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

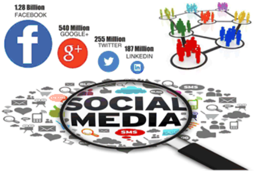
Pin-Yu Chen
IBM Research AI

joint work with
Lingfei Wu (IBM Research AI)
Sijia Liu (IBM Research AI)
Indika Rajapakse (Univ. Michigan Ann Arbor)

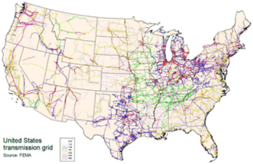
Poster: Tuesday 6:30-9:00 pm, Pacific Ballroom #265

June 10, 2019

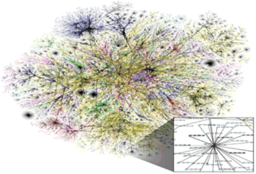
Graph as a Data Representation



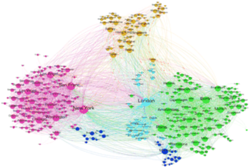
Social Network



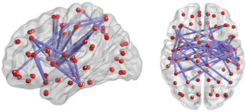
Power Grid



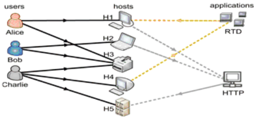
Communication Network



Information System



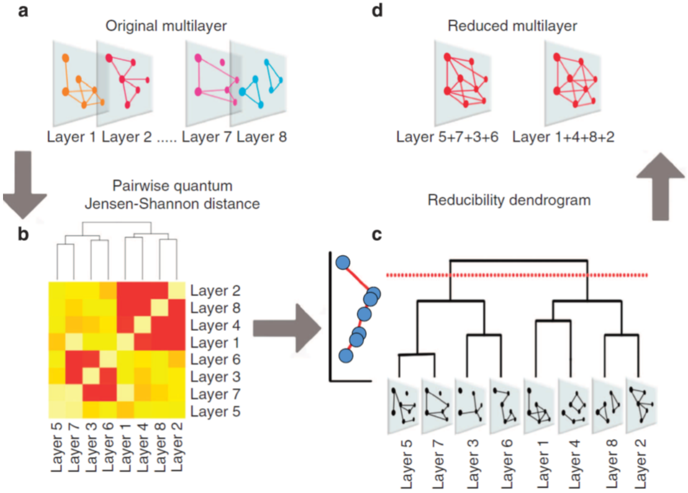
Bio Informatics



Cyber-Physical Sytem

Information-Theoretic Measures between Graphs

- Structural reducibility of multilayer networks (unsupervised learning)
- De Domenico et al., "Structural reducibility of multilayer networks." Nature Communications 6 (2015).



Von Neumann Graph Entropy (VNGE): Introduction

- Quantum information theory: $\mathbf{\Phi}$ is a $n \times n$ density matrix that is symmetric, positive semidefinite, and trace$(\mathbf{\Phi}) = 1$
  $\{\lambda_i\}_{i=1}^n$: eigenvalues of $\mathbf{\Phi}$
- Von Neumann entropy $H = -$trace$(\mathbf{\Phi} \ln \mathbf{\Phi}) = -\sum_{i:\lambda_i>0} \lambda_i \ln \lambda_i$
  $\rightarrow$ Shannon entropy over eigenspectrum $\{\lambda_i\}_{i=1}^n$, since $\sum_i \lambda_i = 1$
  $\Rightarrow$ Generally requires $O(n^3)$ computation complexity for H
- Graph $G = (\mathcal{V}, \mathcal{E}, \mathbf{W}) \in \mathcal{G}$: undirected weighted graphs with nonnegative edge weights. G has $|\mathcal{V}| = n$ nodes and $|\mathcal{E}| = m$ edges.
- $\mathbf{L} = \mathbf{D} - \mathbf{W}$: combinatorial graph Laplacian matrix of G.
  $\mathbf{D} =$ diag$(\{\lambda_i\})$: diagonal degree matrix. $[\mathbf{W}]_{ij} = w_{ij}$: edge weight.
- Von Neumann graph entropy (VNGE): $\mathbf{\Phi} = \mathbf{L}_\mathcal{N} = c \cdot \mathbf{L}$, where
  $c = \frac{1}{\text{trace}(\mathbf{L})} = \frac{1}{\sum_{i \in \mathcal{V}} d_i} = \frac{1}{2\sum_{(i,j)\in\mathcal{E}} w_{ij}}$
- $H \leq \ln(n-1)$, "=" when G is a complete graph with identical edge weight

