

Fabian Bosshard, Fanny Rorri, Laura Jimena Tagle Muñoz
 {fabian.bosshard, fanny.rorri, jimenatagle}@usi.ch

Abstract

Contents

List of Acronyms

1 Introduction

2 Background

2.1	normalized adjacency and multi-hop propagation	1
	distance layers and layer degrees	1
2.2	graph Laplacian	2
2.3	Cheeger inequality	2
2.4	effective resistance	2
	Interpretation	2
	Connection to random walks	2

3 Theoretical analysis

3.1	Dynamic rewiring paradigm	2
3.2	Jacobian sensitivity	3
3.3	Preservation of locality	3

4 Related works

5 Methodology

6 Implementation

6.1	Datasets	4
6.2	Hyperparameters	4
6.3	Experimental setup	4
6.4	Computational requirements	4

7 Results

8 Discussion and conclusion

List of Acronyms

GCN	Graph Convolutional Network
LASER	Locality-Aware Sequential Rewiring

1 Introduction

Locality-Aware Sequential Rewiring (LASER)

2 Background

Let $G = (V, E)$ be an undirected graph with the adjacency matrix $\underline{A} \in \mathbb{R}^{n \times n}$

$$\underline{A}_{uv} = \begin{cases} 1 & (u, v) \in E \\ 0 & (u, v) \notin E \end{cases} \quad (1)$$

The diagonal degree matrix $\underline{D} \in \mathbb{R}^{n \times n}$ is defined by

$$\underline{D}_{uv} = \begin{cases} d_u & u = v \\ 0 & u \neq v \end{cases} \quad (2)$$

i.e. \underline{D} simply places all node degrees on the diagonal.

2.1 normalized adjacency and multi-hop propagation

Definition 1. The symmetrically normalized adjacency matrix is

$$\hat{\underline{A}} = \underline{D}^{-1/2} \underline{A} \underline{D}^{-1/2} \quad (3)$$

or, entrywise,

$$\hat{\underline{A}}_{uv} = \begin{cases} \frac{1}{\sqrt{d_u d_v}} & (u, v) \in E \\ 0 & (u, v) \notin E \end{cases} \quad \square$$

Proposition 1. multi-hop propagation. The entry $(\hat{\underline{A}}^k)_{vu}$ can be computed explicitly as follows:

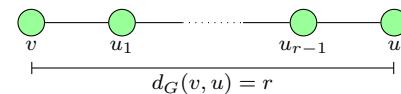
$$(\hat{\underline{A}}^k)_{vu} = \sum_{\pi} \prod_{(x,y) \in E_{\pi}} \frac{1}{\sqrt{d_x d_y}} \quad (4)$$

the sum is over all walks $\pi = (v, \dots, u)$ of length k from v to u and the product is over the edges $E_{\pi} = \{(v, u_1), \dots, (u_{k-1}, u)\}$ on the walk. \square

Corollary 2. Let $v, u \in V$ with $r = d_G(v, u)$, where $d_G(\cdot, \cdot)$ denotes the shortest-path distance. Assume there is exactly one path

$$(v, u_1, \dots, u_{r-1}, u)$$

of length r between v and u :



Then

$$\begin{aligned} (\hat{\underline{A}}^r)_{vu} &= \frac{1}{\sqrt{d_v d_{u_1}}} \cdot \prod_{i=1}^{r-2} \frac{1}{\sqrt{d_{u_i} d_{u_{i+1}}}} \cdot \frac{1}{\sqrt{d_{u_{r-1}} d_u}} \\ &= \frac{1}{\sqrt{d_v d_u}} \prod_{i=1}^{r-1} \frac{1}{d_{u_i}} \end{aligned} \quad (5)$$

distance layers and layer degrees

Definition 2. For $\ell \in \mathbb{N}_0$, we define the distance- $(\ell + 1)$ adjacency matrix $\underline{A}_{\ell} \in \mathbb{R}^{n \times n}$ by

$$(\underline{A}_{\ell})_{uv} = \begin{cases} 1 & d_G(u, v) = \ell + 1 \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

