Fast Incremental von Neumann Graph Entropy Computation: Theory, Algorithm, and Applications

Pin-Yu Chen [1] **Lingfei Wu** [1] **Sijia Liu** [1] **Indika Rajapakse** [2]

Abstract

The von Neumann graph entropy (VNGE) facilitates measurement of information divergence and distance between graphs in a graph sequence. It has been successfully applied to various learning tasks driven by network-based data. While effective, VNGE is computationally demanding as it requires the full eigenspectrum of the graph Laplacian matrix. In this paper, we propose a new computational framework, **F**ast **I**ncremental von **N**eumann **G**raph **E**nt**R**opy (FINGER), which approaches VNGE with a performance guarantee. FINGER reduces the cubic complexity of VNGE to linear complexity in the number of nodes and edges, and thus enables online computation based on incremental graph changes. We also show asymptotic equivalence of FINGER to the exact VNGE, and derive its approximation error bounds. Based on FINGER, we propose efficient algorithms for computing Jensen-Shannon distance between graphs. Our experimental results on different random graph models demonstrate the computational efficiency and the asymptotic equivalence of FINGER. In addition, we apply FINGER to two real-world applications and one synthesized anomaly detection dataset, and corroborate its superior performance over seven baseline graph similarity methods.

1. Introduction

In recent years, graph-based learning has become an active research field (Shuman et al., 2013; Kalofolias, 2016; Luo et al., 2012; Shivanna & Bhattacharyya, 2014; Wang et al., 2016; Kipf & Welling, 2017; Wu et al., 2018a;b; Xu et al., 2018). Its success is rooted in the advanced capability of summarizing and representing phenomenal structural features embedded in graphs. In particular, evaluating similarity between graphs is crucial to network analysis and graph-based anomaly detection (Papadimitriou et al., 2010; Akoglu et al., 2015; Ranshous et al., 2015). For example, Yanardag and Vishwanathan used graph similarity for learning novel graph kernels (Yanardag & Vishwanathan, 2015), and Sharpnack et al. proposed the Lovasz extended scan statistic for anomaly detection in connected graphs (Sharpnack et al., 2013). Koutra et al. proposed DeltaCon, a state-of-the-art similarity algorithm in terms of its scalability and capability of handling weighted graphs using fast belief propagation (Koutra et al., 2016). However, these methods are sensitive to heuristic metrics and presumed models, and thus provide limited understanding on the general notion of variations between graphs. On the other hand, model-agnostic approaches such as graph entropy have been used to quantify the structural complexity of a single graph, which relates to the Shannon entropy of a probability distribution over a function of enumerated subgraphs in a graph (Simonyi, 1995; Shetty & Adibi, 2005; Li & Pan, 2016). However, graph entropy can be computationally demanding due to its use of exhaustive subgraph search.

Different from the aforementioned approaches and inspired by quantum information theory, the von Neumann graph entropy (VNGE) (Braunstein et al., 2006; Passerini & Severini, 2008; 2009) facilitates the measure of (quantum) Jensen-Shannon divergence and distance (Endres & Schindelin, 2003; Briët & Harremoës, 2009) between graphs. It associates with a model-agnostic information measure for quantifying variation between two quantum density matrices. In addition, the VNGE has been shown to be linearly correlated with classical graph entropy measures (Anand & Bianconi, 2009; Anand et al., 2011). The VNGE and the Jensen-Shannon distance have been successfully applied to structural reduction in multiplex networks (De Domenico et al., 2015), depth analysis in image processing (Han et al., 2012; Bai & Hancock, 2014), structure-function analysis in genetic networks (Seaman et al., 2017; Liu et al., 2018b), and network-ensemble comparison (Li et al., 2018). However, despite its effectiveness, the computation of VNGE requires (at most) cubic complexity in the number of nodes, thereby impeding its applicability to machine learning and data mining tasks involving a sequence of large graphs.

[1]IBM Research [2]University of Michigan, Ann Arbor, USA. Correspondence to: Pin-Yu Chen <pin-yu.chen@ibm.com>.

Proceedings of the 36^{th} International Conference on Machine Learning, Long Beach, California, PMLR 97, 2019.

Contributions. To overcome the computational inefficiency of VNGE, we propose a **F**ast **I**ncremental von **N**eumann **G**raph **Ent**ropy (FINGER) framework to approximate VNGE with a performance guarantee, reducing its cubic complexity to linear complexity in the number of nodes and edges. FINGER is a generic tool that applies to both batch and online graph sequences. It enables fast entropy computation when every single graph in a graph sequence is presented (e.g., a snapshot of a dynamic network, or a single-layer connectivity pattern of a multiplex network). For applications where changes in a graph (e.g., addition and deletion of nodes or edges over time) are continuously reported (e.g., streaming graphs), FINGER also allows online computation based on incremental graph changes. We prove that FINGER maintains an approximation guarantee and is asymptotically equivalent to the exact VNGE under some eigenspectrum conditions, which is further validated by different synthetic random graphs. We then apply FINGER to developing efficient algorithms for the computation of Jensen-Shannon distance between graphs. Comparing to the state-of-the-art graph similarity methods and two alternative approximate VNGE, FINGER yields superior and robust performance for anomaly detection in evolving Wikipedia networks and router communication networks, as well as bifurcation analysis in dynamic genomic networks. These applications show the effectiveness and potentials of Jensen-Shannon distance for network learning in a wide range of domains, which has not been rigorously explored owing to its high computation complexity in the absence of FINGER.

The contributions of this paper and the proposed framework (FINGER) are summarized as follows.

- Two types of approximate VNGE reducing its cubic complexity to linear complexity are proposed to support fast and incremental computation of VNGE. We derive their approximation error bounds and show asymptotic equivalence relative to the exact VNGE under mild conditions.
- FINGER achieves nearly 100% reduction in computation time for VNGE of different graphs and enables scalable Jensen-Shannon graph distance computation.
- On two real-world applications (anomaly detection and cellular bifurcation analysis) and one synthesized dataset, FINGER exhibits outstanding and robust performance over 7 baseline and state-of-the-art methods.

Related Work. The VNGE was firstly defined based on the combinatorial graph Laplacian matrix (Braunstein et al., 2006; Passerini & Severini, 2008; 2009; De Domenico et al., 2015; Li et al., 2018). Variants of VNGE and their approximations have been proposed in the literature, including the normalized graph Laplacian matrix (Shi & Malik, 2000) proposed in (Han et al., 2012) and the generalized graph Laplacian matrix of directed graphs (Chung, 2005) proposed in (Ye et al., 2014). However, these alternatives lack approximation justification and are shown to be suboptimal in Section 4. To the best of our knowledge, this paper is the first work that provides fast VNGE computation with a provable approximation analysis.

2. FINGER: Theory and Algorithms

2.1. Background and Preliminaries

Using terminology from quantum statistical mechanics, a density matrix $\mathbf{\Phi}$ describing a quantum system in a mixed state can be cast as a statistical ensemble of several quantum states. The $n \times n$ matrix $\mathbf{\Phi}$ is symmetric, positive semidefinite, and satisfies $\text{trace}(\mathbf{\Phi}) = 1$. The von Neumann entropy of a quantum system is defined as $H = -\text{trace}(\mathbf{\Phi} \ln \mathbf{\Phi})$ (Von Neumann, 1955), where $\ln \mathbf{\Phi}$ denotes matrix logarithm. Let $\{\lambda_i\}_{i=1}^n$ be the sorted eigenvalues of $\mathbf{\Phi}$ such that $0 \leq \lambda_n \leq \ldots \leq \lambda_1$. The definition of von Neumann entropy is equivalent to $H = -\sum_{i=1}^n \lambda_i \ln \lambda_i$, where the convention $0 \ln 0 = 0$ is used due to $\lim_{x \to 0^+} x \ln x = 0$. Moreover, since $\sum_i \lambda_i = 1$ and $\lambda_i \geq 0$ for all i, the von Neumann entropy can be viewed as the Shannon entropy associated with the eigenspectrum $\{\lambda_i\}_{i=1}^n$.

