Locality-Aware Graph Rewiring in GNNs

GDL 2023
Group id: ——
Project id: ——

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

Abstract

1 Introduction

2 Background

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

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

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

$$\underline{\boldsymbol{D}}_{uv} = \begin{cases} d_u & u = v \\ 0 & u \neq v \end{cases} \tag{2}$$

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

2.1 normalized adjacency and multi-hop propagation

Definition 1. The symmetrically normalized adjacency matrix is

$$\boxed{\hat{\underline{\boldsymbol{A}}} = \underline{\boldsymbol{D}}^{-1/2}\underline{\boldsymbol{A}}\underline{\boldsymbol{D}}^{-1/2}} \tag{3}$$

or, entrywise,

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

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

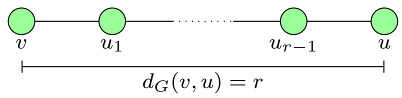
$$(\hat{\underline{\boldsymbol{A}}}^k)_{vu} = \sum_{\pi} \prod_{(x,y)\in\pi} \frac{1}{\sqrt{d_x d_y}} \tag{4}$$

where the sum is over all walks $\pi = (v, \dots, u)$ of length k from v to u. $\triangleleft$

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

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

$\triangleleft$

distance layers and layer degrees

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

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

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

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

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

$$\boxed{\hat{\underline{\boldsymbol{A}}}_\ell = \underline{\boldsymbol{D}}_\ell^{-1/2}\underline{\boldsymbol{A}}_\ell\underline{\boldsymbol{D}}_\ell^{-1/2}} \tag{8}$$

so that

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

Finally, we denote by

$$d_{\min} = \min_{v\in V} d_v \tag{9}$$

the minimum node degree in the graph.

2.2 graph Laplacian

Definition 3. The combinatorial graph Laplacian is

$$\underline{\boldsymbol{L}} = \underline{\boldsymbol{D}} - \underline{\boldsymbol{A}} \tag{10}$$

and the normalized graph Laplacian is

$$\boxed{\hat{\underline{\boldsymbol{L}}} = \underline{\boldsymbol{D}}^{-1/2}\underline{\boldsymbol{L}}\underline{\boldsymbol{D}}^{-1/2} \overset{(10)}{=} \underline{\boldsymbol{D}}^{-1/2}(\underline{\boldsymbol{D}} - \underline{\boldsymbol{A}})\underline{\boldsymbol{D}}^{-1/2} \overset{(3)}{=} \underline{\boldsymbol{I}}_n - \hat{\underline{\boldsymbol{A}}}} \tag{11}$$

It is symmetric and positive semidefinite, and its eigenvalues satisfy

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

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.

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

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

i.e., $(\hat{\underline{\boldsymbol{L}}}\vec{\boldsymbol{f}})_v$ 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

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

which shows that $\underline{\boldsymbol{L}}$ (and hence also $\hat{\underline{\boldsymbol{L}}}$) 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

$$\tfrac{1}{2}h(G)^2 \le \lambda_1 \le 2h(G),$$

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 $1\,\Omega$. 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

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

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

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

$$\underline{\boldsymbol{P}} = \underline{\boldsymbol{D}}^{-1}\underline{\boldsymbol{A}} \tag{15}$$

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

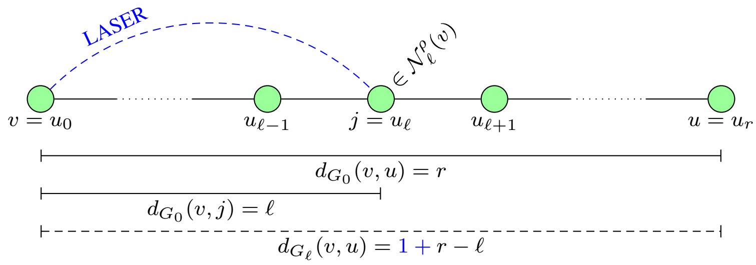
For two nodes u,v, the commute time $\mathtt{CT}(u,v)$ is the expected number of steps for the random walk to start at u, reach v, and return to u again. It can be related to the effective resistance via

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

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 Jacobian sensitivity



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:

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

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

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

Now we plug this into our expression:

$$\begin{aligned}
\frac{(\hat{\underline{\boldsymbol{A}}}_{\ell-1})_{vj}(\hat{\underline{\boldsymbol{A}}}^{r-\ell})_{ju}}{\frac{1}{\sqrt{d_vd_u}}\prod_{s=1}^{r-1}\frac{1}{d_{u_s}}} &= \frac{(\hat{\underline{\boldsymbol{A}}}_{\ell-1})_{vj}}{\frac{1}{\sqrt{d_vd_u}}\prod_{s=1}^{\ell-1}\frac{1}{d_{u_s}}}\cdot\frac{(\hat{\underline{\boldsymbol{A}}}^{r-\ell})_{ju}}{\prod_{s=\ell}^{r-1}\frac{1}{d_{u_s}}}\\
&\overset{(5)}{=} \frac{(\hat{\underline{\boldsymbol{A}}}_{\ell-1})_{vj}}{\frac{1}{\sqrt{d_vd_u}}\prod_{s=1}^{\ell-1}\frac{1}{d_{u_s}}}\cdot\frac{\frac{1}{\sqrt{d_jd_u}}\prod_{s=\ell+1}^{r-1}\frac{1}{d_{u_s}}}{\prod_{s=\ell}^{r-1}\frac{1}{d_{u_s}}}\\
&= \frac{(\hat{\underline{\boldsymbol{A}}}_{\ell-1})_{vj}}{\frac{1}{\sqrt{d_vd_u}}\prod_{s=1}^{\ell-1}\frac{1}{d_{u_s}}}\cdot\frac{1}{\sqrt{d_jd_u}}\cdot\frac{\prod_{s=\ell+1}^{r-1}\frac{1}{d_{u_s}}}{\frac{1}{d_{u_\ell}}\prod_{s=\ell+1}^{r-1}\frac{1}{d_{u_s}}}\\
&= \frac{(\hat{\underline{\boldsymbol{A}}}_{\ell-1})_{vj}}{\frac{1}{\sqrt{d_vd_u}}\prod_{s=1}^{\ell-1}\frac{1}{d_{u_s}}}\cdot\frac{1}{\sqrt{d_jd_u}}\cdot d_{u_\ell}\\
&= \frac{(\hat{\underline{\boldsymbol{A}}}_{\ell-1})_{vj}}{\frac{1}{\sqrt{d_vd_u}}\prod_{s=1}^{\ell-1}\frac{1}{d_{u_s}}}\cdot\sqrt{\frac{d_j}{d_u}}
\end{aligned}$$

So

$$\frac{(\hat{\underline{\boldsymbol{A}}}_{\ell-1})_{vj}(\hat{\underline{\boldsymbol{A}}}^{r-\ell})_{ju}}{\frac{1}{\sqrt{d_vd_u}}\prod_{s=1}^{r-1}\frac{1}{d_{u_s}}} = \frac{(\hat{\underline{\boldsymbol{A}}}_{\ell-1})_{vj}}{\frac{1}{\sqrt{d_vd_u}}\prod_{s=1}^{\ell-1}\frac{1}{d_{u_s}}}\cdot\sqrt{\frac{d_j}{d_u}} \tag{17}$$

Using

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

and

$$\frac{1}{\sqrt{d_vd_u}}\prod_{s=1}^{\ell-1}\frac{1}{d_{u_s}} \le \frac{1}{d_{\min}^{\ell}}$$

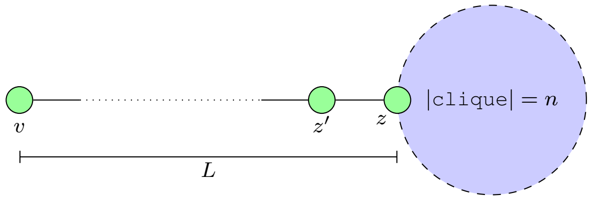
we obtain the bound

$$\frac{(\hat{\underline{\boldsymbol{A}}}_{\ell-1})_{vj}}{\frac{1}{\sqrt{d_vd_u}}\prod_{s=1}^{\ell-1}\frac{1}{d_{u_s}}} \ge \frac{(d_{\min})^{\ell}}{\sqrt{d_{v,\ell-1}d_{j,\ell-1}}} \tag{18}$$

Combining (17) and (18) yields

$$\boxed{\frac{(\hat{\underline{\boldsymbol{A}}}_{\ell-1})_{vj}(\hat{\underline{\boldsymbol{A}}}^{r-\ell})_{ju}}{\frac{1}{\sqrt{d_vd_u}}\prod_{s=1}^{r-1}\frac{1}{d_{u_s}}} \ge \frac{(d_{\min})^{\ell}}{\sqrt{d_{v,\ell-1}d_{j,\ell-1}}}\cdot\sqrt{\frac{d_j}{d_u}}} \tag{19}$$

3.2 Locality awareness



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