(x)$, $l_i(x)$ is monotonically decreasing in x, then we have

$$\mathbb{E}[f(X_i)] - f(\mathbb{E}[X_i]) \leq l_i(0) \cdot \mathrm{var}(X_i) \leq \frac{\log_2 e}{\varphi} \cdot \mathrm{var}(X_i).$$

Then, (6) can be simplified as

$$\sum_{j=1}^{n} q_{ij} f(\lambda_j) - f(\boldsymbol{\phi}(i)) \leq \frac{\log_2 e}{\varphi} \cdot \left(\sum_{j=1}^{n} q_{ij}{\lambda_j}^2 - \boldsymbol{\phi}^2(i)\right) \tag{7}$$

By summing both sides of the (7) over i, we obtain an upper bound on $\Delta\mathcal{H}(G)$ as

$$\Delta\mathcal{H}(G) \leq \frac{\log_2 e}{\varphi} \cdot \sum_{i=1}^{n} \left(\sum_{j=1}^{n} q_{ij}{\lambda_j}^2 - \phi^2(i) \right) = \frac{\log_2 e}{\varphi} \cdot \left(\mathrm{tr}(L^2) - \mathrm{tr}(\Phi^2) \right).$$

Since $L = \Phi - (\Phi P + P^T\Phi)/2$, we then simplify $\mathrm{tr}(L^2) - \mathrm{tr}(\Phi^2)$ by the cyclic property of the trace as follows.

$$\begin{aligned} &\mathrm{tr}(L^2) - \mathrm{tr}(\Phi^2) \\ &= -\mathrm{tr}(\Phi^2 P + \Phi P^T\Phi)/2 - \mathrm{tr}(\Phi P\Phi + P^T\Phi^2)/2 \\ &\quad + \mathrm{tr}(\Phi P\Phi P + \Phi P P^T\Phi + P^T\Phi\Phi P + P^T\Phi P^T\Phi)/4 \\ &= -\mathrm{tr}(\Phi^2 P) - \mathrm{tr}(\Phi^2 P^T) + \mathrm{tr}(\Phi P P^T\Phi)/2 + \mathrm{tr}(\Phi P\Phi P)/2 \\ &= \mathrm{tr}(\Phi^2 P P^T)/2 + \mathrm{tr}(\Phi P\Phi P)/2 \\ &= \mathrm{tr}(M^2). \end{aligned}$$

To continue the development, denote $E_{\rightleftarrows} = \{(u,v)|(u,v) \in E \text{ and } (v,u) \in E\}$ as the bidirectional edge set. According to the definition of the transition matrix P and the diagonal matrix Φ we obtain

$$\mathrm{tr}(\Phi^2 P P^T) = \sum_{u\in V}\sum_{v\in V} P_{uv}^2 \cdot \phi^2(u) = \sum_{(u,v)\in E} \frac{\phi^2(u)}{{d_u^{rmout}}^2}.$$

Similarly, $\mathrm{tr}(\Phi P\Phi P) = \sum_{(u,v)\in E_{\rightleftarrows}} \frac{\phi(u)\cdot\phi(v)}{d_u^{rmout}\cdot d_v^{rmout}}$.

Therefore, $\Delta\mathcal{H}(G) \leq \frac{\log_2 e}{\varphi} \cdot \mathrm{tr}(M^2)$. □

Analysis of upper bound: The expression of the derived upper bound regards to the Perron vector, out degrees and edge directionality (whether the edges are bidirectional or not). Generally, the stationary distribution or Perron vector of general directed graphs has no closed form or a specific distribution. However, on some typical digraph structures such as bidirectional digraph, Eulerian digraph and regular digraph, the bounds in Theorem 1 indeed yield tightness, as stated in Corollaries 1, 2, and 3.

Corollary 1: For any unweighted bidirectional digraph G, $0 \leq \Delta\mathcal{H}(G) \leq \log_2 e$ holds.

Proof: In unweighted bidirectional digraph G, $\boldsymbol{\phi}(u) = d_u / \sum_{u\in G} d_u$ for any $u \in G$, where d_u is either the out-degree or in-degree of u. Denote by $|E| = m$, then $\varphi \geq 1/m$, and it can be easily derived from (5) that $\mathrm{tr}(M^2) = 1/m$. Therefore, $0 \leq \Delta\mathcal{H}(G) \leq m\log_2 e \cdot 1/m = \log_2 e$. □

Since any undirected graph can be viewed as a bidirectional digraph, the conclusion here is consistent with that in [13]. Therefore, our results not only generalize the result in [13], but also cover more complex and asymmetric cases.

Corollary 2: For any Eulerian digraph G, where the in-degree of each node is equal to its out-degree, the inequality $0 \leq \Delta\mathcal{H}(G) \leq \log_2 e \cdot (\frac{1}{2} + \frac{|E_{\rightleftarrows}|}{2m})$ holds, where $m = |E|$, $|E_{\rightleftarrows}|$ denotes the number of bidirectional edges.

Proof Sketch: For an Eulerian digraph G, its Perron vector $\boldsymbol{\phi}$ is proportional to the out-degree sequence [14], i.e., $\boldsymbol{\phi}(v) = d_v^{rmout}/m$, $\varphi \geq 1/m$. □

Corollary 3: For any unweighted regular digraph G with degree d, $0 \leq \Delta\mathcal{H}(G) \leq \frac{\log_2 e}{d}$ holds.

Proof Sketch: In any unweighted regular digraph G with degree d, $\boldsymbol{\phi}(v) = 1/n$ for any $v \in G$, $\mathrm{tr}(M^2) = 1/nd$. □

The detailed proofs of Corollaries 2 and 3 can be referred to Sections 1.1 and 1.2 in Supplementary Material.

Denote by UB the derived upper bound of entropy gap. Due to the specific form of Perron vector in bidirectional digraphs, we derive that $\mathtt{UB} \leq \log_2 e$, which is a tight constant bound regardless of network categories and sizes. Also, a tight bound can be seen in Eulerian digraph such that $\log_2 e/2 \leq \mathtt{UB} \leq \log_2 e$, where the right equality is obtained since the bidirectional digraph is a special Eulerian digraph in which $|E_{\rightleftarrows}| = m$. The upper bound is further sharpened for regular digraph, such that $\mathtt{UB} = \log_2 e/d$.

Although our result is established on the strongly connected and aperiodic digraphs, we will empirically show in Section VI that if G is not strongly connected or aperiodic, RWE is still a fairly good approximation of the von Neumann entropy.

C. Application: Digraph Similarity Measure

As a measure of the structural complexity of graphs, VNE has been applied in a variety of applications, including measuring the importance of an edge, the distance between graphs for graph classification and anomaly detection [7], [10]. We note that in these applications, VNE is primitively used to address the basic task of graph similarity measure.

Entropy based graph similarity measure aims to compare two graphs from a graph sequence using the Jensen-Shannon divergence (JSD). The JSD between the probability distributions P and Q is defined by $JS(P,Q) = H((P+Q)/2) - H(P)/2 - H(Q)/2$, where $H(P) = -\sum_x P(x)\log_2 P(x)$ is the entropy of distribution P.

The square root of $JS(P,Q)$ is known as Jensen-Shannon distance, which has been proved to be a bounded metric on the space of distributions. However, to measure the similarity between high objects such as matrices or graphs, Majtey et al. [38] extend the definition to a quantum Jensen-Shannon divergence, so that the connection with VNE is built as follows.

Definition 4 (Quantum Jensen-Shannon Divergence between Digraphs): The quantum Jensen-Shannon divergence between two weighted, directed graphs $G_1 = (V, E_1, A_1)$ and $G_2 = (V, E_2, A_2)$ on the same node set V is defined as

$$QJS(G_1, G_2) = \mathcal{H}_{vn}(\bar{G}) - \mathcal{H}_{vn}(G_1)/2 - \mathcal{H}_{vn}(G_2)/2,$$

where $\bar{G}$ denotes their average graph with the same node set V such that $\bar{L}/\mathrm{tr}(\bar{L}) = L_1/(2\mathrm{tr}(L_1)) + L_2/(2\mathrm{tr}(L_2))$.