Braunstein, Samuel L., Sibasish Ghosh, and Simone Severini. "The Laplacian of a graph as a density matrix: a basic combinatorial approach to separability of mixed states." Annals of Combinatorics 10.3 (2006): 291-317.
Passerini, Filippo, and Simone Severini. "The von Neumann entropy of networks." (2008).

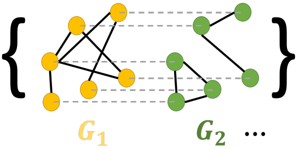
Von Neumann Graph Entropy (VNGE): Introduction

- VNGE characterizes structural complexity of a graph and enables computation of Jensen-Shannon distance (JSdist) between graphs.
- Applications in network learning, computer vision and data science:
  1. Structural reducibility of multilayer networks (hierarchical clustering)
     De Domenico et al., "Structural reducibility of multilayer networks." Nature Communications 6 (2015).
  2. Depth-analysis for image processing
     Han, Lin, et al. "Graph characterizations from von Neumann entropy." Pattern Recognition Letters 33.15 (2012): 1958-1967.
     Bai, Lu, and Edwin R. Hancock. "Depth-based complexity traces of graphs." Pattern Recognition 47.3 (2014): 1172-1186.
  3. Network-ensemble comparison via edge rewiring
     Li, Zichao, Peter J. Mucha, and Dane Taylor. "Network-ensemble comparisons with stochastic rewiring and von Neumann entropy." SIAM Journal on Applied Mathematics, 78(2): 897920 (2018).
  4. Structure-function analysis in genetic networks
     Liu et al., "Dynamic network analysis of the 4D nucleome." bioRxiv, pp. 268318 (2018).
- High consistency with classical Shannon graph entropy that is defined as a probability distribution of a function on subgraphs of G.

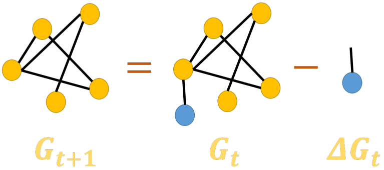
Anand, Kartik, Ginestra Bianconi, and Simone Severini. "Shannon and von Neumann entropy of random networks with heterogeneous expected degree." Physical Review E 83.3 (2011): 036109.
Anand, Kartik, and Ginestra Bianconi. "Entropy measures for networks: Toward an information theory of complex topologies." Physical Review E 80.4 (2009): 045102.
Li, Angsheng, and Yicheng Pan. "Structural Information and Dynamical Complexity of Networks." IEEE Transactions on Information Theory 62.6 (2016): 3290-3339.

Outline

- The main challenge of exact VNGE computation: it generally requires cubic complexity $O(n^3)$ for obtaining the full eigenspectrum
  $\rightarrow$ NOT scalable to large graphs
- Our solution: **FINGER**, a scalable and provably asymptotically correct approximate computation framework of VNGE
- FINGER supports two different data modes: *batch* and *online*



(a) Batch mode: $O(n+m)$



(b) Online mode: $O(\Delta n + \Delta m)$

- New applications:
  1. Anomaly detection in evolving Wikipedia hyperlink networks
  2. Bifurcation detection of cellular networks during cell reprogramming
  3. Synthesized denial of service attack detection in router networks

Efficient VNGE Computation via FINGER

- Recall $H = -\sum_{i=1}^{n} \lambda_i \ln \lambda_i \Rightarrow O(n^3)$ cubic complexity
- FINGER enables fast and incremental computation of H with asymptotic approximation guarantee

Lemma (Quadratic approximation of H)

