# Appendix for PaperID 2093



## A PROOFS

### A.1 Proof of Lemma 4.2

Before we prove Lemma 4.2, we introduce two useful lemmas firstly.

Lemma A.1.

$$\mathbf{Q}^{\mathrm{T}}\mathbf{D}^{-1}\mathbf{Q} = \mathbf{D}_s^{-1}, \tag{16}$$

*where* $\mathbf{Q}$ *is the reconstruction matrix in the configuration-based reconstruction scheme (see Equation* (7)*),* $\mathbf{D}$ *and* $\mathbf{D}_s$ *are degree matrix of the original graph and the summary graph.*

Proof. The $(p,q)$-th entry in $\mathbf{Q}^{\mathrm{T}}\mathbf{D}^{-1}\mathbf{Q}$ is:

$$\mathbf{Q}^{\mathrm{T}}\mathbf{D}^{-1}\mathbf{Q}(p,q) = \sum_i \mathbf{Q}(i,p)\frac{1}{d_i}\mathbf{Q}(i,q) \tag{17}$$

It is easy to see that the result is not zero only when $p = q$ (since a node $i$ cannot belongs to two supernodes $\mathcal{S}_p$ and $\mathcal{S}_q$ simultaneously). And diagonal items are (note that $d_p^{(s)} = \sum_{v_i \in \mathcal{S}_p} d_i$):

$$\begin{aligned}
\mathbf{Q}^{\mathrm{T}}\mathbf{D}^{-1}\mathbf{Q}(p,p) &= \sum_{v_i \in \mathcal{S}_p} \mathbf{Q}(i,p)\frac{1}{d_i}\mathbf{Q}(i,p) \\
&= \sum_{v_i \in \mathcal{S}_p} \frac{d_i}{d_p^{(s)}}\frac{1}{d_i}\frac{d_i}{d_p^{(s)}} \\
&= \sum_{v_i \in \mathcal{S}_p} \frac{d_i}{d_p^{(s)}}\frac{1}{d_p^{(s)}} \\
&= \frac{1}{d_p^{(s)}} = \mathbf{D}_s^{-1}(p,p)
\end{aligned} \tag{18}$$

■

Lemma A.2.

$$\mathbf{R}\mathbf{D}_s^{-c} = \mathbf{D}^{-c}\mathbf{Q} \tag{19}$$

Proof. Suppose $v_i \in \mathcal{S}_p$, then the $(i,p)$-th entry of $\mathbf{R}\mathbf{D}_s^{-c}$ is:

$$\mathbf{R}\mathbf{D}_s^{-c}(i,p) = \left(\frac{d_i}{d_p^{(s)}}\right)^{1-c}\left(d_p^{(s)}\right)^{-c} = \frac{d_i^{1-c}}{d_p^{(s)}}$$

And the $(i,p)$-th entry of $\mathbf{D}^{-c}\mathbf{Q}$ is:

$$\mathbf{D}^{-c}\mathbf{Q}(i,p) = (d_i)^{-c}\frac{d_i}{d_p^{(s)}} = \frac{d_i^{1-c}}{d_p^{(s)}}$$

Thus $\mathbf{R}\mathbf{D}_s^{-c} = \mathbf{D}^{-c}\mathbf{Q}$. ■

**Proof of Lemma 4.2**:

Proof. Denote $\mathcal{K}_\tau(\mathcal{G}) = \left(\mathbf{D}^{-c}\mathbf{A}_r\mathbf{D}^{-1+c}\right)^\tau \mathbf{D}^{1-2c}$ for convenience. Prove by induction. When $\tau = 1$,

$$\begin{aligned}
\mathcal{K}_1(\mathcal{G}_r) &= \mathbf{D}^{-c}\mathbf{A}_r\mathbf{D}^{-1+c}\mathbf{D}^{1-2c} \\
&= \mathbf{D}^{-c}\mathbf{Q}\mathbf{A}_s\mathbf{Q}^{\mathrm{T}}\mathbf{D}^{-c} \qquad \text{(Lemma A.2)} \\
&= \mathbf{R}\mathbf{D}_s^{-c}\mathbf{A}_s\mathbf{D}_s^{-c}\mathbf{R}^{\mathrm{T}} \\
&= \mathbf{R}\,\mathcal{K}_1(\mathcal{G}_s)\,\mathbf{R}^{\mathrm{T}}
\end{aligned}$$