Based on the quantum Jensen-Shannon divergence, we can compute $\sqrt{QJS(G_i, G_{i+1})}$ for $1 \leq i \leq k-1$ in a stream of digraphs $\{G_i = (V, E_i, t_i)\}_{i=1}^{k}$ where t_i is the timestamp of the graph G_i, thus measuring the adjacent graph distance. However, due to high complexity of VNE, $\sqrt{QJS(G_i, G_{i+1})}$ is computationally expensive. We thereby propose a new graph distance measure based on RWE as bellow.

Algorithm 1: RWE-JS (Inc.).

Input: A graph G_k and ΔG_k
Output: RWE-JS$_{\text{incre}}(G_k, G_{k+1})$
1: $m = E(G_k)$
2: $d^{\text{in}} \leftarrow$ the in-degree sequence of G_k
3: $\mathcal{H}_{SI}(G_k) = \log_2 m - \sum_{i=1}^{n} f(d_i^{\text{in}})/m$
4: ΔV_k = set(edge[1] for edge in $E_{+,k} \cup E_{-,k}$)
5: $\Delta d^{\text{in}} \leftarrow$ the modified in-degree sequence of ΔV_k
6: $\Delta m = |E_{+,k}| - |E_{-,k}|$
7: Compute $\mathcal{H}_{SI}(G_{k+1})$ and $\mathcal{H}_{SI}(\hat{G}_k)$ via Lemma 3
8: RWE-JS$_{\text{incre}}(G_k, G_{k+1}) =$
$\sqrt{\mathcal{H}_{SI}(\hat{G}_k) - (\mathcal{H}_{SI}(G_k) + \mathcal{H}_{SI}(G_{k+1}))/2}$
9: **return** RWE-JS$_{\text{incre}}(G_k, G_{k+1})$

Definition 5 (Random Walk based Entropy Distance between Two Graphs): The random walk based entropy distance between two weighted, directed graphs $G_1 = (V, E_1, A_1)$ and $G_2 = (V, E_2, A_2)$ on the same node set V is defined as

$$RWE\text{-}JS(G_1, G_2) = \sqrt{\mathcal{H}_{rw}^1(\hat{G}) - (\mathcal{H}_{rw}^1(G_1) + \mathcal{H}_{rw}^1(G_2))/2}, \tag{8}$$

where $\hat{G}$ denotes the average graph with the same node set V such that $\Phi(\hat{G}) = (\Phi(G_1) + \Phi(G_2))/2$.

In practice, the dynamic graph stream could be efficiently stored by the sequential graph changes $\{\Delta G_k = (E_{+,k} = \bigcup_i (u_i, v_i, 1), E_{-,k} = \bigcup_i (u_i, v_i, -1), t_k)\}_{k=1}^{K-1}$ with the node set fixed, where t_k is the timestamp, $E_{+,k}$ and $E_{-,k}$ denote the set of edge insertions and the set of edge deletions during the time t_k to t_{k+1} respectively. In this scenario, to allow simple incremental update of RWE-JS, we propose to approximate the Perron vector by in-degree sequence to buy more time but at the price of cumulative approximation error. In other words, we regard $\mathcal{H}_{SI}(G)$ as a rough approximation for $\mathcal{H}_{rw}^1(G)$. Based on the original graph G_k and ΔG_k, we can incrementally and efficiently compute the approximated entropy statistics of G_{k+1} via the following lemma. (The proof can be referred to Section 1.3 in the supplementary material).

Lemma 3: Given the original graph G_k, denote d^{in} as the in-degree sequence of G_k, ΔV_k is the set of target nodes covered by $E_{+,k} \cup E_{-,k}$, and the modified in-degree sequence Δd^{in} with regard to ΔV_k, $\mathcal{H}_{SI}(G_{k+1})$ can be efficiently updated by

$$\mathcal{H}_{SI}(G_{k+1}) = \frac{f(m + \Delta m) - f(m) - c_1 + m\mathcal{H}_{SI}(G_k)}{m + \Delta m}, \tag{9}$$

where $m = E(G_k)$, $\Delta m = |E_{+,k}| - |E_{-,k}|$, and $c_1 = \sum_{i \in \Delta V_k} f(d_i^{\text{in}} + \Delta d_i^{\text{in}}) - f(d_i^{\text{in}})$. Besides, for the average graph $\hat{G}_k$ of G_k and G_{k+1}, $\mathcal{H}_{SI}(\hat{G}_k)$ can be efficiently computed by

$$\mathcal{H}_{SI}(\hat{G}_k) = -(m - c_2) f(c_3) - c_3 (f(m) - m\mathcal{H}_{SI}(G_k) - c_4) - c_5,$$

where $c_2 = \sum_{i \in \Delta V_k} d_i^{\text{in}}$, $c_3 = \frac{2m + \Delta m}{2m(m + \Delta m)}$, $c_4 = \sum_{i \in \Delta V_k} f(d_i^{\text{in}})$, and $c_5 = \sum_{i \in \Delta V_k} f(\frac{d_i^{\text{in}}}{2m} + \frac{d_i^{\text{in}} + \Delta d_i^{\text{in}}}{2(m + \Delta m)})$.

The incremental update procedure is summarized in Algorithm 1. We start by computing the directed SI of the original graph G_k (Lines 1–3), which costs $\Theta(n)$ time. Next, using the information of ΔG_k, we compute $\mathcal{H}_{SI}(G_{k+1})$ and $\mathcal{H}_{SI}(\hat{G}_k)$ via Lemma 3 (Lines 4–7), then obtain the incremental version of RWE-JS between two adjacent graphs (Line 8). The total time complexity is only $\Theta(n + |\Delta V_k|)$. Actually, for a given graph stream, we can update the adjacent graph distance via Algorithm 1 iteratively.

V. Two-Dimensional Structural Complexity

We further extend RWE to a two-dimensional description of information flow by incorporating community structures.

A. Structural Complexity of Digraph by a Partition

Suppose the node set V is partitioned into disjoint sets $\mathcal{P} = \{V_1, \ldots, V_L\}$, where V_j is a community or technically called a *module*. In this case, we divide the structural complexity of digraphs into two levels of description: the probability flow of random walks intra- and inter-communities. By summing these two parts weighted by the frequency with which it occurs, we define the *RWE by a partition* as follows.

Definition 6 (RWE of a Directed Network by a Partition): Given a connected and weighted digraph $G(V, E, A)$ with n nodes and a partition $\mathcal{P}$ of V, suppose a random walker with teleportation parameter α on G, and let p_i denote the ergodic node visit frequency at node i within the random walk, the random walk based graph entropy of G by $\mathcal{P}$ is defined as:

$$\mathcal{H}_{rw}^{\mathcal{P}}(G) = -\sum_{j=1}^{L} p_j \sum_{i=1}^{n_j} \frac{p_i}{p_j} \log_2 \frac{p_i}{p_j} - \sum_{j=1}^{L} q_j \log_2 p_j, \tag{10}$$

where q_j is the enter probability of module V_j specified as:

$$q_j = \alpha \cdot \frac{n_j}{n} \cdot \sum_{i \notin V_j} p_i + (1 - \alpha) \cdot \sum_{i \notin V_j} \sum_{k \in V_j} p_i \cdot p_{i \to k}, \tag{11}$$

and $n_j = |V_j|$ is the number of nodes in module V_j, p_j is the ergodic visit frequency at module V_j such that $p_j = \sum_{i=1}^{n_j} p_i$, and $p_{i \to k}$ is the normalized probability that the random walk transfers from node i to its outgoing neighbor k.

As stated in (10), the structural complexity of a module V_j consists of two levels: (a) from a module level, the information of the entire V_j as the arriver of the random walk, and (b) from a node level, the information of each single node $i \in V_j$ as the arriver of the random walk. The key is that we retain unique code for large-scale objects, i.e., the communities or modules to be identified within the directed network, but can reuse the codes associated with the individual nodes within each module. In other words, we can omit the module level code when the random walker keeps in the same module. For (a), the information of V_j as the arriver is $-\log_2 p_j$ with probability q_j, which is the probability that the random walker enters the module V_j, because we only need consider the walkers whose starters are not in V_j. The enter probability in (11) follows since every node outside the module V_j teleports a fraction $\alpha n_j/n$ and guides a fraction $(1 - \alpha) \sum_{k \in V_j} p_{i \to k}$ with weight p_i to nodes inside the module V_j.

