DPGS: Degree-Preserving Graph Summarization (Appendix)

Houquan Zhou*[1], Shenghua Liu[1], Kyuhan Lee[2], Kijung Shin[2], Huawei Shen[1] and Xueqi Cheng[1]

[1]Institute of Computing Technology, Chinese Academy of Sciences
[2]Graduate School of AI, KAIST

October 19, 2020

A Analysis of $\Delta L_E(i,j)$

Assumed that the input graphs are simple and undirected, the error description length $L(D \mid M)$ is:

$$
\begin{aligned}
L(D \mid M) &= \sum_{(i,j)\in E} \ln \frac{1}{\frac{d_i}{D_k} A_S(k,l) \frac{d_j}{D_l}} \\
&= \sum_{(i,j)\in E} \ln \frac{D_k D_l}{A_S(k,l)} - \sum_{(i,j)\in E} \ln(d_i d_j) \,.
\end{aligned}
$$
(A.1)

Since the second term only depends on the input graph, we only consider the first term in the following.

$$
\begin{aligned}
L(D \mid M) &= \sum_{(i,j)\in E} \ln \frac{D_k D_l}{A_S(k,l)} \\
&= 2\sum_{k=1}^{n_s} D_k \ln D_k - \sum_{k=1}^{n_s}\sum_{l=1}^{n_s} A_S(k,l) \ln A_S(k,l) \,.
\end{aligned}
$$
(A.2)

Suppose the current supernode set is $\mathcal{S}$. After merging supernodes S_i and S_j into a new supernode S_k, we get a new supernode set $\mathcal{S}' = \mathcal{S} \setminus \{S_i, S_j\} \cup \{S_k\}$. The connectivity of S_k is the aggregation of S_i and S_j, that is,

$$
A_S(k,l) = \begin{cases} A_S(i,l) + A_S(j,l) & l \neq k \\ A_S(i,i) + A_S(j,j) + 2A_S(i,j) & l = k \,. \end{cases}
$$
(A.3)

Now we expand the $\Delta L_E(i,j)$ as follow (we define $f(x) = x \ln x \,(x > 0)$ and $f(0) = 0$ for simplicity):

$$
\begin{aligned}
\Delta L_E(i,j) &= L(D \mid M') - L(D \mid M) \\
&= 2(f(D_i + D_j) - f(D_i) - f(D_j)) \\
&\quad + 2 \sum_{l\in\mathcal{S}, l\neq i,j} [f(A_S(i,l)) + f(A_S(j,l))] \\
&\quad - 2 \sum_{l\in\mathcal{S}', l\neq k} f(A_S(k,l)) \\
&\quad + f(A_S(i,i)) + f(A_S(i,j)) + 2f(A_S(i,j)) \\
&\quad - f(A_S(k,k)) \,.
\end{aligned}
$$
(A.4)

Note that if S_l is not the common neighbor of supernode S_i and S_j, it makes no contribution to $\Delta L_E(i,j)$. Each common neighbors makes positive contribution to $\Delta L_E(i,j)$ (since $f(x+y) > f(x) + f(y)$ for $x, y \geq 1$). Thus, the more common neighbors supernodes S_i and S_j have, the more likely the merging cost $\Delta L_E(i,j)$ is small.

---

*Email: zhouhouquan18@mails.ucas.ac.cn

B PROOFS

B.1 Proof of Theorem 3.1

Proof. Denote the normalized Laplacian matrix of the original graph and the reconstructed graph as $\mathcal{L}$ and $\mathcal{L}'$. By [?], the squared error between eigenvalues of $\mathcal{L}$ and $\mathcal{L}'$ are bounded by:

$$(B.5) \qquad \sum_{i=1}^{n}(\lambda(i) - \lambda'(i))^2 \leq \|\mathcal{L} - \mathcal{L}'\|_F^2$$

$$\|\mathcal{L} - \mathcal{L}'\|_F^2 = \sum_{i=1}^{n}\sum_{j=1}^{n}|\mathcal{L}(i,j) - \mathcal{L}'(i,j)|^2$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{A(i,j)}{\sqrt{d_i d_j}} - \frac{A'(i,j)}{\sqrt{d_i d_j}}\right)^2$$

$$\leq \sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{A(i,j)}{\sqrt{d_i}} - \frac{A'(i,j)}{\sqrt{d_i}}\right)^2$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n}d_i\left(\frac{A(i,j)}{d_i} - \frac{A'(i,j)}{d_i}\right)^2$$