Suppose the the lemma holds for $\tau = i$, i.e., $\mathcal{K}_i(\mathcal{G}_r) = \mathbf{R}\,\mathcal{K}_i(\mathcal{G}_s)\,\mathbf{R}^{\mathrm{T}}$. For the case $\tau = i+1$,

$$\begin{aligned}
\mathcal{K}_{i+1}(\mathcal{G}_r) &= \mathbf{D}^{-c}\mathbf{A}_r\mathbf{D}^{-1+c}\,\mathcal{K}_i(\mathcal{G}_r) \qquad \text{(By induction hypothesis)} \\
&= \mathbf{D}^{-c}\mathbf{A}_r\mathbf{D}^{-1+c}\mathbf{R}\,\mathcal{K}_i(\mathcal{G}_s)\,\mathbf{R}^{\mathrm{T}} \\
&= \mathbf{D}^{-c}\mathbf{Q}\mathbf{A}_s\mathbf{Q}^{\mathrm{T}}\mathbf{D}^{-1+c}\mathbf{R}\,\mathcal{K}_i(\mathcal{G}_s)\,\mathbf{R}^{\mathrm{T}} \\
&= \mathbf{D}^{-c}\mathbf{Q}\mathbf{A}_s\mathbf{Q}^{\mathrm{T}}\mathbf{D}^{-1}(\mathbf{D}^c\mathbf{R})\,\mathcal{K}_i(\mathcal{G}_s)\,\mathbf{R}^{\mathrm{T}} \\
&\overset{4}{=} \mathbf{D}^{-c}\mathbf{Q}\mathbf{A}_s\mathbf{Q}^{\mathrm{T}}\mathbf{D}^{-1}(\mathbf{Q}\mathbf{D}_s^c)\,\mathcal{K}_i(\mathcal{G}_s)\,\mathbf{R}^{\mathrm{T}} \\
&\text{(Lemma A.1 and A.2.)} \\
&= \mathbf{R}\mathbf{D}_s^{-c}\mathbf{A}_s\mathbf{D}_s^{-1+c}\,\mathcal{K}_i(\mathcal{G}_s)\,\mathbf{R}^{\mathrm{T}} \\
&= \mathbf{R}\,\mathcal{K}_{i+1}(\mathcal{G}_s)\,\mathbf{R}^{\mathrm{T}}
\end{aligned}$$

Applying principal of induction finishes the proof. ■

### A.2 Proof of Theorem 4.5

Proof. Consider $\mathbf{A}_r$ as a low-rank approximation of $\mathbf{A}$, and replace $\mathbf{A}$ by $\mathbf{A}_r$ in the DeepWalk matrix. According to Corollary 4.3:

$$\begin{aligned}
\mathbf{M} &= \log\left(\frac{\mathrm{vol}(\mathcal{G})}{bT}\sum_{\tau=1}^{T}(\mathbf{D}^{-1}\mathbf{A})^\tau\mathbf{D}^{-1}\right) \\
&\approx \log\left(\frac{\mathrm{vol}(\mathcal{G})}{bT}\sum_{\tau=1}^{T}(\mathbf{D}^{-1}\mathbf{A}_r)^\tau\mathbf{D}^{-1}\right) \\
&= \log\left(\frac{\mathrm{vol}(\mathcal{G})}{bT}\mathbf{R}\left(\sum_{\tau=1}^{T}(\mathbf{D}_s^{-1}\mathbf{A}_s)^\tau\mathbf{D}_s^{-1}\right)\mathbf{R}^{\mathrm{T}}\right) \\
&\overset{5}{=} \mathbf{R}\cdot\log\left(\frac{\mathrm{vol}(\mathcal{G})}{bT}\left(\sum_{\tau=1}^{T}(\mathbf{D}_s^{-1}\mathbf{A}_s)^\tau\mathbf{D}_s^{-1}\right)\right)\cdot\mathbf{R}^{\mathrm{T}} \\
&= \mathbf{R}\,\mathbf{M}_s\,\mathbf{R}^{\mathrm{T}},
\end{aligned}$$

where $\mathbf{M}_s$ is the corresponding matrix DeepWalk factorizing on summary graph $\mathcal{G}_s$. Suppose $\mathbf{M}_s$ is factorized into $\mathbf{M}_s = \mathbf{X}_s\mathbf{Y}_s^{\mathrm{T}}$, then $\mathbf{M} \approx (\mathbf{R}\mathbf{X}_s)(\mathbf{R}\mathbf{Y}_s)^{\mathrm{T}}$. That is, embeddings of original graph $\mathcal{G}$ can be approximated by embeddings learned on summary graph $\mathcal{G}_s$ with a matrix $\mathbf{R}$.