For (b), the information for all nodes in V_j is $-\sum_{i=1}^{n_j} \frac{p_i}{p_j} \log_2 \frac{p_i}{p_j}$ weighted by p_j. As a result, the RWE of a digraph G by the partition $\mathcal{P}$ indeed describes the average number of bits needed to encode the information of arrivers of random walk, which fully characterizes the structural complexity of a digraph from a two-level description.

Next, based on the definitions of $\mathcal{H}_{rw}^{\mathcal{P}}(G)$ and $\mathcal{H}_{rw}^{1}(G)$, we obtain the following two basic properties of RWE in digraphs.

Theorem 2 (Locality Property): Given a connected digraph G of order n, suppose a random walker with teleportation parameter α, let $\mathcal{P}_{whole}$ denote the partition containing one module of the whole node set of G, and $\mathcal{P}_{single}$ be the partition such that each module consists of a single node of G, then

$$\mathcal{H}_{rw}^{\mathcal{P}_{single}}(G) = \left(1 - \frac{\alpha}{n}\right) \cdot \mathcal{H}_{rw}^{\mathcal{P}_{whole}}(G). \tag{12}$$

Equation (12) holds based on the fact that for the module V_j that contains a single node i, the enter probability q_j is reduced to $p_i(1-\alpha/n)$. From Theorem 2, we can derive that $\lim_{n\to\infty} \mathcal{H}_{rw}^{\mathcal{P}_{single}}(G) = \mathcal{H}_{rw}^{\mathcal{P}_{whole}}(G)$. This result indicates that if we try to minimize $\mathcal{H}_{rw}^{\mathcal{P}}(G)$ of a digraph G via an agglomerative process, which starts from the trivial partition that each module consists of a single node and merges similar nodes/communities recursively, it is nearly impossible to obtain the single module of the whole graph.

Theorem 3 (RW Entropy Gain): Given a connected digraph G with a partition $\mathcal{P} = \{V_1, \ldots, V_L\}$, we call the difference between $\mathcal{H}_{rw}^{1}(G)$ and $\mathcal{H}_{rw}^{\mathcal{P}}(G)$ the RW entropy gain such that

$$\mathcal{H}_{rw}^{1}(G) - \mathcal{H}_{rw}^{\mathcal{P}}(G) = -\sum_{j=1}^{L} (p_j - q_j) \log_2 p_j. \tag{13}$$

where p_j is the ergodic visit frequency at module V_j and q_j is the enter probability of module V_j defined in (11).

The proof of Theorem 3 can be referred to Section 1.4 in Supplementary Material. Intuitively, the right hand side in (13) describes the cumulative information that each module is the arriver of the random walk excluding the transfer part between modules. On one hand, Theorem 3 builds a bridge between one-dimensional RWE and the RWE of digraph by a partition. On the other hand, RWE gain shows the bits saved to describe the information flow in a two-level manner than we could do with one-dimensional description.

B. Community Detection in Directed Networks

So far, we have associated the structural complexity of a given structure (a set of modules or communities) with a coding quantity. Next, we will introduce a community detection method in digraphs based on the RWE measure.

Recall that we capitalize on the network structures, and the intuition is that the random walker statistically tends to spend a long time within communities. As stated in Theorem 3, (13) shows the bits saved to describe the information flow in the two-level manner description. Clearly, if communities are well separated from each other, transitions of the random walker between communities will be unfrequent, so there is a large RWE gain due to the little enter probability of each community.

Algorithm 2: Global Greedy Merge Method (GGM).

Input: A weighted directed graph $G = (V, E, A)$, ergodic visit frequency p_i for $i \in V$.
Output: A partition $\mathcal{P}$ of node set V

1 Initialize $\mathcal{P} = \{V_1, \ldots, V_n\}$ where V_i is singleton $\{v_i\}$
2 S = {}
3 **for** *all two neighboring modules* V_x, V_y *in* $\mathcal{P}$ **do**
4 Set the union of V_x, V_y as a new module V_z such that $V_z = V_x \cup V_y$
5 Record the change of RWE if merging V_x and V_y: $\Delta\mathcal{H}_{x,y}^{\mathcal{P}}(G) = \sum_{i=1}^{n_x} p_i \log_2 \frac{p_i}{p_x} + \sum_{i=1}^{n_y} p_i \log_2 \frac{p_i}{p_y} - \sum_{i=1}^{n_z} p_i \log_2 \frac{p_i}{p_z} + q_x \log_2 p_x + q_y \log_2 p_y - q_z \log_2 p_z$
6 **if** $\Delta\mathcal{H}_{x,y}^{\mathcal{P}}(G) < 0$ **then**
7 $S[(x, y)] = \Delta\mathcal{H}_{x,y}^{\mathcal{P}}(G)$
8 Find the two modules that bring the largest decrease of RWE after merging: $(x, y) = \arg\min_{(x,y)} S[(x, y)]$
9 Update $\mathcal{P} \leftarrow (\mathcal{P} \setminus \{V_x, V_y\}) \cup V_z$
10 Repeat Lines $2-9$ until no more two neighboring modules V_x, V_y such that $\Delta\mathcal{H}_{x,y}^{\mathcal{P}}(G) < 0$ holds
11 **return** $\mathcal{P}$

Instead, if the partition is not representative of the actual community structure of the graph, transitions between the clusters of the partition will be very frequent and there will be little RWE gain. Based on that, to find the community structures of digraphs, a partition with RWE gain as large as possible is expected. Moreover, Since $\mathcal{H}_{rw}^{1}(G)$ is fixed given a digraph G, herein lies the duality between community detection and the coding problem: we look for a partition $\mathcal{P}$ so as to minimize the RWE of G by $\mathcal{P}$, i.e., $\mathcal{H}_{rw}^{\mathcal{P}}(G)$, which is defined as the two-dimensional RWE of G as follows.

Definition 7 (Two-Dimensional RWE of a Directed Network): Let G be a connected directed graph, the two-dimensional random walk based graph entropy of G is defined as

$$\mathcal{H}_{rw}^{2}(G) = \min_{\mathcal{P}} \{\mathcal{H}_{rw}^{\mathcal{P}}(G)\}, \tag{14}$$

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

Therefore, the community structures in digraph G are equivalent to the modules in the partition corresponding to the two-dimensional RWE of G. However, except for small networks, it is computationally hard to check all possible partitions to find the one that minimizes $\mathcal{H}_{rw}^{\mathcal{P}}(G)$. Instead, we use computational search. By the power method, we first calculate the ergodic node visit frequency at each node. Using these visit frequencies, we explore the space of possible partitions by a global greedy merge (GGM) method, which is a kind of agglomerative process summarized in Algorithm 2. Specifically, Line 1 of the algorithm first assigns each single node to a unique module, and then the enter probability of each module can be derived in a more

TABLE II
COMPUTATIONAL COMPLEXITY COMPARISON BETWEEN GGM AND FIVE CLASSIC ALGORITHMS

Detectors	Description	Time Complexity
GGM	two-dimension RWE optimization	$O(m+n)$
infomap [29]	maps of random walk	$O(m)$
louvain [23]	two-level modularity optimization	$O(n \log n)$
CNM [20]	greedy modularity maximization	$O(n \log^2 n)$
LPA [22]	label propagation	$O(n^2 \log n)$
Danmf [39]	nonnegative matrix factorization	$O(c(n^2 r + nr^2))$

efficient way as follows:

$$q_i = \alpha \cdot \frac{n - n_j}{n} \cdot \sum_{i \in V_j} p_i + (1-\alpha) \cdot \sum_{i \in V_j} \sum_{k \notin V_j} p_i \cdot p_{i \to k},$$

where the intuition is that under the stationary state, the enter probability is equal to the exit probability for each module. In Lines 2–10, we calculate $\mathcal{H}^{\mathcal{P}}_{rw}(G)$ in (10) and repeatedly merge the two modules that give the global maximum decrease of $\mathcal{H}^{\mathcal{P}}_{rw}(G)$ until further merging brings about a larger RWE. The time complexity of this process is $O(m+n)$. For sparse networks where the order of edges is the same order as that of nodes, it is $O(n)$. While for large-scale networks, our algorithm is of complexity $O(m)$. Table II compares the computational complexity of the proposed framework GGM with 5 classic algorithms. As we can see, our framework is efficient in large-scale networks.