We consider the class of undirected weighted simple nonempty graphs with nonnegative edge weights, denoted by $\mathcal{G}$. Let $G = (\mathcal{V}, \mathcal{E}, \mathbf{W}) \in \mathcal{G}$ denote a single graph, where $\mathcal{V}$ and $\mathcal{E}$ denote its node and edge set with cardinality $|\mathcal{V}| = n$ and $|\mathcal{E}| = m$, respectively, and $\mathbf{W}$ is an $n \times n$ matrix with entry $[\mathbf{W}]_{ij} = w_{ij}$ denoting the weight of an edge $(i, j) \in \mathcal{E}$. A graph sequence $\{G_t\}_{t=1}^T$ refers to a set of T graphs indexed by $t \in \{1, \ldots, T\}$ with known node-to-node correspondence, where $G_t \in \mathcal{G}$ for all t. The combinatorial graph Laplacian matrix of G is defined as $\mathbf{L} = \mathbf{S} - \mathbf{W}$ (Luxburg, 2007), where $\mathbf{S} = \text{diag}(s_1, \ldots, s_n)$ is a diagonal matrix and its diagonal entry $s_i = \sum_{j=1}^n w_{ij}$ is the nodal strength (weighted degree) of a node $i \in \mathcal{V}$. Connecting the von Neumann entropy to graphs, the VNGE, denoted by $H(G)$, is defined by replacing $\mathbf{\Phi}$ with $\mathbf{L}_N = c \cdot \mathbf{L}$ (Braunstein et al., 2006; Passerini & Severini, 2008; 2009), where $c = 1/\text{trace}(\mathbf{L})$ is a trace normalization factor. It has been proved in (Passerini & Severini, 2008) that for any $G \in \mathcal{G}$, $H(G) \leq \ln(n-1)$, where the equality holds when G is a complete graph. Note that since computing VNGE requires the entire eigenspectrum $\{\lambda_i\}_{i=1}^n$ of $\mathbf{L}_N$, it incurs full eigenvalue decomposition on $\mathbf{L}_N$ and has cubic complexity $O(n^3)$[12] (Horn & Johnson, 1990), making it computationally infeasible for large graphs.

[1] $f(n) = O(h(n))$, $f(n) = o(h(n))$ and $f(n) = \Omega(h(n))$ mean $\limsup_{n\to\infty} |\frac{f(n)}{h(n)}| < \infty$, $\lim_{n\to\infty} \frac{f(n)}{h(n)} = 0$, and $\limsup_{n\to\infty} |\frac{f(n)}{h(n)}| > 0$, respectively.

[2] For computing all eigenvalues of large matrices, a viable solution is direct methods, possibly with parallel eigensolvers for acceleration. The complexity for computing $\{\lambda_i\}_{i=1}^n$ of $\mathbf{L}_N$ is typically $O(n^2 + \frac{4}{3}n^3)$ (Bai et al., 2000).

In what follows, we propose two types of approximate VNGE ($\widehat{H}$ and $\widetilde{H}$) for the exact VNGE H, where $\widehat{H}$ and $\widetilde{H}$ possess linear computation complexity and satisfy $\widetilde{H} \leq \widehat{H} \leq H$. Depending on the data format and problem setup, $\widehat{H}$ is designed for fast computation of H for a single graph, and $\widetilde{H}$ is designed for online computation of H based on incremental graph changes. Furthermore, we derive approximation error and prove asymptotic equivalence relative to H under mild conditions on the eigenspectrum $\{\lambda_i\}_{i=1}^n$ of $\mathbf{L}_N$. Our proofs are given in the supplementary material.

2.2. Approximation of von Neumann Graph Entropy

Recall that computing $H = -\sum_{i=1}^n \lambda_i \ln \lambda_i$ requires $O(n^3)$ computation complexity. For computation acceleration, we first reduce its computation complexity by using the quadratic approximation of the term $\lambda_i \ln \lambda_i$ in H via Taylor series expansion, leading to the following lemma.

Lemma 1 (Quadratic approximation Q of H)**.** *For any $G \in \mathcal{G}$, the quadratic approximation Q of the von Neumann graph entropy H via Taylor series expansion is equivalent to $Q = 1 - c^2(\sum_{i\in\mathcal{V}} s_i^2 + 2\sum_{(i,j)\in\mathcal{E}} w_{ij}^2)$, where $c = \frac{1}{S}$ and $S = \text{trace}(\mathbf{L}) = \sum_{i\in V} s_i = 2\sum_{(i,j)\in\mathcal{E}} w_{ij}$.*

It is clear from Lemma 1 that Q only depends on the edge weights in $G = (\mathcal{V}, \mathcal{E}, \mathbf{W})$, resulting in linear computation complexity[3] $O(n+m)$, where $|\mathcal{V}| = n$ and $|\mathcal{E}| = m$. We note that higher-order (beyond quadratic) approximation of H is plausible at the price of less computational efficiency and possibly excessive subgraph pattern searching. For example, the cubic approximation of H involves the computation of trace($\mathbf{W}^3$), which relates to the sum of edge weights of every triangle in G. To identify the approximation accuracy and equivalence of Q with respect to H, the following theorem shows the approximation bounds on H in terms of Q and the eigenspectrum $\{\lambda_i\}_{i=1}^n$ of $\mathbf{L}_N$.

Theorem 1 (Approximation bounds on H)**.** *For any $G \in \mathcal{G}$, let $\lambda_{\max}$ and $\lambda_{\min}$ be the largest and smallest positive eigenvalue of $\mathbf{L}_N$, respectively. If $\lambda_{\max} < 1$, then $-Q\frac{\ln \lambda_{\max}}{1-\lambda_{\min}} \leq H \leq -Q\frac{\ln \lambda_{\min}}{1-\lambda_{\max}}$. The bounds become exact and $H = \ln(n-1)$ when G is a complete graph with identical edge weight.*

Note that Theorem 1 excludes the extreme case when $\lambda_{\max} = 1$, as the resulting VNGE is trivial ($H = 0$). The condition $\lambda_{\max} < 1$ holds for any graph $G \in \mathcal{G}$ having a connected subgraph with at least 3 nodes. In addition to the approximation bounds presented in Theorem 1, the corollary below further shows asymptotic equivalence between Q and $\frac{H}{\ln n}$ under mild conditions on $\lambda_{\max}$ and $\lambda_{\min}$.

Corollary 1 (Asymptotic equivalence of Q)**.** *For any $G \in \mathcal{G}$, let n_+ denote the number of positive eigenvalues of $\mathbf{L}_N$. If $n_+ = \Omega(n)$[1] and $\lambda_{\min} = \Omega(\lambda_{\max})$, then $\frac{H}{\ln n} - Q \to 0$ as $n \to \infty$.*

Corollary 1 suggests that the VNGE of large graphs with balanced eigenspectrum (i.e., $\lambda_{\min} = \Omega(\lambda_{\max})$) can be well approximated by Q and a factor $\ln n$. The condition of balanced eigenspectrum holds in regular and homogeneous random graphs (Passerini & Severini, 2008; Du et al., 2010). Furthermore, since n_+ equals to $n - g$, where g is the number of connected components in G (Merris, 1994), the condition $n_+ = \Omega(n)$ holds when $g = o(n)$[1].

