Supplement to "The Multiscale Laplacian Graph Kernel"

1 Proofs

Proposition 1. *Let $\mathcal{G}_1$ and $\mathcal{G}_2$ be two graphs with vertex sets $V_1 = \{v_1 \dots v_{n_1}\}$ and $V_2 = \{v'_1 \dots v'_{n_2}\}$, and let $\{\xi_1, \dots, \xi_p\}$ be an orthonormal basis for the subspace*

$$W = \operatorname{span}\big\{\boldsymbol{\phi}(v_1), \dots, \boldsymbol{\phi}(v_{n_1}), \boldsymbol{\phi}(v'_1), \dots, \boldsymbol{\phi}(v'_{n_2})\big\}.$$

Then, k_{FLG}(as defined in equation (4) of our paper) can be rewritten as

$$k_{FLG}(\mathcal{G}_1, \mathcal{G}_2) = \frac{\big|\big(\frac{1}{2}\overline{S}_1^{-1} + \frac{1}{2}\overline{S}_2^{-1}\big)^{-1}\big|^{1/2}}{|\overline{S}_1|^{1/4}\,|\overline{S}_2|^{1/4}}\ , \tag{1}$$

where $[\overline{S}_1]_{i,j} = \xi_i^\top S_1 \xi_j$ and $[\overline{S}_2]_{i,j} = \xi_i^\top S_2 \xi_j$. In other words, $\overline{S}_1$ and $\overline{S}_2$ are the projections of S_1 and S_2 to W.

Proof. The proposition hinges on the fact that k_{FLG}(as defined in equation (4) in our paper) is invariant to rotation. In particular, if we extend $\{\xi_1, \dots, \xi_p\}$ to an orthonormal basis $\{\xi_1, \dots, \xi_m\}$ for the whole of $\mathbb{R}^m$, let $O = [\xi_1, \dots, \xi_m]$ (the change of basis matrix) and set $\tilde{S}_1 = O^\top S_1 O$, and $\tilde{S}_2 = O^\top S_2 O$, then (4) can equivalently be written as

$$k_{\text{FLG}}(\mathcal{G}_1, \mathcal{G}_2) = \frac{\big|\big(\frac{1}{2}\tilde{S}_1^{-1} + \frac{1}{2}\tilde{S}_2^{-1}\big)^{-1}\big|^{1/2}}{|\tilde{S}_1|^{1/4}\,|\tilde{S}_2|^{1/4}}\ . \tag{2}$$

However, in the $\{\xi_1, \dots, \xi_m\}$ basis $\tilde{S}_1$ and $\tilde{S}_2$ take on a special form. Writing S_1 in the outer product form

$$S_1 = \sum_{a,b=1}^{n_1} \boldsymbol{\phi}(v_a)\,[L_1^{-1}]_{a,b}\,\boldsymbol{\phi}(v_b)^\top + \gamma I$$

and considering that for $i > p$, $\langle \boldsymbol{\phi}(v_a), \xi_i \rangle = 0$ shows that $\tilde{S}_1$ splits into a direct sum $\tilde{S}_1 = \overline{S}_1 \oplus \widehat{S}_1$ of two matrices: a $p \times p$ matrix $\overline{S}_1$ whose (i,j) entry is

$$\xi_i^\top S_1 \xi_j = \sum_{a,b=1}^{n_1} \langle \xi_i, \boldsymbol{\phi}(v_{1,a}) \rangle\,[L_1^{-1}]_{a,b} \langle \boldsymbol{\phi}(v_{1,b}), \xi_j \rangle + \gamma \delta_{i,j}, \tag{3}$$

where $\delta_{i,j}$ is the Kronecker delta; and an $(n-p) \times (n-p)$ dimensional matrix $\widehat{S}_1 = \gamma I_{n-p}$ (where I_{n-p} denotes the $n-p$ dimensional identity matrix). Naturally, $\tilde{S}_2$ decomposes into $\overline{S}_2 \oplus \widehat{S}_2$ in an analogous way.