The result of GGM method can be further refined by a local movement (LM) policy. Since in GGM process, nodes assigned to the same module are forced to move jointly and can never be separated, this may lead to a local minima. Based on the result of GGM, we allow every single node i to move locally to the module that induces the maximum decrease of RWE among all neighboring modules of node i (ties are broken uniformly at random), otherwise, it stays in its original module of GGM. Practically, we make multiple independent runs with different random sequences of nodes, and select the one that has minimum final RWE value. Part of the efficiency of the algorithm results from the following proposition.

Proposition 1 (Decrease in RWE by A Partition of Single Node Merging): The decrease in RWE by a partition obtained by moving an isolated node v to a module V_j is

$$\begin{aligned}\Delta\mathcal{H}^{\mathcal{P}}_{v,j} = &-p_j \cdot \log_2 \frac{p_j}{p_j + p_v} - p_v \cdot \log_2 \frac{p_v}{p_j + p_v} + q_j \cdot \log_2 p_j \\ &- q_{V_j \cup \{v\}} \cdot \log_2 (p_j + p_v) + \left(1 - \frac{\alpha}{n}\right) \cdot p_v \cdot \log_2 p_v,\end{aligned}$$

where p_v and p_j are the ergodic frequency at node v and module V_j respectively, $q_{V_j \cup \{v\}}$ is the enter probability of the new module by merging V_j and node v, which can be easily computed by

$$\begin{aligned}q_{V_j \cup \{v\}} = &\, q_j + \alpha \cdot \frac{n - n_j}{n} \cdot p_v - \frac{\alpha}{n} \cdot (p_j + p_v) \\ &+ (1-\alpha) \cdot \left[\sum_{k \notin V_j} p_v \cdot p_{v \to k} - \sum_{i \in V_j} p_i \cdot p_{i \to v} \right].\end{aligned}$$

The proof of Proposition 1 is based on the definition in (10), which can be easily derived. A similar expression is used to evaluate the change of RWE by a partition when the node v is moved from its module. In our LM policy, we therefore evaluate the change of RWE by a partition by removing it from its original module and then by merging it with a neighboring module.

VI. EXPERIMENTS AND EVALUATIONS

In this section, we conduct extensive experiments to evaluate the performance of RWE from various perspectives.

Datasets and Settings: The real-world datasets [40], [41], [42] listed in Table 1 of supplementary material contains static graphs, temporal graphs and digraphs with ground-truth communities, and presents statistics like average clustering coefficient (CLC), community numbers and snapshots of temporal networks. For each static graph, we remove self-loops and multiple edges. We regard every temporal graph as a stream of directed weighted edges with timestamps. For convenience, we partition the edges of temporal networks into several groups to generate snapshots where each snapshot is within a certain time interval. The experiments are performed on a server with Intel(R) Xeon(R) CPU 2.10 GHz, 512 GB RAM and implemented in Python.

A. One-Dimensional Structural Complexity Characterization

We first show the performance of one-dimensional RWE for characterizing the structural complexity of digraphs.

1) Approximation Accuracy: To evaluate the accuracy of RWE as an approximation for the von Neumann entropy in digraphs, we measure the RWE, exact VNE and directed SI as a competitor in 6 real-world digraphs (including both the original graph and the largest strongly connected component of each graph). The results are shown in Table III.

As we can see, compared to SI, RWE has a lower normalized entropy gap regardless of the network types or sizes, who even offers 42× reduction (in original WV) over SI. This demonstrates that RWE is an accurate approximation for VNE in digraphs. Therefore, such is not consistent with the case in undirected graphs, in which, however, SI is a fairly good approximation for VNE [13]. Besides, we show the result of approximation accuracy in directed weighted graphs in Table IV. Besides the 4 real-world weighted digraphs, we also choose two unweighted graphs (Wiki-Vote, P2P-G05), where the weight of each edge is set uniformly at random in the range [1,10]. The results also demonstrate a better approximation for VNE in weighted digraphs than SI.

2) Analysis of Upper Bound: Moreover, to illustrate the tightness of the derived upper bound, Fig. 3 presents the normalized derived upper bound (UB) and the normalized entropy gap in two special digraphs with 1,000 nodes. Besides, we also report the trends of true VNE value as corresponding references. For the bidirectional Barab á si-Albert (BA) graphs, we set in-degrees in {10,15,...,300}. While for the regular digraphs, the in-degrees are set from 10 to 999 (the complete graph). The observations are as follows. First, for both two models, the normalized entropy gap are all below 1%, demonstrating the high accuracy for

TABLE III
ACCURACY COMPARISON FOR APPROXIMATING THE VON NEUMANN ENTROPY IN 6 REAL-WORLD DIRECTED NETWORKS (EP IS SHORT FOR ENTROPY GAP)

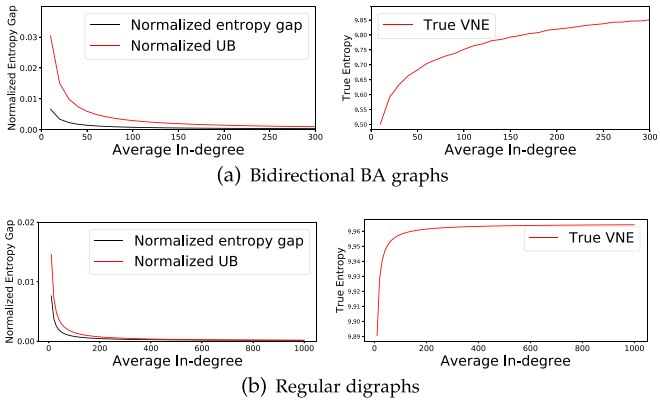
Datasets	Lake		Wiki-Vote		P2P-G05		Cora		HepTh		P2P-G30	
Graph size	Original	LSCC	Original	LSCC	Original	LSCC	Original	LSCC	Original	LSCC	Original	LSCC
	183	22	7,115	1,300	8,846	3,234	23,166	3,991	27,770	7,464	36,682	8,490
True VNE	4.8450	2.4745	11.8101	9.5681	12.9401	10.2687	12.8029	7.9766	13.3320	6.3568	15.0969	11.8946
RWE	4.9706	2.8322	11.8372	9.6276	12.9618	10.4082	12.9127	8.2627	13.3703	6.5515	15.1087	12.0501
SI	6.2471	4.2571	10.6667	9.9622	12.4198	10.8397	12.6845	11.1393	13.0295	11.5811	14.6200	12.5245
RWE normalized EP (%)	**2.5929**	**14.4523**	**0.2298**	**0.6121**	**0.1674**	**1.3584**	**0.8578**	**3.5874**	**0.2872**	**3.0630**	**0.0777**	**1.3074**
SI normalized EP (%)	28.9410	72.0375	9.6816	4.1189	4.0211	5.5604	0.9253	39.6504	2.2690	82.1848	3.1589	5.2955

The bold values represent the optimal performance data in the corresponding comparative experiments.

TABLE IV
ACCURACY COMPARISON FOR APPROXIMATING THE VON NEUMANN ENTROPY IN 6 REAL-WORLD WEIGHTED DIRECTED NETWORKS (EP IS SHORT FOR ENTROPY GAP)

Datasets	Cattle		Highschool		Florida-Eco		Advogato		Wiki-Vote[1,10]		P2P-G05[1,10]	
Graph size	Original	LSCC	Original	LSCC	Original	LSCC	Original	LSCC	Original	LSCC	Original	LSCC
	28	20	70	67	128	103	6,539	3,140	7,115	1,300	8,846	3,234
True VNE	3.6935	2.6093	5.4850	3.2895	5.1186	1.4878	11.4475	9.5341	11.8033	9.5699	12.9308	10.0999
RWE	3.8924	2.9417	5.6080	3.4907	5.1944	2.0005	11.4921	9.5986	11.8328	9.6328	12.9551	10.2582
SI	4.4497	4.0554	5.7403	5.7354	3.7059	2.7044	10.6228	10.3866	10.6627	9.9550	12.3658	10.7889
RWE normalized EP (%)	**5.3856**	**12.7381**	**2.2415**	**6.1182**	**1.4805**	**34.4629**	**0.3900**	**0.6769**	**0.2495**	**0.6571**	**0.1880**	**1.5667**
SI normalzed EP (%)	20.4749	55.4200	4.6549	74.3550	-27.5987	81.7757	-7.2040	8.9418	-9.6636	4.0234	-4.3694	6.8212

The bold values represent the optimal performance data in the corresponding comparative experiments.



