Supplementary Material

A. Proof of Lemma 1

For any real x such that $0 < x < 1$, it is easy to show that the Taylor series expansion of $-x \ln x$ at 1 is $\sum_{z=1}^{\infty} \frac{(-1)^z}{z} x(x-1)^z$. Applying this result to the term $-\lambda_i \ln \lambda_i$ in H and taking the quadratic approximation of the series expansion gives

$$Q = \sum_{i=1}^{n} \lambda_i (1 - \lambda_i) = 1 - \sum_{i=1}^{n} \lambda_i^2 \tag{S1}$$

since by definition $\sum_{i=1}^{n} \lambda_i = \text{trace}(\mathbf{L}_N) = 1$. The term $\sum_{i=1}^{n} \lambda_i^2$ in (S1) can be expressed as

$$\begin{aligned}
\sum_{i=1}^{n} \lambda_i^2 &= \text{trace}(\mathbf{L}_N^2) && \text{(S2)}\\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} [\mathbf{L}_N]_{ij} [\mathbf{L}_N]_{ji} && \text{(S3)}\\
&\overset{(a)}{=} \sum_{i=1}^{n} \sum_{j=1}^{n} [\mathbf{L}_N]_{ij}^2 \\
&\overset{(b)}{=} c^2 \left(\sum_{i=1}^{n} [\mathbf{L}]_{ii}^2 + \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} [\mathbf{L}]_{ij}^2 \right) && \text{(S4)}\\
&\overset{(c)}{=} c^2 \left(\sum_{i \in \mathcal{V}} s_i^2 + 2 \sum_{(i,j) \in \mathcal{E}} w_{ij}^2 \right), && \text{(S5)}
\end{aligned}$$

where (a) is due to the matrix symmetry of $\mathbf{L}_N$, (b) is due to the definition that $\mathbf{L}_N = c \cdot \mathbf{L}$, and (c) is due to the definition of $\mathbf{L}$ such that $[\mathbf{L}]_{ii} = s_i$, and $[\mathbf{L}]_{ij} = w_{ij}$ when $(i,j) \in \mathcal{E}$ and $[\mathbf{L}]_{ij} = 0$ otherwise. Furthermore, define

$$S = \text{trace}(\mathbf{L}) = \sum_{i=1}^{n} [\mathbf{L}]_{ii} = \sum_{i \in \mathcal{V}} s_i = 2 \sum_{(i,j) \in \mathcal{E}} w_{ij}. \tag{S6}$$

Using the relation $c = \frac{1}{\text{trace}(\mathbf{L})}$, we obtain the expression $Q = 1 - c^2 \left(\sum_{i \in \mathcal{V}} s_i^2 + 2 \sum_{(i,j) \in \mathcal{E}} w_{ij}^2 \right)$, where $c = \frac{1}{S}$ and $S = \sum_{i \in V} s_i = 2 \sum_{(i,j) \in \mathcal{E}} w_{ij}$.

B. Proof of Theorem 1

The assumption $\lambda_{\max} < 1$ implies $0 < \lambda_i \leq \lambda_{\max} < 1$ for all nonzero eigenvalues λ_i. Following the definition of H, we can rewrite H as

$$\begin{aligned}
H &= -\sum_{i=1}^{n} \lambda_i \ln \lambda_i && \text{(S7)}\\
&= -\sum_{i:\lambda_i > 0} \lambda_i \ln \lambda_i && \text{(S8)}\\
&= -\sum_{i:\lambda_i > 0} \lambda_i (1 - \lambda_i) \frac{\ln \lambda_i}{1 - \lambda_i}. && \text{(S9)}
\end{aligned}$$

Since for all $\lambda_i > 0$, $\ln \lambda_{\min} \leq \ln \lambda_i \leq \ln \lambda_{\max} < 0$ and $0 < 1 - \lambda_{\max} \leq 1 - \lambda_i \leq 1 - \lambda_{\min} < 1$, we obtain the relation

$$\frac{-\ln \lambda_{\max}}{1 - \lambda_{\min}} \leq \frac{-\ln \lambda_i}{1 - \lambda_i} \leq \frac{-\ln \lambda_{\min}}{1 - \lambda_{\max}}. \tag{S10}$$

Using $Q = \sum_{i=1}^{n} \lambda_i (1 - \lambda_i) = \sum_{i:\lambda_i > 0} \lambda_i (1 - \lambda_i)$ in (S1) and applying (S10) to (S9) yields

$$-Q \frac{\ln \lambda_{\max}}{1 - \lambda_{\min}} \leq H \leq -Q \frac{\ln \lambda_{\min}}{1 - \lambda_{\max}}. \tag{S11}$$

