Bridging the Gap between von Neumann Graph Entropy and Structural Information: Theory and Applications

Xuecheng Liu
Shanghai Jiao Tong University
liuxuecheng@sjtu.edu.cn

Luoyi Fu
Shanghai Jiao Tong University
yiluofu@sjtu.edu.cn

Xinbing Wang
Shanghai Jiao Tong University
xwang8@sjtu.edu.cn

ABSTRACT

The von Neumann graph entropy is a measure of graph complexity based on the Laplacian spectrum. It has recently found applications in various learning tasks driven by networked data. However, it is computational demanding and hard to interpret using simple structural patterns. Due to the close relation between Lapalcian spectrum and degree sequence, we conjecture that the structural information, defined as the Shannon entropy of the normalized degree sequence, might be a good approximation of the von Neumann graph entropy that is both scalable and interpretable.

In this work, we thereby study the difference between the structural information and von Neumann graph entropy named as *entropy gap*. Based on the knowledge that the degree sequence is majorized by the Laplacian spectrum, we for the first time prove the entropy gap is between 0 and $\log_2 e$ in any undirected unweighted graphs. Consequently we certify that the structural information is a good approximation of the von Neumann graph entropy that achieves provable accuracy, scalability, and interpretability simultaneously. We further study two entropy based applications which can benefit from the bounded entropy gap and structural information: network design and graph similarity measure. We combine greedy method and pruning strategy to develop fast algorithm for the network design, and propose a novel graph similarity measure with a fast incremental algorithm for graph streams. Our experimental results on graphs of various scales and types show that the very small entropy gap readily applies to a wide range of graphs and weighted graphs. As an approximation of the von Neumann graph entropy, the structural information is the only one that achieves both high efficiency and high accuracy among the prominent methods. It is at least two orders of magnitude faster than SLaQ [40] with comparable accuracy. Our structural information based methods also exhibit superior performance in two entropy based applications.

ACM Reference Format:
Xuecheng Liu, Luoyi Fu, and Xinbing Wang. 2021. Bridging the Gap between von Neumann Graph Entropy and Structural Information: Theory and Applications. In *WWW '21: Proceedings of The Web Conference 2021, April 19–23, 2021, Ljubljana, Slovenia.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/1122445.1122456

WWW '21, April 19–23, 2021, Ljubljana, Slovenia

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

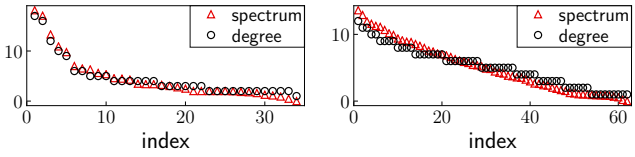
Evidence has rapidly grown in the past few years that graphs are ubiquitous in our daily life; online social networks, metabolic networks, transportation networks, and collaboration networks are just a few examples that could be represented precisely by graphs. One important issue in graph analysis is to measure the complexity of these graphs [4, 28] which refers to the level of organization of the structural features such as the scaling behavior of degree distribution, community structure, etc. In order to capture the inherent structural complexity of graphs, many entropy based graph measures [5, 13, 21, 28, 36, 37] are proposed, each of which is a specific form of the Shannon entropy for different types of distributions extracted from the graphs.

As one of the aforementioned entropy based graph complexity measures, the von Neumann graph entropy defined as the Shannon entropy of the spectrum of the trace rescaled Laplacian matrix of a graph (see Definition 3.1), is of special interests to scholars and practitioners [2, 7, 8, 12, 15, 22, 30, 40]. This spectral based entropy measure distinguishes between different graph structures. For instance, it is maximal for complete graphs, minimal for graphs with only single edge, and takes on intermediate values for ring graphs. Actually, the entropy measure originates from quantum information theory and is used to describe the mixedness of a quantum system. It is Braunstein et al. that first use the von Neumann entropy to measure the complexity of graphs by viewing each pure state of a quantum system as one of the edges of a graph [5].

Built upon the Laplacian spectra, the von Neumann graph entropy is a natural choice to capture the graph complexity since the Laplacian spectra is well-known to contain rich information about the multi-scale structure of graphs [17, 20]. As a result, it has recently found applications in downstream tasks of complex network analysis and pattern recognition. For example, the von Neumann graph entropy facilitates the measure of graph similarity via Jensen-Shannon divergence, which could be used to compress multilayer networks [15] and detect anomalies in graph streams [7]. As another example, the von Neumann graph entropy could be used to measure edge centrality [30] and design entropy-driven networks [33].

1.1 Motivations

However, despite the popularity received in applications, the main obstacle encountered in practice is the computational inefficiency of the exact von Neumann graph entropy. Indeed, as the spectral based entropy measure, the von Neumann graph entropy suffers from computational inefficiency since the computational complexity of the graph spectrum is cubic in the number of nodes. Meanwhile, the existing approximation approaches [7, 8, 40] such as quadratic approximation fail to capture the presence of non-trivial structural



(a) Zachary's karate club **(b) Dolphins**

Figure 1: The close relation between Laplacian spectra and degree sequence in two representative real-world graphs. Both the Laplacian spectra and degree sequence are sorted in non-increasing order. The x-axis represents the index of the sorted sequences, and the y-axis represents the value of Laplacian spectrum and degree.

patterns that seem to interpret the spectral based entropy measure. Therefore, *there is a strong desire to find a good approximation that achieves accuracy, scalability, and interpretability simultaneously.*

Instead of starting from scratch, we are inspired by the well-known knowledge that there is a close relationship between the combinatorial characteristics of a graph and the algebraic properties of its associated matrices [9]. To illustrate, we plot the Laplacian spectrum and degree sequence together in a same figure for two representative real-world graphs. As shown in Fig. 1, the sorted spectrum sequence and the sorted degree sequence almost coincide with each other. The similar phenomenon can also be observed in larger scale free graphs, which indicates that it is possible to reduce the approximation of the von Neumann graph entropy to the time-efficient computation of simple node degree statistics. Therefore, we ask without hesitation the first research question,

RQ1: *Does there exist some non-polynomial function ϕ such that $\sum_{i=1}^{n} \phi\left(d_i / \sum_{j=1}^{n} d_j\right)$ is close to the von Neumann graph entropy?*

where d_i is the degree of the node i in a graph of order n.

We emphasize on the non-polynomial property of the function ϕ since most of previous works that are based on polynomial approximations fail to fulfill the interpretability. The challenges from scalability and interpretability are translated directly into two requirements on the function ϕ to be determined. First, the explicit expression of ϕ must exist and keep simple to ensure the interpretability of the sum over degree statistics. Second, the function ϕ should be graph-agnostic to meet the scalability requirement, that is, ϕ should be independent from the graph to be analyzed. One natural choice yielded by the entropy nature of the graph complexity measure for the non-polynomial function ϕ is $\phi(x) = -x \log_2 x$. The sum $-\sum_{i=1}^{n} \left(d_i / \sum_{j=1}^{n} d_j\right) \log_2 \left(d_i / \sum_{j=1}^{n} d_j\right)$ has been named as one-dimensional *structural information* by Li et al. [28] in a connected graph since it has an entropy form and captures the information of a classic random walker in a graph. We extend this notion to arbitrary undirected graphs. Following the question **RQ1**, we raise the second research question,

RQ2: *Is the structural information an accurate proxy of the von Neumann graph entropy?*

To address the second question, we conduct to our knowledge a first study of the difference between structural information and von Neumann graph entropy, which we name as *entropy gap*.

1.2 Contributions

To study the entropy gap, we are based on a fundamental relationship between Laplacian spectrum $\boldsymbol{\lambda}$ and degree sequence $\mathbf{d}$ in undirected graphs: $\mathbf{d}$ is majorized by $\boldsymbol{\lambda}$. In other words, there is a doubly stochastic matrix P such that $P\boldsymbol{\lambda} = \mathbf{d}$. Leveraging the majorization and classic Jensen's inequality, we prove that the entropy gap is no less than 0 in arbitrary undirected graphs. By exploiting the Jensen's gap [29] which is an inverse version of the classic Jensen's inequality, we further prove that the entropy gap is no more than $\log_2 e$ in arbitrary unweighted undirected graphs. The constant lower and upper bounds on the entropy gap are further sharpened using more advanced knowledge about the Lapalcian spectrum and degree sequence, such as the Grone-Merris majorization [1]. We also apply the similar technique to bound the entropy gap in weighted graphs.

In a nutshell, our paper makes the following contributions:

- **Theory and interpretability:** Inspired by the close relation between Laplacian spectrum and degree sequence, we for the first time bridge the gap between the von Neumann graph entropy and structural information by proving that the entropy gap is between 0 and $\log_2 e$ in any unweighted graph. To the best of our knowledge, the constant bounds on the approximation error in unweighted graphs are sharper than that of any existing approaches with provable accuracy, such as FINGER [7]. Therefore, the answers to both **RQ1** and **RQ2** are YES! Besides, the structural information provides a simple geometric interpretation of the von Neumann graph entropy as a measure of degree heterogeneity. Thus, the structural information is a good approximation of the von Neumann graph entropy that achieves provable accuracy, scalability, and interpretability simultaneously.
- **Applications and efficient algorithms:** Using the structural information as a proxy of the von Neumann graph entropy with bounded error (entropy gap), we develop fast algorithms for two entropy based applications: network design and graph similarity measure. For the network design aiming to maximize the von Neumann entropy, we combine greedy method and pruning strategy to speed up the searching process. For the graph similarity measure, we propose a new distance measure based on structural information and Jensen-Shannon divergence. We further show that the proposed measure is a pseudometric and devise fast incremental algorithm to compute the similarity between adjacent graphs in a graph stream.
- **Extensive experiments and evaluations:** We use 3 random graph models, 9 real-world static graphs, and 2 real-world temporal graphs to evaluate the properties of the entropy gap and proposed algorithms. The results show that the entropy gap is small in a wide range of graphs, including the weighted graphs. And it is insensitive to the change of graph size. Compared with prominent methods for approximating the von Neumann graph entropy, the structural information is superior in both accuracy and computational speed. It is at least 2 orders of magnitude faster than the accurate SLaQ [40] algorithm with comparable accuracy. Our proposed algorithms based on structural information also exhibit superb performance in two entropy based applications.

Roadmap: The remainder of this paper is organized as follows. We review two related issues in Section 2. In Section 3 we introduce

Table 1: Comparison of methods for approximating the von Neumann graph entropy in terms of fulfilled (✓) and missing (✗) properties.

	[7]	[40]	[8]	Structural Information (Ours)
Provable accuracy	✓	✗	✗	✓
Scalability	✓	✓	✗	✓
Interpretability	✗	✗	✗	✓

the definitions of the von Neumann graph entropy, structural information, and the notion of entropy gap. Section 4 shows the close relationship between von Neumann graph entropy and structural information by bounding the entropy gap. Section 5 presents efficient algorithms for two graph entropy based applications. Section 6 provides experimental results. Section 7 offers some conclusions and directions for future research.

2 RELATED WORK

We review two main issues arised from the broad applications [2, 6, 11, 15, 26, 30, 31, 33] of the von Neumann graph entropy: computation and interpretation.