Denote the normalized i-th row of A and A' as $\tilde{A}(i)$ and $\tilde{A}'(i)$, then $\tilde{A}(i)$ and $\tilde{A}'(i)$ can be seen as two distributions. Moreover, with the following inequality [?]:

$$(B.6) \qquad D_{KL}(p\|q) = \sum_{i} p_i \log_2 \frac{p_i}{q_i} \geq \frac{1}{2\ln 2}\|p - q\|_1^2,$$

we have[1]

$$\|\mathcal{L} - \mathcal{L}'\|_F^2 = \sum_{i=1}^{n} d_i \cdot \|\tilde{A} - \tilde{A}'\|_1^2$$

$$\leq 2\ln 2 \sum_{i=1}^{n} d_i \cdot D_{KL}(\tilde{A}\|\tilde{A}')$$

$$= 2 \cdot \sum_{i=1}^{n} d_i \cdot \sum_{j=1}^{n} \frac{A(i,j)}{d_i} \ln \frac{A(i,j)}{A'(i,j)}$$

$$= 2 \cdot \sum_{i=1}^{n}\sum_{j=1}^{n} A(i,j) \ln \frac{A(i,j)}{A'(i,j)}$$

$$= 2 \cdot L(D \mid M)$$

Together, we have

$$(B.7) \qquad \sum_{i=1}^{n}(\lambda(i) - \lambda'(i))^2 \leq 2 \cdot L(D \mid M)$$

∎

---

[1] The base of logarithm of Inequality (??) is 2, thus the factor $\ln 2$ vanished since we change the base from 2 to e.

B.2 Proof of Theorem 3.2

Proof. The key tool is Jensen's inequality. Suppose $f(x)$ is a convex function, then

$$(B.8) \qquad f(\frac{\sum_i \lambda_i x_i}{\sum_i \lambda_i}) \leq \frac{\sum_i \lambda_i f(x_i)}{\sum_i \lambda_i},$$

where λ_i is positive weights.

In the following proof, the inequality is applied on $f(x) = -\ln x$.

$$(B.9) \qquad \begin{aligned} \Delta L_E(i,j) = {}& 2\left(D_i \ln \frac{D_i + D_j}{D_i} + D_j \ln \frac{D_i + D_j}{D_j} \right) \\ & + 2 \sum_{l \neq i,j} \left(A_S(i,l) \ln \frac{A_S(i,l)}{A_S(i,l) + A_S(j,l)} + A_S(j,l) \ln \frac{A_S(j,l)}{A_S(i,l) + A_S(j,l)} \right) \\ & + A_S(i,i) \ln \frac{A_S(i,i)}{A_S(i,i) + A_S(j,j) + 2A_S(i,j)} + A_S(j,j) \ln \frac{A_S(j,j)}{A_S(i,i) + A_S(j,j) + 2A_S(i,j)} \\ & + 2A_S(i,j) \ln \frac{A_S(i,j)}{A_S(i,i) + A_S(j,j) + 2A_S(i,j)} \end{aligned}$$

Let's focused on the part related to i in Equation (??), since i and j are symmetric.

$$\begin{aligned} \Delta L_E(i,j)_i = {}& 2D_i \ln \frac{D_i + D_j}{D_i} \\ & + 2 \sum_{l \neq i,j} \left(A_S(i,l) \ln \frac{A_S(i,l)}{A_S(i,l) + A_S(j,l)} \right) \\ & + A_S(i,i) \ln \frac{A_S(i,i)}{A_S(i,i) + A_S(j,j) + 2A_S(i,j)} \\ & + A_S(i,j) \ln \frac{A_S(i,j)}{A_S(i,i) + A_S(j,j) + 2A_S(i,j)} \end{aligned}$$

Apply Jensen's inequality on the last two row

$$\begin{aligned} \geq {}& 2D_i \ln \frac{D_i + D_j}{D_i} \\ & + 2 \sum_{l \neq i,j} \left(A_S(i,l) \ln \frac{A_S(i,l)}{A_S(i,l) + A_S(j,l)} \right) \\ & + 2(A_S(i,i) + A_S(i,j)) \cdot \left(-\ln \frac{A_S(i,i) + A_S(j,j) + 2A_S(i,j)}{A_S(i,i) + A_S(i,j)} \right) \\ = {}& 2D_i \left[\sum_{l \neq i,j} \frac{A_S(i,l)}{D_i} \cdot \left(-\ln \frac{A_S(i,l) + A_S(j,l)}{A_S(i,l)} \frac{D_i}{D_i + D_j} \right) \right. \\ & \left. + \frac{A_S(i,i) + A_S(i,j)}{D_i} \cdot \left(-\ln \frac{A_S(i,i) + A_S(j,j) + 2A_S(i,j)}{A_S(i,i) + A_S(i,j)} \frac{D_i}{D_i + D_j} \right) \right] \end{aligned}$$

Apply Jensen's inequality again

$$\begin{aligned} \geq {}& 2D_i \cdot \left(-\ln \left[\sum_{l \neq i,j} \frac{A_S(i,l) + A_S(j,l)}{D_i + D_j} + \frac{A_S(i,i) + A_S(j,j) + 2A_S(i,j)}{D_i + D_j} \right] \right) \\ = {}& 2D_i \cdot (-\ln 1) \\ = {}& 0 \end{aligned}$$

Similarly, $\Delta L_E(i,j)_j \geq 0$. Together, $\Delta L_E(i,j) \geq 0$. ∎