where $d_G(u, v)$ is the shortest-path distance. The corresponding **layer degree** of a node v at distance level ℓ is

$$d_{v,\ell} = \sum_{u \in V} (\underline{\mathbf{A}}_\ell)_{vu}, \quad (7)$$

i.e. the number of nodes at graph distance $\ell + 1$ from v . Let $\underline{\mathbf{D}}_\ell$ be the diagonal matrix with $(\underline{\mathbf{D}}_\ell)_{vv} = d_{v,\ell}$. The **normalized distance- $(\ell + 1)$ adjacency** is

$$\hat{\underline{\mathbf{A}}}_\ell = \underline{\mathbf{D}}_\ell^{-1/2} \underline{\mathbf{A}}_\ell \underline{\mathbf{D}}_\ell^{-1/2} \quad (8)$$

so that

$$(\hat{\underline{\mathbf{A}}}_\ell)_{uv} = \begin{cases} \frac{1}{\sqrt{d_{u,\ell} d_{v,\ell}}} & d_G(u, v) = \ell + 1 \\ 0 & \text{otherwise} \end{cases}$$

Finally, we denote by

$$d_{\min} = \min_{v \in V} d_v \quad (9)$$

the **minimum node degree** in the graph.

2.2 graph Laplacian

Definition 3. The **combinatorial graph Laplacian** is

$$\underline{\mathbf{L}} = \underline{\mathbf{D}} - \underline{\mathbf{A}} \quad (10)$$

and the **normalized graph Laplacian** is

$$\hat{\underline{\mathbf{L}}} = \underline{\mathbf{D}}^{-1/2} \underline{\mathbf{L}} \underline{\mathbf{D}}^{-1/2} \stackrel{(10)}{=} \underline{\mathbf{D}}^{-1/2} (\underline{\mathbf{D}} - \underline{\mathbf{A}}) \underline{\mathbf{D}}^{-1/2} \stackrel{(3)}{=} \underline{\mathbf{I}}_n - \hat{\underline{\mathbf{A}}} \quad (11)$$

It is symmetric and positive semidefinite, and its eigenvalues satisfy

$$0 = \lambda_0 \leq \lambda_1 \leq \dots \leq \lambda_{n-1}$$

λ_1 is called the **spectral gap**. The number of zero eigenvalues (i.e., the multiplicity of the 0 eigenvalue) equals the number of connected components of the graph. \square

To understand Definition 3, consider a function $f: V \rightarrow \mathbb{R}$. Denote by $\vec{f} \in \mathbb{R}^n$ the vector whose v -th entry is $f(v)$. Then

$$(\hat{\underline{\mathbf{L}}} \vec{f})_v = f(v) - \frac{1}{\sqrt{d_v}} \sum_{(u,v) \in E} \frac{f(u)}{\sqrt{d_u}} \quad (12)$$

i.e., $(\hat{\underline{\mathbf{L}}} \vec{f})_v$ is the value at v minus a degree-normalized average of the neighbors. This is why the Laplacian is often viewed as a **discrete second derivative** on the graph: it measures how much f at v deviates from its neighborhood. Another important identity is the quadratic form

$$\vec{f}^\top \underline{\mathbf{L}} \vec{f} = \frac{1}{2} \sum_{(u,v) \in E} (f(u) - f(v))^2 \quad (13)$$

which shows that $\underline{\mathbf{L}}$ (and hence also $\hat{\underline{\mathbf{L}}}$) is positive semidefinite, since the right-hand side is always nonnegative. Moreover, (13) is small exactly when f varies slowly across edges, so the Laplacian encodes the **smoothness** of functions on the graph.

2.3 Cheeger inequality

The **Cheeger inequality** relates the spectral gap λ_1 to the **Cheeger constant** $h(G)$, which measures how difficult it is to separate the graph into two large pieces. It states, in particular, that

$$\frac{1}{2} h(G)^2 \leq \lambda_1 \leq 2h(G),$$

so a larger spectral gap implies that the graph is more “well-connected”.