Approximate computation of the von Neumann graph entropy: In an effort to overcome the computational inefficiency of the von Neumann graph entropy, past works have resorted to various numerical approximations. Chen et al. [7] first compute a quadratic approximation of the entropy via Taylor expansion, then derive two finer approximations with accuracy guarantee by spectrum-based and degree-based rescaling, respectively. Before Chen's work, the Taylor expansion is widely adopted to give computationally efficient approximations [45], but there is no theoretical guarantee on the approximation accuracy. Following Chen's work, Choi et al. [8] propose several more complex quadratic approximations based on advanced polynomial approximation methods whose superiority are verified through experiments.

Besides, there is a trend to approximate spectral sums using stochastic trace estimation based approximations [19], the merit of which is the provable error-bounded estimation of the spectral sums. For example, Kontopoulou et al. [22] propose three randomized algorithms based on Taylor series, Chebyshev polynomials, and random projection matrices to approximate the von Neumann entropy of density matrices. As another example, based on the stochastic Lanczos quadrature technique [41], Tsitsulin et al. [40] propose an efficient and effective approximation technique called SLaQ to estimate the von Neumann entropy and other spectral descriptors for web-scale graphs. However, the approximation error bound of SLaQ for the von Neumann graph entropy is not provided. The disadvantages of such stochastic approximations are also obvious; their computational efficiency depends on the number of random vectors used in stochastic trace estimation, and they are not suitable for applications like anomaly detection in graph streams and entropy-driven network design.

The comparison of methods for approximating the von Neumann graph entropy is presented in Table 1. One of the common drawbacks of the aforementioned methods is the lack of interpretability, that is, none of these methods provide enough evidence to interpret this spectral based entropy measure in terms of structural patterns. By contrast, as a good proxy of the von Neumann graph entropy, the structural information offers us the intuition that the spectral based entropy measure is closely related to the degree heterogeneity of graphs.

Spectral descriptor of graphs and its structural counterpart: Researchers in spectral graph theory have always been interested in establishing a connection between combinatorial characteristics of a graph and the algebraic properties of its associated matrices. For example, the algebraic connectivity (also known as Fiedler eigenvalue), defined as the second smallest eigenvalue of a graph Laplacian matrix, has been used to measure the robustness [20] and synchronizability [46] of graphs. The magnitude of the algebraic connectivity has also been found to reflect how well connected the overall graph is [17]. As another example, the Fiedler vector, defined as the eigenvector corresponding to the Fiedler eigenvalue of a graph Laplacian matrix, has been found to be a good sign of the bi-partition structure of a graph [14]. However, there are some other spectral descriptors that have found applications in graph analytics, but require more structural interpretations, such as the heat kernel trace [39, 44] and von Neumann graph entropy.

Simmons et al. [38] suggest to interpret the von Neumann graph entropy as the centralization of graphs, which is very similar to our interpretation using structural information. They derive both upper and lower bounds on the von Neumann graph entropy in terms of graph centralization under some hard assumptions on the range of the von Neumann graph entropy. Therefore, their results cannot be directly converted to accuracy guaranteed approximations of the von Neumann graph entropy for arbitrary simple graphs. By constrast, our work shows that the structural information is an accurate, scalable, and interpretable proxy of the von Neumann graph entropy for arbitrary simple graphs. Besides, the techniques used in our proof are also quite different from [38].

3 PRELIMINARIES

In this paper, we study the undirected graph $G = (V, E, A)$ with positive edge weights, where $V = \{1, \ldots, n\}$ is the node set, E is the edge set, and $A \in \mathbb{R}_+^{n\times n}$ is the symmetric weight matrix with positive entry A_{ij} denoting the weight of an edge $(i, j) \in E$. If the node pair $(i, j) \notin E$, then $A_{ij} = 0$. If graph G is unweighted, the weight matrix $A \in \{0, 1\}^{n\times n}$ is called the adjacency matrix of G. The degree of node $i \in V$ in graph G is defined as $d_i = \sum_{j=1}^{n} A_{ij}$. The Laplacian matrix of graph G is defined as $L = D - A$ where $D = \mathrm{diag}(d_1, \ldots, d_n)$ is the degree matrix. Let $\{\lambda_i\}_{i=1}^{n}$ be the sorted eigenvalues of L such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n = 0$, which is called Laplacian spectrum. We define $\mathrm{vol}(G) = \sum_{i=1}^{n} d_i$ as the volume of graph G, then $\mathrm{vol}(G) = \mathrm{tr}(L) = \sum_{i=1}^{n} \lambda_i$ where $\mathrm{tr}(\cdot)$ is the trace operator. For the convenience of delineation, we define a special function $f(x) \triangleq x \log_2 x$ on the support $[0, \infty)$ where $f(0) \triangleq \lim_{x\downarrow 0} f(x) = 0$ by convention. In the following, we present formal definitions of the von Neumann graph entropy, structural information, and the entropy gap. Slightly different from the one-dimensional structural information proposed by Li et al. [28], our definition of structural information does not require the graph G to be connected.

Definition 3.1 (von Neumann graph entropy). The von Neumann graph entropy of an undirected graph $G = (V, E, A)$ is defined as

$\mathcal{H}_{\text{vn}}(G) = -\sum_{i=1}^{n} f(\lambda_i/\text{vol}(G))$, where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n = 0$ are the eigenvalues of the Laplacian matrix $L = D - A$ of the graph G, and $\text{vol}(G) = \sum_{i=1}^{n} \lambda_i$ is the volume of G.

Definition 3.2 (Structural information). The structural information of an undirected graph $G = (V, E, A)$ is defined as $\mathcal{H}_1(G) = -\sum_{i=1}^{n} f(d_i/\text{vol}(G))$, where d_i is the degree of node i in G and $\text{vol}(G) = \sum_{i=1}^{n}$ is the volume of G.

Definition 3.3 (Entropy gap). The entropy gap of an undirected graph $G = (V, E, A)$ is defined as $\Delta\mathcal{H}(G) = \mathcal{H}_1(G) - \mathcal{H}_{\text{vn}}(G)$.

The von Neumann graph entropy and structural information are well-defined for all the undirected graphs except for the graphs with empty edge set, in which $\text{vol}(G) = 0$. When $E = \varnothing$, we take it for granted that $\mathcal{H}_1(G) = \mathcal{H}_{\text{vn}}(G) = 0$.

4 APPROXIMATION ERROR ANALYSIS

In this section we bound the entropy gap in the undirected graphs of order n. Since the nodes with degree 0 have no contribution to structural information and von Neumann graph entropy, without loss of generality we assume that $d_i > 0$ for any node $i \in V$.

4.1 Bounds on the Approximation Error

We first provide the additive approximation errors in Theorem 4.1, Corollary 4.5, and Corollary 4.6, then obtain the multiplicative approximation error in Theorem 4.7.

Theorem 4.1 (Bounds on the absolute approximation error). *For any undirected graph $G = (V, E, A)$, the inequality*

$$0 \leq \Delta\mathcal{H}(G) \leq \frac{\log_2 e}{\delta} \cdot \frac{\text{tr}(A^2)}{\text{vol}(G)} \tag{1}$$

holds, where $\delta = \min\{d_i | d_i > 0\}$ is the minimum positive degree.

Before proving Theorem 4.1, we introduce two techniques: majorization and Jensen's gap. The former one is a preorder of the vector of reals, while the latter is an inverse version of the Jensen's inequality, whose definitions are presented as follows.

Definition 4.2 (Majorization [32]). For a vector $\mathbf{x} \in \mathbb{R}^d$, we denote by $\mathbf{x}^{\downarrow} \in \mathbb{R}^d$ the vector with the same components, but sorted in descending order. Given $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$, we say that $\mathbf{x}$ majorizes $\mathbf{y}$ (written as $\mathbf{x} \succ \mathbf{y}$) 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 $\mathbf{x}^T\mathbf{1} = \mathbf{y}^T\mathbf{1}$.

Lemma 4.3 (Jensen's gap [29]). *Let X be a one-dimensional random variable with mean μ and support Ω. Let $\psi(x)$ be a twice differentiable function on Ω and define function $h(x) = \frac{\psi(x)-\psi(\mu)}{(x-\mu)^2} - \frac{\psi'(\mu)}{x-\mu}$, then $\mathbb{E}[\psi(X)] - \psi(\mathbb{E}[X]) \leq \sup_{x\in\Omega}\{h(x)\} \cdot \text{var}(X)$. Additionally, if $\psi'(x)$ is convex, then $h(x)$ is monotonically increasing in x, and if $\psi'(x)$ is concave, then $h(x)$ is monotonically decreasing in x.*

Lemma 4.4. *The function $f(x) = x\log_2 x$ is convex, its first order derivative $f'(x) = \log_2 x + \log_2 e$ is concave.*

Proof. The second order derivative $f''(x) = (\log_2 e)/x > 0$, thus $f(x) = x\log_2 x$ is convex. □

We can see that the majorization characterizes the degree of concentration between two vectors, $\mathbf{x} \succ \mathbf{y}$ means that the entries of $\mathbf{y}$ are more concentrated on its mean $\mathbf{y}^T\mathbf{1}/\mathbf{1}^T\mathbf{1}$ than the entires of $\mathbf{x}$. An equivalent definition of the majorization [32] using linear algebra says that $\mathbf{x} \succ \mathbf{y}$ if and only if there exists a doubly stochastic matrix P such that $P\mathbf{x} = \mathbf{y}$. As a famous example of the majorization, the Schur-Horn theorem [32] says that the diagonal elements of a positive semidefinite Hermitian matrix are majorized by its eigenvalues. Since $\mathbf{x}^T L\mathbf{x} = \sum_{(i,j)\in E} A_{ij}(x_i - x_j)^2 \geq 0$ for any vector $\mathbf{x} \in \mathbb{R}^n$, the Laplacian matrix L is a positive semidefinite symmetric matrix whose diagonal elements form the degree sequence $\mathbf{d}$ and eigenvalues form the spectrum $\boldsymbol{\lambda}$. Therefore, $\boldsymbol{\lambda} \succ \mathbf{d}$ implying that there exists some doubly stochastic matrix $P = (p_{ij}) \in [0,1]^{n\times n}$ such that $P\boldsymbol{\lambda} = \mathbf{d}$.

Using the fact that $P\boldsymbol{\lambda} = \mathbf{d}$ and the convexity of $f(x)$ in Lemma 4.4, we can now proceed to prove Theorem 4.1.

Proof of Theorem 4.1. For each $i \in V$, we define a discrete random variable X_i with probability mass function $\sum_{j=1}^{n} p_{ij}\delta_{\lambda_j}(x)$, where $\delta_a(x)$ is the Kronecker delta function. Then the expectation $\mathbb{E}[X_i] = \sum_{j=1}^{n} p_{ij}\lambda_j = d_i$ and the variance $\text{var}(X_i) = \sum_{j=1}^{n} p_{ij}(\lambda_j - d_i)^2 = \sum_{j=1}^{n} p_{ij}\lambda_j^2 - d_i^2$.