2.3. FINGER-$\widehat{H}$: Approximate von Neumann Graph Entropy $\widehat{H}$ Using Q and $\lambda_{\max}$

Based on the derived lower bound of H in Theorem 1, we propose the first type of approximate VNGE $\widehat{H}$ using Q and $\lambda_{\max}$ for any $G \in \mathcal{G}$, which is defined as

$$\widehat{H}(G) = -Q \ln \lambda_{\max}. \tag{1}$$

Comparing to the lower bound $-Q\frac{\ln \lambda_{\max}}{1-\lambda_{\min}}$ in Theorem 1, $\widehat{H}$ is a looser lower bound on H since $1 - \lambda_{\min} < 1$. Here we use $1 - \lambda_{\min} \approx 1$ when approximating H, since trace($\mathbf{L}_N$) $= \sum_{i=1}^n \lambda_i = 1$ and hence $\lambda_{\min}$ is negligible, especially for large graphs.

More importantly, since $\lambda_{\max}$ is the largest eigenvalue of $\mathbf{L}_N$ and by definition $\mathbf{L}_N$ has $n + m$ nonzero entries, the computation of $\lambda_{\max}$ only requires $O(m+n)$ operations via power iteration methods (Horn & Johnson, 1990; Wu et al., 2017; Liao et al., 2019), leading to the same complexity as Q. Consequently, by only acquiring $\lambda_{\max}$ instead of the entire eigenspectrum $\{\lambda_i\}_{i=1}^n$, the computation of $\widehat{H}$ has linear complexity $O(m + n)$, resulting in significant computation reduction when compared with the exact VNGE H, which requires cubic complexity[2] $O(n^3)$. In addition to computational efficiency, the following corollary shows that the approximation error of $\widehat{H}$, defined as $H - \widehat{H}$, decays at the rate of $\ln n$ under the same conditions as in Corollary 1. We note that the $o(\ln n)$ approximation error rate is nontrivial since $H \leq \ln(n-1)$ for any $G \in \mathcal{G}$ (Passerini & Severini, 2008; Du et al., 2010).

Corollary 2 ($o(\ln n)$ approximation error of $\widehat{H}$)**.** *For any $G \in \mathcal{G}$, if $n_+ = \Omega(n)$ and $\lambda_{\min} = \Omega(\lambda_{\max})$, then the scaled approximation error (SAE) $\frac{H-\widehat{H}}{\ln n} \to 0$ as $n \to \infty$, implying $H - \widehat{H} = o(\ln n)$.*

2.4. FINGER-$\widetilde{H}$: Approximate von Neumann Graph Entropy $\widetilde{H}$ Using Q and $s_{\max}$

The proxy $\widehat{H}$ in Section 2.3 enables fast computation of VNGE for a single graph. As the exact online update of the eigenvalue $\lambda_{\max}$ in $\widehat{H}$ based on incremental graph changes is challenging, we propose the second type of approximate

[3]The complexity becomes $O(n^2)$ when $m = O(n^2)$ (i.e., dense graphs). In sparse graphs m could be $O(n)$.

VNGE $\widetilde{H}$ using Q and the largest nodal strength $s_{\max} = \max_{i\in\mathcal{V}} s_i$ in a graph, which allows simple incremental update of $\widetilde{H}$ based on graph changes but at the price of larger approximation error than that of $\widehat{H}$. The approximate VNGE H is defined as

$$\widetilde{H}(G) = -Q\ln(2c\cdot s_{\max}), \tag{2}$$

where c is the trace normalization constant. Using the definition $\mathbf{L}_N = c\cdot\mathbf{L}$ and the upper bound on the largest eigenvalue of $\mathbf{L}$ in (Anderson Jr & Morley, 1985), we obtain $\widetilde{H} \le \widehat{H} \le H$ since $\lambda_{\max} \le 2c\cdot s_{\max}$, implying $\widetilde{H}$ is a looser lower bound on H when compared with $\widehat{H}$. Nonetheless, the following corollary shows the approximation error of $\widetilde{H}$ also decays at the same rate $o(\ln n)$ as $\widehat{H}$.

Corollary 3 ($o(\ln n)$ approximation error of $\widetilde{H}$)**.** *For any $G\in\mathcal{G}$, if $n_+ = \Omega(n)$ and $\lambda_{\min} = \Omega(\lambda_{\max})$, then the scaled approximation error (SAE) $\frac{H-\widetilde{H}}{\ln n} \to 0$ as $n\to\infty$, implying $H-\widetilde{H} = o(\ln n)$.*

To enable incremental computation of VNGE using $\widetilde{H}$, let $G = (\mathcal{V},\mathcal{E},\mathbf{W})$ and $G' = (\mathcal{V}',\mathcal{E}',\mathbf{W}')$ be any two graphs from a graph sequence. Without loss of generality we assume G and G' have a common node set $\mathcal{V}_c$ with $|\mathcal{V}_c| = n$ nodes[4]. In particular, the graph $\Delta G = (\Delta\mathcal{V},\Delta\mathcal{E},\Delta\mathbf{W})$ with $|\Delta\mathcal{V}| = \Delta n$ and $|\Delta\mathcal{E}| = \Delta m$ is introduced to represent the changes made from converting G to G', denoted by $G' = G\oplus\Delta G$[5]. The terms $\{\Delta s_i\}_{i\in\Delta\mathcal{V}}$ and $\{\Delta w_{ij}\}_{(i,j)\in\Delta\mathcal{E}}$ denote the nodal strengths and edge weights of ΔG, respectively, and $\Delta S = \sum_{i\in\Delta\mathcal{V}}\Delta s_i$. Let Q' be the quadratic approximation of $H(G')$. The theorem below shows that Q' can be efficiently updated based on Q of $H(G)$, the values of $s_{\max}$ and c from G, and ΔG, yielding competent complexity $O(\Delta n+\Delta m)$.

Theorem 2 (Incremental update of Q')**.** *For any $G,G'\in\mathcal{G}$ such that $G' = G\oplus\Delta G$, given Q, G and ΔG, the term Q' can be efficiently updated by incremental graph changes as $Q' = \frac{Q-1}{(1+c\Delta S)^2} - \left(\frac{c}{1+c\Delta S}\right)^2\Delta Q + 1$, where $\Delta Q = 2\sum_{i\in\Delta\mathcal{V}} s_i\Delta s_i + \sum_{i\in\Delta\mathcal{V}}\Delta s_i^2 + 4\sum_{(i,j)\in\Delta\mathcal{E}} w_{ij}\Delta w_{ij} + 2\sum_{(i,j)\in\Delta\mathcal{E}}\Delta w_{ij}^2$, and $\Delta c = \frac{-c^2\Delta S}{1+c\Delta S}$.*

Furthermore, by the definition of $\widetilde{H}$ in (2), $\widetilde{H}(G\oplus\Delta G)$ can be efficiently updated by

$$\widetilde{H}(G\oplus\Delta G) = -Q'\ln[2(c+\Delta c)(s_{\max}+\Delta s_{\max})] \tag{3}$$

given Q, $s_{\max}$ and c from G, and graph changes ΔG, where Δc is defined in Theorem 2, and $\Delta s_{\max}$ is the maximum value of 0 and $\max_{i\in\Delta\mathcal{V}}(s_i+\Delta s_i)-s_{\max}$. The computation complexity of $\widetilde{H}(G\oplus\Delta G)$ is $O(\Delta n+\Delta m)$ since the incremental update formula of Q' in Theorem 2 and the computation of $\Delta s_{\max}$ only take $O(\Delta n+\Delta m)$ operations.

[4] If G and G' have different nodes, the set $\mathcal{V}_c$ can be constructed by the set union $\mathcal{V}_c = \mathcal{V}\cup\mathcal{V}'$.