When G is a complete graph with identical edge weight $x > 0$, it can be shown that the eigenvalues of $\mathbf{L}$ have 1 eigenvalue at 0 and $n - 1$ identical eigenvalues at nx (Merris, 1994). Since the trace normalization constant $c = \frac{1}{\text{trace}(\mathbf{L})} = \frac{1}{(n-1)nx}$, the eigenvalues of $\mathbf{L}_N = c \cdot \mathbf{L}$ are $\lambda_n = 0$ and $\lambda_i = \frac{nx}{(n-1)nx} = \frac{1}{n-1}$ for all $1 \leq i \leq n-1$, which implies $H = \ln(n-1)$. It is easy to see that in this case $Q = 1 - \frac{1}{n-1} = 1 - \lambda_{\min} = 1 - \lambda_{\max}$ and $-\ln \lambda_{\max} = -\ln \lambda_{\min} = \ln(n-1)$. Consequently, the bounds in (S11) become exact and $H = \ln(n-1)$ when G is a complete graph with identical edge weight.

C. On the condition $\lambda_{\max} < 1$ in Theorem 1

Here we show that the condition $\lambda_{\max} < 1$ is always satisfied with any graph $G \in \mathcal{G}$ having a connected subgraph with at least 3 nodes. By definition, $\lambda_{\max} \leq 1$ since it is the largest eigenvalue of the scaled matrix $\mathbf{L}_N = \mathbf{L}/\text{trace}(\mathbf{L})$. Since any connected subgraph with at least 3 nodes will contribute to at least 2 positive eigenvalues of $\mathbf{L}_N$ (Van Mieghem, 2010; Chen & Hero, 2013) and all eigenvalues of $\mathbf{L}_N$ sum to 1, we have $\lambda_{\max} < 1$.

D. Proof of Corollary 1

Since $\sum_{i=1}^{n} \lambda_i = 1$, the condition $\lambda_{\min} = \Omega(\lambda_{\max})$ implies $\lambda_{\max}$ and $\lambda_{\min}$ are of the same order $\frac{1}{n_+}$, where n_+ is the number of positive eigenvalues of $\mathbf{L}_N$. When the condition $n_+ = \Omega(n)$ also holds, then $\lambda_{\max} = \frac{a}{n}$ and $\lambda_{\min} = \frac{b}{n}$ for some constants a, b such that $a \geq b > 0$, and we obtain

$$\lim_{n \to \infty} -\frac{1}{\ln n} \cdot \frac{\ln \lambda_{\max}}{1 - \lambda_{\min}} = \lim_{n \to \infty} \frac{1}{\ln n} \cdot \frac{\ln n - \ln a}{1 - \frac{b}{n}} = 1. \tag{S12}$$

Similarly,

$$\lim_{n\to\infty} -\frac{1}{\ln n}\cdot\frac{\ln\lambda_{\min}}{1-\lambda_{\max}} = 1. \tag{S13}$$

Taking the limit of $\frac{H}{\ln n}$ and applying (S12) and (S13) to the bounds in (S11), we obtain

$$\lim_{n\to\infty}\frac{H}{\ln n} - Q = 0, \tag{S14}$$

which completes the proof.

E. Proof of Corollary 2

Following the proof of Corollary 1, if $n_+ = \Omega(n)$ and $\lambda_{\min} = \Omega(\lambda_{\max})$, then $\lambda_{\max} = \frac{a}{n}$ and $\lambda_{\min} = \frac{b}{n}$ for some constants a, b such that $a \geq b > 0$. We have

$$\begin{aligned}
\lim_{n\to\infty}\frac{H-\widehat{H}}{\ln n} &= \lim_{n\to\infty}\frac{H}{\ln n} - Q + Q - \frac{\widehat{H}}{\ln n} && \text{(S15)}\\
&\overset{(a)}{=} \lim_{n\to\infty} Q - \frac{\widehat{H}}{\ln n} && \text{(S16)}\\
&\overset{(b)}{=} \lim_{n\to\infty} Q - Q\cdot\frac{\ln n - \ln a}{\ln n} && \text{(S17)}\\
&= 0, && \text{(S18)}
\end{aligned}$$

where (a) uses (S14) and (b) uses the definition of $\widehat{H}$ in (1) and $\lambda_{\max} = \frac{a}{n}$. This implies the approximation error $H - \widehat{H}$ decays with $\ln n$. That is, $H - \widehat{H} = o(\ln n)$.