First, we express the entropy gap in terms of the Lapalcian spectrum and the degree sequence. Since

$$\begin{aligned}\mathcal{H}_1(G) &= -\sum_{i=1}^{n}\left(\frac{d_i}{\text{vol}(G)}\right)\log_2\left(\frac{d_i}{\text{vol}(G)}\right)\\ &= -\frac{1}{\text{vol}(G)}\left(\sum_{i=1}^{n} f(d_i) - \sum_{i=1}^{n} d_i\log_2(\text{vol}(G))\right)\\ &= \log_2(\text{vol}(G)) - \frac{\sum_{i=1}^{n} f(d_i)}{\text{vol}(G)},\end{aligned} \tag{2}$$

and similarly $\mathcal{H}_{\text{vn}}(G) = \log_2(\text{vol}(G)) - \sum_{i=1}^{n} f(\lambda_i)/\text{vol}(G)$, we have

$$\Delta\mathcal{H}(G) = \mathcal{H}_1(G) - \mathcal{H}_{\text{vn}}(G) = \frac{\sum_{i=1}^{n} f(\lambda_i) - \sum_{i=1}^{n} f(d_i)}{\text{vol}(G)}. \tag{3}$$

Second, we use Jensen's inequality to prove $\Delta\mathcal{H}(G) \geq 0$. Since $f(x)$ is convex, $f(d_i) = f(\mathbb{E}[X_i]) \leq \mathbb{E}[f(X_i)]$ for any $i \in \{1, \ldots, n\}$. By summing over i, we have

$$\sum_{i=1}^{n} f(d_i) \leq \sum_{i=1}^{n}\mathbb{E}[f(X_i)] = \sum_{i=1}^{n}\sum_{j=1}^{n} p_{ij}f(\lambda_j) = \sum_{j=1}^{n} f(\lambda_j).$$

Therefore, $\Delta\mathcal{H}(G) \geq 0$ for any undirected graphs.

Finally, we use Jensen's gap to prove $\Delta\mathcal{H}(G) \leq \frac{\log_2 e}{\delta}\frac{\text{tr}(A^2)}{\text{vol}(G)}$. Apply the Jensen's gap to X_i and $f(x)$,

$$\mathbb{E}[f(X_i)] - f(\mathbb{E}[X_i]) \leq \sup_{x\in[0,\text{vol}(G)]}\{h_i(x)\} \cdot \text{var}(X_i), \tag{4}$$

where

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

Since $f'(x)$ is concave, $h_i(x)$ is monotonically decreasing in x. Therefore, $\sup_{x\in[0,\text{vol}(G)]}\{h_i(x)\} = h_i(0)$. Since

$$h_i(0) = \frac{f(0) - f(d_i)}{d_i^2} + \frac{f'(d_i)}{d_i} = \frac{\log_2 e}{d_i} \leq \frac{\log_2 e}{\delta},$$

the inequality in (4) can be simplified as

$$\sum_{j=1}^{n} p_{ij} f(\lambda_j) - f(d_i) \leq \frac{\log_2 e}{\delta} \cdot \left(\sum_{j=1}^{n} p_{ij} \lambda_j^2 - d_i^2 \right). \tag{5}$$

By summing both sides of the inequality (5) over i, we get an upper bound UB on $\sum_{j=1}^{n} f(\lambda_j) - \sum_{i=1}^{n} f(d_i)$ as

$$\begin{aligned} \text{UB} &= \frac{\log_2 e}{\delta} \cdot \sum_{i=1}^{n} \left(\sum_{j=1}^{n} p_{ij} \lambda_j^2 - d_i^2 \right) = \frac{\log_2 e}{\delta} \cdot \left(\sum_{j=1}^{n} \lambda_j^2 - \sum_{i=1}^{n} d_i^2 \right) \\ &= \frac{\log_2 e}{\delta} \cdot \Big(\text{tr}(L^2) - \text{tr}(D^2) \Big) \\ &= \frac{\log_2 e}{\delta} \cdot \Big(\text{tr}(A^2) - \text{tr}(AD) - \text{tr}(DA) \Big) = \frac{\log_2 e}{\delta} \cdot \text{tr}(A^2) \end{aligned}$$

As a result, $\Delta\mathcal{H}(G) = \frac{\sum_{i=1}^{n} f(\lambda_i) - \sum_{i=1}^{n} f(d_i)}{\text{vol}(G)} \leq \frac{\log_2 e}{\delta} \frac{\text{tr}(A^2)}{\text{vol}(G)}$. □

To illustrate the tightness of the bounds in Theorem 4.1, we further derive bounds on the entropy gap for unweighted graphs, especially the regular graphs. Via multiplicative error analysis, we show that the structural information converges to the von Neumann graph entropy as graph size grows.

Corollary 4.5 (Constant bounds on the entropy gap). *For any unweighted, undirected graph G, $0 \leq \Delta\mathcal{H}(G) \leq \log_2 e$ holds.*

Proof. In unweighted graph G, $\text{tr}(A^2) = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} A_{ji} = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} = \sum_{i=1}^{n} d_i = \text{vol}(G)$ and $\delta \geq 1$, therefore $0 \leq \Delta\mathcal{H}(G) \leq \frac{\log_2 e}{\delta} \frac{\text{tr}(A^2)}{\text{vol}(G)} = \frac{\log_2 e}{\delta} \leq \log_2 e$. □

Corollary 4.6 (Entropy gap of regular graphs). *For any unweighted, undirected, regular graph G_d of degree d, the inequality $0 \leq \Delta\mathcal{H}(G_d) \leq \frac{\log_2 e}{d}$ holds.*

Proof sketch. In any unweighted, regular graph G_d, $\delta = d$. □

Theorem 4.7 (Convergence of the multiplicative approximation error). *For almost all unweighted graphs G of order n, $\frac{\mathcal{H}_1(G)}{\mathcal{H}_{\text{vn}}(G)} - 1 \geq 0$ and decays to 0 at the rate of $o(1/\log_2(n))$.*

Proof. Dairyko et al. [10] proved that for almost all unweighted graphs G of order n, $\mathcal{H}_{\text{vn}}(G) \geq \mathcal{H}_{\text{vn}}(K_{1,n-1})$ where $K_{1,n-1}$ stands for the star graph. Since $\mathcal{H}_{\text{vn}}(K_{1,n-1}) = \log_2(2n-2) - \frac{n}{2n-2} \log_2 n = 1 + \frac{1}{2} \log_2 n + o(1)$, $\frac{\mathcal{H}_1(G)}{\mathcal{H}_{\text{vn}}(G)} - 1 = \frac{\Delta\mathcal{H}(G)}{\mathcal{H}_{\text{vn}}(G)} \leq \frac{\log_2 e}{\mathcal{H}_{\text{vn}}(K_{1,n-1})} = o(\frac{1}{\log_2 n})$. □

4.2 Sharpened Bounds on the Entropy Gap

Though the constant bounds on the entropy gap is tight enough for applications, we can still sharpen the bounds on the entropy gap in unweighted graphs using more advanced majorizations.

Theorem 4.8 (Sharpened lower bound on entropy gap). *For any unweighted, undirected graph G, $\Delta\mathcal{H}(G)$ is lower bounded by $(f(d_{\max}+1) - f(d_{\max}) + f(\delta-1) - f(\delta))/\text{vol}(G)$ where $d_{\max}$ is the maximum degree and δ is the minimum positive degree.*

Proof. The proof is based on the advanced majorization [18]: $\boldsymbol{\lambda} \succ (d_1 + 1, d_2, \ldots, d_n - 1)$ holds on any unweighted, undirected graph G where $d_1 \geq d_2 \geq \cdots \geq d_n$ is the sorted degree sequence of G. Similar to the proof of Theorem 4.1, we have $\sum_{i=1}^{n} f(\lambda_i) \geq f(d_1+1) + f(d_n-1) + \sum_{i=2}^{n-1} f(d_i)$. Then the sharpened upper bound follows from the equation (3) since $d_1 = d_{\max}$ and $d_n = \delta$. □

Theorem 4.9 (Sharpened upper bound on entropy gap). *For any unweighted, undirected graph $G = (V, E)$, $\Delta\mathcal{H}(G)$ is upper bounded by $\min\{\log_2 e, b_1, b_2\}$ where $b_1 = \frac{\sum_{i=1}^{n} f(d_i^*)}{\text{vol}(G)} - \frac{\sum_{i=1}^{n} f(d_i)}{\text{vol}(G)}$ and $b_2 = \log_2(1 + \sum_{i=1}^{n} d_i^2/\text{vol}(G)) - \frac{\sum_{i=1}^{n} f(d_i)}{\text{vol}(G)}$. Here $(d_1^*, \ldots, d_n^*)$ is the conjugate degree sequence of G where $d_k^* = |\{i | d_i \geq k\}|$.*

Proof. We first prove $\Delta\mathcal{H}(G) \leq b_1$ using the Grone-Merris majorization [1]: $(d_1^*, \ldots, d_n^*) \succ \boldsymbol{\lambda}$. Similar to the proof of Theorem 4.1, we have $\sum_{i=1}^{n} f(d_i^*) \geq \sum_{i=1}^{n} f(\lambda_i)$, thus $b_1 \geq \frac{\sum_{i=1}^{n} f(\lambda_i) - \sum_{i=1}^{n} f(d_i)}{\text{vol}(G)} = \Delta\mathcal{H}(G)$. We then prove $\Delta\mathcal{H}(G) \leq b_2$. Since

$$\frac{\sum_{i=1}^{n} f(\lambda_i)}{\text{vol}(G)} = \sum_{i=1}^{n} \left(\frac{\lambda_i}{\sum_{j=1}^{n} \lambda_j} \right) \log_2 \lambda_i \leq \log_2 \left(\frac{\sum_{i=1}^{n} \lambda_i^2}{\sum_{j=1}^{n} \lambda_j} \right)$$

and $\frac{\sum_{i=1}^{n} \lambda_i^2}{\sum_{i=1}^{n} \lambda_i} = \frac{\text{tr}(L^2)}{\text{vol}(G)} = 1 + \frac{\sum_{i=1}^{n} d_i^2}{\text{vol}(G)}$, $\Delta\mathcal{H}(G) = \frac{\sum_{i=1}^{n} f(\lambda_i) - f(d_i)}{\text{vol}(G)} \leq b_2$. □

5 APPLICATIONS AND ALGORITHMS

As a measure of the structural complexity of a graph, the von Neumann entropy has been applied in a variety of applications. For example, the von Neumann graph entropy is exploited to measure the importance of an edge [30]. As another example, the von Neumann graph entropy can also be used to measure the distance between graphs for graph classification and anomaly detection [2, 7]. In addition, the von Neumann graph entropy is used in the context of network embedding [11] to learn low-dimensional feature representations of nodes. We observe that, in these applications, the von Neumann graph entropy is used to address the following primitive tasks:

- **Entropy-based network design**: Change the existing graph to a new graph such that the entropy requirement is attained with minimal perturbations on the existing graph. For example, Minello et al. [33] use the von Neumann entropy to explore the potential network growth model via experiments.
- **Graph similarity measure**: Compute a real positive number to reveal the similarity between two graphs. For example, Domenico et al. [15] use the von Neumann graph entropy to compute the Jensen-Shannon distance between graphs for the purpose of compressing multilayer networks.

To resolve both tasks, it requires computing the von Neumann graph entropy exactly. To reduce the computational cost and preserve the interpretability, we can use the accurate proxy, structural information, to approximately solve these tasks.