[5] The notation $\oplus$ denotes set additions $\mathcal{V}' = \mathcal{V}\uplus\Delta V$, $\mathcal{E}' = \mathcal{E}\uplus\Delta\mathcal{E}$ and matrix addition $\mathbf{W}' = \mathbf{W}+\Delta\mathbf{W}$.

Algorithm 1 FINGER-JSdist (Fast)

Input: Two graphs G and G' from a graph sequence
Output: JSdist(G,G')
1. Obtain $\overline{G} = \frac{G\oplus G'}{2}$ and compute $\widehat{H}(G)$, $\widehat{H}(G')$, and $\widehat{H}(\overline{G})$ via FINGER-$\widehat{H}$ from (1)
2. JSdist$(G,G') = \left(\widehat{H}(\overline{G}) - \frac{1}{2}[\widehat{H}(G)+\widehat{H}(G')]\right)^{1/2}$

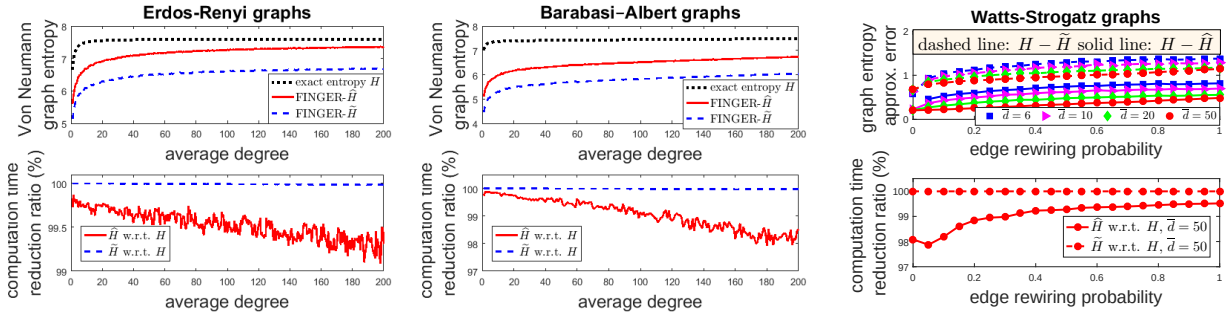
Algorithm 2 FINGER-JSdist (Incremental)

Input: A graph G, graph changes ΔG, and $\widetilde{H}(G)$
Output: JSdist$(G, G\oplus\Delta G)$
1. Compute $\widetilde{H}(G\oplus\frac{\Delta G}{2})$ and $\widetilde{H}(G\oplus\Delta G)$ via FINGER-$\widetilde{H}$ from (3) and Theorem 2
2. JSdist$(G, G\oplus\Delta G) =$
$\left(\widetilde{H}(G\oplus\frac{\Delta G}{2}) - \frac{1}{2}[\widetilde{H}(G)+\widetilde{H}(G\oplus\Delta G)]\right)^{1/2}$

2.5. Fast and Incremental Algorithms for Jensen-Shannon Distance between Graphs

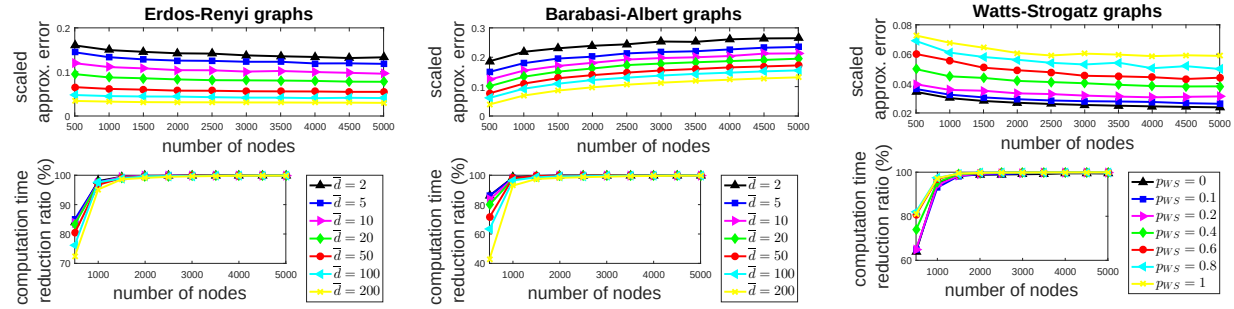
As summarized in Algorithms 1 and 2, one major utility of VNGE[6] is the computation of Jensen-Shannon distance (JSdist) between any two graphs from a graph sequence. Consider two graphs $G = (\mathcal{V}_c,\mathcal{E},\mathbf{W})\in\mathcal{G}$ and $G' = (\mathcal{V}_c,\mathcal{E}',\mathbf{W}')\in\mathcal{G}$, and let $\overline{G} = (\mathcal{V}_c,\overline{\mathcal{E}},\overline{\mathbf{W}}) = \frac{G\oplus G'}{2}$ denote their averaged graph such that $\overline{\mathbf{W}} = \frac{\mathbf{W}+\mathbf{W}'}{2}$. Then the Jensen-Shannon divergence between G and G' can be computed by JSdiv$(G,G') = H(\overline{G}) - \frac{1}{2}[H(G)+H(G')]$ (De Domenico et al., 2015). Furthermore, the Jensen-Shannon distance between G and G' is defined as JSdist$(G,G') = \sqrt{\text{JSdiv}(G,G')}$, which has been proved to be a valid distance metric in (Endres & Schindelin, 2003; Briët & Harremoës, 2009). The exact computation of JSdist requires $O(n^3)$ computation complexity by the definition of H, where $|\mathcal{V}_c| = n$, which is computationally cumbersome for large graphs. To overcome its computational inefficiency, we apply the developed FINGER-$\widehat{H}$ and FINGER-$\widetilde{H}$ to the computation of JSdist. If each graph G_t in a graph sequence $\{G_t\}_{t=1}^T$ is given, then FINGER-JSdist (Fast) allows fast computation of JSdist and features linear computation complexity inherited from $\widehat{H}$. If a graph sequence is presented by sequential graph changes $\{\Delta G_t\}_{t=1}^{T-1}$ such that $G_{t+1} = G_t\oplus\Delta G_t$, then FINGER-JSdist (Incremental) allows online computation of JSdist relative to the incremental graph changes. Their superior performance will be discussed in Section 4.

[6] Codes: https://github.com/pinyuchen/FINGER



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

Figure 1. Performance evaluation of von Neumann graph entropy approximation in different random graph models with $n = 2000$ nodes under varying average degree $\overline{d}$ and edge rewiring probability p_{WS}. The approximation error of FINGER decays as $\overline{d}$ increases or p_{WS} decreases. FINGER achieves nearly 100% speed-up relative to the exact entropy computation.



(a) ER model (b) BA model (c) WS model ($\overline{d} = 50$)

Figure 2. Scaled approximation error (SAE) and computation time reduction ratio (CTRR) of $\widehat{H}$ via FINGER for different random graph models and varying number of nodes n. The SAE of ER and WS graphs validates the $o(\ln n)$ approximation error analysis in Corollaries 2 and 3. The CTRR attains nearly 100% speed-up relative to H for moderate-size graphs ($n \geq 2000$).

3. Experiments

In this section we conducted intensive experiments on the VNGE of three kinds of synthetic random graphs to study the effects of graph size, average degree, and graph regularity on the approximation error of FINGER and its computational efficiency. The three random graph models are: (i) Erdos-Renyi (ER) model (Erdös & Rényi, 1959) – every node pair is connected independently with probability p_{ER}; (ii) Barabasi-Albert (BA) model (Barabási & Albert, 1999) – the degree distribution follows a power-law distribution; and (iii) Watts-Strogatz (WS) model (Watts & Strogatz, 1998) – an initially regular ring network with independent edge rewiring probability p_{WS} for simulating small-world networks. The parameter p_{WS} controls the regularity of graph connectivity, and smaller p_{WS} gives more regular graphs.

