Appendix A
Proofs

A. Proof of Lemma 1

Before we prove Lemma 1, we introduce two useful lemmas firstly.

Lemma 2.

$$\mathbf{Q}^{\mathrm{T}}\mathbf{D}^{-1}\mathbf{Q} = \mathbf{D}_s^{-1}\,, \tag{19}$$

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:

$$\mathbf{Q}^{\mathrm{T}}\mathbf{D}^{-1}\mathbf{Q}(p,q) = \sum_i \mathbf{Q}(i,p)\frac{1}{d_i}\mathbf{Q}(i,q) \tag{20}$$

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 $d_p^{(s)} = \sum_{v_i\in\mathcal{S}_p} d_i$):

$$\begin{aligned}\mathbf{Q}^{\mathrm{T}}\mathbf{D}^{-1}\mathbf{Q}(p,p) &= \sum_{v_i\in\mathcal{S}_p} \mathbf{Q}(i,p)\frac{1}{d_i}\mathbf{Q}(i,p)\\ &= \sum_{v_i\in\mathcal{S}_p} \frac{d_i}{d_p^{(s)}}\frac{1}{d_i}\frac{d_i}{d_p^{(s)}}\\ &= \sum_{v_i\in\mathcal{S}_p} \frac{d_i}{d_p^{(s)}}\frac{1}{d_p^{(s)}}\\ &= \frac{1}{d_p^{(s)}} = \mathbf{D}_s^{-1}(p,p)\end{aligned} \tag{21}$$

■

Lemma 3.

$$\mathbf{R}\mathbf{D}_s^{-c} = \mathbf{D}^{-c}\mathbf{Q} \tag{22}$$

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 1:

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 $\tau = 1$,

$$\begin{aligned}\mathcal{K}_1(\mathcal{G}_r) &= \mathbf{D}^{-c}\mathbf{A}_r\mathbf{D}^{-1+c}\mathbf{D}^{1-2c}\\ &= \mathbf{D}^{-c}\mathbf{Q}\mathbf{A}_s\mathbf{Q}^{\mathrm{T}}\mathbf{D}^{-c} \qquad \text{(Lemma 3)}\\ &= \mathbf{R}\mathbf{D}_s^{-c}\mathbf{A}_s\mathbf{D}_s^{-c}\mathbf{R}^{\mathrm{T}}\\ &= \mathbf{R}\ \ \mathcal{K}_1(\mathcal{G}_s)\ \ \mathbf{R}^{\mathrm{T}}\end{aligned}$$

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 $\tau = i+1$,

$$\begin{aligned}\mathcal{K}_{i+1}(\mathcal{G}_r) &= \mathbf{D}^{-c}\mathbf{A}_r\mathbf{D}^{-1+c}\ \ \mathcal{K}_i(\mathcal{G}_r) \qquad \text{(By induction hypothesis)}\\ &= \mathbf{D}^{-c}\mathbf{A}_r\mathbf{D}^{-1+c}\mathbf{R}\ \ \mathcal{K}_i(\mathcal{G}_s)\ \ \mathbf{R}^{\mathrm{T}}\\ &= \mathbf{D}^{-c}\mathbf{Q}\mathbf{A}_s\mathbf{Q}^{\mathrm{T}}\mathbf{D}^{-1+c}\mathbf{R}\ \ \mathcal{K}_i(\mathcal{G}_s)\ \ \mathbf{R}^{\mathrm{T}}\\ &= \mathbf{D}^{-c}\mathbf{Q}\mathbf{A}_s\mathbf{Q}^{\mathrm{T}}\mathbf{D}^{-1}(\mathbf{D}^{c}\mathbf{R})\ \ \mathcal{K}_i(\mathcal{G}_s)\ \ \mathbf{R}^{\mathrm{T}}\\ &\overset{6}{=} \mathbf{D}^{-c}\mathbf{Q}\mathbf{A}_s\mathbf{Q}^{\mathrm{T}}\mathbf{D}^{-1}(\mathbf{Q}\mathbf{D}_s^{c})\ \ \mathcal{K}_i(\mathcal{G}_s)\ \ \mathbf{R}^{\mathrm{T}}\end{aligned}$$

(Lemma 2 and 3.)

$$\begin{aligned}&= \mathbf{R}\mathbf{D}_s^{-c}\mathbf{A}_s\mathbf{D}_s^{-1+c}\ \ \mathcal{K}_i(\mathcal{G}_s)\ \ \mathbf{R}^{\mathrm{T}}\\ &= \mathbf{R}\ \ \mathcal{K}_{i+1}(\mathcal{G}_s)\ \ \mathbf{R}^{\mathrm{T}}\end{aligned}$$

Applying principal of induction finishes the proof. ■

B. Further details in proof of Thm. 1 and 2

Proposition 1. $\log(\mathbf{R}\mathbf{M}\mathbf{R}^{\mathrm{T}}) = \mathbf{R}\log(\mathbf{M})\mathbf{R}^{\mathrm{T}}$*, 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:

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

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

■

Proposition 2. $\sigma(\mathbf{R}\ \mathbf{M}) = \mathbf{R}\cdot\sigma(\mathbf{M})$*, 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:

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

■

Appendix B
Details of DPGS algorithm

In GELSumm, we use a graph summarization method DPGS, which minimizes the total description length of both the model cost and configuration-based reconstruction error.

$$L(M,D) = L(M) + L(D \mid M) \tag{23}$$

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

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

The error part is measured by KL divergence:

$$\begin{aligned} L(D \mid M) &= \mathrm{KL}(A \| A_r) \\ &= \sum_{i,j} A(i,j) \ln \frac{A(i,j)}{A_r(i,j)} - A(i,j) + A_r(i,j) \\ &= \sum_{i,j} A(i,j) \ln \frac{A(i,j)}{A_r(i,j)}, \end{aligned} \tag{24}$$

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

For model part, we encode the number of

$$L(M) = \mathrm{L}_{\mathbb{N}}(n_s) + n\,\mathrm{L}_{\mathbb{N}}(n_s) + \sum_{i=1}^{n} \mathrm{L}_{\mathbb{N}}(d_i) + L(A_S)\,, \tag{25}$$

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

In each iteration of the algorithm, nodes pair are sampled and merged. The complete process is described in Algorithm 4.

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