5.1 Entropy-based network design

Network design aims to minimally perturb the network to fulfill some goals. Consider a goal to maximize the von Neumann entropy of a graph, it helps to understand how different structural patterns

influence the entropy value. The entropy-based network design problem is formulated as follows,

Problem 1 (MaxEntropy). *Given an unweighted, undirected graph $G = (V, E)$ of order n and an integer budget k, find a set F of non-existing edges of G whose addition to G creates the largest increase of the von Neumann graph entropy and $|F| \leq k$.*

Due to the spectral nature of the von Neumann graph entropy, it is not easy to find an effective strategy to perturb the graph, especially in the scenario where there are exponential number of combinations for the subset F. If we use the structural information as a proxy of the von Neumann entropy, Problem 1 reduces to maximize $\mathcal{H}_1(G')$ where $G' = (V, E \cup F)$ such that $|F| \leq k$. To further alleviate the computational pressure rooted in the exponential size of the search space for F, we adopt the greedy method in which the new edges are added one by one until either the structural information attains its maximum value $\log_2 n$ or k new edges have already been added. We denote the graph with l new edges as $G_l = (V, E_l)$, then $G_0 = G$. Now suppose that we have G_l whose structural information is less than $\log_2 n$, then we want to find a new edge $e_{l+1} = (u, v)$ such that $\mathcal{H}_1(G_{l+1})$ is maximized, where $G_{l+1} = (V, E_l \cup \{e_{l+1}\})$. Since $\mathcal{H}_1(G_{l+1})$ can be rewritten as

$$\log_2(2|E_l| + 2) - \frac{1}{2|E_l| + 2}\left(f(d_u + 1) + f(d_v + 1) + \sum_{i \neq u, v} f(d_i)\right),$$

the edge e_{l+1} maximizing $\mathcal{H}_1(G_{l+1})$ should also minimize the edge centrality $EC(u, v) = f(d_u + 1) - f(d_u) + f(d_v + 1) - f(d_v)$, where d_i is the degree of node i in G_l.

We present the pseudocode of our fast algorithm EntropyAug in Algorithm 1, which leverages the pruning strategy to accelerate the process of finding a single new edge that creates a largest increase of the von Neumann entropy. EntropyAug starts by initiating an empty set F used to contain the node pairs to be found and an entropy value $\mathcal{H}$ used to record the maximum structural information in the graph evolution process (line 1). In each following iteration, it sorts the set of nodes V in non-decreasing degree order (line 3). Note that the edge centrality $EC(u, v)$ has a nice monotonic property: $EC(u_1, v_1) \leq EC(u_2, v_2)$ if $\min\{d_{u_1}, d_{v_1}\} \leq \min\{d_{u_2}, d_{v_2}\}$ and $\max\{d_{u_1}, d_{v_1}\} \leq \max\{d_{u_2}, d_{v_2}\}$. With the sorted list of nodes V_s, the monotonicity of $EC(u, v)$ can be translated into $EC(V_s[i_1], V_s[j_1]) \leq EC(V_s[i_2], V_s[j_2])$ if the indices satisfy $i_1 < j_1$, $i_2 < j_2$, $i_1 < i_2$, and $j_1 < j_2$. Thus, using the two pointers {head, tail} and a threshold T, it can prune the search space and find the desired non-adjacent node pair as fast as possible (line 4-12). It then adds the non-adjacent node pair minimizing $EC(u, v)$ into F and update the graph G (line 13). The structural information of the updated graph is computed to determine whether F is the optimal subset till current iteration (line 14-15).

Algorithm 1: EntropyAug

Input: The graph $G = (V, E)$ of order n, the budget k
Output: A set of node pairs

1 $F \leftarrow \varnothing, \mathcal{H} \leftarrow 0$;
2 **while** $|F| < k$ **do**
3 V_s: list $\leftarrow$ sort V in non-decreasing degree order;
4 head $\leftarrow 0$, tail $\leftarrow |V_s| - 1$, $T \leftarrow +\infty$;
5 **while** head $<$ tail **do**
6 **for** i = head + 1, head + 2, ..., tail **do**
7 **if** $EC(V_s[\text{head}], V_s[i]) \geq T$ **then**
8 tail $\leftarrow i - 1$; **break**;
9 **if** $(V_s[\text{head}], V_s[i]) \notin E$ **then**
10 $u \leftarrow V_s[\text{head}]$, $v \leftarrow V_s[i]$, $T \leftarrow EC(u, v)$;
11 tail $\leftarrow i - 1$; **break**;
12 head $\leftarrow$ head + 1;
13 $E \leftarrow E \cup \{(u, v)\}$, $F \leftarrow F \cup \{(u, v)\}$;
14 **if** $\mathcal{H}_1(G) > \mathcal{H}$ **then** $\mathcal{H} \leftarrow \mathcal{H}_1(G)$, $F^* \leftarrow F$;
15 **if** $\mathcal{H} = \log_2 n$ **then break**;
16 **return** F^*.

5.2 Graph Similarity Measure

Entropy based graph similarity measure aims to compare graphs using Jensen-Shannon divergence. The Jensen-Shannon divergence, as a symmetrized and smoothed version of the Kullback-Leibler divergence, is defined formally in the following Definition 5.1.

Definition 5.1 (Jensen-Shannon divergence). Let P and Q be two probability distributions on the same support set $\Omega_N = \{1, \ldots, N\}$. The Jensen-Shannon divergence between P and Q is defined as

$$\mathcal{D}_{\text{JS}}(P, Q) = H((P + Q)/2) - H(P)/2 - H(Q)/2,$$

where $H(P) = -\sum_{i=1}^{N} p_i \log p_i$ is the entropy of the distribution P.

Endres et al. [16] prove that $\sqrt{\mathcal{D}_{\text{JS}}(P, Q)}$ is a bounded metric on the space of distributions over Ω_N with its maximum value $\sqrt{\log 2}$ being attained when $\min\{p_i, q_i\} = 0$ for every $i \in \Omega_N$. Since the von Neumann graph entropy is an entropy of the spectrum based distribution, Lamberti et al. [25] define a quantum Jensen-Shannon distance between two graphs which is closely related to the von Neumann graph entropy in the following Definition 5.2.

Definition 5.2 (Quantum Jensen-Shannon distance). The quantum Jensen-Shannon distance between two weighted, undirected graphs $G_1 = (V, E_1, A_1)$ and $G_2 = (V, E_2, A_2)$ is defined as $\mathcal{D}_{\text{QJS}}(G_1, G_2) = \sqrt{\mathcal{H}_{\text{vn}}(\overline{G}) - (\mathcal{H}_{\text{vn}}(G_1) + \mathcal{H}_{\text{vn}}(G_2))/2}$, where $\overline{G} = (V, E_1 \cup E_2, \overline{A})$ is an weighted graph with $\overline{A} = A_1/2\text{vol}(G_1) + A_2/2\text{vol}(G_2)$.

Based on the quantum Jensen-Shannon distance, we consider the following problem that can be applied in anomaly detection and multiplex network compression,

Problem 2. *Compute the quantum Jensen-Shannon distance between adjacent graphs in a stream of graphs $\{G_k = (V, E_k, t_k)\}_{k=1}^{K}$ where t_k is the timestamp of the graph G_k and $t_k < t_{k+1}$.*

As a distance measure between graphs, $\mathcal{D}_{\text{QJS}}$ is typically required to be a pseudometric [39], that is, it should be symmetric and satisfy triangle inequality. However, to the best of our knowledge, it is still an open problem whether $\mathcal{D}_{\text{QJS}}$ fulfills the triangle inequality [25]. Meanwhile, the quantum Jensen-Shannon distance inherits the computational inefficiency from the von Neumann graph entropy. Therefore, to solve Problem 2 efficiently we propose a new distance measure based on structural information as a surrogate for $\mathcal{D}_{\text{QJS}}$.

Algorithm 2: IncreSim

Input: G_1 and $\{\Delta G_k\}_{k=1}^{K-1}$
Output: $\{\mathcal{D}_{\mathrm{SI}}(G_k, G_{k+1})\}_{k=1}^{K-1}$

1 $d \leftarrow$ the degree sequence of the graph G_1;
2 $m \leftarrow \sum_{i=1}^{n} d_i/2$;
3 $\mathcal{H}_1(G_1) \leftarrow \log_2(2m) - \frac{1}{2m}\sum_{i=1}^{n} f(d_i)$;
4 **for** $k = 1, \ldots, K-1$ **do**
5 $\quad \Delta d \leftarrow$ the degree sequence of the signed graph ΔG_k;
6 $\quad \Delta m \leftarrow \sum_{i \in V_k} \Delta d_i/2$;
7 $\quad$ Compute a, b, y, z in Lemma 5.6 via iterating over V_k;
8 $\quad$ Compute $\mathcal{H}_1(G_{k+1})$ and $\mathcal{H}_1(\bar{G}_k)$ based on Lemma 5.6;
9 $\quad \mathcal{D}_{\mathrm{SI}}(G_k, G_{k+1}) \leftarrow \sqrt{\mathcal{H}_1(\bar{G}_k) - (\mathcal{H}_1(G_k) + \mathcal{H}_1(G_{k+1}))/2}$;
10 $\quad m \leftarrow m + \Delta m$;
11 $\quad$ **foreach** $i \in V_k$ **do** $d_i \leftarrow d_i + \Delta d_i$;
12 **return** $\{\mathcal{D}_{\mathrm{SI}}(G_k, G_{k+1})\}_{k=1}^{K-1}$

Definition 5.3 (Structural information distance). The structural information distance between two weighted, undirected graphs $G_1 = (V, E_1, A_1)$ and $G_2 = (V, E_2, A_2)$ is defined as $\mathcal{D}_{\mathrm{SI}}(G_1, G_2) = \sqrt{\mathcal{H}_1(\bar{G}) - (\mathcal{H}_1(G_1) + \mathcal{H}_1(G_2))/2}$, where $\bar{G} = (V, E_1 \cup E_2, \bar{A})$ is an weighted graph with $\bar{A} = A_1/2\mathrm{vol}(G_1) + A_2/2\mathrm{vol}(G_2)$.

It is a little surprising to find that $\mathcal{D}_{\mathrm{SI}}$ is a pseudometric, the details of which are stated in Theorem 5.4.

Theorem 5.4 (Properties of the distance measure $\mathcal{D}_{\mathrm{SI}}$). *The distance measure $\mathcal{D}_{\mathrm{SI}}(G_1, G_2)$ is a pseudometric on the space of undirected graphs:*

- *$\mathcal{D}_{\mathrm{SI}}$ is symmetric, i.e., $\mathcal{D}_{\mathrm{SI}}(G_1, G_2) = \mathcal{D}_{\mathrm{SI}}(G_2, G_1)$;*
- *$\mathcal{D}_{\mathrm{SI}}$ is non-negative, i.e., $\mathcal{D}_{\mathrm{SI}}(G_1, G_2) \geq 0$ where the equality holds if and only if $\frac{d_{i,1}}{\sum_{k=1}^{n} d_{k,1}} = \frac{d_{i,2}}{\sum_{k=1}^{n} d_{k,2}}$ for every node $i \in V$ where $d_{i,j}$ is the degree of node i in G_j;*
- *$\mathcal{D}_{\mathrm{SI}}$ obeys the triangle inequality, i.e.,*

$$\mathcal{D}_{\mathrm{SI}}(G_1, G_2) + \mathcal{D}_{\mathrm{SI}}(G_2, G_3) \geq \mathcal{D}_{\mathrm{SI}}(G_1, G_3);$$

- *$\mathcal{D}_{\mathrm{SI}}$ is upper bounded by 1, i.e., $\mathcal{D}_{\mathrm{SI}}(G_1, G_2) \leq 1$ where the equality holds if and only if $\min\{d_{i,1}, d_{i,2}\} = 0$ for every node $i \in V$ where $d_{i,j}$ is the degree of node i in G_j.*

To establish a connection between $\mathcal{D}_{\mathrm{SI}}$ and $\mathcal{D}_{\mathrm{QJS}}$, we study their extreme values and present the results in Theorem 5.5.

