A PROOFS

A.1 DERIVATION OF EQUIVALENT EFFICIENT OPTIMIZATION OBJECTIVE

We first demonstrate that our loss function is an upper bound of the negative log-likelihood.

$$\begin{aligned}
-\log p(\mathcal{G}) &\leq KL(p(\boldsymbol{Z}|\mathcal{G};\theta_{enc})||q(\boldsymbol{Z}|\mathcal{G};\theta_{dec}) - \log p(\mathcal{G}) \\
&= -\mathbb{E}_{p(\boldsymbol{Z}|\mathcal{G};\theta_{enc})} \log q(\boldsymbol{Z}|\mathcal{G};\theta_{dec}) - \log p(\mathcal{G}) + \mathbb{E}_{p(\boldsymbol{Z}|\mathcal{G};\theta_{enc})} \log p(\boldsymbol{Z}|\mathcal{G};\theta_{enc}) \\
&= -\mathbb{E}_{p(\boldsymbol{Z}|\mathcal{G};\theta_{enc})} \log q(\mathcal{G},\boldsymbol{Z};\theta_{dec}) + \mathbb{E}_{p(\boldsymbol{Z}|\mathcal{G};\theta_{enc})} \log p(\boldsymbol{Z}|\mathcal{G};\theta_{enc}) \\
&= -\mathbb{E}_{p(\boldsymbol{Z}|\mathcal{G};\theta_{enc})} \log q(\mathcal{G}|\boldsymbol{Z};\theta_{dec}) - \mathbb{E}_{p(\boldsymbol{Z}|\mathcal{G};\theta_{enc})} \log p_\theta(\boldsymbol{Z}) + \mathbb{E}_{p(\boldsymbol{Z}|\mathcal{G};\theta_{enc})} \log p(\boldsymbol{Z}|\mathcal{G};\theta_{enc}) \\
&= -\mathbb{E}_{p(\boldsymbol{Z}|\mathcal{G};\theta_{enc})} \log q(\mathcal{G}|\boldsymbol{Z};\theta_{dec}) + D_{\mathrm{KL}}(p(\boldsymbol{Z}|\mathcal{G})||p_\theta(\boldsymbol{Z})) \\
&= \mathcal{L}_{rec} + \mathcal{L}_{diff}
\end{aligned} \tag{18}$$

The second term in $\mathcal{L}_{diff}$ is the negative entropy of the latent variable and has a simply expression w.r.t. the posterior variance:

$$\mathbb{E}_{p(\boldsymbol{Z}|\mathcal{G};\theta_{enc})} \log p(\boldsymbol{Z}|\mathcal{G};\theta_{enc}) = -H(\boldsymbol{\mu}(\mathcal{G};\theta_{enc}) + \boldsymbol{\Sigma}^{\frac{1}{2}}(\mathcal{G};\theta_{enc})\mathcal{E}), \tag{19}$$

with $\mathcal{E} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_{NF\times NF})$. For simplicity, we drop $\mathcal{G}, \theta_{enc}$ and omit the constant that is irrelevant of the model, we have

$$\begin{aligned}
H(\boldsymbol{\mu} + \boldsymbol{\Sigma}\mathcal{E}) &= \frac{\log|\boldsymbol{\Sigma}|}{2} - \frac{1}{2}\mathbb{E}_{\boldsymbol{x}\sim\mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) \\
&= \frac{\log|\boldsymbol{\Sigma}|}{2} - \frac{1}{2}\mathbb{E}_{\boldsymbol{x}\sim\mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})} Tr((\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})) \\
&= \frac{\log|\boldsymbol{\Sigma}|}{2} - \frac{1}{2}\mathbb{E}_{\boldsymbol{x}\sim\mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})} Tr(\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^T) \\
&= \frac{\log|\boldsymbol{\Sigma}|}{2} - \frac{1}{2} Tr(\mathbb{E}_{\boldsymbol{x}\sim\mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^T) \\
&= \frac{\log|\boldsymbol{\Sigma}|}{2}
\end{aligned} \tag{20}$$

Using Eq.9 as the first term in $\mathcal{L}_{diff}$, thus one can efficiently estimate $\mathcal{L}_{diff}$.