2.4 effective resistance

Definition 4. **effective resistance**. View each edge $(u, v) \in E$ as an electrical resistor of resistance 1Ω . The resulting network has a well-defined resistance between any two nodes.

For two nodes $s, t \in V$, the **effective resistance** $R(s, t)$ is defined as the voltage difference needed to send one unit of electrical current from s to t . It can be computed as

$$R(s, t) = (\vec{e}_s - \vec{e}_t)^\top \underline{\mathbf{L}}^\dagger (\vec{e}_s - \vec{e}_t) \quad (14)$$

where $\underline{\mathbf{L}}^\dagger$ is the Moore–Penrose pseudoinverse of $\underline{\mathbf{L}}$ and \vec{e}_v is the standard basis vector of vertex v . \square

Interpretation If the graph offers many short, parallel paths between s and t , then current can flow easily, so $R(s, t)$ is small. If there are few or long paths, the current is “bottlenecked” and $R(s, t)$ is large. Thus, effective resistance measures how “well-connected” two nodes are inside the global geometry of the graph.

Connection to random walks A **random walk** on G is the Markov chain that, from a node v , moves to a uniformly random neighbor of v . Its transition matrix is

$$\underline{\mathbf{P}} = \underline{\mathbf{D}}^{-1} \underline{\mathbf{A}} \quad (15)$$

so $\underline{\mathbf{P}}_{vu} = 1/d_v$ if $(v, u) \in E$.

For two nodes u, v , the **commute time** $\text{CT}(u, v)$ is the expected number of steps for the random walk to start at u , reach v , and return to u again. It can be related to the effective resistance via

$$\text{CT}(u, v) = 2|E|R(u, v) \quad (16)$$

giving a geometric interpretation of how “far apart” two nodes are in terms of random-walk behavior, i.e. two nodes have small commute time exactly when they have small effective resistance.

3 Theoretical analysis

3.1 Dynamic rewiring paradigm

The main contribution of [1] is the following idea: Instead of applying a single rewired graph $\mathcal{R}(G)$, consider a sequence of rewired graphs

$$G = G_0 \xrightarrow{\mathcal{R}} \dots \xrightarrow{\mathcal{R}} G_{\ell-1} \xrightarrow{\mathcal{R}} G_\ell \xrightarrow{\mathcal{R}} \dots \xrightarrow{\mathcal{R}} G_L \quad (17)$$

where G_ℓ is called the ℓ -snapshot, since we think of G_ℓ as the input graph evolved along a dynamical process for ℓ iterations.

The variant to derive the theoretical results in [1] is a Graph Convolutional Network (GCN)-LASER architecture

$$\vec{x}_v^{(t)} = \text{ReLU} \left(\sum_{u \in \mathcal{N}_1(v) \cup \{v\}} \frac{1}{\sqrt{d_v d_u}} \cdot \underline{\mathbf{W}}_0^{(t)} \vec{x}_u^{(t-1)} + \sum_{\ell=1}^L \sum_{u \in \mathcal{N}_{\ell+1}^{\rho}(v)} \frac{1}{\sqrt{d_{v,\ell} d_{u,\ell}}} \cdot \underline{\mathbf{W}}_\ell^{(t)} \vec{x}_u^{(t-1)} \right) \quad (18)$$

where $\underline{\mathbf{W}}_\ell^{(t)}$, for $\ell = 0, \dots, L$, are learnable weight matrices and

$$\mathcal{N}_{\ell+1}^{\rho}(v) \subset \mathcal{N}_{\ell+1}(v) := \{u \in V \mid d_{G_0}(u, v) = \ell + 1\} \quad (19)$$

is a ρ -fraction of nodes at distance exactly $\ell + 1$ from v with the lowest connectivity score μ in the original graph $G = G_0$. So, for the ℓ -th snapshot G_ℓ , LASER considers $\rho \cdot d_{v,\ell}$ artificial edges from v to nodes at distance $d_{G_0}(u, v) = \ell + 1$.