Theorem 5.5 (Connection between $\mathcal{D}_{\mathrm{QJS}}$ and $\mathcal{D}_{\mathrm{SI}}$). *Both $\mathcal{D}_{\mathrm{QJS}}(G_1, G_2)$ and $\mathcal{D}_{\mathrm{SI}}(G_1, G_2)$ attain the same maximum value of 1 under the identical condition that $\min\{d_{i,1}, d_{i,2}\} = 0$ for every node $i \in V$ where $d_{i,j}$ is the degree of node i in G_j.*

In order to compute the structural information distance between adjacent graphs in the graph stream $\{G_k = (V, E_k, t_k)\}_{k=1}^{K}$, we first compute the structural information $\mathcal{H}_1(G_k)$ for each $k \in \{1, \ldots, K\}$, which takes $\Theta(Kn)$ time. Then we compute the structural information of $\bar{G}_k$ whose adjacent matrix $\bar{A}_k = A_k/2\mathrm{vol}(G_k) + A_{k+1}/2\mathrm{vol}(G_{k+1})$ for each $k \in \{1, \ldots, K-1\}$. Since the degree of node i in $\bar{G}_k$ is $\bar{d}_{i,k} = \frac{d_{i,k}}{2\mathrm{vol}(G_k)} + \frac{d_{i,k+1}}{2\mathrm{vol}(G_{k+1})}$ and $\sum_{i=1}^{n} \bar{d}_{i,k} = 1$, the structural information of $\bar{G}_k$ is $\mathcal{H}_1(\bar{G}_k) = -\sum_{i=1}^{n} f(\bar{d}_{i,k})$ which takes $\Theta(n)$ time for each k. Therefore, the total computational cost is $\Theta((2K-1)n)$.

In practice, the graph stream is fully dynamic such that it would be more efficient to represent the graph stream as a stream of edge insertions and deletions over time, rather than a sequence of graphs. Suppose that the graph stream is represented as an initial base graph $G_1 = (V, E_1, t_1)$ and a sequence of graph changes $\{\Delta G_k = (V_k, E_{+,k}, E_{-,k}, t_k)\}_{k=1}^{K-1}$ where t_k is the timestamp of the set $E_{+,k}$ of edge insertions and the set $E_{-,k}$ of edge deletions, and V_k is the subset of nodes covered by $E_{+,k} \cup E_{-,k}$. We can view the graph change ΔG_k as a signed network where the edge in $E_{+,k}$ has positive weight $+1$ and the edge in $E_{-,k}$ has negative weight -1. The degree of node $i \in V_k$ in the graph change ΔG_k refers to $\sum_{j \in V_k} \mathbb{I}\{(i,j) \in E_{+,k}\} - \mathbb{I}\{(i,j) \in E_{-,k}\}$. Using the information about previous graph G_k and current graph change ΔG_k, we can compute the entropy statistics of the current graph G_{k+1} incrementally and efficiently via the following lemma, whose proof can be found in the appendix.

Lemma 5.6. *Using the degree sequence d of the graph G_k, the structural information $\mathcal{H}_1(G_k)$, and the degree sequence Δd of the signed graph ΔG_k, the structural information of the graph G_{k+1} can be efficiently computed as*

$$\mathcal{H}_1(G_{k+1}) = \frac{f(2(m+\Delta m)) - a - f(2m) + 2m\mathcal{H}_1(G_k)}{2(m+\Delta m)},$$

where $m = \sum_{i=1}^{n} d_i/2$, $\Delta m = \sum_{i \in V_k} \Delta d_i/2$, and $a = \sum_{i \in V_k} f(d_i + \Delta d_i) - f(d_i)$. Moreover, the structural information of the averaged graph $\bar{G}_k$ between G_k and G_{k+1} can be efficiently computed as

$$\mathcal{H}_1(\bar{G}_k) = -b - (2m - y)f(c) - c(f(2m) - 2m\mathcal{H}_1(G_k) - z),$$

where $y = \sum_{i \in V_k} d_i$, $z = \sum_{i \in V_k} f(d_i)$, $c = \frac{2m+\Delta m}{4m(m+\Delta m)}$, and $b = \sum_{i \in V_k} f\left(\frac{d_i}{4m} + \frac{d_i + \Delta d_i}{4(m+\Delta m)}\right)$.

The pseudocode of our fast algorithm IncreSim for computing the structural information distance in a graph stream is shown in Algorithm 2. It starts by computing the structural information of the base graph G_1 (line 1-3), which takes $\Theta(n)$ time. In each following iteration, it first computes the value of a, b, c, y, z (line 5-7), then calculates the structural information distance between two adjacent graphs (line 8-9), finally updates the edge count m and the degree sequence d (line 10-11). The time cost of each iteration is $\Theta(|V_k|)$, consequently the total time complexity is $\Theta(n + \sum_{k=1}^{K-1} |V_k|)$.

6 EXPERIMENTS AND EVALUATIONS

We conduct extensive experiments over both synthetic and real-world datasets to answer the following questions:

Q1. **Universality** of the entropy gap over arbitrary simple graphs: Is the entropy gap close to 0 for a wide range of graphs? Is the structural information a good proxy of the von Neumann graph entropy for a wide range of graphs?
Q2. **Sensitivity** of the entropy gap to graph properties: How do graph properties affect the value of entropy gap?
Q3. **Accuracy** of the approximation: As a proxy of the von Neumann graph entropy, is the structural information more accurate than its prominent competitors?

Table 2: Real-world datasets used in our experiments.

Name	#Nodes	#Edges	Category	Statistics
Static graphs without timestamps				Avg. degree
Zachary (ZA)	34	78	Friendship	4.59
Dolphins (DO)	62	159	Animal	5.13
Jazz (JA)	198	2,742	Contact	27.70
Skitter (SK)	1,696,415	11,095,298	Internet	13.08
Brightkite (BK)	58,228	214,078	Friendship	7.35
Caida (CA)	26,475	53,381	Internet	4.03
YouTube (YT)	1,134,890	2,987,624	Friendship	5.27
LiveJournal (LJ)	3,997,962	34,681,189	Friendship	17.35
Pokec (PK)	1,632,803	22,301,964	Friendship	27.32
Dynamic graphs with timestamps				#Snapshots
Wiki-IT (WK)	1,204,009	34,826,283	Hyperlink	100
Facebook (FB)	61,096	788,135	Friendship	29

Q4. **Speed** of the computation: Is the computation of the structural information faster than its prominent competitors?

Q5. **Extensibility** of the entropy gap to weighted graphs: Is the entropy gap sensitive to the change of edge weights? Is the entropy gap still close to 0 for weighted graphs?

Q6. **Performance analysis (Appendix A)**: What is the performance of EntropyAug (Algorithm 1) in maximizing the von Neumann graph entropy? What is the performance of IncreSim (Algorithm 2) in analyzing graph streams? Can the structural information distance be further used to detect anomalies in a graph stream?

6.1 Experimental Settings

Datasets: We consider both synthetic graphs and real-world graphs. The synthetic graphs are generated from three well-known random graph models: Erdös-Rényi (ER) model, Barabási-Albert (BA) model [3], and Watts-Strogatz (WS) model [43]. The real-world graphs [24, 27, 42] used in our experiments are listed in Table 2, which contain both static graphs with varying size and average degree, and temporal graphs with varying size and time span. In every static graph, we ignore the direction and weight of all edges and remove both self-loops and multiple edges. We treat every temporal graph as a stream of undirected weighted edges with timestamps. For the convenience of analysis, we partition these edges into several groups where each group is within a certain time interval.

Hardwares: The experiments have been performed on a server with Intel(R) Xeon(R) CPU 2.40 GHz (32 virtual cores) and 256GB RAM, averaging 10 runs for random algorithms and random inputs unless stated otherwise.

Implementation: All of the proposed algorithms and baselines are implemented in Python.

Reproducibility: The code and datasets used in the paper are available at https://github.com/xuecheng27/WWW21-Structural-Information.

6.2 Q1. Universality (Fig. 2)

To evaluate the universality of the entropy gap, we measure the structural information and exact von Neumann entropy on a set of synthetic graphs with 2,000 nodes. For the ER and BA models, we generate graphs with average degree in $\{2, 4, \ldots, 200\}$. For the WS model, we generate graphs with edge rewiring probability in $\{0, 1/20, \ldots, 1\}$ for each average degree in $\{6, 10, 20, 50\}$. We additionally measure the sharpened lower and upper bounds of the entropy gap. The results are shown in Fig. 2.

The observations are three fold. First, **the entropy gap is close to 0 for a wide range of graphs**. The entropy gap of each synthetic graph is no more than 0.2, whereas the exact von Neumann entropy is greater than 10. Second, **the entropy gap is negatively correlated with the average degree**. Dense graph tends to have very small entropy gap. Third, **the structural information is linearly correlated with the von Neumann graph entropy**, with only few exceptions. There is no exception for the ER synthetic graphs. For the BA synthetic graphs, the exceptions are those graphs with extremely small average degree. For the WS synthetic graphs, the exceptions are those graphs with extremely small edge rewiring probability.

6.3 Q2. Sensitivity (Fig. 2, Fig. 3)

To evaluate the sensitivity of the entropy gap to graph properties such as average degree, graph size, and rewiring probability, we further measure the entropy gap of 10 synthetic graphs with average degree in $\{500, 1000, \ldots, 5000\}$ for each random model. The average degree is chosen from $\{2, 5, 10, 20, 50, 100\}$ for ER and BA models, and the edge rewiring probability is chosen from $\{0, 0.1, 0.2, 0.4, 0.8, 1\}$ for the WS model.

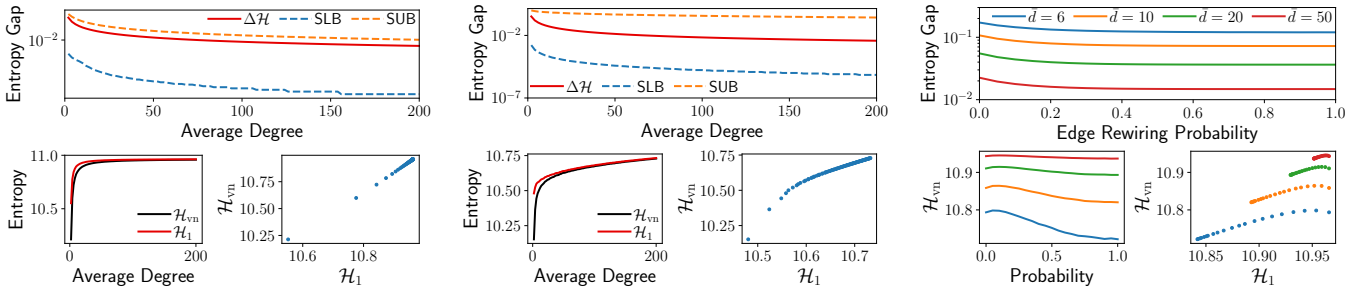
The observations from Fig. 2 and Fig. 3 are three fold. First, the entropy gap decreases as the average degree increases for all the three random graph models. Second, the entropy gap decreases as the edge rewiring probability increases for the WS model. Third, **the entropy gap is nearly insensitive to the change of graph size**.