The quadratic approximation of the von Neumann graph entropy H via Taylor expansion is equivalent to $Q = 1 - c^2(\sum_{i \in \mathcal{V}} d_i^2 + 2 \cdot \sum_{(i,j) \in \mathcal{E}} w_{ij}^2)$

- d_i: degree (sum of edge weights) of node i
- w_{ij} : edge weight of edge (i,j)
- $c = \frac{1}{2 \sum_{(i,j) \in \mathcal{E}} w_{ij}}$
- $O(n+m)$ linear complexity. $|\mathcal{V}| = n$, $|\mathcal{E}| = m$.
- Q can be incremental updated given graph changes ΔG $\Rightarrow O(\Delta n + \Delta m)$ complexity

Approximate VNGE with Asymptotic Guarantees

- Let $\lambda_{\max}$ ($\lambda_{\min}$) be the largest (smallest) positive eigenvalue in $\{\lambda_i\}$
- Approx. VNGE for *batch* graph sequence: $\widehat{H}(G) = -Q \ln \lambda_{\max}$
- Approx. VNGE for *online* graph sequence: $\widetilde{H}(G) = -Q \ln(2c \cdot d_{\max})$
- Relation: $\widetilde{H} \leq \widehat{H} \leq H$

Theorem ($o(\ln n)$ approximation error with balanced eigenspectrum)

If the number of positive eigenvalues $n_+ = \Omega(n)$ and $\lambda_{\min} = \Omega(\lambda_{\max})$, the scaled approximation error (SAE) $\frac{H-\widehat{H}}{\ln n} \to 0$ and $\frac{H-\widetilde{H}}{\ln n} \to 0$ as $n \to \infty$.

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

- Computing $\lambda_{\max}$ only requires $O(n+m)$ operations via power iteration $\Rightarrow O(n+m)$ linear complexity for $\widehat{H}$.

Theorem (Incremental update of $\widetilde{H}$ with $O(\Delta n + \Delta m)$ complexity)

The VNGE $\widetilde{H}(G \oplus \Delta G)$ can be updated by $\widetilde{H}(G \oplus \Delta G) = F(\widetilde{H}(G), \Delta G)$

Numerical Validation on Synthetic Random Graphs

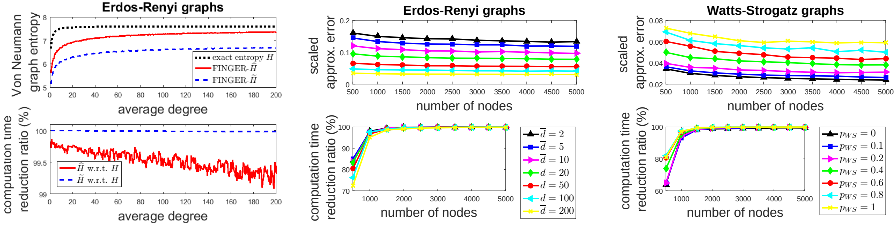


Figure: Scaled approximation error (SAE) and computation time reduction ratio

- scaled approximation error (SAE) $= \frac{H - H_{\text{approx}}}{\ln n}$
- computation time reduction ratio $= \frac{Time_H - Time_{H_{\text{approx}}}}{Time_H}$
- almost 100% speed-up ($O(n^3)$ v.s. $O(n+m)$)
- approximation error decreases as average degree increases
- regular (random) graphs have smaller (larger) approximation error