3.2 Jacobian sensitivity

Let $\underline{\mathbf{J}}^{(t)}(v, u) := \partial \bar{\mathbf{x}}_v^{(t)} / \partial \bar{\mathbf{x}}_u^{(0)}$ be the Jacobian of features after t layers of message passing on G using a GCN as described in [2], and similarly let $\tilde{\mathbf{J}}^{(t)}(v, u)$ be the Jacobian after t layers using LASER, as defined in Equation 18.

Proposition 3. Let $v, u \in V$ with $r = d_{G_0}(v, u)$, where G_0 is the original graph before applying LASER. Assume there is exactly one path $\pi \subset G_0$ of length r connecting v and u (see Figure 1).

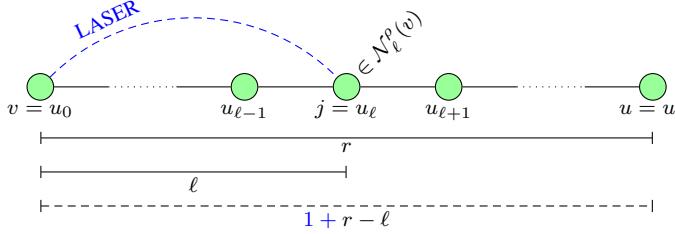


Figure 1. Unique shortest path $\pi \subset G_0$ between v and u of length r with an intermediate node j at distance ℓ from v . The dashed blue line indicates the edge added by LASER.

Assume LASER adds an edge between v and a node $j \in V_\pi$ with $d_{G_0}(v, j) = \ell < r$. Furthermore, assume that all products of weight matrices have unit norm. Then for all $m \leq r$ we have

$$\left\| \mathbb{E} [\tilde{\mathbf{J}}^{(r-\ell+1)}(v, u)] \right\| \geq \frac{(d_{\min})^\ell \sqrt{\frac{d_j}{d_u}}}{\sqrt{d_{v,\ell-1} d_{j,\ell-1}}} \left\| \mathbb{E} [\underline{\mathbf{J}}^{(m)}(v, u)] \right\| \quad (20)$$

where $d_{\min} = \min_{v \in V} d_v$ is the minimum degree in G_0 . \square

Proof In the following, the expectations are with respect to ReLU' which is assumed to have probability of success p for all paths in the computational graph as in [3, 4]. The Jacobian after m layers of GCN is

$$\left\| \mathbb{E} [\underline{\mathbf{J}}^{(m)}(v, u)] \right\| = p \left\| \prod_{k=1}^m \underline{\mathbf{W}}^{(k)} \right\| (\hat{\mathbf{A}}^m)_{vu} = p(\hat{\mathbf{A}}^m)_{vu}$$

where we used the unit-norm assumption on the weight matrices and $\hat{\mathbf{A}}$ denotes the normalized adjacency matrix from Definition 1.

For $m < r$, the expression above vanishes, satisfying (20), so we consider $m = r$ from now on.

With the same assumptions, the Jacobian after $r - \ell + 1$ layers of LASER is

$$\begin{aligned} \left\| \mathbb{E} [\tilde{\mathbf{J}}^{(r-\ell+1)}(v, u)] \right\| &= p \left\| \prod_{k=2}^{r-\ell+1} \underline{\mathbf{W}}_0^{(k)} \underline{\mathbf{W}}_{\ell-1}^{(1)} \right\| (\hat{\mathbf{A}}_{\ell-1})_{vj} (\hat{\mathbf{A}}^{r-\ell})_{ju} \\ &= p(\hat{\mathbf{A}}_{\ell-1})_{vj} (\hat{\mathbf{A}}^{r-\ell})_{ju} \end{aligned}$$

As we can see in Figure 1, we have a unique path

$$(v = u_0, u_1, \dots, u_{r-1}, u_r = u)$$

and we assume $j = u_\ell$. From Corollary 2, for the full path from v to u :

$$(\hat{\mathbf{A}}^r)_{vu} = \frac{1}{\sqrt{d_v d_u}} \prod_{s=1}^{r-1} \frac{1}{d_{u_s}}$$

For the sub-path from $j = u_\ell$ to $u = u_r$ of length $r - \ell$, the same reasoning gives

$$(\hat{\mathbf{A}}^{r-\ell})_{ju} = \frac{1}{\sqrt{d_j d_u}} \prod_{s=\ell+1}^{r-1} \frac{1}{d_{u_s}}.$$