6.4 Q3. Accuracy (Fig. 4)

To evaluate the accuracy of the structural information as an approximation of the von Neumann graph entropy, we measure the structural information, exact von Neumann entropy (when the graph size is small), and three prominent approximations (as competitors) in 9 real-world static graphs. The competitors are 1) FINGER-$\widehat{H}$ [7] defined as $\widehat{\mathcal{H}}_{\mathrm{F}}(G) = -Q\log_2(\lambda_{\max}/\mathrm{tr}(L))$ where $Q = 1 - \mathrm{tr}(L^2)/\mathrm{tr}^2(L)$, 2) FINGER-$\widetilde{H}$ [7] defined as $\widetilde{\mathcal{H}}_{\mathrm{F}}(G) = -Q\log_2(2d_{\max}/\mathrm{tr}(L))$, and 3) SLaQ [40]. The results in Fig. 4 show that **the structural information is an accurate approximation of the von Neumann graph entropy**. The approximation error of structural information is obviously much smaller than $\widehat{\mathcal{H}}_{\mathrm{F}}$ and $\widetilde{\mathcal{H}}_{\mathrm{F}}$. And it is comparable to the approximation error of SLaQ with only few exceptions such as YT and SK where the structural information is slightly better.

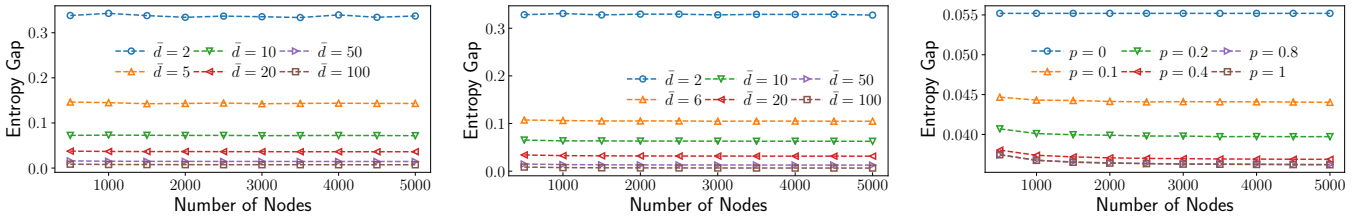
6.5 Q4. Speed (Fig. 5)

To evaluate computational speed of the structural information, we measure the running time of structural information and its three competitors in 9 real-world static graphs. The results in Fig. 5 show that **the computation of structural information is fast**. It is about 2 orders of magnitude faster than $\widehat{\mathcal{H}}_{\mathrm{F}}$, at least 2 orders of magnitude faster than SLaQ, and comparable to $\widetilde{\mathcal{H}}_{\mathrm{F}}$. Combining



(a) ER model (b) BA model (c) WS model

Figure 2: The structural information, von Neumann graph entropy, and entropy gap of synthetic graphs generated from three random graph models with $2,000$ nodes, varying average degree, and edge rewiring probability.



(a) ER model (b) BA model (c) WS model ($\bar{d} = 20$)

Figure 3: Effects of input graph properties on the entropy gap for three random graph models.

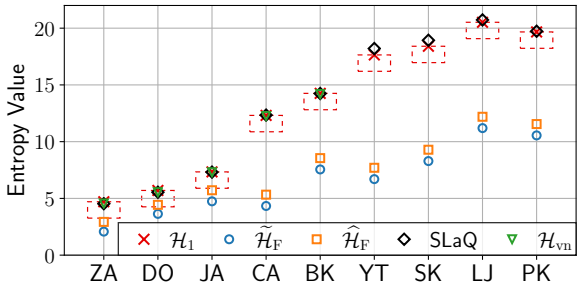


Figure 4: Structural information is an accurate proxy of von Neumann graph entropy. The exact von Neumann graph entropy lies in the red dotted box whose height is $\log_2 e$.

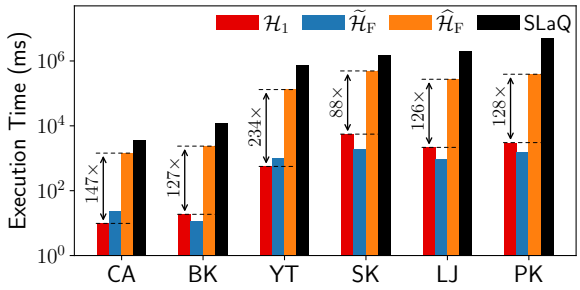


Figure 5: The computation of structural information is fast.

Fig. 4 and Fig. 5, we conclude that **the structural information is the only one that achieves both high efficiency and high accuracy among the prominent methods**.

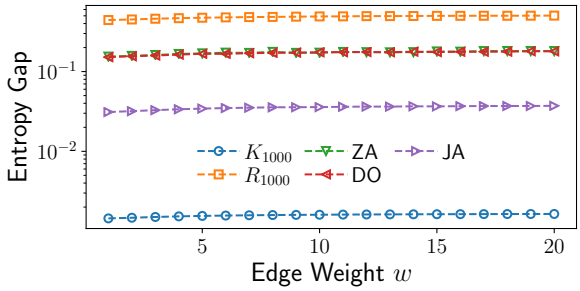


Figure 6: The entropy gap is insensitive to the edge weights.

6.6 Q5. Extensibility (Fig. 6)

To evaluate the extensibility of the entropy gap to weighted graphs, we measure the entropy gap of synthetic weighted graphs. Specifically, we choose 3 real-world graphs (ZA, DO, JA) with small size, a complete graph K_{1000} and ring graph R_{1000} each with 1000 nodes. The weight of each edge is set uniformly at random in the range $[1, w]$. We repeat the experiments for each $w \in \{1, 2, \ldots, 20\}$. The results in Fig. 6 show that **the entropy gap is insensitive to the change of edge weights in these graphs**. Therefore, it is of high probability that the entropy gap is still very small for a wide range of weighted graphs.

7 CONCLUSIONS AND FUTURE WORK

In this work, we suggest to use the structural information as a proxy of the von Neumann graph entropy such that provable accuracy, scalability, and interpretability are achieved at the same time. Since experimental results show the entropy gap is insensitive to the

graph size, we can estimate the entropy gap of a very large graph using small graphs generated from the same generative random graph model. We believe that our idea also provides new insights into approximations of graph spectral descriptors: besides function approximation, we can try to approximate the graph spectrum using simple and easily available graph statistics, such as the degree sequence.

There are multiple tangible research fronts we can pursue. First, in some access limited scenarios such as the World Wide Web, the complete degree sequence is often not available, therefore we need to develop sampling-based methods to estimate the structural information. Second, both the von Neumann graph entropy and the structural information can be viewed as a function on the edge set. Their properties such as the submodularity and monotonicity is under exploration. Last, the approximation of von Neumann entropy defined on the eigenvalues of normalized Laplacian matrix is still in its infancy.

APPENDIX

A ADDITIONAL EXPERIMENTS (Q6)

A.1 Performance of EntropyAug (Fig. 7)

To evaluate the performance of EntropyAug (Algorithm 1) in maximizing the von Neumann graph entropy, we measure the running time and dynamics of von Neumann graph entropy for EntropyAug and two competitors in three small real-world graphs ZA, DO, and JA. The two baselines are 1) "random" referring to the random addition of k non-existing edges, and 2) "algebraic" [17] referring to greedy addition of k non-existing edges that leads to the largest increase of the algebraic connectivity λ_{n-1}. We believe the "algebraic" algorithm is a competent competitor since maximizing λ_{n-1} would make the Laplacian spectrum concentrated on its mean, thereby maximizing the von Neumann entropy. The results in Fig. 7 show that **EntropyAug is the only one that achieves both high efficiency and large increments of von Neumann graph entropy**.

A.2 Performance of IncreSim (Fig. 8)

To evaluate the performance of IncreSim (Algorithm 2) and its relation with the VEO score, we measure the distance between two adjacent graphs in two real-world temporal graphs. We choose three methods (IncreSim, VEO score, and deltaCon) along with two simple measures (the number of added edges and the number of deleted edges). The VEO score [35] between two adjacent graphs G_t and G_{t+1} is defined as $1 - \frac{2(|V_t \cap V_{t+1}| + |E_t \cap E_{t+1}|)}{|V_t| + |V_{t+1}| + |E_t| + |E_{t+1}|}$, which measures the change rate of edge set and node set. The deltaCon [23] is a prominent method to measure graph similarity based on fast belief propagation. The results are shown in Fig. 8.

The observations are two fold. First, **the structural information distance is linearly correlated with the VEO score**, indicating that the structural information distance is not dominated by only local information, but rather a global measure on the graphs. For the FB temporal graph, the Pearson correlation coefficient and Spearman rank-order correlation coefficient of $\mathcal{D}_{\mathrm{SI}}$ with the VEO score are $(0.95, 0.97)$ respectively, which is much higher than $(0.70, 0.77)$ with deltaCon. For the WK temporal graph, the two correlation coefficients of $\mathcal{D}_{\mathrm{SI}}$ with the VEO score are $(0.96, 0.96)$ respectively, which is also much higher than $(-0.14, 0.00)$ with deltaCon. Second, **all of the three methods effectively captures the dynamics of graph streams**. For the FB temporal graph, the trend of the three distance measures are similar. For the WK temporal graph, we can see that the distance measure changes dramatically in the beginning, then gradually turns to be flat, which implies that the structure of WK temporal graph gradually becomes stable.

A.3 Performance of Structural Information Distance in Anomaly Detection (Fig. 9)

We further evaluate the effectiveness of the structural information distance in detecting the distributed denial-of-service (DDoS) attacks in a graph stream. We first generate 10 synthetic graphs $\mathcal{G} = \{G_t\}_{t=1}^{10}$ from the BA model, each of which has 100 nodes and average degree $\bar{d} = 4$. We believe that the synthetic graph stream $\mathcal{G}$ is a good representative of the real-world scale-free graph streams. Then we model the DDoS attack with strength k as follows: (1) Randomly select a graph G_{t^*} from $\mathcal{G}$. (2) Transform G_{t^*} into an anomalous graph G'_{t^*}. Specifically, we first randomly select a target node v, then randomly select k source nodes $\mathcal{S} = \{s_i\}_{i=1}^{k}$. Finally, we connect the target node v with the source node s_i for each $i \in \{1, \ldots, k\}$. (3) Generate the anomalous graph stream $\mathcal{G}'$ via replacing the graph G_{t^*} from $\mathcal{G}$ with G'_{t^*}.