Since $\widetilde{H} \leq \widehat{H} \leq H$, the approximation error (AE) is defined as $H - \widehat{H}$ and $H - \widetilde{H}$, respectively. The scaled approximation error (SAE) is defined as $\frac{\text{AE}}{\ln n}$, which is a proper scaling according to our error analysis in Section 2, and it also makes a fair comparison of graphs with different number of nodes. The computation time reduction ratio (CTRR) is defined as $\frac{\text{Time}(H) - \text{Time}(X)}{\text{Time}(H)}$, where $X \in \{\widehat{H}, \widetilde{H}\}$ and $\text{Time}(Y)$ denotes the computation time for $Y \in \{H, \widehat{H}, \widetilde{H}\}$. All experiments (including Section 4) were conducted by Matlab R2016 on a 16-core machine with 128 GB RAM. The results in this section are averaged over 10 random trials. Additional results are reported in the supplementary material.

The effect of average degree $\overline{d}$ and graph regularity parameter p_{WS}. Figures 1 (a) and 1 (b) display the exact and the two approximate VNGE of ER and BA graphs and the corresponding CTRR under varying $\overline{d}$. When fixing the number of nodes n, both $\widehat{H}$ and $\widetilde{H}$ better match H as $\overline{d}$ increases, suggesting their AE decays with $\overline{d}$. Comparing their CTRR, the computation of $\widehat{H}$ and $\widetilde{H}$ enjoys at least 97% speed-up relative to H. The drastic reduction in computation time can be explained by the efficient linear complexity of FINGER, as opposed to the high complexity in computing the entire eigenspectrum for calculating H. The CTRR of $\widehat{H}$ slightly decays with $\overline{d}$ due to the growing number of nonzero entries (edges) in $\mathbf{L}_N$, resulting in increasing operations for computing $\lambda_{\max}$. Although the AE of $\widehat{H}$ is always smaller than that of $\widetilde{H}$ due to the fact that $\widetilde{H} \leq \widehat{H} \leq H$, the CTRR of $\widetilde{H}$ has nearly 100% speed-up relative to H by simply requiring the information of $s_{\max}$ instead of $\lambda_{\max}$ from a graph.

Figure 1 (c) displays the AE and CTRR of $\widehat{H}$ and $\widetilde{H}$ under

varying edge rewiring probability p_{WS} and different average degree $\overline{d} \in \{6, 10, 20, 50\}$ of WS model. Similar to ER and BA graphs, when fixing n and p_{WS}, the AE of $\widehat{H}$ and $\widetilde{H}$ decays as $\overline{d}$ increases. When n and $\overline{d}$ are fixed, smaller p_{WS} yields less AE for both $\widehat{H}$ and $\widetilde{H}$, suggesting that FINGER attains better approximation when graphs are more regular. Since the curves of CTRR for different $\overline{d}$ in WS model have similar behavior, here we only report the results when $\overline{d} = 50$. Consistent with the observations in ER and BA graphs, in WS graphs the CTRR of $\widehat{H}$ and $\widetilde{H}$ achieves nearly 100% improvement relative to H, and $\widetilde{H}$ attains slightly better CTRR than $\widehat{H}$ at the price of larger AE.

The effect of graph size n. Figure 2 displays the SAE of FINGER under the three random graph models when varying the number of nodes n. Since the results of $\widehat{H}$ and $\widetilde{H}$ are similar, we show the SAE of $\widehat{H}$ in Figure 2 and report the SAE of $\widetilde{H}$ in the supplementary material. By the fact that ER and WS graphs have balanced eigenspectrum (Van Mieghem, 2010), for ER and WS models the SAE of both $\widehat{H}$ and $\widetilde{H}$ decays as n increases, which verifies the $o(\ln n)$ approximation error as stated in Corollaries 2 and 3. On the other hand, the SAE of BA graphs is observed to grow logarithmically in n due to the existence of extreme eigenvalues (imbalanced eigenspectrum) (Van Mieghem, 2010; Goh et al., 2001). Similar to the observations from fixed-size graphs, for a fixed n the SAE decays with $\overline{d}$ and graph regularity in all cases. In addition, the CTRR attains nearly 100% speed-up relative to H for moderate-size graphs ($n \geq 2000$).

4. Applications

Here we apply FINGER to the computation of Jensen-Shannon (JS) distance between graphs (Section 2.5) in two applications and one synthesized dataset and demonstrate its outstanding performance over seven baseline and state-of-the-art methods in terms of efficiency and effectiveness.

Anomaly detection in evolving Wikipedia hyperlink networks. Wikipedia is an online encyclopedia that allows editing and referencing between articles. By viewing an article as a node and a hyperlink as an edge, the evolution of Wikipedia forms a graph sequence $\{G_t\}_{t=1}^{T}$ over time. Table 1 summarizes four evolving Wikipedia networks of different language settings collected in (Mislove, 2009; Preusse et al., 2013), where each graph $G_t = (\mathcal{V}_t, \mathcal{E}_t, \mathbf{W}_t)$ corresponds to a monthly snapshot of a hyperlink network. These datasets are presented in terms of addition and deletion of nodes or edges with timestamps (i.e., continuous graph changes $\{\Delta G_t\}_{t=1}^{T-1}$), which directly applies to incremental JS distance computation via FINGER (Algorithm 2). Fast JS distance computation via FINGER (Algorithm 1) can also be applied by computing $G_{t+1} = G_t \oplus \Delta G_t$ to obtain $\{G_t\}_{t=1}^{T}$. The task of anomaly detection is to identify noticeable changes (relative to the bulk network) in the consecutive monthly snapshots of these massive Wikipedia hyperlink networks.

Bifurcation detection in dynamic genomic networks. The genome-wide chromosome conformation capture (Hi-C) contact maps (Beloqui et al., 2009) for studying cell reprogramming from human fibroblasts to skeletal muscle can be viewed as a graph sequence consisting of 12 sampled spatial measurements, in which the cell reprogramming undergoes a space-time bifurcation at the 6th measurement as verified in (Liu et al., 2018a). The task is to identify this bifurcation instance based on the dynamic Hi-C contact maps. Additional descriptions of this dataset are given in the supplementary material.

Evaluation. We note that there are two major differences between these two applications: (i) unweighted v.s. weighted graphs and (ii) with v.s. without ground truth.

In the Wikipedia case (unweighted graphs), our main goal is to use these large datasets to demonstrate the efficient computation of JS distance via FINGER owing to its linear complexity. Additionally, since there are no labels for verifying the detected changes, we conduct an ex post facto correlation analysis using an explicit and explainable anomaly metric – the vertex/edge overlapping (VEO) score (Papadimitriou et al., 2010). VEO is a properly normalized metric reflecting topological differences between two unweighted graphs, defined as $1 - \frac{2(|\mathcal{V}_t \cap \mathcal{V}_{t+1}| + |\mathcal{E}_t \cap \mathcal{E}_{t+1}|)}{|\mathcal{V}_t| + |\mathcal{V}_{t+1}| + |\mathcal{E}_t| + |\mathcal{E}_{t+1}|}$, which is between $[0, 1]$ and relates to the SorensenDice coefficient (Dice, 1945; Sørensen, 1948) for comparing the similarity of two samples. In the Wikipedia experiments, a high VEO score directly pinpoints the month when articles are edited by a relatively significant amount. Consequently, VEO can be used as an anomaly proxy for ex post facto analysis in our setting.