Now we calculate the ratio

$$\begin{aligned} \frac{\left\| \mathbb{E} [\tilde{\mathbf{J}}^{(r-\ell+1)}(v, u)] \right\|}{\left\| \mathbb{E} [\underline{\mathbf{J}}^{(r)}(v, u)] \right\|} &= \frac{(\hat{\mathbf{A}}_{\ell-1})_{vj} (\hat{\mathbf{A}}^{r-\ell})_{ju}}{\frac{1}{\sqrt{d_v d_u}} \prod_{s=1}^{r-1} \frac{1}{d_{u_s}}} \\ &= \frac{(\hat{\mathbf{A}}_{\ell-1})_{vj}}{\frac{1}{\sqrt{d_v d_u}} \prod_{s=1}^{\ell-1} \frac{1}{d_{u_s}}} \cdot \frac{(\hat{\mathbf{A}}^{r-\ell})_{ju}}{\prod_{s=\ell}^{r-1} \frac{1}{d_{u_s}}} \\ &\stackrel{(5)}{=} \frac{(\hat{\mathbf{A}}_{\ell-1})_{vj}}{\frac{1}{\sqrt{d_v d_u}} \prod_{s=1}^{\ell-1} \frac{1}{d_{u_s}}} \cdot \frac{\frac{1}{\sqrt{d_j d_u}} \prod_{s=\ell+1}^{r-1} \frac{1}{d_{u_s}}}{\prod_{s=\ell}^{r-1} \frac{1}{d_{u_s}}} \\ &= \frac{(\hat{\mathbf{A}}_{\ell-1})_{vj}}{\frac{1}{\sqrt{d_v d_u}} \prod_{s=1}^{\ell-1} \frac{1}{d_{u_s}}} \cdot \frac{\frac{1}{\sqrt{d_j d_u}} \prod_{s=\ell+1}^{r-1} \frac{1}{d_{u_s}}}{\frac{1}{d_{u_\ell}}} \\ &= \frac{(\hat{\mathbf{A}}_{\ell-1})_{vj}}{\frac{1}{\sqrt{d_v d_u}} \prod_{s=1}^{\ell-1} \frac{1}{d_{u_s}}} \cdot \frac{d_{u_\ell}}{\sqrt{d_j d_u}} \\ &= \frac{(\hat{\mathbf{A}}_{\ell-1})_{vj}}{\frac{1}{\sqrt{d_v d_u}} \prod_{s=1}^{\ell-1} \frac{1}{d_{u_s}}} \cdot \sqrt{\frac{d_j}{d_u}} \end{aligned}$$

Using

$$(\hat{\mathbf{A}}_{\ell-1})_{vj} \stackrel{(8)}{=} \frac{1}{\sqrt{d_{v,\ell-1} d_{j,\ell-1}}}$$

and

$$\frac{1}{\sqrt{d_v d_u}} \prod_{s=1}^{\ell-1} \frac{1}{d_{u_s}} \leq \frac{1}{(d_{\min})^\ell}$$

we obtain

$$\frac{(\hat{\mathbf{A}}_{\ell-1})_{vj}}{\frac{1}{\sqrt{d_v d_u}} \prod_{s=1}^{\ell-1} \frac{1}{d_{u_s}}} \geq \frac{(d_{\min})^\ell}{\sqrt{d_{v,\ell-1} d_{j,\ell-1}}}$$

which, combined with the previous equation, gives

$$\frac{\left\| \mathbb{E} [\tilde{\mathbf{J}}^{(r-\ell+1)}(v, u)] \right\|}{\left\| \mathbb{E} [\underline{\mathbf{J}}^{(r)}(v, u)] \right\|} \geq \frac{(d_{\min})^\ell}{\sqrt{d_{v,\ell-1} d_{j,\ell-1}}} \cdot \sqrt{\frac{d_j}{d_u}}$$

concluding the proof. \square

Remark Compared to the bound stated in [1], our derivation yields an additional multiplicative factor $\sqrt{d_j/d_u}$ in (20), which arises from the explicit evaluation of the subpath contribution $(\hat{\mathbf{A}}^{r-\ell})_{ju}$. This factor is generally non-unit unless further degree assumptions are imposed, but does not affect the qualitative conclusion on the benefit of LASER for alleviating over-squashing. \blacktriangleleft