F. Proof of Corollary 3

Let $\mu_{\max}$ denote the largest eigenvalue of the graph Laplacian matrix $\mathbf{L}$ of a graph $G \in \mathcal{G}$. Then it is known that $\frac{n}{n-1}s_{\max} \leq \mu_{\max} \leq 2s_{\max}$, where the lower bound is proved in (Fiedler, 1973) and the upper bound is proved in (Anderson Jr & Morley, 1985). These bounds suggest that $\mu_{\max}$ has asymptotically the same order as $s_{\max}$. Moreover, since by definition $\mathbf{L}_N = c\cdot\mathbf{L}$, it holds that $\lambda_{\max} = c\cdot\mu_{\max}$ and hence $\lambda_{\max} = O(c\cdot s_{\max})$. Following the proof of Corollary 1, if $n_+ = \Omega(n)$ and $\lambda_{\min} = \Omega(\lambda_{\max})$, then $\lambda_{\max} = \frac{a}{n}$ and $\lambda_{\min} = \frac{b}{n}$ for some constants a, b such that $a \geq b > 0$, and $2c\cdot s_{\max} = \frac{\gamma}{n}$ for some $\gamma > 0$ since $\lambda_{\max} = O(c\cdot s_{\max})$. Similar to the proof of Corollary 2, we have

$$\begin{aligned}
\lim_{n\to\infty}\frac{H-\widetilde{H}}{\ln n} &= \lim_{n\to\infty}\frac{H}{\ln n} - Q + Q - \frac{\widetilde{H}}{\ln n} && \text{(S19)}\\
&\overset{(a)}{=} \lim_{n\to\infty} Q - \frac{\widetilde{H}}{\ln n} && \text{(S20)}\\
&\overset{(b)}{=} \lim_{n\to\infty} Q - Q\cdot\frac{\ln n - \ln\gamma}{\ln n} && \text{(S21)}\\
&= 0, && \text{(S22)}
\end{aligned}$$

where (a) uses (S14) and (b) uses the definition of $\widetilde{H}$ in (2) and $2c\cdot s_{\max} = \frac{\gamma}{n}$. This implies the approximation error $H - \widetilde{H}$ decays with $\ln n$. That is, $H - \widetilde{H} = o(\ln n)$.

G. Proof of Theorem 2

Let $\mathbf{L}$ and $\mathbf{L}'$ denote the graph Laplacian matrix of G and G', respectively, and let $\mathbf{L}_N = c\cdot\mathbf{L}$ and $\mathbf{L}'_{\mathcal{N}} = c'\cdot\mathbf{L}'$ be the corresponding trace-normalized matrices. Since $S = \text{trace}(\mathbf{L}) = 2\sum_{(i,j)\in\mathcal{E}} w_{ij}$ and $\Delta S = 2\sum_{(i,j)\in\Delta\mathcal{E}}\Delta w_{ij}$, it is easy to show that $\text{trace}(\mathbf{L}') = S + \Delta S = 1/c'$. We have

$$c' - c = \frac{1}{S+\Delta S} - \frac{1}{S} = \frac{-\Delta S}{(S+\Delta S)S} = -cc'\Delta S \tag{S23}$$

since $c' = 1/\text{trace}(\mathbf{L}')$ and $c = 1/\text{trace}(\mathbf{L})$. This then implies $c' = \frac{c}{1+c\Delta S}$ and

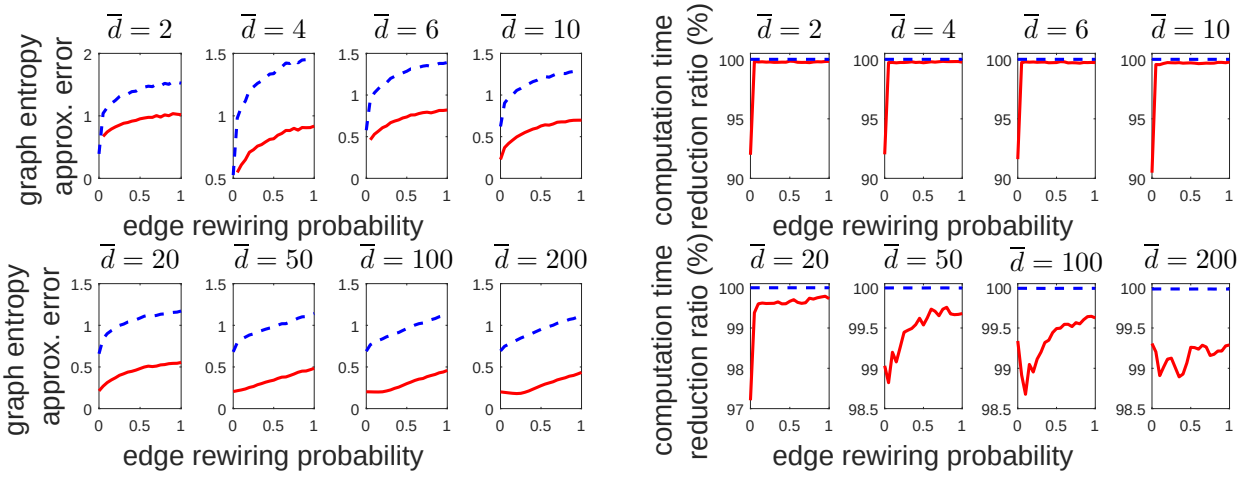
$$\Delta c = c' - c = \frac{-c^2\Delta S}{1+c\Delta S}. \tag{S24}$$