We use graph distance measure to rank the anomalous graph in a graph stream. Suppose that the distance between G_t and G_{t+1} is $\theta_{t,t+1}$, then the anomalous score for G_t is $\frac{\theta_{t-1,t}+\theta_{t,t+1}}{2}$. We rank the graphs according to their anomalous scores in descending order. Then we use the rank of the true synthetic anomalous graph to measure the effectiveness of the graph distance measure in detecting DDoS attacks. We choose four candidates for the graph similarity measure: $\mathcal{D}_{\mathrm{SI}}$, $\mathcal{D}_{\mathrm{QJS}}$, VEO score, and deltaCon. And we repeat the random DDoS attacks for 100 times for each attack strength $k \in \{5, 10, 20, 30, 40\}$. The results are shown in Fig. 9.

The observations are two fold. First, $\mathcal{D}_{\mathrm{SI}}$ and $\mathcal{D}_{\mathrm{QJS}}$ have similar behavior in analyzing graph streams. Their trends in analyzing the synthetic graph stream $\mathcal{G}$ is nearly identical. Second, the structural information distance $\mathcal{D}_{\mathrm{SI}}$ is very suitable for detecting DDoS attacks in a graph stream. The structural information distance $\mathcal{D}_{\mathrm{SI}}$ behaves better than the other competitors for the attack strength $k \in \{20, 30, 40, 50\}$. When $k \in \{5, 10\}$, the performance of all the distance measures are mainly affected by the properties of the original normal graph stream.

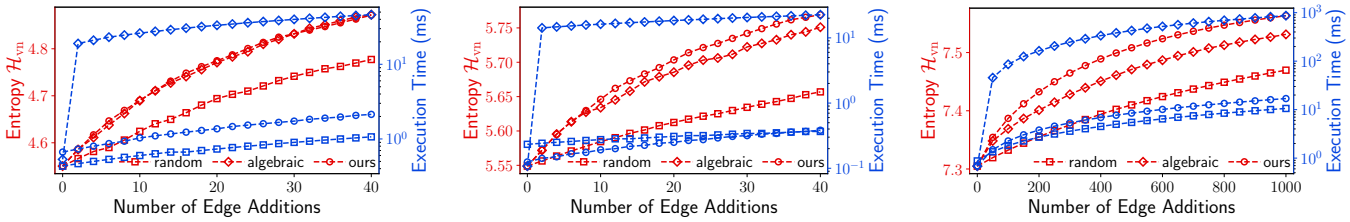
B PROOF OF THEOREM 5.4

We are going to establish a close relation between $\mathcal{D}_{\mathrm{SI}}$ and the Jensen-Shannon divergence $\mathcal{D}_{\mathrm{JS}}$, then the pseudometric properties of $\mathcal{D}_{\mathrm{SI}}$ simply follow from the metric properties of $\sqrt{\mathcal{D}_{\mathrm{JS}}}$.

The structural information $\mathcal{H}_1(G_j) = -\sum_{i=1}^{n} f\left(\frac{d_{i,j}}{\mathrm{vol}(G_j)}\right) = H(P_j)$ where $P_j = \left(\frac{d_{1,j}}{\mathrm{vol}(G_j)}, \ldots, \frac{d_{n,j}}{\mathrm{vol}(G_j)}\right)$ is a distribution on the set V.

In the graph $\bar{G} = (V, E_1 \cup E_2, \bar{A})$, the degree $\bar{d}_i$ of node i is

$$\bar{d}_i = \sum_{j=1}^{n} \bar{A}_{ij} = \sum_{j=1}^{n} \frac{A_{ij,1}}{2\mathrm{vol}(G_1)} + \frac{A_{ij,2}}{2\mathrm{vol}(G_2)} = \frac{d_{i,1}}{2\mathrm{vol}(G_1)} + \frac{d_{i,2}}{2\mathrm{vol}(G_2)}.$$



(a) Zachary (ZA) (b) Dolphins (DO) (c) Jazz (JA)

Figure 7: Compared with the other two methods, our structural information based method is the only one that achieves both high efficiency and large increments of von Neumann graph entropy.

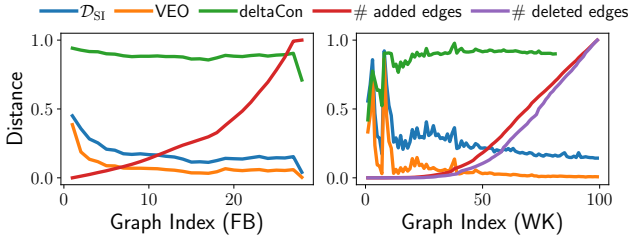


Figure 8: Distance between adjacent graphs in graph streams. The number of added/deleted edges is divided by the total number of added/deleted edges.

Then the volume of $\bar{G}$ is $\mathrm{vol}(\bar{G}) = \sum_{i=1}^{n} \bar{d}_i = 1$. Therefore the structural information of $\bar{G}$ is

$$\mathcal{H}_1(\bar{G}) = -\sum_{i=1}^{n} f\left(\frac{\bar{d}_i}{\mathrm{vol}(\bar{G})}\right) = -\sum_{i=1}^{n} f\left(\frac{d_{i,1}}{2\mathrm{vol}(G_1)} + \frac{d_{i,2}}{2\mathrm{vol}(G_2)}\right),$$

which is equivalent to the entropy of the distribution $(P_1 + P_2)/2$. As a result, $\mathcal{D}_{SI}(G_1, G_2) = \sqrt{\mathcal{D}_{JS}(P_1, P_2)}$.

C PROOF OF THEOREM 5.5

As shown in Theorem 5.4, the claim is true for $\mathcal{D}_{SI}$. It remains to prove that $\mathcal{D}_{QJS}(G_1, G_2) \le 1$, and if $\min\{d_{i,1}, d_{i,2}\} = 0$ for every node $i \in V$ then $\mathcal{D}_{QJS}(G_1, G_2) = 1$.

We prove $\mathcal{D}_{QJS}(G_1, G_2) \le 1$ using the inequality [31, 34] for the von Neumann entropy: if $\rho = \sum_i p_i \rho_i$ is a mixture of density matrix ρ_i with p_i a set of positive real numbers such that $\sum_i p_i = 1$, then $\mathcal{H}_{vn}(\sum_i p_i \rho_i) \le \sum_i p_i \mathcal{H}_{vn}(\rho_i) + H(\{p_i\})$. Here the density matrix ρ_i can be viewed as the scaled Laplacian matrix $\tilde{L}_i \triangleq L_i/\mathrm{tr}(L_i)$ of the graph G_i. Then

$$\begin{aligned}\mathcal{D}_{QJS}(G_1, G_2) &= \sqrt{\mathcal{H}_{vn}(\bar{G}) - (\mathcal{H}_{vn}(G_1) + \mathcal{H}_{vn}(G_2))/2} \\ &= \sqrt{\mathcal{H}_{vn}(\tilde{L}_1 + \tilde{L}_2) - (\mathcal{H}_{vn}(\tilde{L}_1) + \mathcal{H}_{vn}(\tilde{L}_2))/2} \\ &\le \sqrt{\mathcal{H}_{vn}(\tilde{L}_1 + \tilde{L}_2) - \mathcal{H}_{vn}((\tilde{L}_1 + \tilde{L}_2)/2) + 1} = 1.\end{aligned}$$

We denote by S_j the set of singletons in the graph G_j for $j \in \{1, 2\}$. Since $\min\{d_{i,1}, d_{i,2}\} = 0$ for every node $i \in V$, we have $S_1 \cup S_2 = V$ which implies that $(V \backslash S_1) \cap (V \backslash S_2) = \varnothing$ by the De Morgan's laws. Therefore, the node set V can be partitioned into three disjoint subsets $V \backslash S_1$, $V \backslash S_2$, and $S_1 \cap S_2$. Notice that one singleton contributes one eigenvalue of 0 to the Laplacian spectrum, and the Laplacian spectrum of a graph is composed of the Laplacian spectrum of its each connected components. We denote by $\lambda_{j,1}, \ldots, \lambda_{j,n-s_j}, 0, \ldots, 0$ the Laplacian spectrum of G_j, where $s_j = |S_j|$ for $j \in \{1, 2\}$. It follows that $\sum_{i=1}^{n-s_j} \lambda_{j,i} = \mathrm{vol}(G_j)$. Since $\bar{A} = A_1/2\mathrm{vol}(G_1) + A_2/2\mathrm{vol}(G_2)$, $\bar{L} = \bar{L}_1/2\mathrm{vol}(G_1) + \bar{L}_2/2\mathrm{vol}(G_2)$. Then the Laplacian spectrum of $\bar{G}$ is composed of Laplacian spectrum of G_j divided by $2\mathrm{vol}(G_j)$ for $j \in \{1, 2\}$ and zeros. As a result,

$$\begin{aligned}\mathcal{D}^2_{QJS}(G_1, G_2) &= -\sum_{j=1}^{2}\sum_{i=1}^{n-s_j} f\left(\frac{\lambda_{j,i}}{2\mathrm{vol}(G_j)}\right) + \frac{1}{2}\sum_{j=1}^{2}\sum_{i=1}^{n-s_j} f\left(\frac{\lambda_{j,i}}{\mathrm{vol}(G_j)}\right) \\ &= \sum_{j=1}^{2}\sum_{i=1}^{n-s_j} \frac{\lambda_{j,i}}{2\mathrm{vol}(G_j)} \log_2 2 = \sum_{j=1}^{2} \frac{\mathrm{vol}(G_j)}{2\mathrm{vol}(G_j)} = 1.\end{aligned}$$

D PROOF OF LEMMA 5.6

Denote by $\tilde{d}$ the degree sequence of G_{k+1}, then

$$\begin{aligned}\mathcal{H}_1(G_{k+1}) &= -\sum_{i=1}^{n} f\left(\frac{\tilde{d}_i}{2(m+\Delta m)}\right) = \frac{f(2(m+\Delta m)) - \sum_{i=1}^{n} f(\tilde{d}_i)}{2(m+\Delta m)} \\ &= \frac{f(2(m+\Delta m)) - \sum_{i\in V_k} f(d_i + \Delta d_i) - \sum_{i\in \bar{V}_k} f(d_i)}{2(m+\Delta m)} \\ &= \frac{f(2(m+\Delta m)) - a - \sum_{i=1}^{n} f(d_i)}{2(m+\Delta m)} \\ &= \frac{f(2(m+\Delta m)) - a - f(2m) + 2m\mathcal{H}_1(G_k)}{2(m+\Delta m)}.\end{aligned}$$

The structural information $\mathcal{H}_1(\bar{G}_k)$ is equal to

$$-\sum_{i=1}^{n} f\left(\frac{d_i}{4m} + \frac{\tilde{d}_i}{4(m+\Delta m)}\right) = -b - \sum_{i\in\bar{V}_k} f\left(\frac{2m+\Delta m}{4m(m+\Delta m)} d_i\right)$$

$$= -b - \sum_{i\in\bar{V}_k} c d_i(\log_2 c + \log_2 d_i) = -b - f(c)\sum_{i\in\bar{V}_k} d_i - c\sum_{i\in\bar{V}_k} f(d_i)$$

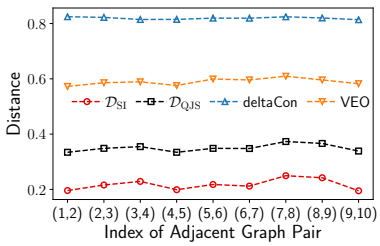
$$= -b - f(c)(2m - y) - c(f(2m) - 2m\mathcal{H}_1(G_k) - z)$$

ACKNOWLEDGMENTS

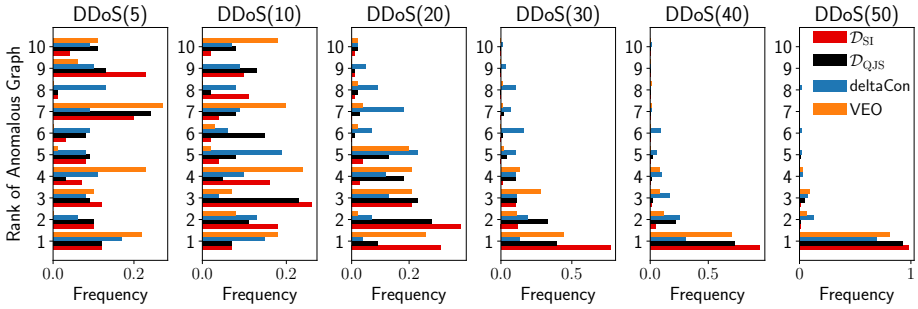
We would like to thank anonymous reviewers for useful feedback. This work was supported by National Key R&D Program of China 2018YFB2100302, and NSF China under Grant (No. 62020106005, 61960206002, 61822206, 61829201, 61832013, 62041205, 61532012), and Tencent AI Lab Rhino-Bird Focused Research Program JR202034.