(a) Bidirectional BA graphs

(b) Regular digraphs

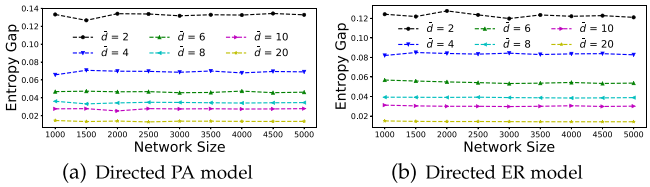
Fig. 3. Derived upper bound (normalized by the true VNE) and the real entropy gap (normalized by the true VNE) of bidirectional BA graphs and regular digraphs with 1,000 nodes, varying average in-degree.

TABLE V
DERIVED UPPER BOUND, REAL ENTROPY GAP AND NORMALIZED ENTROPY GAP OF 10 GENERAL DIGRAPHS

Network	Entropy gap	UB	Normalized entropy gap	In-degree power law exponent γ
Macaques	0.15	0.77	0.0411	12.79
Taro	0.25	0.48	0.0601	28.4
Bison	0.08	6.22	0.0189	5.43
Physicians	0.14	441	0.0263	3.8
Stelzl	0.12	2.59	0.0132	2.61
Wiki-Vote	0.06	201	0.0061	3.86
Cattle	0.36	241	0.1449	3.39
Florida-Eco	0.06	16.44	0.0113	5.86
Flight	0.02	2.65	0.0023	4.48
Email	0.02	10.66	0.0027	6.02

approximating VNE. Second, the normalized UB are almost below 3%, and are closer to the normalized entropy gap with increasing average in-degree, which indeed yields tightness. Third, compared to the case in bidirectional digraphs, the bound is much sharper in regular digraphs under the same average in-degree, which is consistent with the theoretical analysis in Section IV-B. Fourth, for the regular digraphs, as predicted, the true VNE increases with the increasing average degree, and reaches the maximum value when $d_{\text{avg}}^{\text{in}} = 999$, i.e. the complete graph. In this case, the normalized entropy gap decreases to 1.4486×10^{-4} and the normalized UB is 1.4493×10^{-4}. According to the theoretical analysis, both two values will tend to 0 when the complete graph size $n \to \infty$.

While for general digraphs, due to the uncertain distribution of directed edges (including both location and directionality), the order of the minimum positive entry φ of the Perron vector varies greatly, leading to tight or loose bounds for different digraphs. We select 10 directed networks of different categories with varying in-degree power law exponent γ, and Table V presents the entropy gap, upper bound and the normalized entropy gap (normalized by the true VNE). As we predicted, the upper bound varies a lot. However, we observe that networks with tighter bound, such as 0.77, 0.48 and 2.65 in Macaques, Taro and Flight respectively, tend to have a larger γ (basically larger than 4). The reason may be that a larger in-degree power law exponent means there are generally less nodes with large in-degree, so that the network will have a more balanced Perron vector, leading to a larger order of φ. While for networks with more large-degree nodes, due to the existence of more hubs, φ will be as small as possible, then a looser upper bound is derived. The Stelzl network is an exception ($\gamma \leq 4$, but still have a small upper bound 2.59), we find the CLC is only 0.0058, which means a very less degree to which nodes in Stelzl tend to cluster together, avoiding that the random walk is often trapped in a few nodes. Despite the observation of upper bound, this does



(a) Directed PA model (b) Directed ER model

Fig. 4. Effects of average in-degree and graph size on the entropy gap for two models.

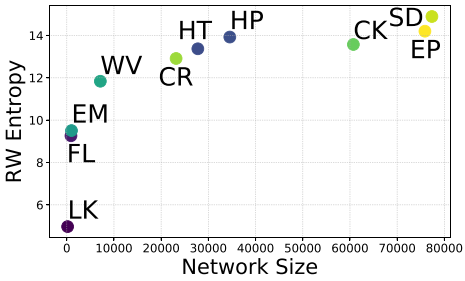


Fig. 5. RWEs of 10 real-world digraphs with various sizes.

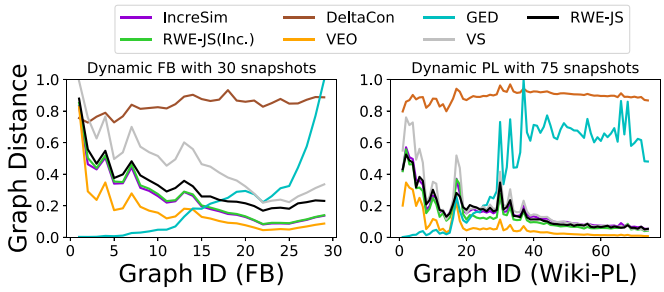


Fig. 6. Graph distance between adjacent graphs in graph streams. The GED distance is normalized to lie in [0, 1].

TABLE VI
COMPUTATION TIME (SECONDS) FOR DYNAMICS OF GRAPH STREAM

Datasets	RWE-JS	RWE-JS (Inc.)	IncreSim	VEO	GED	DeltaCon	VS
FB	0.3317	**0.0338**	0.1432	0.0503	0.0913	403.28	1.6016
Wiki-PL	30.151	**0.5075**	2.8879	3.1631	13.030	30341.7	250.35

RWE-JS (Inc.) attains the best time efficiency.
The bold values represent the optimal performance data in the corresponding comparative experiments.

not conflict with the fact that that our proposed RWE indeed yields small normalized entropy gap in approximating VNE practically. Table V shows that the normalized entopy gap has a minimum value of 0.23% and only a maximum value of 14.49%, and the average value is 3.26%. These results indeed demonstrate that our method is a good approximation of VNE.

3) Entropy Gap Sensitivity: To evaluate the sensitivity of entropy gap to digraph properties such as average in-degree and graph size, we further measure the digraph entropy gap of two synthetic digraph models: directed preferential attachment (PA) model [43] and Erdős Rényi (ER) model, in which we vary the graph size from 1000 to 5000, and the average in-degree is in {2, 4, 6, 8, 10, 20}.

The observations from Fig. 4 are two fold. First, the digraph entropy gap decreases as the average in-degree increases for both two models. Second, **the entropy gap is nearly insensitive to the change of graph size.**

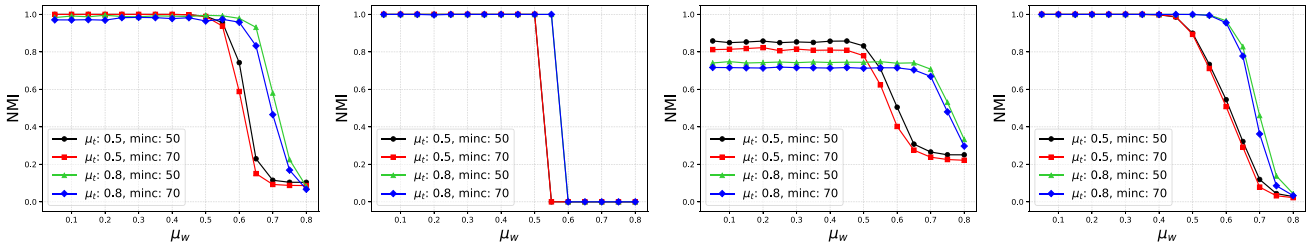
4) Case Study: We further analyze RWE of directed networks by selecting 10 real-world digraphs with varying sizes and showing their RWEs in Fig. 5. The observations are as follows. First, in general, **a larger network size usually brings about a higher RWE value**. The reason is that generally as the size of the network increases, the quantities and ways of connections between nodes usually become more, then from an average view of any single node, the uncertainty of information flow or random walker transfers to the next node will be large, hence the average information that each node is traversed by the random walk will be larger. However, second, this trend is not necessarily hold: **the RWE of a larger network may sometimes be lower than that of a small network**. For instance, the CK network, almost twice the size of HP, but has a lower RWE than that of HP. To analyze the reasons behind, we take a deeper look at their inherent structures, and find that the clustering coefficient (CLC) of CK is 0.5332, which is much larger than 0.1457 of HP. Besides, the two diameters are 13 versus 14. Both evidence indicate that the nodes inside CK tend to cluster together more than those in HP, which ensures that the uncertainty of the information flow or random walk in the network is lower, leading to a lower RWE. Third, **the RWE of networks of similar size may vary greatly**. Take SD and EP as an example. They have almost the same number of nodes, but the RWE of SD is explicitly larger than that of EP, implying more uncertainties of information flow contained in the structure of SD. A similar reason can be found that the CLC of SD is much lower than that of EP, i.e., 0.0555 versus 0.1378.