3.3 Preservation of locality

Proposition 4. Following [4], we compare how much a rewiring perturbs the shortest-path distances by means of the Frobenius norm

$$\|\mathcal{D}_G - \mathcal{D}_{G'}\|_F, \quad (\mathcal{D}_G)_{uv} = d_G(u, v).$$

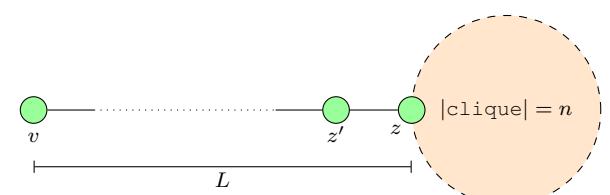


Figure 2. Lollipop graph with a chain of length L attached to a clique of size n . \blacktriangleleft

Setup. Let G be a lollipop graph consisting of a path of length L attached to a clique of size n . Let v be the endpoint of the path, z the junction between path and clique, and let the other $n - 1$ clique nodes be “interior” nodes.

Spectral rewiring: lower bound Consider a spectral rewiring \mathcal{R} that adds one edge between the pair with largest effective resistance, which in this graph is (v, u) for some interior clique node u . Before rewiring,

$$d_G(v, u) = L + 1, \quad d_G(v, z) = L.$$

After adding (v, u) we have

$$d_{\mathcal{R}(G)}(v, u) = 1, \quad d_{\mathcal{R}(G)}(v, w) \leq 2 \text{ for all interior clique } w,$$

Hence, for each of the $n - 1$ interior clique nodes

$$|d_G(v, w) - d_{\mathcal{R}(G)}(v, w)| \geq (L + 1) - 2 = L - 1,$$

and for z

$$|d_G(v, z) - d_{\mathcal{R}(G)}(v, z)| \geq L - 2.$$

Restricting the Frobenius sum to these n pairs yields

$$\|\mathcal{D}_G - \mathcal{D}_{\mathcal{R}(G)}\|_F^2 \geq (n - 1)(L - 1)^2 + (L - 2)^2, \quad (21)$$

and therefore

$$\|\mathcal{D}_G - \mathcal{D}_{\mathcal{R}(G)}\|_F \geq \sqrt{(n - 1)(L - 1)^2 + (L - 2)^2}. \quad (22)$$

LASER: upper bound Now consider LASER with a single snapshot restricted to distance-2 edges. On the path, every node except the penultimate one z' has at most two distance-2 neighbours; z' has the whole clique at distance 2. If we choose ρ such that

$$\rho = \frac{k}{2L}, \quad (23)$$

then the **expected** number of new edges incident to path nodes is k . Assuming that at most k such edges are added, we obtain uniform bounds on the distance changes.

Path–path pairs. Each new edge connects two nodes at distance 2, so using it on a shortest path can shorten that path by at most 1. Any path between two nodes x, y on the path can traverse at most k new edges, hence

$$|d_G(x, y) - d_{\text{LASER}}(x, y)| \leq k \quad \text{for all path nodes } x, y.$$

There are at most L^2 ordered path–path pairs, so their contribution to the Frobenius norm satisfies

$$\sum_{x, y \text{ on path}} (d_G(x, y) - d_{\text{LASER}}(x, y))^2 \leq L^2 k^2. \quad (24)$$

Path–clique pairs. In the worst case, one of the k new edges connects z' directly to some clique node. Then any shortest path between a path node x and a clique node c can use at most k shortcuts along the path and one shortcut to enter the clique. Thus

$$|d_G(x, c) - d_{\text{LASER}}(x, c)| \leq k + 1 \quad \text{for all } x \text{ on the path and } c \text{ in the clique.}$$

