

Name1 Surname1, Name2 Surname2, Name3 Surname3

{name1.surname1, name2.surname2, name3.surname3}@usi.ch

## Abstract

Concise and self-contained description of your project, motivation and main findings.

## GENERAL NOTES

The report should be written as an article intended to present the findings of your work. Your aim should be to be clear and objective, substantiating your claims with references or empirical/theoretical evidence. We are well aware of the fact that carrying out machine learning experiments might be difficult and that often the final performance might be disappointing. For this reason, you will not be evaluated solely on quantitative aspect of your work, but mainly on the quality of your analysis and report. The length of the report should be between 4 and 8 pages (without considering references).

## 1 Introduction

Here you should clarify the context of your project and the problem you are dealing with. You should also make a brief summary of the main results and contributions (i.e., if you tried to replicate the results of an existing paper you should say if you were successful or not). The introduction should help the reader to follow along for the rest of the paper.

## 2 basic notions

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,

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

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

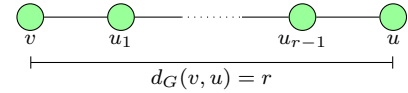
$$(\hat{\underline{A}}^k)_{vu} = \sum_{\pi} \prod_{(x,y) \in \pi} \frac{1}{\sqrt{d_x d_y}} \quad (4)$$

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

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

$\triangleleft$

**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  $\ell$  is

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

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

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

so that

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

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

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

and the **normalized graph Laplacian** is

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

It is symmetric and positive semidefinite, and its eigenvalues satisfy

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

$\lambda_1$  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.  $\mathcal{L}_0$

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

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

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

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

which shows that  $\underline{L}$  (and hence also  $\hat{\underline{L}}$ ) 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  $\lambda_1$  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  $1 \Omega$ . 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

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

where  $\underline{L}^\dagger$  is the Moore–Penrose pseudoinverse of  $\underline{L}$  and  $\vec{e}_v$  is the standard basis vector of vertex  $v$ .  $\mathcal{L}_0$

**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

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

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

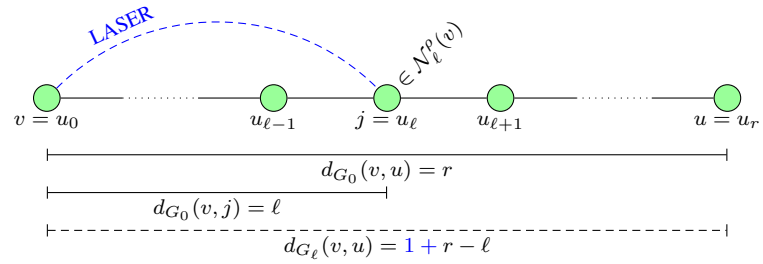
For two nodes  $u, v$ , the **commute time**  $\text{CT}(u, v)$  is the expected number of steps for the random walk to start at  $u$ , reach  $v$ , and return to  $u$  again. It can be related to the effective resistance via

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

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 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  $u$ :

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

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

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

Now we plug this into our expression:

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

So

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

Using

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

and

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

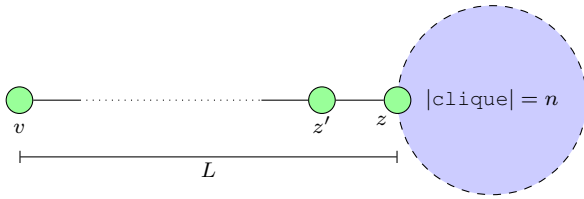
we obtain the bound

$$\frac{(\hat{\mathbf{A}}_{\ell-1})_{vj}}{\frac{1}{\sqrt{d_v d_u}} \prod_{s=1}^{\ell-1} \frac{1}{d_{u_s}}} \geq \frac{(d_{\min})^{\ell}}{\sqrt{d_{v,\ell-1}d_{j,\ell-1}}} \quad (18)$$

Combining (17) and (18) yields

$$\frac{(\hat{\mathbf{A}}_{\ell-1})_{vj}(\hat{\mathbf{A}}_{\ell-1})_{ju}}{\frac{1}{\sqrt{d_v d_u}} \prod_{s=1}^{\ell-1} \frac{1}{d_{u_s}}} \geq \frac{(d_{\min})^{\ell}}{\sqrt{d_{v,\ell-1}d_{j,\ell-1}}} \cdot \sqrt{\frac{d_j}{d_u}} \quad (19)$$

### 3.2 Locality awareness



## 4 Related works

Give a brief summary of (some) existing methods that are related to your project. For instance, you can refer to Gilmer et al. [1], or simply [1], for introducing Message Passing Neural Networks. In this section it is important to provide readers references to the current state of the art and the foundations of the presented method.

N.B.: When referencing a different approach, it is not necessary to provide a detailed description, only one/two brief sentences are enough. The interested readers can eventually read the referenced work.

## 5 Methodology

*You can change the name of this section as you see fit.*

In this section you should give a description of the methodological aspects of your work, for instance how you modified an existing method to perform a particular task or to overcome a particular limitation. If your project is about reproducibility, here you should describe the method presented in the original paper.

## 6 Implementation

This section should be structured as follows (from the Reproducibility challenge template):

Briefly describe what you did and which resources did you use. E.g. Did you use author's code, did you re-implement parts of the pipeline, how much time did it take to produce the results, what hardware you were using and how long it took to train/evaluate.

### 6.1 Datasets

Describe the datasets you used and how you obtained them.

### 6.2 Hyperparameters

Describe how you set the hyperparameters and what was the source for their value (e.g. paper, code or your guess).

## 6.3 Experimental setup

Explain how you ran your experiments, e.g. the CPU/GPU resources and provide the link to your code and notebooks.

## 6.4 Computational requirements

Provide information on computational requirements for each of your experiments. For example, the number of CPU/GPU hours and memory requirements. You'll need to think about this ahead of time, and write your code in a way that captures this information so you can later add it to this section.

## 7 Results

In this section you should report the results of your work (e.g., the outcome of an empirical analysis). You should be objective and support your statements with empirical evidence.

**Table 1.** Experimental results (average of 3 runs).

Methods	MAE	MSE
Baseline1	21.23 ± 1.65	841.36 ± 12.65
Baseline2	15.45 ± 1.02	652.38 ± 09.89
Method	12.03 ± 0.35	324.13 ± 05.56

Use figures, plots and tables (like Table 1) to present your results in a nice and readable way.

## 8 Discussion and conclusion

Here you can express your judgments and draw your conclusions based on the evidences produced on the previous sections.

Try to summarize the achievements of your project and its limits, suggesting (when appropriate) possible extensions and future works.

## References

- [1] Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural message passing for quantum chemistry. In *International conference on machine learning*, pages 1263–1272. PMLR, 2017.