5) Performance of RWE-JS: We evaluate the performance of RWE-JS in terms of both efficiency an effectiveness, and compare with five prominent graph similarity baselines: VEO [44], GED [45], IncreSim [13], DeltaCon [46] and VS [44].

Efficiency. Dynamics of Graph Stream: We first measure the distance between two adjacent graphs in 2 real-world dynamic digraphs. The distance metrics are shown in Fig. 6, and the computation time is reported in Table VI .

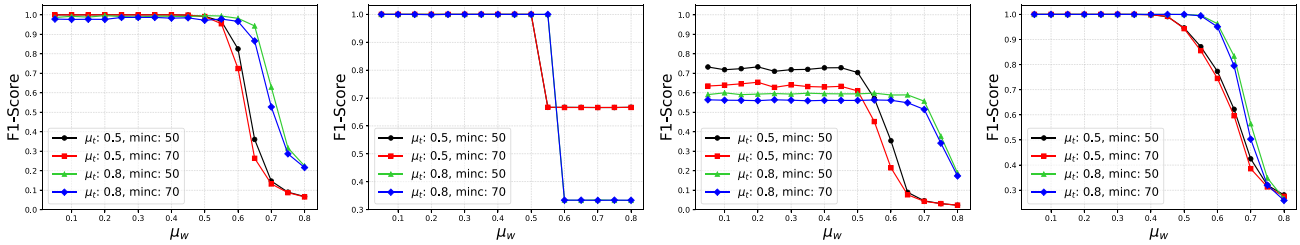
The observations are two folds. First, all the six methods except GED could capture the dynamics of graph streams. For both two temporal graphs, the distance measure changes dramatically in the beginning. For the Wiki graph, it gradually tends to be flat, which implies that the structure of Wiki-PL gradually tends to be stable. However, as for the FB graph, the distance measure has an upward trend in the end, indicating that FB is still changing constantly. Second, RWE-JS (Inc.) has the least computation time, which is roughly 5 times faster than IncreSim, 25 times faster than GED, 60 times faster than RWE-JS and 500 times faster than VS in the large-scale Wiki graph. Despite that, compared to DeltaCon and VS, the complexity of RWE-JS is even more acceptable in practice.

Effectiveness. Anomaly Detection in Graph Stream: To further validate the effectiveness of RWE-JS in detecting anomaly graph in a graph stream, we mimic the DDoS attacks in Caida



(a) GGM (b) Infomap (c) SI (d) Modularity

Fig. 7. NMI of four algorithms in directed LFR benchmarks with varying parameters μ_t, μ_w and community size range.



(a) GGM (b) Infomap (c) SI (d) Modularity

Fig. 8. F1 score of four algorithms in directed LFR benchmarks with varying parameters μ_t, μ_w and community size range.

TABLE VII
ACCURACY FOR ANOMALY DETECTION IN THE DYNAMIC AS-CAIDA

DDoS attack (X%)	RWE-JS	RWE-JS (Inc.)	IncreSim	VEO	GED	DeltaCon	VS
1%	**43%**	34%	34%	7%	13%	11%	16%
2%	**76%**	37%	36%	23%	9%	13%	35%
3%	**78%**	58%	37%	36%	12%	35%	38%
5%	**91%**	**91%**	54%	34%	19%	90%	58%

RWE-JS attains the best detection rate and robustness.
The bold values represent the optimal performance data in the corresponding comparative experiments.

autonomous system networks, where we extract the first 9 snapshots as a digraph stream. We model the anomalous events by first selecting one graph from the graph stream randomly, and then connecting $X\%$ source nodes to a target node (The weight is added by 1 in case the edge already exists) to mimic increasing incoming traffic of the target. The task is to detect the anomalous graph by comparing the graph distance metric between adjacent graphs.

Specifically, for any graph G_k, by calculating the dissimilarity score as the average graph distance with G_{k-1} and G_{k+1}, we rank the graphs by their dissimilarity score in descending manner, and repeat 100 random instances of DDoS attack for each $X = \{1, 2, 3, 5\}$. Table VII shows the results that the anomalous graph appears in the Top-1 ranking based on 7 different measures. Firstly, RWE-JS consistently attains the best detection performance. Secondly, when X is small, it is much harder for detection as the attack becomes stealthier, but the performance of RWE-JS is more sensible than the baselines, indicating the robustness of RWE-JS. Thirdly, since VEO and GED are both insensitive to the edge weight change, their detection rate is much lower than other methods.

B. Community Detection in Directed Networks

We then perform experiments to find community structures in digraphs by two-dimensional RWE. To evaluate the results with underlying communities, we adopt two commonly used metrics: NMI (Normalized Mutual Information) [47] and F1 score. NMI measures the quality of clustering and is a normalization of the Mutual Information (MI) score to scale the results between 0 (no mutual information) and 1 (perfect correlation). F1 score combines the precision and recall scores to measure the accuracy of community detection (The larger the F1 score, the higher the accuracy.). And we select 3 other algorithms for community detection in digraphs: ▷ *Infomap:* Find the community structures using maps to describe the dynamics across the links and nodes [29]. ▷ *SI:* Find the community structures with the minimum directed two-dimensional SI [2] via an agglomerative process. ▷ *Modularity:* Find the community structures with maximum directed modularity [28].

1) Results in LFR Benchmarks: First, we conduct extensive experiments on directed LFR (Lancichinetti-Fortunato-Radicchi) benchmarks [48] with varying community sizes and mixing parameters (we introduce the LFR benchmark in Section 5 of Supplementary Material.), i.e., μ_t and μ_w, referring to topology and link weight respectively. For all the LFR benchmark used here, we set the network size $n = 1000$, and the mixing parameter μ_w is varying from 0.05 to 0.8. In each plot, we show four curves, corresponding to two topological mixing parameter μ_t (0.5 and 0.8) and, for a given μ_t, to two different

TABLE VIII
PERFORMANCE COMPARISONS AMONG 5 DIGRAPH BASED AND 5 UNDIRECTED GRAPH BASED COMMUNITY DETECTION METHODS IN FOUR REAL-WORLD NETWORKS

Networks	Statistics	GGM+LM	GGM	Infomap	SI	Modularity	Louvain	CNM	LPA	Danmf	CommunityGAN
Flight network	NMI	**0.83**	0.82	0.81	0.75	0.67	0.56	0.46	0.52	0.65	0.62
	F1	**0.91**	0.90	0.89	0.22	0.85	0.84	0.80	0.82	0.85	0.23
E-mail	NMI	**0.66**	0.64	0.55	0.45	0.44	0.56	0.48	0.18	0.50	0.33
	F1	**0.70**	0.64	**0.70**	0.62	0.62	0.68	0.63	0.53	**0.70**	0.24
WiKi	NMI	**0.48**	**0.48**	0.46	0.47	0.39	0.39	0.32	0.46	0.25	0.18
	F1	0.84	0.83	**0.85**	0.83	0.83	0.82	0.81	0.80	0.80	0.43
GPCN citation	NMI	**0.54**	0.51	0.47	0.46	0.39	0.40	0.37	0.48	0.15	0.17
	F1	0.46	0.45	0.48	0.50	**0.52**	0.51	0.51	0.46	0.49	0.29

The bold values represent the optimal performance data in the corresponding comparative experiments.

ranges for the community sizes (50 to 100 nodes and 70 to 100 nodes, respectively). Other parameters are set to default as mentioned in the original implementation.