In the genome case (weighted graphs), the ground-truth bifurcation instance was verified. Moreover, unlike the Wikipedia case, the genome dataset contains nonnegative edge weights indicating cell interaction strengths. Therefore, in this case VEO is not an appropriate anomaly proxy because by definition it is insensitive to edge weight changes.

Comparative methods. We compare the proposed method with the following baseline methods:

- DeltaCon (Koutra et al., 2016): DeltaCon uses the idea of fast belief propagation to compute graph similarity and outputs a similarity score $\text{Sim}_{\text{DC}} \in [0, 1]$. We use $1 - \text{Sim}_{\text{DC}}$ as the anomaly score.
- RMD (Koutra et al., 2016): RMD is the Matusita distance deduced from DeltaCon, which is defined as $\frac{1}{\text{Sim}_{\text{DC}}} - 1$.
- λ distance (Bunke et al., 2007; Wilson & Zhu, 2008): The Euclidean distance between two sets of top k eigenvalues of a matrix. Here we use the weight matrix $\mathbf{W}$ (Adj.) and the graph Laplacian matrix $\mathbf{L}$ (Lap.), and set $k = 6$.

Table 1. Summary of four evolving Wikipedia hyperlink networks.

Datasets (graph sequence)	maximum # of nodes	maximum # of edges	# of graphs
Wikipedia - simple English (sEN)	100,312 (0.1 M)	746,086 (0.7 M)	122
Wikipedia - English (EN)	1,870,709 (1.8 M)	39,953,145 (39 M)	75
Wikipedia - French (FR)	2,212,682 (2.2 M)	24,440,537 (24 M)	121
Wikipedia - German (GE)	2,166,669 (2.1 M)	31,105,755 (31 M)	127

Table 2. Computation time (seconds) and Pearson correlation coefficient (PCC) between the anomaly proxy and different methods. FINGER attains the best PCC and time efficiency. The Spearman's rank correlation analysis is given in Table S1 of the supplement.

Datasets		FINGER -JS (Fast)	FINGER -JS (Inc.)	DeltaCon	RMD	λ dist. (Adj.)	λ dist. (Lap.)	GED	VNGE -NL	VNGE -GL
Wiki (sEN)	PCC	**0.5593**	0.3382	0.1596	0.1718	0.1871	-0.0095	-0.2036	0.2065	0.2462
	time	26.065	**0.7438**	44.952	44.952	150.16	99.905	1.666	13.574	30.483
Wiki (EN)	PCC	**0.9029**	0.5583	-0.2411	-0.1167	-0.0175	-0.1759	-0.3429	-0.0442	0.1519
	time	603.98	**13.975**	1846.1	1846.1	4417.7	2898.3	47.299	335.66	858.22
Wiki (FR)	PCC	**0.8183**	0.592	-0.1503	-0.1203	0.0133	-0.1877	-0.4915	0.0552	0.2349
	time	1038.6	**23.667**	2804.5	2804.5	6664.5	4411.4	83.398	474.42	1129.1
Wiki (GE)	PCC	**0.6764**	0.4619	-0.2035	-0.1542	0.0182	-0.3814	-0.4677	0.2194	0.2679
	time	1457.3	**32.647**	4184.1	4184.1	9462.5	6013.7	115.923	716.31	1674.6

• GED (Bunke et al., 2007): graph edit distance (GED) for undirected unweighted graphs is the number of operations (node/edge additions and removals) required to convert a graph G_t to another graph G_{t+1}.
• VNGE-NL (Han et al., 2012) / VNGE-GL (Ye et al., 2014): Two VNGE heuristics using the normalized/generalized graph Laplacian matrix. Unlike FINGER, they lack approximation error guarantee.

Wikipedia results. We compute the dissimilarity metrics of each method and compare them with the anomaly proxy in terms of the Pearson correlation coefficient (PCC). A higher PCC suggests a better match to the anomaly proxy for detecting abnormal monthly edit changes relative to the bulk network. The PCC and computation time of each method are reported in Table 2. For illustration, the dissimilarity metrics of Wikipedia-EN are shown in Figure 3. The plots of the other Wikipedia networks are given in the supplementary material. The statistics of the anomaly proxy meet the intuition that in the earlier stage the monthly evolution of Wikipedia is more drastic, and in the later stage it becomes stable (i.e., less anomalous) since the changes are subtle relative to the entire network. In Table 2, FINGER-JSdist (Fast) attains the best PCC (0.9029) and competitive computation time. This suggests that the computation of JS distance can be made efficient by FINGER, and its ex post facto analysis is highly correlated with the anomaly proxy. For example, in Figure 3 their top 10 flagged anomalies have 9 months in common. On the other hand, the other dissimilarity metrics are either implicitly defined, unnormalized or lacking approximation guarantees, making the detected anomalies less explainable. FINGER-JSdist (Incremental) has the least

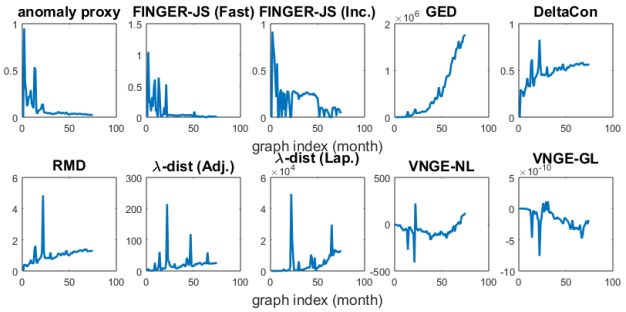


Figure 3. Dissimilarity (anomaly) metrics of consecutive monthly Wikipedia-English hyperlink networks. The ex post facto analysis shows FINGER-JSdist (Fast) is highly correlated with the anomaly proxy (0.9029 PCC in Table 2 and 0.7973 SRCC in Table S1). FINGER-JSdist (Incremental) has efficient computation time and attains the second best PCC and SRCC among all methods.

computation time by leveraging online computation, and it achieves the second best PCC due to looser approximation error of $\widetilde{H}$ than $\widehat{H}$. Nonetheless, FINGER-JSdist (Incremental) is roughly 3 times faster than GED, 20 times faster than VNGE-GL, 50 times faster than FINGER-JSdist (Fast), 100 times faster than DeltaCon, RMD and VNGE-NL, and 200-300 times faster than λ distance. In addition to PCC, we also report the rank correlation coefficients in the supplementary material to show the high correlation between FINGER and the anomaly proxy.

As discussed in the "Evaluation" paragraph, the main purpose of the Wikipedia experiments (without ground truths) is to show the efficiency in fast JS distance computation of large real-world graphs, enabled by FINGER. Additionally, our ex post facto analysis shows high correlation of FIN-

Table 3. Detection rate on synthesized anomalous events in the dynamic communication networks.

DoS attack ($X\%$)	FINGER -JS (Fast)	FINGER -JS (Inc.)	DeltaCon	RMD	λ dist. (Adj.)	λ dist. (Lap.)	GED	VNGE -NL	VNGE -GL
1 %	**24 %**	10%	14%	14%	10%	**24%**	14%	22%	22%
3 %	**75%**	62%	58%	58%	12%	23%	36%	39%	39%
5 %	**90%**	77%	**90%**	**90%**	12%	28%	41%	67%	67%
10 %	**91%**	**91%**	**91%**	**91%**	**91%**	**91%**	81%	**91%**	**91%**

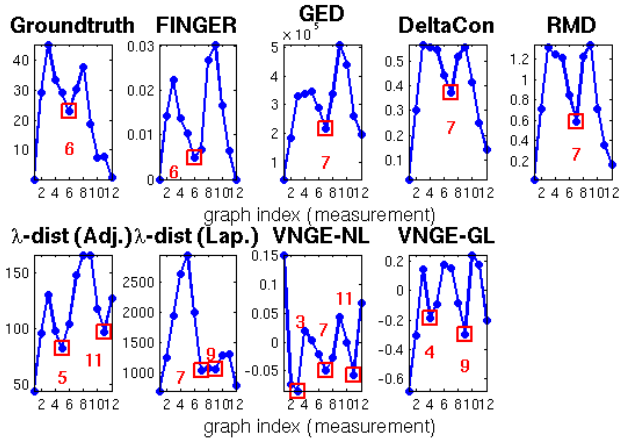


Figure 4. Bifurcation detection of cell reprogramming in dynamic genomic networks via the temporal difference score (TDS) of different methods (y-axis). The red squares indicate the detected bifurcation points. Among all the compared methods, FINGER-JSdist (Algorithm 1) is the only method that correctly detects the ground-truth bifurcation point (index 6), and its TDS resembles the shape of the ground-truth statistic.