Using the expression of quadratic approximation for VNGE in Lemma 1 and the relation that $G' = G \oplus \Delta G$, we have

$$\begin{aligned}
&Q - Q'\\
&= (c+\Delta c)^2\left(\sum_{i\in\mathcal{V}}(s_i+\Delta s_i)^2 + 2\sum_{(i,j)\in\mathcal{E}}(w_{ij}+\Delta w_{ij})^2\right)\\
&\quad - c^2\left(\sum_{i\in\mathcal{V}} s_i^2 + 2\sum_{(i,j)\in\mathcal{E}} w_{ij}^2\right) && \text{(S25)}\\
&= (2\Delta c + \Delta c^2)\left(\sum_{i\in\mathcal{V}} s_i^2 + 2\sum_{(i,j)\in\mathcal{E}} w_{ij}^2 + \Delta Q\right)\\
&\quad + c^2\Delta Q, && \text{(S26)}
\end{aligned}$$

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 we use the convention $\Delta s_i = 0$ and $\Delta w_{ij} = 0$ when there are no changes made in the nodal strength of node i and in the weight of edge (i,j) from G to G', respectively. Since $Q = 1 - c^2\left(\sum_{i\in\mathcal{V}} s_i^2 + 2\sum_{(i,j)\in\mathcal{E}} w_{ij}^2\right)$, replacing $\sum_{i\in\mathcal{V}} s_i^2 + 2\sum_{(i,j)\in\mathcal{E}} w_{ij}^2$ with $\frac{1-Q}{c^2}$ in (S26) and using the relation $c' = c + \Delta c$ yields

$$Q' = \left(\frac{c'}{c}\right)^2 Q - c'^2\Delta Q - \frac{2\Delta c + \Delta c^2}{c^2}. \tag{S27}$$



(a) Approximation error (b) CTRR

Figure S1. Approximation error and computation time reduction ratio (CTRR) of FINGER under different average degree $\overline{d}$ of WS model. The red solid line and blue dashed line refer to the results of $\widehat{H}$ and $\widetilde{H}$, respectively. Both $\widehat{H}$ and $\widetilde{H}$ achieve at least 97% speed-up relative to the computation of H in all cases. It is observed that $\widetilde{H}$ has larger approximation error than $\widehat{H}$ but better CTRR.

Using the result from (S24) that $\frac{c'}{c} = \frac{1}{1+c\Delta S}$, we can further simplify (S27) as

$$Q' = \frac{Q}{(1+c\Delta S)^2} - \left(\frac{c}{1+c\Delta S}\right)^2 \Delta Q - \frac{1}{(1+c\Delta S)^2} + 1 \tag{S28}$$

$$= \frac{Q-1}{(1+c\Delta S)^2} - \left(\frac{c}{1+c\Delta S}\right)^2 \Delta Q + 1, \tag{S29}$$

which completes the proof.

H. Finite-size analysis and asymptotic equivalence of JS distance using FINGER

Beyond asymptotic analysis, we believe our results can provide new insights to finite-size analysis, especially based on the facts that: (i) our entropy inequality $\widetilde{H} \leq \widehat{H} \leq H$ is a finite-size result; (ii) The VNGE approximation error rate $o(\ln n)$ is in fact optimal in n for any finite-size analysis, since Theorem 1 shows that the rate is tight for complete graphs with identical edge weights.