Figs. 7 and 8 show the results of our analysis. Each point of every curve is averaged over 100 realizations of the benchmark (the corresponding standard deviation results can be referred to Section 4 in Supplementary Material). First, as one can see, considering both NMI and F1 score, the GGM method performs fairly well until $\mu_w = 0.6$ for all settings, which has a much wider range of good performance than all the baseline algorithms. For example, the NMI values of Infomap under $\mu_t = 0.8$ are already 0, but those of GGM are almost 1. This demonstrates that although both methods are based on two-dimensional information description, where GGM method describes the average information of random walker, while the Infomap method focuses on the length of two-dimensional coding, GGM can better describe the information flow in a digraph. Second, compared to SI, who worsens if communities are topologically more mixed (higher μ_t), GGM is insensitive to the mixing parameter μ_t, since RWE is not directly defined on the incoming or outcoming connections of graphs, but the probability flow of random walk as a proxy of information flow. Although the Modularity algorithm has a similar results for $\mu_t = 0.8$ as GGM, we can still see a slightly higher value when μ_w is larger than 0.6, which demonstrates the advantages of capturing the information flow of digraphs. Third, similar to Infomap, the performance of GGM for $\mu_t = 0.8$ decays a little later than that for $\mu_t = 0.5$, the reason behind is that both two methods focus on the information flow within networks, a higher μ_t means a more obvious flow between different communities that can be found.

2) Results in Real-World Networks: We next pick 4 real-world networks with ground-truth communities. Besides EM [40] and WK [41], we also collect the flight information all over the world from Jan. to June 2021, and construct the flight network between airports, with countries as communities and the flow as edge weights. Moreover, we construct a citation network (GPCN) of geosciences during 2017 to 2020, where different venues are considered as communities.

Table VIII not only presents the results of the above four algorithms, but also the refined version of GGM by local single node movements, i.e., GGM+LM, as well as 3 well-known community detection methods (Louvain [23], CNM [20], LPA [22]) and 2 deep learning methods (Danmf [39], CommunityGAN [49]) exactly designed for undirected graphs by ignoring the edge directions. The observations are two fold. First, the GGM+LM algorithm performs the best among ten methods in almost all the cases, especially for FL and EM networks, which demonstrates that the RWE based approach has advantages to identify important aspects of structures in networks where links represent patterns of movement among nodes. Second, as we can see, moving nodes locally can indeed enhance the performance of GGM.

VII. CONCLUSION AND FUTURE WORK

In this paper, we propose a random walk based graph entropy to capture both one-dimensional and two-dimensional structural complexity of directed networks. We suggest to use one-dimensional RWE as a proxy of the von Neumann entropy in digraphs such that provable accuracy, scalability, and interpretability are achieved simultaneously, and the experimental results also show the capability of well capturing the structural complexity of real-world networks. Based on the two-dimensional RWE, we design an efficient community detection algorithm for digraphs. Extensive results in LFR benchmarks and four real-world networks demonstrate the superior performance to the baselines.

There are multiple tangible research directions we can pursue. First, although the derived upper bound of digraph entropy gap yields tightness in some digraphs, for general ones where the Perron vector has no closed form, a more sharpened bound is expected if possible. Second, for some access limited scenarios such as the World Wide Web, there is need to develop sampling-based methods to estimate the RWE. Last, since there may also be substructures within communities, the higher dimension structural complexity of directed networks is still in its infancy.

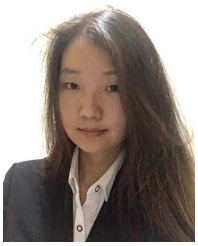
REFERENCES

[1] D. Bonchev and G. A. Buck, "Quantitative measures of network complexity," in *Complexity in Chemistry, Biology, and Ecology*, Boston, MA, USA: Springer, 2005, pp. 191–235.
[2] A. Li and Y. Pan, "Structural information and dynamical complexity of networks," *IEEE Trans. Inf. Theory*, vol. 62, no. 6, pp. 3290–3339, Jun. 2016.
[3] N. Rashevsky, "Life, information theory, and topology," *Bull. Math. Biophys.*, vol. 17, no. 3, pp. 229–235, 1955.
[4] C. Raychaudhury, S. Ray, J. Ghosh, A. Roy, and S. Basak, "Discrimination of isomeric structures using information theoretic topological indices," *J. Comput. Chem.*, vol. 5, no. 6, pp. 581–588, 1984.