Jensen-Shannon Distance between Graphs using FINGER

- Two graphs G and $\widetilde{G}$ of the same node set $\mathcal{V}$.
- KL divergence $D_{KL}(G|\widetilde{G}) = \text{trace}(\mathbf{L}_{\mathcal{N}}(G) \cdot [\ln \mathbf{L}_{\mathcal{N}}(G) - \ln \mathbf{L}_{\mathcal{N}}(\widetilde{G})])$ (not symmetric)
- Let $\overline{G} = \frac{G \oplus \widetilde{G}}{2}$ denote the averaged graph of G and $\widetilde{G}$, where $\mathbf{L}_{\mathcal{N}}(\overline{G}) = \frac{\mathbf{L}_{\mathcal{N}}(G) + \mathbf{L}_{\mathcal{N}}(\widetilde{G})}{2}$.
- The Jensen-Shannon divergence is defined as $\text{DIV}_{JS}(G, \widetilde{G}) = \frac{1}{2} D_{KL}(G|\widetilde{G}) + \frac{1}{2} D_{KL}(\widetilde{G}|G) = H(\overline{G}) - \frac{1}{2}[H(G) + H(\widetilde{G})]$ (symmetric)
- The Jensen-Shannon distance is defined as $\text{JSdist}(G, \widetilde{G}) = \sqrt{\text{DIV}_{JS}}$, which is proved to be a valid distance metric.

Briet, Jop, and Peter Harremos. "Properties of classical and quantum Jensen-Shannon divergence." Physical review A 79.5 (2009): 052311.

FINGER Algorithms for Jensen-Shannon Distance

- Jensen-Shannon distance computation via FINGER-$\widehat{H}$ (batch mode):

  Input: Two graphs G and $\widetilde{G}$

  Output: JSdist$(G, \widetilde{G})$

  1. Obtain $\overline{G} = \frac{G \oplus \widetilde{G}}{2}$ and compute $\widehat{H}(G)$, $\widehat{H}(\widetilde{G})$, and $\widehat{H}(\overline{G})$ via FINGER (Fast)
  2. JSdist$(G, \widetilde{G})$= $\widehat{H}(\overline{G}) - \frac{1}{2}[\widehat{H}(G) + \widehat{H}(\widetilde{G})]$

  $\Rightarrow O(n+m)$ complexity inherited from $\widehat{H}$

- Jensen-Shannon distance computation via FINGER-$\widetilde{H}$ (online mode):

  Input: Graph G and its changes ΔG, Approx VNGE $\widetilde{H}(G)$ of 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 (Inc.)
  2. JSdist$(G, G \oplus \Delta G)$= $\widetilde{H}(G \oplus \frac{\Delta G}{2}) - \frac{1}{2}[\widetilde{H}(G) + \widetilde{H}(G \oplus \Delta G)]$

  $\Rightarrow O(\Delta n + \Delta m)$ complexity inherited from $\widetilde{H}$

- $o(\sqrt{\ln n})$ approximation guarantee of JSdist via FINGER (see paper)

Application I: Anomaly Detection in Wikipedia Networks

- Compare dissimilarity metrics of consecutive graphs via FINGER and other baseline methods:
  1. DeltaCon & RMD
  2. λ distance (6 leading eigenvalues) & graph edit distance (GED)
  3. VNGE-NL & VNGE-GL
  4. divergence based on degree distribution

Table: 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

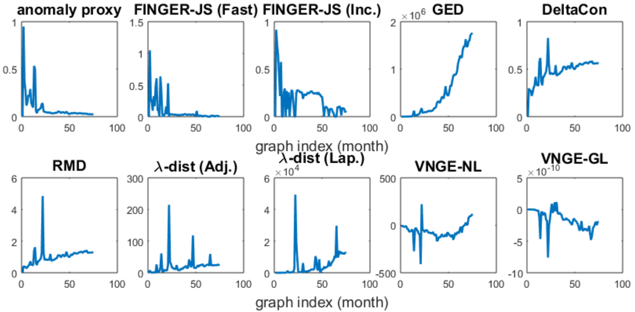
- Node: article. Edge: existence of hyperlinks.
  Graph: monthly hyperlink network.
- Anomaly proxy : vextex/edge overlapping dissimilarity

VEO $(G, \widetilde{G}) = 1 - \frac{2(|\mathcal{V} \cap \widetilde{\mathcal{V}}| + |\mathcal{E} \cap \widetilde{\mathcal{E}}|)}{|\mathcal{V}| + |\widetilde{\mathcal{V}}| + |\mathcal{E}| + |\widetilde{\mathcal{E}}|}$

Application I: Anomaly Detection in Wikipedia Networks

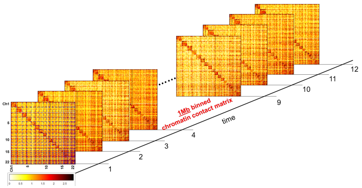
Table: Computation time (sec.) and Pearson correlation coefficient (PCC) of anomaly proxy and different methods. FINGER attains the best PCC and efficiency.