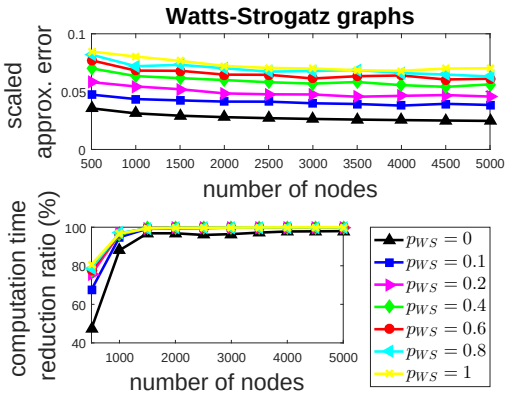
Furthermore, based on the asymptotic equivalence results of VNGE, it is straightforward to establish asymptotic equivalence of JS distance using FINGER as described in Algorithms 1 and 2. Let JS denote the exact JS distance and $\mathsf{JS}_{\text{FINGER}}$ denote the approximate JS distance using the VNGE computation from FINGER (either $\widehat{H}$ or $\widetilde{H}$). Using Corollaries 2 and 3, the properly scaled absolute approximation error (SAAE) of JS distance, $\frac{|\mathsf{JS}-\mathsf{JS}_{\text{FINGER}}|}{\sqrt{\ln n}}$, converges to 0 as $n \to \infty$, which proves $|\mathsf{JS} - \mathsf{JS}_{\text{FINGER}}| = o(\sqrt{\ln n})$ and $\frac{\mathsf{JS}_{\text{FINGER}}}{\sqrt{\ln n}}$ is asymptotically a distance metric.

I. Additional experimental results on synthetic random graphs

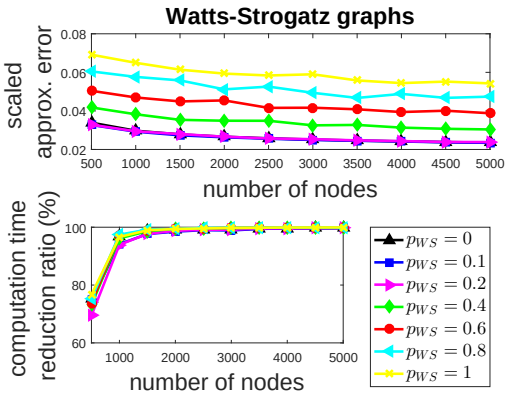
The effect of average degree $\overline{d}$ on Watts-Strogatz graphs. Figure S1 displays the approximation error and computation time reduction ratio (CTRR) of FINGER-$\widehat{H}$ and FINGER-$\widetilde{H}$ under different average degree $\overline{d}$ of WS model, which is defined as $H - \widehat{H}$ and $H - \widetilde{H}$, respectively. It can be observed that when fixing $\overline{d}$, the approximation error decays with the edge rewiring probability for both $\widehat{H}$ and $\widetilde{H}$. In addition, for the same edge rewiring probability, larger $\overline{d}$ yields less approximation error. Using FINGER, both $\widehat{H}$ and $\widetilde{H}$ achieve at least 97% speed-up relative to the computation of H in all cases. The approximate VNGE $\widetilde{H}$ always attains better CTRR than $\widehat{H}$ but at the price of larger approximation error due to the fact that $\widetilde{H} \leq \widehat{H} \leq H$.

Figure S2 displays the scaled approximation error (SAE) and computation time reduction ratio of $\widehat{H}$ via FINGER for WS model under varying number of nodes n and two different settings of the average degree $\overline{d}$. Their behaviors are similar to the case of $\overline{d} = 50$ as displayed in Figure 2 (c).

The effect of graph size n on FINGER-$\widetilde{H}$. In comparison to $\widehat{H}$ via FINGER in Figure 2, Figure S3 displays the SAE and CTRR of $\widetilde{H}$ for the three different random graph models and varying number of nodes n. Consistent with the findings in Section 3, the SAE of $\widetilde{H}$ for ER and WS graphs obeys the $o(\ln n)$ approximation error analysis as established in Corollary 3 since they have balanced eigenspectrum. On the other hand, the SAE of BA graphs grows logarithmically with n due to imbalanced eigenspectrum.

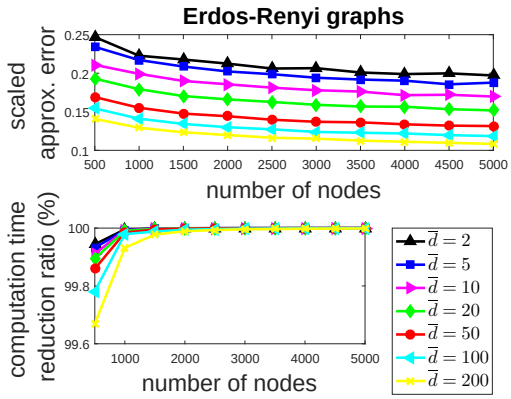


(a) WS model ($\overline{d} = 20$)