GER with an explainable anomaly proxy. Beyond efficiency, we use the next two sets of experiments (with ground truths) to demonstrate the effectiveness of FINGER.

Bifurcation detection results. Using the ground-truth statistic provided by (Liu et al., 2018a), we compare the performance of detecting the critical bifurcation point by each method. Let $\theta_{t,t'}$ denote a dissimilarity metric between two graphs G_t and $G_{t'}$ from $\{G_t\}_{t=1}^T$. For each method, the temporal difference score (TDS) proposed in (Liu et al., 2018a) is used for bifurcation detection, which is defined as $\text{TDS}(t) = \frac{1}{2}[\theta_{t,t-1} + \theta_{t,t+1}]$ when $t \in \{2, \ldots, T-1\}$, and $\text{TDS}(1) = \theta_{1,2}$ and $\text{TDS}(T) = \theta_{T,T-1}$. The measurement(s) corresponding to a local minimum in TDS is detected as a bifurcation instance. The ground-truth statistic and TDS of each method are shown in Figure 4. Among all the compared methods, FINGER-JSdist (Algorithm 1) is the only method that correctly detects the bifurcation point (index 6), and its TDS based on JS distance also resembles the shape of the ground-truth statistic.

Synthesized anomaly detection results. For further validation, we use another real-world dynamic peering network dataset at the autonomous system (AS) level (the Oregon-1 dataset (Leskovec et al., 2005)) to synthesize anomalous connectivity patterns that mimic the denial-of-service (DoS) attacks. Here each graph represents the router connectivity over a certain time period, leading to 9 such graphs. We synthesize anomalous events by first selecting one graph from the first 8 graphs at random, and then connecting $X\%$ of nodes to a randomly chosen node in the selected graph. This synthesized connection pattern mimics that of the DoS attack, in which multiple nodes (e.g., a botnet) aim to connect to the target node simultaneously. The task is to detect this synthesized anomalous event by comparing the dissimilarity metric between consecutive graphs. Table 3 reports the detection rate of different methods, where the detection rate is defined as the fraction of 100 random instances in which the anomalous event appears in the top-2 ranking based on a dissimilarity metric. Tested on $X = \{1, 3, 5, 10\}\%$, FINGER-JS (Fast) consistently attains the best detection rate among all methods, suggesting the stability and superiority of the proposed method. On the other hand, the compared methods are not as robust as FINGER. Notably, when X is small (i.e., the more challenging case for detection as the attack becomes stealthier), the detection performance of FINGER is more sensible than other methods. As X becomes large, which means the DoS attack pattern is more apparent, the detection performance becomes similar.

5. Conclusion

In this paper, we proposed FINGER, a novel framework for efficiently computing von Neumann graph entropy (VNGE). FINGER reduces the computation of VNGE from cubic complexity to linear complexity for a given graph, and allows online computation based on incremental graph changes. In addition to bounded approximation error, our theory shows that FINGER is guaranteed to have asymptotic equivalence to the exact VNGE under mild conditions, which has been validated by extensive experiments on three different random graph models. The high efficiency of FINGER also leads to scalable network learning algorithms for computing Jensen-Shannon distance between graphs. Furthermore, we use two domain-specific applications and one synthesized dataset to corroborate the efficiency and effectiveness of FINGER compared to 7 baseline graph similarity methods. The results demonstrate the power of FINGER in tackling large network analysis and (unsupervised) learning problems in different domains. Our future work includes extension to directed graphs and negative edge weights.

Acknowledgment

Pin-Yu Chen, Lingfei Wu and Sijia Liu acknowledge the support from MIT-IBM Watson AI Lab. Indika Rajapakse is supported in part by the Lifelong Learning Machines program from DARPA/MTO.

References

Akoglu, L., Tong, H., and Koutra, D. Graph based anomaly detection and description: a survey. *Data Mining and Knowledge Discovery*, 29(3):626–688, 2015.

Anand, K. and Bianconi, G. Entropy measures for networks: Toward an information theory of complex topologies. *Physical Review E*, 80(4):045102, 2009.

Anand, K., Bianconi, G., and Severini, S. Shannon and von Neumann entropy of random networks with heterogeneous expected degree. *Physical Review E*, 83(3):036109, 2011.

Anderson Jr, W. N. and Morley, T. D. Eigenvalues of the Laplacian of a graph. *Linear and Multilinear Algebra*, 18 (2):141–145, 1985.

Bai, L. and Hancock, E. R. Depth-based complexity traces of graphs. *Pattern Recognition*, 47(3):1172–1186, 2014.

Bai, Z., Demmel, J., Dongarra, J., Ruhe, A., and van der Vorst, H. *Templates for the solution of algebraic eigenvalue problems: a practical guide*. SIAM, 2000.

Barabási, A.-L. and Albert, R. Emergence of scaling in random networks. *Science*, 286(5439):509–512, October 1999.

Beloqui, A., Guazzaroni, M.-E., Pazos, F., Vieites, J. M., Godoy, M., Golyshina, O. V., Chernikova, T. N., Waliczek, A., Silva-Rocha, R., Al-ramahi, Y., et al. Reactome array: forging a link between metabolome and genome. *Science*, 326(5950):252–257, 2009.

Braunstein, S. L., Ghosh, S., and Severini, S. The Laplacian of a graph as a density matrix: a basic combinatorial approach to separability of mixed states. *Annals of Combinatorics*, 10(3):291–317, 2006.

Briët, J. and Harremoës, P. Properties of classical and quantum Jensen-Shannon divergence. *Physical Review A*, 79(5):052311, 2009.

Bunke, H., Dickinson, P. J., Kraetzl, M., and Wallis, W. D. *A graph-theoretic approach to enterprise network dynamics*, volume 24. Springer Science & Business Media, 2007.

Chen, P.-Y. and Hero, A. O. Node removal vulnerability of the largest component of a network. In *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 587–590, 2013.

Chen, P.-Y., Choudhury, S., and Hero, A. O. Multi-centrality graph spectral decompositions and their application to cyber intrusion detection. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4553–4557, 2016.

Chung, F. Laplacians and the Cheeger inequality for directed graphs. *Annals of Combinatorics*, 9(1):1–19, 2005.

De Domenico, M., Nicosia, V., Arenas, A., and Latora, V. Structural reducibility of multilayer networks. *Nature Communications*, 6, 2015.

Del Vecchio, D., Abdallah, H., Qian, Y., and Collins, J. J. A blueprint for a synthetic genetic feedback controller to reprogram cell fate. *Cell Systems*, 2017.

Dice, L. R. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.

Du, W., Li, X., Li, Y., and Severini, S. A note on the von Neumann entropy of random graphs. *Linear Algebra and its Applications*, 433(11-12):1722–1725, 2010.

Endres, D. M. and Schindelin, J. E. A new metric for probability distributions. *IEEE Transactions on Information theory*, 49(7):1858–1860, 2003.