A.2 INTRACTABILITY OF CLASSICAL SCORE-MATCHING OBJECTIVES

We demonstrates the intractability of classical score-matching objectives. We can not directly optimize the classic score-matching objectives, due to the intractability of SGM prior $p_\theta(\boldsymbol{Z})$. The explicit score-matching objectives has the form:

$$\mathcal{L}_{ESM} = \mathbb{E}_{t\sim\mathcal{U}(0,T)}\left[\mathbb{E}_{q(\boldsymbol{Z}_t)}[||s_\theta(\boldsymbol{Z}_t,t) - \nabla\log q_t(\boldsymbol{Z}_t)||_2^2]\right], \tag{21}$$

with the estimation $s_\theta(\boldsymbol{Z}_t, t)$. But $\nabla\log q_t(\boldsymbol{Z}_t)$ depends on $p_\theta(\boldsymbol{Z}_0)$ and thus intractable:

$$q_\theta(\boldsymbol{Z}_t) = \int q(\boldsymbol{Z}_t|\boldsymbol{Z}_0)q_\theta(\boldsymbol{Z}_0)d\boldsymbol{Z}_0. \tag{22}$$

One can develop the connection between explicit score-matching objective with the denosing score-matching objective which is the standard objective in SGMs, and we provide a detailed derivation here:

$$\begin{aligned}
\mathcal{L}_{ESM} &= \mathbb{E}_{t\sim\mathcal{U}(0,T)}\left[\mathbb{E}_{q(\boldsymbol{Z}_t)}||s_\theta(\boldsymbol{Z}_t,t)||_2^2 - 2\mathbb{E}_{q(\boldsymbol{Z}_t)}\langle s_\theta(\boldsymbol{Z}_t,t), \nabla\log q_t(\boldsymbol{Z}_t)\rangle\right] + C_1 \\
&= \mathbb{E}_{t\sim\mathcal{U}(0,T)}\left[||s_\theta(\boldsymbol{Z}_t,t)||_{L_2}^2 - 2\mathbb{E}_{q(\boldsymbol{Z}_t)}\langle s_\theta(\boldsymbol{Z}_t,t), \frac{\nabla q_t(\boldsymbol{Z}_t)}{q_t(\boldsymbol{Z}_t)}\rangle\right] + C_1,
\end{aligned} \tag{23}$$

where

$$C_1 = \mathbb{E}_{t\sim\mathcal{U}(0,T)}\mathbb{E}_{q(\boldsymbol{Z}_t)}||\frac{\nabla q_t(\boldsymbol{Z}_t)}{q_t(\boldsymbol{Z}_t)}||_2^2 \tag{24}$$

The latent space is assumed to be continuous with positive density function, and thus:

$$\begin{aligned}
\mathbb{E}_{q(\boldsymbol{Z}_t)}\langle s_\theta(\boldsymbol{Z}_t,t), \frac{\nabla_{\boldsymbol{Z}_t}\cdot q_t(\boldsymbol{Z}_t)}{q_t(\boldsymbol{Z}_t)}\rangle &= \int \langle s_\theta(\boldsymbol{Z}_t,t), \nabla_{\boldsymbol{Z}_t}\cdot q_t(\boldsymbol{Z}_t)\rangle d\boldsymbol{Z}_t \\
&= \int \langle s_\theta(\boldsymbol{Z}_t,t), \nabla_{\boldsymbol{Z}_t}\cdot \int q_t(\boldsymbol{Z}_t|\boldsymbol{Z}_0)q(\boldsymbol{Z}_0)d\boldsymbol{Z}_0\rangle d\boldsymbol{Z}_t \\
&= \int \langle s_\theta(\boldsymbol{Z}_t,t), \int \nabla_{\boldsymbol{Z}_t}\cdot q_t(\boldsymbol{Z}_t|\boldsymbol{Z}_0)q(\boldsymbol{Z}_0)d\boldsymbol{Z}_0\rangle d\boldsymbol{Z}_t \\
&= \int\int \langle s_\theta(\boldsymbol{Z}_t,t), \nabla_{\boldsymbol{Z}_t}\cdot q_t(\boldsymbol{Z}_t|\boldsymbol{Z}_0)q(\boldsymbol{Z}_0)\rangle d\boldsymbol{Z}_0 d\boldsymbol{Z}_t \\
&= \int\int \langle s_\theta(\boldsymbol{Z}_t,t), q(\boldsymbol{Z}_0)q(\boldsymbol{Z}_t|\boldsymbol{Z}_0)\nabla_{\boldsymbol{Z}_t}\cdot \log q_t(\boldsymbol{Z}_t|\boldsymbol{Z}_0)\rangle d\boldsymbol{Z}_0 d\boldsymbol{Z}_t \\
&= \int\int q(\boldsymbol{Z}_0)q(\boldsymbol{Z}_t|\boldsymbol{Z}_0)\langle s_\theta(\boldsymbol{Z}_t,t), \nabla_{\boldsymbol{Z}_t}\cdot \log q_t(\boldsymbol{Z}_t|\boldsymbol{Z}_0)\rangle d\boldsymbol{Z}_t d\boldsymbol{Z}_0 \\
&= \mathbb{E}_{q(\boldsymbol{Z}_0)}\mathbb{E}_{q(\boldsymbol{Z}_t|\boldsymbol{Z}_0)}\langle s_\theta(\boldsymbol{Z}_t,t), \nabla_{\boldsymbol{Z}_t}\cdot \log q_t(\boldsymbol{Z}_t|\boldsymbol{Z}_0)\rangle.
\end{aligned} \tag{25}$$

Plugging the Eq.25 into $\mathcal{L}_{ESM}$, we have:

$$\begin{aligned}
\mathcal{L}_{ESM} &= \mathbb{E}_{t\sim\mathcal{U}(0,T)}\left[||s_\theta(\boldsymbol{Z}_t,t)||^2_{L_2} - 2\mathbb{E}_{q(\boldsymbol{Z}_0)}\mathbb{E}_{q(\boldsymbol{Z}_t|\boldsymbol{Z}_0)}\langle s_\theta(\boldsymbol{Z}_t,t), \nabla_{\boldsymbol{Z}_t}\cdot \log q_t(\boldsymbol{Z}_t|\boldsymbol{Z}_0)\rangle\right] + C_1 \\
&= \mathbb{E}_{t\sim\mathcal{U}(0,T)}\left[\mathbb{E}_{q(\boldsymbol{Z}_t,\boldsymbol{Z}_0)}||s_\theta(\boldsymbol{Z}_t,t) - \nabla_{\boldsymbol{Z}_t}\cdot \log q_t(\boldsymbol{Z}_t|\boldsymbol{Z}_0)||^2_2\right] + C_2 \\
&= \mathcal{L}_{DSM} + C_2
\end{aligned} \tag{26}$$

As a result, we can calculate the difference between explicit score-matching and denosing score-matching objectives.

$$C_2 = \mathbb{E}||\nabla \log q_t(\boldsymbol{Z}_t)||^2_2 - \mathbb{E}||\nabla_{\boldsymbol{Z}_t}\cdot \log q_t(\boldsymbol{Z}_t|\boldsymbol{Z}_0)||^2_2 \tag{27}$$

If one directly diffuse the observed data with prefixed diffusion coefficient, then C_2 is indeed a constant. But if the diffusion is performed on the learned latent space with prior $p_\theta(\boldsymbol{Z}_0)$, then C_2 will depend on the learnable parameter θ, even if the forward diffusion is fixed:

$$q_\theta(\boldsymbol{Z}_t) = \int q(\boldsymbol{Z}_t|\boldsymbol{Z}_0)q_\theta(\boldsymbol{Z}_0)d\boldsymbol{Z}_0 \tag{28}$$

and thus the denosing score matching is also intractable. But the objectives we use, as Eq.9, do not include the term $q_\theta(\boldsymbol{Z}_t)$ and thus tractable.