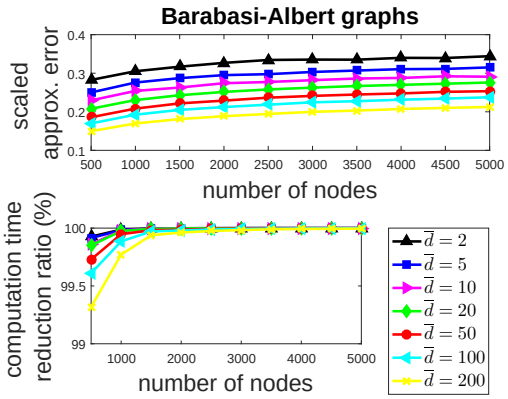


(b) WS model ($\overline{d} = 100$)

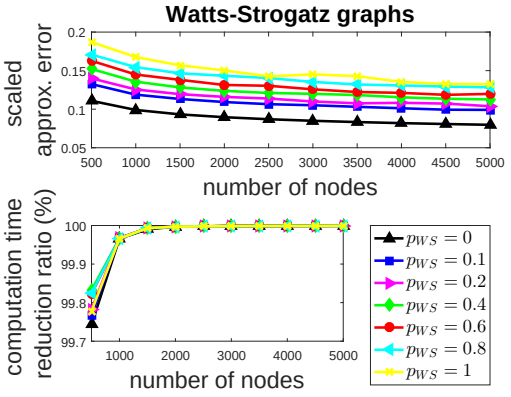
Figure S2. Scaled approximation error (SAE) and computation time reduction ratio (CTRR) of $\widehat{H}$ via FINGER for WS model under varying number of nodes n. Their behaviors are similar to the case of $\overline{d} = 50$ as displayed in Figure 2 (c).



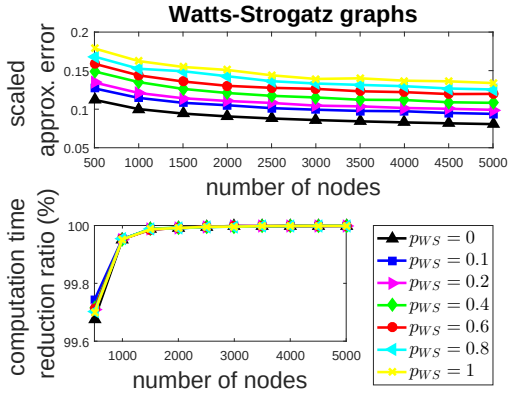
(a) ER model



(b) BA model

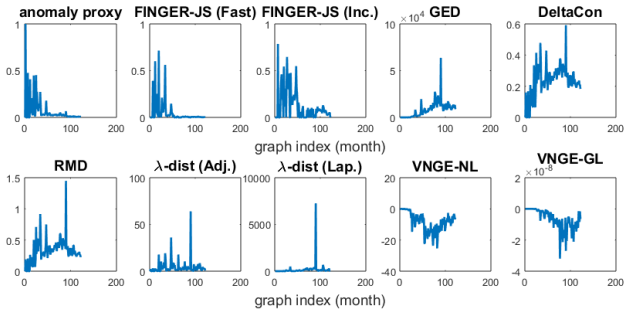


(c) WS model ($\overline{d} = 50$)

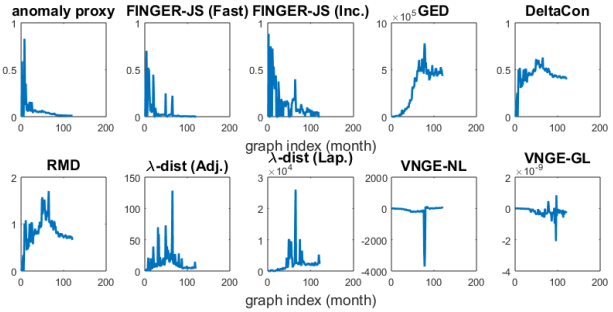


(d) WS model ($\overline{d} = 100$)

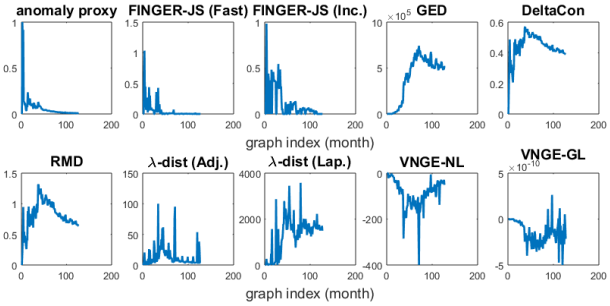
Figure S3. Scaled approximation error (SAE) and computation time reduction ratio (CTRR) of $\widetilde{H}$ via FINGER for different random graph models and varying number of nodes n. The SAE of ER and WS graphs validates the $o(\ln n)$ approximation error analysis in Corollary 3, whereas the SAE of BA graphs grows logarithmically with n due to imbalanced eigenspectrum. The CTRR attains nearly 100% speed-up relative to H for moderate-size graphs ($n \geq 1500$).