$$\mathbf{E} \approx \mathbf{R}\cdot\mathbf{E}_s \tag{20}$$

■

### A.3 Proof of Theorem 4.6

Proof. Consider $\mathbf{A}_r$ as a low-rank approximation of $\mathbf{A}$, and replace $\mathbf{A}$ by $\mathbf{A}_r$ in the $k$-th layer of GCN. According to Corollary 4.4:

$$\begin{aligned}
\mathbf{E}^{(k+1)} &= \sigma(\tilde{\mathbf{D}}^{-\frac{1}{2}}\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-\frac{1}{2}}\mathbf{E}^{(k)}\mathbf{W}^{(k)}) \\
&\approx \sigma(\tilde{\mathbf{D}}^{-\frac{1}{2}}\tilde{\mathbf{A}}_r\tilde{\mathbf{D}}^{-\frac{1}{2}}\mathbf{E}^{(k)}\mathbf{W}^{(k)}) \\
&= \sigma(\mathbf{R}(\mathbf{D}_s^{-\frac{1}{2}}\mathbf{A}_s\mathbf{D}_s^{-\frac{1}{2}})(\mathbf{R}^{\mathrm{T}}\mathbf{E}^{(k)})\mathbf{W}) \\
&\overset{6}{=} \mathbf{R}\cdot\sigma((\mathbf{D}_s^{-\frac{1}{2}}\mathbf{A}_s\mathbf{D}_s^{-\frac{1}{2}})(\mathbf{R}^{\mathrm{T}}\mathbf{E}^{(k)})\mathbf{W}^{(k)})
\end{aligned}$$

Let $\mathbf{E}_t^{(k)} = \mathbf{R}^{\mathrm{T}}\mathbf{E}^{(k)}$. Note that $\mathbf{R}^{\mathrm{T}}\mathbf{R} = \mathbf{I}$, we have

$$\begin{aligned}
\mathbf{E}_t^{(k)} &= \mathbf{R}^{\mathrm{T}}\mathbf{E}^{(k)} \\
&\approx \sigma((\mathbf{D}_s^{-\frac{1}{2}}\mathbf{A}_s\mathbf{D}_s^{-\frac{1}{2}})\mathbf{E}_t^{(k-1)}\mathbf{W}^{(k-1)})
\end{aligned} \tag{21}$$

[4] $\mathbf{D}^{-c}\mathbf{Q} = \mathbf{R}\mathbf{D}_s^{-c} \Rightarrow \mathbf{Q} = \mathbf{D}^c\mathbf{R}\mathbf{D}_s^{-c} \Rightarrow \mathbf{Q}\mathbf{D}_s^c = \mathbf{D}^c\mathbf{R}$.

[5] This equation holds since each row of $\mathbf{R}$ contains exactly one non-zero value "1" (see Eq. (12)). Thus we can take it out of the log function. See appendix for details.

[6] This equation holds since each row of $\mathbf{R}$ contains only one non-zero value (see Eq. (14)). That's why we can take $\mathbf{R}$ out of the $\sigma$ function. See next subsection for details.

**Algorithm 3** Configuration-based summarization method

**Input:** $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, summarization ratio $r$
**Output:** $\mathcal{G}_s = (\mathcal{V}_s, \mathcal{E}_s)$
$\mathcal{G}_s \leftarrow \mathcal{G}, \mathcal{V}_s \leftarrow \mathcal{V}, \mathcal{E}_s \leftarrow \mathcal{E}$
$n_s \leftarrow r * |\mathcal{V}|$
**while** $|\mathcal{V}_s| > n_s$ **do**
  Update LSH
  Divide supernodes into disjoint groups by LSH
  **for** each group $g$ **do**
    MergeGroup($g$)
  **end for**
**end while**
**return** $\mathcal{G}_s$

Note that approximate equation (21) is a GCN convolution layer on the summary graph $\mathcal{G}_s$.

By optimizing a GCN network on summary graph $\mathcal{G}_s$ with initial feature $\mathbf{X}_s \coloneqq \mathbf{R}^{\mathrm{T}}\mathbf{X}$, we can get exact embedding solution denoted as $\mathbf{E}_s^{(k)}$. Then we have

$$\mathbf{E}_s^{(k)} \approx \mathbf{E}_t^{(k)} = \mathbf{R}^{\mathrm{T}}\mathbf{E}^{(k)}. \tag{22}$$

And then given $\mathbf{E}_s^{(k)}$ and $\mathbf{R}$, we can solve original embedding $\mathbf{E}^{(k)}$ using a **least-square approximation**, that is:

$$\mathbf{E}^{(k)} \approx (\mathbf{R}^{\mathrm{T}})^{\dagger}\mathbf{E}_s^{(k)} = \mathbf{R}\,\mathbf{E}_s^{(k)} \tag{23}$$

Therefore, we have $\mathbf{E} = \mathbf{E}^{(K)} \approx \mathbf{R}\,\mathbf{E}_s^{(K)} = \mathbf{R}\,\mathbf{E}_s$. ∎