Recall that for any pair of square matrices M_1 and M_2, $|\,M_1 \oplus M_2\,| = |M_1| \cdot |M_2|$ and $(M_1 \oplus M_2)^{-1} = M_1^{-1} \oplus M_2^{-1}$. Applying this to (2) then gives

$$\begin{aligned}
k_{\text{FLG}}(\mathcal{G}_1, \mathcal{G}_2) &= \frac{\big|\big(\big(\frac{1}{2}\overline{S}_1^{-1} + \frac{1}{2}\overline{S}_2^{-1}\big) \oplus \gamma^{-1} I_{n-p}\big)^{-1}\big|^{1/2}}{\big|\overline{S}_1 \oplus \gamma I_{n-k}\big|^{1/4}\,\big|\overline{S}_2 \oplus \gamma I_{n-k}\big|^{1/4}} \\
&= \frac{\big|\,\big(\frac{1}{2}\overline{S}_1^{-1} + \frac{1}{2}\overline{S}_2^{-1}\big)^{-1} \oplus \gamma I_{n-p}\,\big|^{1/2}}{\big|\overline{S}_1 \oplus \gamma I_{n-k}\big|^{1/4}\,\big|\overline{S}_2 \oplus \gamma I_{n-k}\big|^{1/4}} \\
&= \frac{\gamma^{(n-p)/2}}{\gamma^{(n-p)/4}\,\gamma^{(n-p)/4}} \frac{\big|\,\big(\frac{1}{2}\overline{S}_1^{-1} + \frac{1}{2}\overline{S}_2^{-1}\big)^{-1}\,\big|^{1/2}}{\big|\overline{S}_1\big|^{1/4}\ \big|\overline{S}_2\big|^{1/4}} \\
&= \frac{\big|\,\big(\frac{1}{2}\overline{S}_1^{-1} + \frac{1}{2}\overline{S}_2^{-1}\big)^{-1}\,\big|^{1/2}}{\big|\overline{S}_1\big|^{1/4}\ \big|\overline{S}_2\big|^{1/4}}
\end{aligned}$$

□

Proposition 2. *Let $\mathcal{G}_1$ and $\mathcal{G}$ be as in Proposition 1, $\overline{V} = \{\overline{v}_1, \ldots, \overline{v}_{n_1+n_2}\}$ be the union of their vertex sets (where it is assumed that the first n_1 vertices are $\{v_1, \ldots, v_{n_1}\}$ and the second n_2 vertices are $\{v'_1, \ldots, v'_{n_2}\}$), and define the joint Gram matrix $K \in \mathbb{R}^{(n_1+n_2)\times(n_1+n_2)}$ as*

$$K_{i,j} = \kappa(\overline{v}_i, \overline{v}_j) = \boldsymbol{\phi}(\overline{v}_i)^\top \boldsymbol{\phi}(\overline{v}_j).$$

Let $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p$ be (a maximal orthonormal set of) the non-zero eigenvalue eigenvectors of K with corresponding eigenvalues $\lambda_1, \ldots, \lambda_p$. Then the vectors

$$\xi_i = \frac{1}{\sqrt{\lambda_i}} \sum_{\ell=1}^{n_1+n_2} [\boldsymbol{u}_i]_\ell \, \boldsymbol{\phi}(\overline{v}_\ell) \tag{4}$$

form an orthonormal basis for W. Moreover, defining $Q = [\lambda_1^{1/2}\boldsymbol{u}_1, \ldots, \lambda_p^{1/2}\boldsymbol{u}_p] \in \mathbb{R}^{p\times p}$ and setting $Q_1 = Q_{1:n_1,:}$ and $Q_2 = Q_{n_1+1:n_2,:}$ (the first n_1 and remaining n_2 rows of Q, respectively), the matrices $\overline{S}_1$ and $\overline{S}_2$ appearing in (5) can be computed as

$$\overline{S}_1 = Q_1^\top L_1^{-1} Q_1 + \gamma I, \qquad \overline{S}_2 = Q_2^\top L_2^{-1} Q_2 + \gamma I. \tag{5}$$

Proof. For $i \neq j$,

$$\begin{aligned}
\xi_i^\top \xi_j &= \frac{1}{\sqrt{\lambda_i \lambda_j}} \sum_{k=1}^{n_1+n_2} \sum_{\ell=1}^{n_1+n_2} [\boldsymbol{u}_i]_k \underbrace{\boldsymbol{\phi}(\overline{v}_k)^\top \boldsymbol{\phi}(\overline{v}_\ell)}_{\kappa(\overline{v}_k, \overline{v}_\ell)} [\boldsymbol{u}_j]_\ell \\
&= (\lambda_i \lambda_j)^{-1/2} \, \boldsymbol{u}_i^\top K \boldsymbol{u}_j = (\lambda_j/\lambda_i)^{1/2} \boldsymbol{u}_i^\top \boldsymbol{u}_j \\
&= 0,
\end{aligned}$$

while for $i = j$, $\xi_i^\top \xi_j = \lambda_i^{-1} \, \boldsymbol{u}_i^\top K \boldsymbol{u}_i = \boldsymbol{u}_i^\top \boldsymbol{u}_i = 1$, showing that $\{\xi_1, \ldots, \xi_p\}$ is an orthonormal set.