(a) Dissimilarity (anomaly) metrics of Wikipedia-sEN



(b) Dissimilarity (anomaly) metrics of Wikipedia-FR



(c) Dissimilarity (anomaly) metrics of Wikipedia-GE

Figure S4. Anomaly detection in consecutive monthly Wikipedia hyperlink networks via different dissimilarity metrics. The corresponding computation time and Pearson correlation coefficient are reported in Table 2. Similar to the observations in Figure 3, FINGER-JSdist (Fast) best aligns with the anomaly proxy in all datasets. FINGER-JSdist (Incremental) has efficient computation time but less consistency (second best PCC among all methods).

Fixing n, larger average degree or more graph regularity leads to less approximation error. Comparing to $\widehat{H}$, the CTRR of $\widetilde{H}$ attains nearly 100% speed-up relative to H for relatively small-size graphs ($n \geq 1500$).

J. Implementation details for VNGE-NL and VNGE-GL

We note that in the Wikipedia application, we omit the edge direction for all methods except VNGE-GL since the resulting performance is almost identical. The implementation of VNGE-GL indeed considers the edge direction. We also note that in these two applications, the Jensen-Shannon distances of VNGE-NL and VNGE-GL are ineffective. Therefore, we use the consecutive difference of their approximate VNGE as the anomaly score, and take the absolute value of the anomaly score for anomaly ranking.

K. Additional results for anomaly detection in evolving Wikipedia hyperlink networks

Additional Wikipedia network plots. The plots of dissimilarity (anomaly) metrics of different methods in Section 4 for consecutive monthly hyperlink networks of Wikipedia-sEN, Wikipedia-FR, and Wikipedia-GE are shown in Figure S4. Their performance in terms of the computation time and Pearson correlation coefficient are reported in Table 2. Similar to the observations in Figure 3, FINGER-JSdist (Fast) best aligns with the anomaly proxy in all datasets. FINGER-JSdist (Incremental) has efficient computation time but less consistency (still attains second best PCC among all methods).

Rank correlation coefficients. In addition to PCC, we further use the Spearman's rank correlation coefficient (SRCC) to evaluate the consistency of each method with the anomaly proxy in this task. The results are summarized in Table S1. Similar to the results using PCC, FINGER-JS (Fast) attains the best SRCC among all the compared methods in the four Wikipedia networks. This result again confirms that JS distance via FINGER indeed learns the similar notion of anomaly as indicated by the anomaly proxy.

L. Addition descriptions for bifurcation detection of cell reprogramming in dynamic genomic networks

Genome architecture is important in studying cell development, but its dynamics and role in determining cell iden-

Table S1. Performance comparison of Spearman's rank correlation coefficient (SRCC) between the anomaly proxy and each method in the Wikipedia application. FINGER attains the best SRCC across all datasets.

Datasets	FINGER -JS (Fast)	FINGER -JS (Inc.)	DeltaCon	RMD	λ dist. (Adj.)	λ dist. (Lap.)	GED	VNGE -NL	VNGE -GL
Wiki (sEN)	**0.5055**	0.3849	0.4518	0.4518	0.4208	0.0402	-0.1355	-0.0542	0.2231
Wiki (EN)	**0.7973**	0.5039	-0.4620	-0.4620	-0.3014	-0.5981	-0.7759	-0.1823	0.4840
Wiki (FR)	**0.7026**	0.4563	0.2652	0.2652	0.4297	-0.4355	-0.6125	-0.4792	0.3938
Wiki (GE)	**0.6591**	0.4930	0.3167	0.3167	0.3707	-0.4343	-0.5695	-0.0156	0.2606