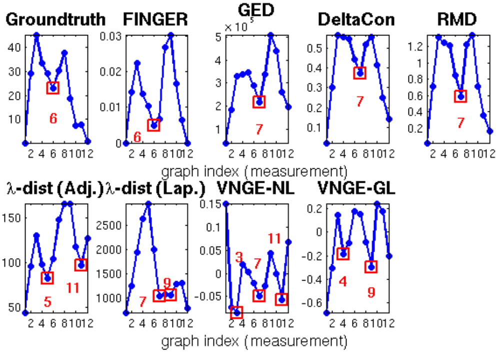
Datasets		FINGER -JS (Fast)	FINGER -JS (Inc.)	DeltaCon	RMD	λ dist. (Adj.)	λ dist. (Lap.)	GED	VNGE -NL	VNGE -GL
Wiki	PCC	**0.5593**	0.3382	0.1596	0.1718	0.1871	-0.0095	-0.2036	0.2065	0.2462
(sEN)	time	26.065	**0.7438**	44.952	44.952	150.16	99.905	1.666	13.574	30.483
Wiki	PCC	**0.9029**	0.5583	-0.2411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



Application II: Detection of Bifurcation Time Instance in Dynamic Cellular Networks

- Genome-wide chromosome conformation capture contact maps among 3K cells with 12 observations
- Cellular reprogramming from human fibroblasts to skeletal muscle at some critical time instance (index 6) - Liu et al., iScience (2018)
- Temporal difference score TDS(G_t)= $\frac{dist(G_t, G_{t-1}) + dist(G_t, G_{t+1})}{2}$





Application III: Synthesized Attacks in Router Networks

- Connectivity pattern of 9 real-world autonomous system level router communication graph
- Synthesize the connectivity pattern of distributed denial of service (DoS) attacks by randomly selecting one graph and then connecting $X\%$ of nodes to a randomly chosen node in the selected graph

Table: Average detection rate on synthesized anomalous events

DoS attack ($X\%$)	FINGER -JS (Fast)	FINGER -JS (Inc.)	DeltaCon	RMD	λ dist. (Adj.)	λ dist. (Lap.)	GED	VNGE -NL	VNGE -GL	VEO	Cosine distance	Bhattacharyya distance	Hellinger distance
1 %	**24 %**	10%	14%	14%	10%	**24%**	14%	22%	22%	14%	12%	10%	12%
3 %	**75%**	62%	58%	58%	12%	23%	36%	39%	39%	36%	35%	14%	16%
5 %	**90%**	77%	**90%**	**90%**	12%	28%	41%	67%	67%	41%	37%	37%	34%
10 %	**91%**	**91%**	**91%**	**91%**	**91%**	**91%**	81%	**91%**	**91%**	46%	46%	67%	71%

- FINGER consistently outperforms other dissimilarity metrics for different X
- When X is small (difficult case for detection), JSdist via FINGER is more sensible than other methods
- When X is large (easy case), the performance becomes similar

Conclusion and Future Work

- An efficient framework (FINGER) for fast and incremental computation of von Newman Graph Entropy and Jensen-Shannon graph distance
- For batch graph mode, FINGER features linear complexity $O(n+m)$. For online graph mode, FINGER features incremental complexity $O(\Delta n + \Delta m)$. Both modes have asymptotic approximation guarantee.
- New applications in anomaly detection and bifurcation detection
- Code: https://github.com/pinyuchen/FINGER
- Future work:
  1. stochastic computation of Jensen-Shannon distance via sampling
  2. extension to directed graphs, and graphs with negative weights
  3. applications involving graph distance: e.g., brain networks, traffic networks, unsupervised and active learning
- Contact: pin-yu.chen at ibm.com; pinyuchenTW (Twitter)
- Poster: Tuesday 6:30-9:00 pm, Pacific Ballroom #265