At the same time, $p = \text{rank}(K) = \dim(W)$ and $\xi_1, \ldots, \xi_p \in W$, proving that $\{\xi_1, \ldots, \xi_p\}$ is an orthonormal basis for W.

To derive the form of $\overline{S}_1$, simply plug (4) into (3):

$$\begin{aligned}
\xi_i^\top S_1 \xi_j &= \frac{1}{\sqrt{\lambda_i \lambda_j}} \sum_{k=1}^{n_1} \sum_{\ell=1}^{n_1} \sum_{a,b=1}^{n} [\boldsymbol{u}_i]_k \underbrace{\boldsymbol{\phi}(\overline{v}_k)^\top \boldsymbol{\phi}(\overline{v}_a)}_{\kappa(\overline{v}_k, \overline{v}_a)} [L_1^{-1}]_{a,b} \underbrace{\boldsymbol{\phi}(\overline{v}_b)^\top \boldsymbol{\phi}(\overline{v}_\ell)}_{\kappa(\overline{v}_b, \overline{v}_\ell)} [\boldsymbol{u}_j]_\ell + \gamma \delta_{i,j} \\
&= (\lambda_i \lambda_j)^{-1/2} \, \boldsymbol{u}_i^\top K L^{-1} K \boldsymbol{u}_j + \gamma \delta_{i,j} \\
&= (\lambda_i \lambda_j)^{1/2} \, \boldsymbol{u}_i^\top L^{-1} \boldsymbol{u}_j + \gamma \delta_{i,j},
\end{aligned}$$

and similarly for $\overline{S}_2$. □

2 Experiments

2.1 Datasets

We used the following datasets in our experiments: MUTAG, PTC, ENZYMES, PROTEINS, NCI1 and NCI109. MUTAG is a dataset of 188 mutagenic aromatic and heteroaromatic compounds. PTC is a dataset of 344 chemical compounds that reports their carcinogenicity for male and female rats. ENZYMES is a dataset of protein tertiary structures consisting of 600 enzymes from the BRENDA enzyme database with 3 discrete node labels. PROTEINS is a dataset of 1113 compounds where the nodes are seconadary structure elements with 3 discrete node labels representing helix, sheet and turn. NCI1 and NCI109 are two datasets(4110, 4127 nodes) of chemical compounds made available by the National Cancer Institute, screened for activity against non-small cell lung cancer and ovarian cancer cell lines, respectively. All six datasets are endowed with discrete node labels.

2.2 Parameter Settings

In our experiments, we found the optimal settings for the number of levels to be 2 or 3 and the radius size to be 2 or 3 for each dataset. As can be seen from the average number of nodes and average diameter values in Table 1, the

Table 1: Summary of the datasets used in our experiments

Dataset	Size	Labels	Nodes	Edges	Diameter	Classes
MUTAG	188	7	17.9	39.6	8.2	2 (125 vs 63)
PTC	344	19	25.6	51.9	8.9	2 (192 vs 152)
ENZYMES	600	3	32.6	124.3	10.9	6 (100 each)
PROTEINS	1113	3	39.1	145.6	11.6	2 (663 vs 450)
NCI1	4110	37	29.9	64.6	13.3	2 (2057 vs 2053)
NCI109	4127	38	29.7	64.3	13.1	2 (2079 vs 2048)

Table 2: Runtime for Computing Gram Matrix

Method	MUTAG	PTC	ENZYMES	PROTEINS	NCI1	NCI109
WL	2.0s	5.2s	12.0s	30.1s	1min 22.4s	1min 23.1s
WL-Edge	2.1s	5.4s	12.99s	41.1s	1min 34.8	1min 34.7
SP	0.10s	0.4s	0.9s	1min 53.6s	35.0s	35.2s
Graphlet	1min 17.6s	3min 3.7s	7min 22.6s	11min 10.7s	40min 41.1s	41min 0.10s
p–RW	4min 9.3s	70min 54.4s	38min 25.0s	34min 8.0s	$>$24hrs	$>$24hrs
MLG	0.86s	1min 11.18s	36.65s	16min 19.8s	22min 12.6s	23min 40.3

graphs in each dataset are small enough that a 2 or 3 level deep MLG kernel is sufficient to effectively characterize the similarity between graphs.

Across all datasets, the optimal η and γ parameters were set to 0.01 and 0.1 or 0.01 respectively. In general, these two parameters can be set through cross validation over a small set of values. For two graphs G and $\hat{G}$, that are reasonably similar with only slight differences(ex: $\hat{G}$ is similar to G in degree distribution, connectivity, etc), increasing the η and/or γ value will have the effect of artificially increasing the value of $k_{FLG}(G, \hat{G})$, smoothing out their differences. Of course, this sort of smoothing is not desirable for all pairs of graphs, particularly graphs that belong to different classes, so typically the optimal η and γ values will be small, often between 0.01 and 1.