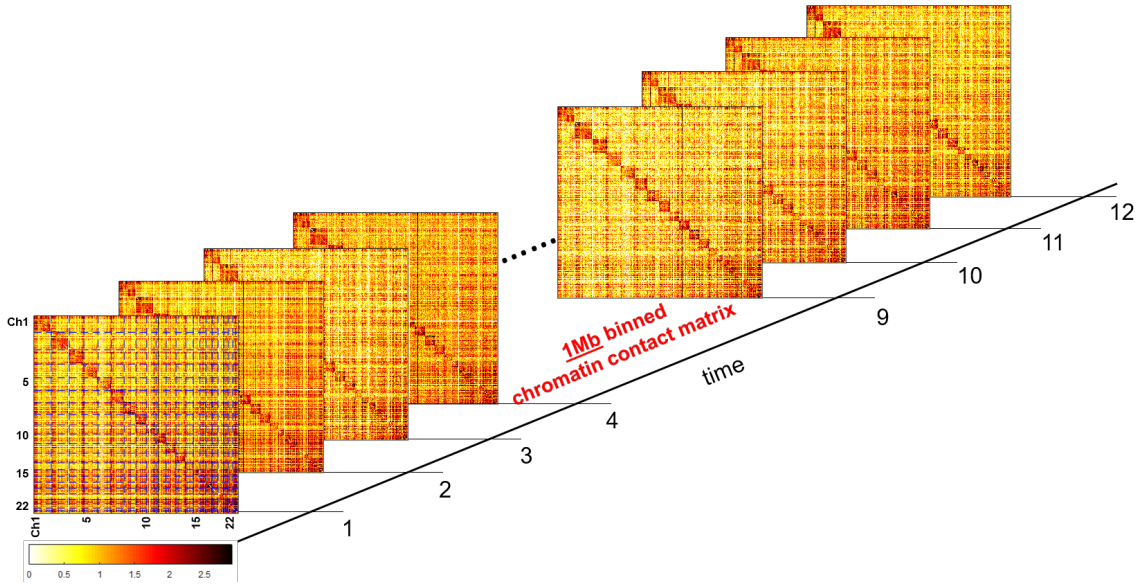


Figure S5. Chromatin contact matrix from Hi-C over a time course of 12 samples, which correspond to -48 hour (hr), 0 hr, 8 hr, , 80 hr over 6 days.

tity are not well understood. Myogenic differentiation 1 (MYOD1) is a master transcription factor that directly converts human fibroblasts to myogenic cells as studied in (Weintraub et al., 1989; Weintraub, 1993). Very recently, Liu et al. (Liu et al., 2018a) studied the chromatin contact map (genome-wide structure) through chromosome conformation capture (Hi-C) during the conversion of human fibroblasts to myogenic cells. To understand cell reprogramming, one major question is detecting when the phase transition occurs for cell identity conversion. Liu et al. conducted experiments and constructed a 1Mb binned chromatin contact matrix (namely, Hi-C matrix) of dimension 2894 over a 6-day time course, leading to 12 sampled measurements. It was found that there exists a bifurcation point at the 6th sample (the measurement at 32 hour), suggesting that the cell reprogramming can be interpreted as a genome-wide dynamic system (Del Vecchio et al., 2017) (i.e., a graph sequence) as displayed in Figure S5, where the bifurcation occurs when a small structure change made to the cellular system causes a significant system-wide change for genome.

Liu et al. further used complex graph analysis techniques involving the temporal difference score (TDS) and multiple graph centrality features (Chen et al., 2016) to construct a representative statistic for expressing the states of the studied dynamic genomic contact network as displayed in Figure 4, which is used in this paper as the ground-truth statistic for comparing the performance of detecting bifurcation point using different dissimilarity and distance metrics. In particular, given the TDS of a graph dissimilarity method over measurements, a bifurcation point is defined as the saddle point of the TDS curve excluding the first and last measurements (i.e., $t = 1$ and $t = T$). The detected bifurcation point(s) of each method is displayed in Figure 4.

Table S2. 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	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%

M. Additional results using VEO as a baseline

As the VEO score only applies to unweighted undirected graphs, we omit the edge weights in the bifurcation dataset and find that VEO incorrectly detects graph index 8 as a bifurcation instance. In addition, for the synthesized anomaly detection task, VEO only attains $\{46, 41, 36, 14\}\%$ detection rate when the DoS attack fraction $X = \{10, 5, 3, 1\}\%$, respectively, as given in Table S2.

N. Additional results using degree distribution as dissimilarity metric

For the synthesized anomalous event detection task, in addition to the dissimilarity metrics in Table 3, we also compare the performance of some distance metrics defined on degree distributions – the cosine distance, the Bhattacharyya distance and the Hellinger distance. We exclude the Kullback-Leibler divergence as the degree distributions of two graphs usually do not have a common support. On the synthesized dataset, Table S2 shows that their performance is not competitive to FINGER and other dissimilarity metrics.