## A.4 Further details in proof of Thm. 4.5 and 4.6

Proposition A.3. $\log(\mathbf{R}\mathbf{M}\mathbf{R}^{\mathrm{T}}) = \mathbf{R}\log(\mathbf{M})\mathbf{R}^{\mathrm{T}}$, *where* $\mathbf{R}$ *is defined in Eq.* (12)*, and* $\mathbf{M}$ *is a* $n_s \times n_s$ *matrix.*

Proof. Suppose $v_i \in \mathcal{S}_p, v_j \in \mathcal{S}_q$, the $(i,j)$-th entry of $\log(\mathbf{R}\mathbf{M}\mathbf{R}^{\mathrm{T}})$ is:

$$\begin{aligned}\log(\mathbf{R}\mathbf{M}\mathbf{R}^{\mathrm{T}})(i,j) &= \log(\mathbf{R}(i,p)\cdot\mathbf{M}(p,q)\cdot\mathbf{R}(j,q))\\ &= \log(1\cdot\mathbf{M}(p,q)\cdot 1)\\ &= \log(\mathbf{M}(p,q))\\ \mathbf{R}\log(\mathbf{M})\mathbf{R}^{\mathrm{T}}(i,j) &= \mathbf{R}(i,p)\log\mathbf{M}(p,q)\mathbf{R}(j,q)\\ &= \log(\mathbf{M}(p,q))\\ &= \log(\mathbf{R}\mathbf{M}\mathbf{R}^{\mathrm{T}})(i,j)\end{aligned}$$

∎

Proposition A.4. $\sigma(\mathbf{R}\,\mathbf{M}) = \mathbf{R}\cdot\sigma(\mathbf{M})$, *where* $\mathbf{R}$ *is defined in Eq.* (14)*, and* $\mathbf{M}$ *is a* $n_s \times d$ *matrix.* $\sigma(\cdot)$ *is ReLU function.*

Proof. Suppose $v_i \in \mathcal{S}_p$, the $i$-th row of $\sigma(\mathbf{R}\,\mathbf{M})$ is:

$$\begin{aligned}\sigma(\mathbf{R}\mathbf{M})(i,:) &= \sigma\Big(\frac{d_i}{d_p^{(s)}}\mathbf{M}(p,:)\Big)\\ &= \frac{d_i}{d_p^{(s)}}\sigma(\mathbf{M}(p,:))\\ &= \mathbf{R}\,\sigma(\mathbf{M})(i,:)\end{aligned}$$

∎

# B DETAILS OF CONFIGURATION-BASED SUMMARIZATION METHOD

In GELSumm, we use a graph summarization method which minimizes the total description length of both the model cost and configuration-based reconstruction error. As described in Alg. 3, it tries to reduce the total description length defined in Eq. (24) by merging nodes successively.

$$L(M, D) = L(M) + L(D \mid M) \tag{24}$$

Here $L(M, D)$ is the total description length which consists of two parts, model part $L(M)$ and error part $L(D \mid M)$.

The error part is measured by KL divergence:

$$\begin{aligned}L(D \mid M) &= \mathrm{KL}(A \| A_r)\\ &= \sum_{i,j} A(i,j)\ln\frac{A(i,j)}{A_r(i,j)} - A(i,j) + A_r(i,j)\\ &= \sum_{i,j} A(i,j)\ln\frac{A(i,j)}{A_r(i,j)},\end{aligned} \tag{25}$$

where $A_r$ is the reconstructed adjacency matrix under configuration-based reconstruction scheme (see Eq. (8)).

For model part, we encode the number of

$$L(M) = \mathrm{L}_{\mathbb{N}}(n_s) + n\,\mathrm{L}_{\mathbb{N}}(n_s) + \sum_{i=1}^{n}\mathrm{L}_{\mathbb{N}}(d_i) + L(A_S)\,, \tag{26}$$

where $\mathrm{L}_{\mathbb{N}}(n)$ is the encoding length for an integer $n$.

The algorithm greedily merges node pair to reduce the total description length (Eq. (24)). Moreover, the LSH (Locality Sensitive Hashing) technique is applied to speed up the algorithm. Specifically, nodes are separated into different groups by LSH according to their neighborhoods, since nodes sharing neighbors are more likely to reduce the objective function. Then, nodes pair are (see Algorithm 4). This process continues until we get the expected summarization ratio.

**Algorithm 4** MergeGroup

**Input:** $g \subset V_S$
$times \leftarrow \log_2 |g|$
$nskip \leftarrow 0$
**while** $nskip < times$ and $|g| \geq 1$ **do**
  $pairs \leftarrow$ Sample $\log_2 |g|$ node pairs from $g$
  $u, v \leftarrow \arg\max_{(i,j)\in pairs} gain(i,j)$
  **if** $gain(u,v) > 0$ **then**
    Merge $u$ and $v$
    $nskip \leftarrow 0$
  **else**
    $nskip \leftarrow nskip + 1$
  **end if**
**end while**