REFERENCES

[1] Hua Bai. 2011. The Grone-Merris conjecture. *Trans. Amer. Math. Soc.* 363, 8 (Aug 2011), 4463–4474.



(a) Distances between adjacent graphs in a synthetic graph stream $\mathcal{G}$ from BA model with 100 nodes and $\bar{d} = 4$.



(b) Rank of the anomalous graph under DDoS attack in the synthetic graph stream.

Figure 9: Structural information distance is well suited for detecting DDoS attacks in a graph stream.

[2] Lu Bai and Edwin R. Hancock. 2011. Graph clustering using the Jensen-Shannon kernel. In *Computer Analysis of Images and Patterns*. Springer, Berlin, 394–401.

[3] Albert-László Barabási and Réka Albert. 1999. Emergence of scaling in random networks. *Science* 286, 5439 (1999), 509–512.

[4] Danail Bonchev and Gregory A. Buck. 2005. *Quantitative Measures of Network Complexity*. Springer, Boston, 191–235.

[5] Samuel L. Braunstein, Sibasish Ghosh, and Simone Severini. 2006. The Laplacian of a graph as a density matrix: a basic combinatorial approach to separability of mixed states. *Annals of Combinatorics* 10 (2006), 291–317.

[6] Jop Briët and Peter Harremoës. 2009. Properties of classical and quantum Jensen-Shannon divergence. *Phys. Rev. A* 79 (May 2009), 052311. Issue 5.

[7] Pin-Yu Chen, Lingfei Wu, Sijia Liu, and Indika Rajapakse. 2019. Fast incremental von Neumann graph entropy computation: theory, algorithm, and applications. In *ICML*. Long Beach, California, USA, 1091–1101.

[8] Hayoung Choi, Jinglian He, Hang Hu, and Yuanming Shi. 2020. Fast computation of von Neumann entropy for large-scale graphs via quadratic approximations. *Linear Algebra Appl.* 585 (2020), 127–146.

[9] Fan R. K. Chung. 1997. *Spectral Graph Theory*. American Mathematical Society.

[10] Michael Dairyko, Leslie Hogben, Jephian C.-H. Lin, Joshua Lockhart, David Roberson, Simone Severini, and Michael Young. 2017. Note on von Neumann and Rényi entropies of a graph. *Linear Algebra Appl.* 521 (2017), 240–253.

[11] George Dasoulas, Giannis Nikolentzos1, Kevin Scaman, Aladin Virmaux, and Michalis Vazirgiannis. 2020. Ego-based entropy measures for structural representations. *arXiv preprint arXiv:2003.00553* (2020).

[12] Manlio De Domenico and Jacob Biamonte. 2016. Spectral entropies as information-theoretic tools for complex network comparison. *Phys. Rev. X* 6 (Dec 2016), 041062. Issue 4.

[13] Matthias Dehmer. 2008. Information processing in complex networks: Graph entropy and information functionals. *Appl. Math. Comput.* 201 (2008), 82–94.

[14] Chris H.Q. Ding, Xiaofeng He, Hongyuan Zha, Ming Gu, and Horst D. Simon. 2001. A min-max cut algorithm for graph partitioning and data clustering. In *ICDM*. 107–114.

[15] Manlio D. Domenico, Vincenzo Nicosia, Alexandre Arenas, and Vito Latora. 2015. Structural reducibility of multilayer networks. *Nature Communications* 6, 6864 (2015).

[16] Dominik M. Endres and Johannes E. Schindelin. 2003. A new metric for probability distributions. *IEEE Transactions on Information Theory* 49, 7 (2003), 1858–1860.

[17] Arpita Ghosh and Stephen Boyd. 2006. Growing well-connected graphs. In *IEEE CDC*.

[18] Robert D. Grone. 1995. Eigenvalues and the degree sequences of graphs. *Linear and Multilinear Algebra* 39, 1–2 (1995), 133–136.

[19] Insu Han, Dmitry Malioutov, Haim Avron, and Jinwoo Shin. 2017. Approximating spectral sums of large-scale matrices using stochastic Chebyshev approximations. *SIAM Journal on Scientific Computing* 39, 4 (2017), A1558–A1585.

[20] Almerima Jamakovic and Steve Uhlig. 2007. On the relationship between the algebraic connectivity and graph's robustness to node and link failures. In *2007 Next Generation Internet Networks*. 96–102.

[21] E. Konstantinova and A. A. Paleev. 1990. Sensitivity of topological indices of polycyclic graphs. *Vychisl. Sistemy* 136 (1990), 38–48.

[22] Eugenia-Maria Kontopoulou, Gregory-Paul Dexter, Wojciech Szpankowski, Ananth Grama, and Petros Drineas. 2020. Randomized linear algebra approaches to estimate the von Neumann entropy of density matrices. *IEEE Transactions on Information Theory* 66, 8 (2020), 5003–5021.

[23] Danai Koutra, Neil Shah, Joshua T. Vogelstein, Brian Gallagher, and Christos Faloutsos. 2016. DeltaCon: principled massive-graph similarity function with attribution. *ACM Trans. Knowl. Discov. Data* 10, 3 (Feb. 2016).

[24] Jérôme Kunegis. 2013. The KONECT Project: Koblenz Network Collection. http://konect.cc/networks.

[25] P. W. Lamberti, A. P. Majtey, A. Borras, M. Casas, and A. Plastino. 2008. Metric character of the quantum Jensen-Shannon divergence. *Phys. Rev. A* 77 (May 2008), 052311. Issue 5.

[26] P. W. Lamberti, M. Portesi, and J. Sparacino. 2008. Natural metric for quantum information theory. *arXiv preprint arXiv:0807.0583* (2008).

[27] Jure Leskovec and Andrej Krevl. 2014. SNAP Datasets: Stanford Large Network Dataset Collection. http://snap.stanford.edu/data.

[28] Angsheng Li and Yicheng Pan. 2016. Structural information and dynamical complexity of networks. *IEEE Transactions on Information Theory* 62, 6 (Jun 2016), 3290–3339.

[29] Jiangang Liao and Arthur Berg. 2019. Sharpening Jensen's inequality. *The American Statistician* 73, 3 (2019), 278–281.

[30] Joshua Lockhart, Giorgia Minello, Luca Rossi, Simone Severini, and Andrea Torsello. 2016. Edge centrality via the Holevo quantity. In *Structural, Syntactic, and Statistical Pattern Recognition*. Springer, 143–152.

[31] A. P. Majtey, P. W. Lamberti, and D. P. Prato. 2005. Jensen-Shannon divergence as a measure of distinguishability between mixed quantum states. *Phys. Rev. A* 72 (Nov 2005), 052310. Issue 5.

[32] Albert W. Marshall, Ingram Olkin, and Barry C. Arnold. 2011. *Inequalities: Theory of Majorization and Its Applications* (second ed.). Springer Series in Statistics.

[33] Giorgia Minello, Luca Rossi, and Andrea Torsello. 2018. On the von Neumann entropy of graphs. *Journal of Complex Networks* 7, 4 (11 2018), 491–514.

[34] Michael A. Nielsen and Isaac L. Chuang. 2010. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press.

[35] Panagiotis Papadimitriou, Ali Dasdan, and Hector Garcia-Molina. 2010. Web graph similarity for anomaly detection. *Journal of Internet Services and Applications* 1 (2010), 19–30.

[36] N. Rashevsky. 1955. Life, information theory, and topology. *Bull. Math. Biophys.* 17, 3 (1955), 229–235.

[37] C. Raychaudhury, S. K. Ray, J. J. Ghosh, and A. B. Basak. 1984. Discrimination of isomeric structures using information theoretic topological indices. *J. Comput. Chem.* 5, 6 (1984), 581–588.

[38] David E Simmons, Justin P Coon, and Animesh Datta. 2018. The von Neumann Theil index: characterizing graph centralization using the von Neumann index. *Journal of Complex Networks* 6, 6 (Jan 2018), 859–876.

[39] Anton Tsitsulin, Davide Mottin, Panagiotis Karras, Alexander Bronstein, and Emmanuel Müller. 2018. NetLSD: Hearing the shape of a graph. In *ACM SIGKDD*. 2347–2356.

[40] Anton Tsitsulin, Marina Munkhoeva, and Bryan Perozzi. 2020. Just SLaQ when you approximate: accurate spectral distances for web-scale graphs. In *Proceedings of The Web Conference 2020*. 2697–2703.

[41] Shashanka Ubaru, Jie Chen, and Yousef Saad. 2017. Fast estimation of tr(f(A)) via stochastic Lanczos quadrature. *SIAM J. Matrix Anal. Appl.* 38, 4 (2017), 1075–1099.

[42] Bimal Viswanath, Alan Mislove, Meeyoung Cha, and Krishna P. Gummadi. 2009. On the Evolution of User Interaction in Facebook. In *WOSN*. 37–42.

[43] Duncan J. Watts and Steven H. Strogatz. 1998. Collective dynamics of 'small-world' networks. *Nature* 393 (1998), 440–442.

[44] Bai Xiao, Edwin R. Hancock, and Richard C. Wilson. 2009. Graph characteristics from the heat kernel trace. *Pattern Recognition* 42, 11 (2009), 2589–2606.

[45] Cheng Ye, Richard C. Wilson, César H. Comin, Luciano da F. Costa, and Edwin R. Hancock. 2014. Approximate von Neumann entropy for directed graphs. *Phys. Rev. E* 89 (May 2014), 052804. Issue 5.

[46] Wenwu Yu, Guanrong Chen, and Ming Cao. 2011. Consensus in directed networks of agents with nonlinear dynamics. *IEEE Trans. Automat. Control* 56, 6 (2011), 1436–1441.