2.3 Runtime

In table (2), we display the wall clock time for how long it took to compute each kernel with the optimal parameter settings. Our implementation was in C++ and takes advantage of multithreading, where as the competing methods were provided by their authors in Matlab. Some of the kernels implemented in Matlab do perform parallelized computations, but the extent to which each method uses parallelism is unclear so a direct comparison of runtimes might not be fair.

Algorithm 1 The fast MLG algorithm (high level overview)

INPUT: A collection of graphs $\mathfrak{G} = \{\mathcal{G}_1, \ldots, \mathcal{G}_M\}$ with joint vertex set $\overline{V}$, a base kernel $\kappa \colon \overline{V} \times \overline{V} \to \mathbb{R}$, a system of nested subgraphs $v \in N_1(v) \subseteq N_2(v) \subseteq \ldots \subseteq N_L(v)$ for each vertex $v \in \overline{V}$, and smoothing parameters η and γ.

Sample $\{\tilde{v}_1, \ldots, \tilde{v}_{\tilde{N}}\}$ from $\overline{V}$
Compute the subsampled Gram matrix $\tilde{K}$, with elements $\tilde{K}_{i,j} = \kappa(\tilde{v}_i, \tilde{v}_j)$
Compute a basis $\{\xi_1, \ldots, \xi_{\tilde{P}}\}$ for the approximate joint feature space $\tilde{W}$ from K
For each ($v \in \overline{V}$) { $\boldsymbol{\phi}(v) \leftarrow$ the projection of $\phi_\kappa(v)$ to $\tilde{W}_\kappa$ }

For ($\ell = 1$ **to** L) {
  For each ($v \in \overline{V}$) {
    $\mathcal{L}_v \leftarrow$ the Laplacian of $G_\ell(v)$
    $U_v \leftarrow [\boldsymbol{\phi}(w_1), \ldots, \boldsymbol{\phi}(w_{|G_\ell(v)|})]^\top$, where $(w_1, \ldots, w_{|G_\ell(v)|})$ are the vertices of $G_\ell(v)$
    $\overline{S}_v \leftarrow U_v(\mathcal{L}_v + \eta I)^{-1} U_v^\top + \gamma I$
  }
  Sample $\{\tilde{v}_1, \ldots, \tilde{v}_{\tilde{N}}\}$ from V
  Compute the subsampled Gram matrix $\tilde{K} \in \mathbb{R}^{\tilde{N} \times \tilde{N}}$, with elements

$$\tilde{K}_{i,j} = k_{\text{FLG}}(G_\ell(\tilde{v}_i), G(\tilde{v}_j)) = \frac{\left|\left(\frac{1}{2}\overline{S}_{\tilde{v}_1}^{-1} + \frac{1}{2}\overline{S}_{\tilde{v}_2}^{-1}\right)^{-1}\right|^{1/2}}{|\overline{S}_{\tilde{v}_1}|^{1/4}\, |\overline{S}_{\tilde{v}_2}|^{1/4}}$$

  Compute a basis $\{\xi_1, \ldots, \xi_{\tilde{P}}\}$ for the approximate joint feature space $\tilde{W}_\kappa$ from K
  For each ($v \in \overline{V}$) { $\boldsymbol{\phi}(v) \leftarrow$ the projection of $\phi_\kappa(v)$ to $\tilde{W}_\kappa$ }
}

For ($i = 1$ **to** M) {
  $\mathcal{L}_i \leftarrow$ the Laplacian of $\mathcal{G}_i$
  $U_i \leftarrow [\boldsymbol{\phi}(v_1), \ldots, \boldsymbol{\phi}(v_{|V_i|})]^\top$, where $(v_1, \ldots, v_{|V_i|})$ are the vertices of $\mathcal{G}_i$
  $\overline{S}_i \leftarrow U_i(\mathcal{L}_i + \eta I)^{-1} U_i^\top + \gamma I$
}

Compute the MLG Gram matrix $G \in \mathbb{R}^{M \times M}$, with elements

$$G_{i,j} = \mathfrak{K}(\mathcal{G}_i, \mathcal{G}_j) = \frac{\left|\left(\frac{1}{2}\overline{S}_i^{-1} + \frac{1}{2}\overline{S}_j^{-1}\right)^{-1}\right|^{1/2}}{|\overline{S}_i|^{1/4}\, |\overline{S}_j|^{1/4}}$$

OUTPUT: the MLG Gram matrix G