Erdös, P. and Rényi, A. On random graphs, I. *Publicationes Mathematicae (Debrecen)*, 6:290–297, 1959.

Fiedler, M. Algebraic connectivity of graphs. *Czechoslovak Mathematical Journal*, 23(98):298–305, 1973.

Goh, K.-I., Kahng, B., and Kim, D. Spectra and eigenvectors of scale-free networks. *Physical Review E*, 64(5):051903, 2001.

Han, L., Escolano, F., Hancock, E. R., and Wilson, R. C. Graph characterizations from von Neumann entropy. *Pattern Recognition Letters*, 33(15):1958–1967, 2012.

Horn, R. A. and Johnson, C. R. *Matrix Analysis*. Cambridge University Press, 1990.

Kalofolias, V. How to learn a graph from smooth signals. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, pp. 920929, 2016.

Kipf, T. N. and Welling, M. Semi-supervised classification with graph convolutional networks. *ICLR*, 2017.

Koutra, D., Shah, N., Vogelstein, J. T., Gallagher, B., and Faloutsos, C. DeltaCon: Principled massive-graph similarity function with attribution. *ACM Transactions on Knowledge Discovery from Data*, 10(3):28, 2016.

Leskovec, J., Kleinberg, J., and Faloutsos, C. Graphs over time: densification laws, shrinking diameters and possible explanations. In *ACM International Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 177–187, 2005.

Li, A. and Pan, Y. Structural information and dynamical complexity of networks. *IEEE Transactions on Information Theory*, 62(6):3290–3339, 2016.

Li, Z., Mucha, P. J., and Taylor, D. Network-ensemble comparisons with stochastic rewiring and von neumann entropy. *SIAM Journal on Applied Mathematics*, 78(2): 897–920, 2018.

Liao, R., Zhao, Z., Urtasun, R., and Zemel, R. S. Lanczosnet: Multi-scale deep graph convolutional networks. In *International Conference on Learning Representations (ICLR)*, 2019.

Liu, S., Chen, H., Ronquist, S., Seaman, L., Ceglia, N., Meixner, W., Chen, P.-Y., Higgins, G., Baldi, P., Smale, S., et al. Genome architecture mediates transcriptional control of human myogenic reprogramming. *iScience*, 6: 232–246, 2018a.

Liu, S., Chen, P.-Y., Hero, A., and Rajapakse, I. Dynamic network analysis of the 4d nucleome. *bioRxiv*, pp. 268318, 2018b.

Luo, D., Huang, H., Nie, F., and Ding, C. H. Forging the graphs: A low rank and positive semidefinite graph learning approach. In *Advances in Neural Information Processing Systems*, pp. 2960–2968, 2012.

Luxburg, U. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, December 2007.

Merris, R. Laplacian matrices of graphs: a survey. *Linear Algebra and its Applications*, 197-198:143–176, 1994.

Mislove, A. E. *Online social networks: measurement, analysis, and applications to distributed information systems*. PhD thesis, Rice University, 2009.

Papadimitriou, P., Dasdan, A., and Garcia-Molina, H. Web graph similarity for anomaly detection. *Journal of Internet Services and Applications*, 1(1):19–30, 2010.

Passerini, F. and Severini, S. The von Neumann entropy of networks. *arXiv preprint arXiv:0812.2597*, 2008.

Passerini, F. and Severini, S. Quantifying complexity in networks: the von Neumann entropy. *International Journal of Agent Technologies and Systems (IJATS)*, 1(4):58–67, 2009.

Preusse, J., Kunegis, J., Thimm, M., Staab, S., and Gottron, T. Structural dynamics of knowledge networks. In *International AAAI Conference on Weblogs and Social Media*, 2013.

Ranshous, S., Shen, S., Koutra, D., Harenberg, S., Faloutsos, C., and Samatova, N. F. Anomaly detection in dynamic networks: a survey. *Wiley Interdisciplinary Reviews: Computational Statistics*, 7(3):223–247, 2015.

Seaman, L., Chen, H., Brown, M., Wangsa, D., Patterson, G., Camps, J., Omenn, G. S., Ried, T., and Rajapakse, I. Nucleome analysis reveals structure-function relationships for colon cancer. *Molecular Cancer Research*, pp. molcanres–0374, 2017.

Sharpnack, J. L., Krishnamurthy, A., and Singh, A. Near-optimal anomaly detection in graphs using lovasz extended scan statistic. In *Advances in Neural Information Processing Systems*, pp. 1959–1967, 2013.

Shetty, J. and Adibi, J. Discovering important nodes through graph entropy the case of enron email database. In *Proceedings of the 3rd international workshop on Link discovery*, pp. 74–81. ACM, 2005.

Shi, J. and Malik, J. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(8): 888–905, 2000.

Shivanna, R. and Bhattacharyya, C. Learning on graphs using orthonormal representation is statistically consistent. In *Advances in Neural Information Processing Systems*, pp. 3635–3643, 2014.

Shuman, D., Narang, S., Frossard, P., Ortega, A., and Vandergheynst, P. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Process. Mag.*, 30(3):83–98, 2013.

Simonyi, G. Graph entropy: A survey. *Combinatorial Optimization*, 20:399–441, 1995.

Sørensen, T. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on danish commons. *Kongelige Danske Videnskabernes Selskab*, 5: 1–34, 1948.

Van Mieghem, P. *Graph Spectra for Complex Networks*. Cambridge University Press, 2010.

Von Neumann, J. *Mathematical foundations of quantum mechanics*. Number 2. Princeton university press, 1955.

Wang, Y., Wang, Y.-X., and Singh, A. Graph connectivity in noisy sparse subspace clustering. In *Artificial Intelligence and Statistics*, pp. 538–546, 2016.

Watts, D. J. and Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, June 1998.

Weintraub, H. The myod family and myogenesis: redundancy, networks, and thresholds. *Cell*, 75(7):1241–1244, 1993.

Weintraub, H., Tapscott, S. J., Davis, R. L., Thayer, M. J., Adam, M. A., Lassar, A. B., and Miller, A. D. Activation of muscle-specific genes in pigment, nerve, fat, liver, and fibroblast cell lines by forced expression of myod. *Proceedings of the National Academy of Sciences*, 86(14): 5434–5438, 1989.

Wilson, R. C. and Zhu, P. A study of graph spectra for comparing graphs and trees. *Pattern Recognition*, 41(9): 2833–2841, 2008.

Wu, L., Romero, E., and Stathopoulos, A. Primme_SVDS: A high-performance preconditioned svd solver for accurate large-scale computations. *SIAM Journal on Scientific Computing*, 39(5):S248–S271, 2017.

Wu, L., Chen, P.-Y., Yen, I. E.-H., Xu, F., Xia, Y., and Aggarwal, C. Scalable spectral clustering using random binning features. In *ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 2506–2515, 2018a.

Wu, L., Yen, I. E.-H., Xu, F., Ravikumar, P., and Witbrock, M. D2KE: From distance to kernel and embedding. *arXiv preprint arXiv:1802.04956*, 2018b.

Xu, K., Wu, L., Wang, Z., Feng, Y., Witbrock, M., and Sheinin, V. Graph2seq: Graph to sequence learning with attention-based neural networks. *arXiv preprint arXiv:1804.00823*, 2018.

Yanardag, P. and Vishwanathan, S. A structural smoothing framework for robust graph comparison. In *Advances in Neural Information Processing Systems*, pp. 2134–2142, 2015.

Ye, C., Wilson, R. C., Comin, C. H., Costa, L. d. F., and Hancock, E. R. Approximate von Neumann entropy for directed graphs. *Physical Review E*, 89(5):052804, 2014.