There are nL such ordered pairs, giving

$$\sum_{\substack{x \text{ on path} \\ c \text{ in clique}}} (d_G(x, c) - d_{\text{LASER}}(x, c))^2 \leq nL(k + 1)^2. \quad (25)$$

Distances between clique nodes are not increased by adding local edges from the path and are at most mildly decreased; their contribution can be bounded and is dominated by the two terms above. Hence we obtain the global upper bound

$$\|\mathcal{D}_G - \mathcal{D}_{\text{LASER}}\|_F^2 \leq L^2 k^2 + nL(k + 1)^2, \quad (26)$$

and therefore

$$\|\mathcal{D}_G - \mathcal{D}_{\text{LASER}}\|_F \leq \sqrt{L^2 k^2 + nL(k + 1)^2}. \quad (27)$$

Comparison and corrected parameter constraints A sufficient condition for LASER to be more locality-preserving than the spectral rewiring is

$$\|\mathcal{D}_G - \mathcal{D}_{\text{LASER}}\|_F \leq \|\mathcal{D}_G - \mathcal{D}_{\mathcal{R}(G)}\|_F.$$

Using (22) and (27), this is guaranteed if

$$L^2 k^2 + nL(k + 1)^2 \leq (n - 1)(L - 1)^2 + (L - 2)^2. \quad (28)$$

Rearranging (28) and solving for n gives

$$d_{\mathcal{R}(G)}(v, z) \leq 2, \quad n \geq \frac{k^2 L^2 + 2L - 3}{L^2 - Lk^2 - 2Lk - 3L + 1}, \quad (29)$$

provided the denominator is positive. A sufficient condition for the denominator to be positive is

$$k^2 < \frac{L}{4}, \quad L \geq 8, \quad (30)$$

which we verified analytically for $k \in \mathbb{N}$. Using the relation $\rho = k/(2L)$ from (23), the constraint (30) is equivalent to

$$\rho^2 < \frac{1}{16L} \iff \rho < \frac{1}{4\sqrt{L}}. \quad (31)$$

In summary, on lollipop graphs with $L \geq 8$, sufficiently small sampling rate $\rho < 1/(4\sqrt{L})$ and n satisfying (29), we obtain

$$\|\mathcal{D}_G - \mathcal{D}_{\text{LASER}}\|_F \leq \|\mathcal{D}_G - \mathcal{D}_{\mathcal{R}(G)}\|_F, \quad (32)$$

i.e. LASER perturbs the shortest-path distances less than the spectral effective-resistance-based rewiring.

Compared to the original derivation, we note that: (i) the path–clique bound should be interpreted as a worst-case inequality $\leq k + 1$ rather than an equality; (ii) the algebraic manipulation leading to the bound on n in (29) fixes a sign error in the denominator; and (iii) the correct constraint on ρ is (31), which differs by a factor of 2 from the value stated in the paper.

4 Related works

5 Methodology

6 Implementation

6.1 Datasets

6.2 Hyperparameters

6.3 Experimental setup

6.4 Computational requirements

7 Results

8 Discussion and conclusion

References

- [1] Federico Barbero, Ameya Velingker, Amin Saberi, Michael Bronstein, and Francesco Di Giovanni. Locality-aware graph-rewiring in gnns. 2024. URL <https://arxiv.org/abs/2310.01668>.
- [2] Thomas N. Kipf and Max Welling. Semi-Supervised Classification with Graph Convolutional Networks. In *International Conference on Learning Representations*, 2017.
- [3] Keyulu Xu, Chengtao Li, Yonglong Tian, Tomohiro Sonobe, Ken-ichi Kawarabayashi, and Stefanie Jegelka. Representation learning on graphs with jumping knowledge networks. In *International Conference on Machine Learning*, pages 5453–5462. PMLR, 2018.

-
- [4] Francesco Di Giovanni, Lorenzo Giusti, Federico Barbero, Giulia Luise, Pietro Lio, and Michael Bronstein. On over-squashing in message passing neural networks: The impact of width, depth, and topology. *arXiv preprint arXiv:2302.02941*, 2023.