[5] 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. Combinatorics*, vol. 10, no. 3, pp. 291–317, 2006.
[6] M. Dehmer and A. Mowshowitz, "A history of graph entropy measures," *Inf. Sci.*, vol. 181, no. 1, pp. 57–78, 2011.
[7] L. Bai and E. R. Hancock, "Graph clustering using the Jensen-Shannon kernel," in *Proc. Int. Conf. Comput. Anal. Images, Patterns*, 2011, pp. 394–401.
[8] M. De Domenico, V. Nicosia, A. Arenas, and V. Latora, "Structural reducibility of multilayer networks," *Nature Commun.*, vol. 6, no. 1, pp. 1–9, 2015.
[9] M. De Domenico and J. Biamonte, "Spectral entropies as information-theoretic tools for complex network comparison," *Phys. Rev. X*, vol. 6, no. 4, 2016, Art. no. 041062.
[10] P.-Y. Chen, L. Wu, S. Liu, and I. Rajapakse, "Fast incremental Von Neumann graph entropy computation: Theory, algorithm, and applications," in *Proc. 36th Int. Conf. Mach. Learn.*, 2019, pp. 1091–1101.
[11] H. Choi, J. He, H. Hu, and Y. Shi, "Fast computation of Von Neumann entropy for large-scale graphs via quadratic approximations," *Linear Algebra Appl.*, vol. 585, pp. 127–146, 2020.
[12] A. Tsitsulin, M. Munkhoeva, and B. Perozzi, "Just slaq when you approximate: Accurate spectral distances for web-scale graphs," in *Proc. Int. Conf. World Wide Web*, 2020, pp. 2697–2703.
[13] X. Liu, L. Fu, and X. Wang, "Bridging the gap between Von Neumann graph entropy and structural information: Theory and applications," in *Proc. Int. Conf. World Wide Web*, 2021, pp. 3699–3710.
[14] F. Chung, "Laplacians and the cheeger inequality for directed graphs," *Ann. Combinatorics*, vol. 9, no. 1, pp. 1–19, 2005.
[15] C. Ye, R. C. Wilson, C. H. Comin, L. D. F. Costa, and E. R. Hancock, "Approximate Von Neumann entropy for directed graphs," *Phys. Rev. E*, vol. 89, no. 5, 2014, Art. no. 052804.
[16] E. Nasiri, K. Berahmand, and Y. Li, "A new link prediction in multiplex networks using topologically biased random walks," *Chaos, Solitons Fractals*, vol. 151, 2021, Art. no. 111230.
[17] Y. Li, J. Fan, Y. Wang, and K.-L. Tan, "Influence maximization on social graphs: A survey," *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 10, pp. 1852–1872, Oct. 2018.
[18] C. Zhang, Q. Guo, L. Fu, X. Gan, and X. Wang, "ITE: A structural entropy based approach for source detection," in *Proc. IEEE Conf. Comput. Commun.*, 2021, pp. 1–10.
[19] M. E. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Phys. Rev. E*, vol. 69, no. 2, 2004, Art. no. 026113.
[20] A. Clauset, M. E. Newman, and C. Moore, "Finding community structure in very large networks," *Phys. Rev. E*, vol. 70, no. 6, 2004, Art. no. 066111.
[21] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, "Uncovering the overlapping community structure of complex networks in nature and society," *Nature*, vol. 435, no. 7043, pp. 814–818, 2005.
[22] U. N. Raghavan, R. Albert, and S. Kumara, "Near linear time algorithm to detect community structures in large-scale networks," *Phys. Rev. E*, vol. 76, no. 3, 2007, Art. no. 036106.
[23] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *J. Stat. Mechanics: Theory Experiment*, vol. 2008, no. 10, 2008, Art. no. P10008.
[24] Y. Yao, Y. Pan, S. Wang, H. Wang, and W. Nazeer, "Flow-based clustering on directed graphs: A structural entropy minimization approach," *IEEE Access*, vol. 8, pp. 152579–152591, 2020.
[25] J. Yang, J. McAuley, and J. Leskovec, "Community detection in networks with node attributes," in *Proc. IEEE 13th Int. Conf. Data Mining*, 2013, pp. 1151–1156.
[26] V. Fionda, L. Palopoli, S. Panni, and S. E. Rombo, "A technique to search for functional similarities in protein-protein interaction networks," *Int. J. Data Mining Bioinf.*, vol. 3, no. 4, pp. 431–453, 2009.
[27] M. Rostami, M. Oussalah, K. Berahmand, and V. Farrahi, "Community detection algorithms in healthcare applications: A systematic review," *IEEE Access*, vol. 11, pp. 30247–30272, 2023.
[28] A. Arenas, J. Duch, A. Fernández, and S. Gómez, "Size reduction of complex networks preserving modularity," *New J. Phys.*, vol. 9, no. 6, 2007, Art. no. 176.
[29] M. Rosvall and C. T. Bergstrom, "Maps of random walks on complex networks reveal community structure," *Proc. Nat. Acad. Sci.*, vol. 105, no. 4, pp. 1118–1123, 2008.
[30] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2012.
[31] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," *Comput. Netw. ISDN Syst.*, vol. 30, no. 1–7, pp. 107–117, 1998.
[32] J. Von Neumann, *Mathematische Grundlagen Der Quantenmechanik*, vol. 38. Berlin, Germany: Springer, 2013.
[33] S. L. Braunstein, S. Ghosh, T. Mansour, S. Severini, and R. C. Wilson, "Some families of density matrices for which separability is easily tested," *Phys. Rev. A*, vol. 73, no. 1, 2006, Art. no. 012320.
[34] G. H. Golub and C. Greif, "An arnoldi-type algorithm for computing page rank," *BIT Numer. Math.*, vol. 46, no. 4, pp. 759–771, 2006.
[35] P. Berkhin, "A survey on pagerank computing," *Internet Math.*, vol. 2, no. 1, pp. 73–120, 2005.
[36] A. W. Marshall, I. Olkin, and B. C. Arnold, *Inequalities: Theory of Majorization and Its Applications*, 2nd ed. Berlin, Germany: Springer, 2011.
[37] J. Liao and A. Berg, "Sharpening Jensen's inequality," *Amer. Statistician*, vol. 73, no. 3, pp. 278–281, 2018.
[38] A. Majtey, P. Lamberti, and D. Prato, "Jensen-Shannon divergence as a measure of distinguishability between mixed quantum states," *Phys. Rev. A*, vol. 72, no. 5, 2005, Art. no. 052310.
[39] F. Ye, C. Chen, and Z. Zheng, "Deep autoencoder-likenonnegative matrix factorization for community detection," in *Proc. 27th ACM Int. Conf. Inf. Knowl. Manage.*, 2018, pp. 1393–1402.
[40] J. Leskovec and A. Krevl, "SNAP datasets: Stanford large network dataset collection," Jun. 2014. [Online]. Available: http://snap.stanford.edu/data
[41] J. Kunegis, "Konect: The koblenz network collection," in *Proc. Int. Conf. World Wide Web*, 2013, pp. 1343–1350.
[42] B. Viswanath, A. Mislove, M. Cha, and K. P. Gummadi, "On the evolution of user interaction in facebook," in *Proc. 2nd ACM Workshop Online Social Netw.*, 2009, pp. 37–42.
[43] E. Pesce, "Networkssimulator," Sep. 2015. [Online]. Available: https://github.com/emanuelepesce/NetworksSimulator
[44] P. Papadimitriou, A. Dasdan, and H. Garcia-Molina, "Web graph similarity for anomaly detection," *J. Internet Serv. Appl.*, vol. 1, no. 1, pp. 19–30, 2010.
[45] H. Bunke, P. J. Dickinson, M. Kraetzl, and W. D. Wallis, *A Graph-Theoretic Approach to Enterprise Network Dynamics*, vol. 24. Berlin, Germany: Springer, 2007.
[46] D. Koutra, N. Shah, J. T. Vogelstein, B. Gallagher, and C. Faloutsos, "Deltacon: Principled massive-graph similarity function with attribution," *ACM Trans. Knowl. Discov. Data*, vol. 10, no. 3, pp. 1–43, 2016.
[47] L. Danon, A. Diaz-Guilera, J. Duch, and A. Arenas, "Comparing community structure identification," *J. Stat. Mechanics: Theory Experiment*, vol. 2005, no. 09, 2005, Art. no. P09008.
[48] A. Lancichinetti and S. Fortunato, "Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities," *Phys. Rev. E*, vol. 80, no. 1, 2009, Art. no. 016118.
[49] Y. Jia, Q. Zhang, W. Zhang, and X. Wang, "CommunityGAN: Community detection with generative adversarial nets," in *Proc. World Wide Web Conf.*, 2019, pp. 784–794.

Chong Zhang received the B.E. degree in telecommunications engineering from Xidian University, Xi'an, China, in 2018, and the Ph.D. degree from the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai, China, in 2023. He is currently an Assistant Professor with the School of Computer Science and Technology, Xi'an Jiaotong University, Xi'an. His research interests include the area of social networks and data mining.

Cheng Deng received the B.E. degree in computer science and technology from the Hunan University, Changsha, China, in 2019. He is currently working toward the Ph.D. degree in computer science with the Shanghai Jiao Tong University, Shanghai, China. His research interests include the area of knowledge graph and social networks natural language processing.



Luoyi Fu (Member, IEEE) received the B.E. degree in electronic engineering and the Ph.D. degree in computer science and engineering from Shanghai Jiao Tong University, Shanghai, China, in 2009, and 2015, respectively. She is currently an Associate Professor with the Department of Computer Science and Engineering, Shanghai Jiao Tong University. Her research include the area of social networking and Big Data, scaling laws analysis in wireless networks, connectivity analysis and random graphs. She is a Member of the Technical Program Committees of several conferences including ACM MobiHoc 2018-2023, IEEE INFOCOM 2018-2023.

Xinbing Wang (Senior Member, IEEE) received the B.S. degree (with Hons.) from the Department of Automation, Shanghai Jiaotong University, Shanghai, China, in 1998, the M.S. degree from the Department of Computer Science and Technology, Tsinghua University, Beijing, China, in 2001, and the Ph.D. degree, major with the Department of electrical and Computer Engineering, minor with the Department of Mathematics, North Carolina State University, Raleigh, NC, USA, in 2006. He is currently a Professor with the Department of Electronic Engineering, Shanghai Jiaotong University. Dr. Wang has been an Associate Editor for IEEE/ACM TRANSACTIONS ON NETWORKING and IEEE TRANSACTIONS ON MOBILE COMPUTING, and the Member of the Technical Program Committees of several conferences including ACM MobiCom 2012, 2018-2019, ACM MobiHoc 2012-2023, IEEE INFOCOM 2009-2023.

Guihai Chen (Fellow, IEEE) received the B.S. degree from Nanjing University, Nanjing, China, the M.E. degree from Southeast University, Nanjing, and the Ph.D. degree from The University of Hong Kong, Hong Kong. He visited the Kyushu Institute of Technology, Fukuoka, Japan, in 1998, as a Research Fellow, and The University of Queensland, Brisbane, QLD, Australia, in 2000, as a Visiting Professor. From 2001 to 2003, he was a Visiting Professor with Wayne State University, Detroit, MI, USA. He is currently a Distinguished Professor and a Deputy Chair with the Department of Computer Science, Shanghai Jiao Tong University, Shanghai. He has authored or coauthored more than 200 papers in peer-reviewed journals and refereed conference proceedings in the areas of wireless sensor networks, high-performance computer architecture, peer-to-peer computing, and performance evaluation. He is a Member of the IEEE Computer Society. He was on technical program committees of numerous international conferences.

Lei Zhou received the B.S. degree and the M.S. degree from the Ocean University of China, Shandong, China, in 2002 and 2005 respectively, and the Ph.D. degree from the Department of Atmospheric and Oceanic Sciences, University of Maryland, College Park, MD, USA. He is currently a Professor with the School of Oceanography, Shanghai Jiao Tong University, Shanghai, China. His research interests include the area of tropical air sea interaction, ocean atmospheric dynamics, and intraseasonal oscillation.

Chenghu Zhou received the B.S. degree from the Department of Geography, Nanjing University, Nanjing, China, in 1984, and the M.S. and Ph.D. degrees from the Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China. He is an Academician of the Chinese Academy of Sciences and an Academician of International Eurasian Academy of Sciences. He is currently the Director of the State Key Laboratory of Resources and Environmental Information System, Beijing. His research interests include the area of spatial data mining, geographic system modeling, hydrology and water resources, geographic information system, and remote sensing application.