A.3 Detailed Explanation of Self-Guidance

Here we present a continuous-time setting of self-guidance mechanism: After embedding, we have latent $\boldsymbol{Z}_0$ with pseudo label $\boldsymbol{c}$, with joint distribution $p(\boldsymbol{Z},\boldsymbol{c})$. In the forward SDE, only the latent $\boldsymbol{Z}$ is diffused:

$$d\boldsymbol{Z} = f(t)\boldsymbol{Z}dt + g(t)d\mathbf{w}, \quad \boldsymbol{Z}\sim p(\boldsymbol{Z}|\boldsymbol{c}) \tag{29}$$

By designing the drift coefficient $f(t)$, the resulting noise $\boldsymbol{Z}_T$ is determined entirely by a independent Brownian motion and forget the initial distribution $\boldsymbol{Z}_0$. For example, we can design:

$$\int_0^T f(s)ds = \infty \tag{30}$$

So the reverse process start with a known prior with label $\boldsymbol{c}$. To reverse the process, we need the reverse SDE with the following form:

$$d\hat{\boldsymbol{Z}} = (f(t)\hat{\boldsymbol{Z}} - g(t)^2\nabla_{\hat{\boldsymbol{Z}}}\log q_t(\hat{\boldsymbol{Z}}|\boldsymbol{c}))dt + g(t)d\bar{\mathbf{w}}. \tag{31}$$

In other word, we need to estimate the conditional score function $\nabla_{\hat{\boldsymbol{Z}}}\log q_t(\hat{\boldsymbol{Z}}|\boldsymbol{c})$ to generate sample with label $\boldsymbol{c}$. Using probability chain rule, one can decompose the conditional score function as:

$$\nabla_{\hat{\boldsymbol{Z}}}\log q_t(\hat{\boldsymbol{Z}}|\boldsymbol{c}) = \nabla_{\hat{\boldsymbol{Z}}}\log q_t(\boldsymbol{c}|\hat{\boldsymbol{Z}}) - \nabla_{\hat{\boldsymbol{Z}}}\log q_t(\hat{\boldsymbol{Z}}) \tag{32}$$

and estimate each part. Real-world graph sets usually do not have explicit label, and thus our model leverages pseudo label to improve the sample quality and cover the whole distribution of target unlabeled graph set by performing self-guidance, as described in section 3.2.

B ADDITIONAL DETAILS AND EXPERIMENTS

B.1 IMPLEMENTATION DETAILS AND HYPERPARAMETERS

Implementation Details of Generic Graph Generation In generic graph generation tasks, we follow the standard setting of existing works (You et al., 2018; Liu et al., 2019; Niu et al., 2020b; Jo et al., 2022). Specifically, we report the result means of 30 runs, 3 different runs for 10 independently trained models on Ego-small and Community-small datasets. And due to the costly training procedure of GraphVAE and EDP-GNN, we report the result means of 3 different runs on Enzymes and Grid datasets. As for the baseline methods implementation, we follow the hyperparameter settings that provided by original works. For SGGM with SLD, we first initilize the node features with the one-hot embedding of the degrees, and then pretrain a variational graph autoencoder. The pretrained autoencoder will initialize a hierarchical pseudo label set by iteratively clustering on all the latent embeddings of the training dataset. The number of classes and hierarchies are detailed in Tab.4 and the hierarchical psuedo label set will be further utilized in SGGM for self-guided diffusion generation with periodic update. We perform the grid search to choose the proper numbers for the hierarchies and classes. After we train SGGM with SLD, we the best MMD with the lowest average of three graph statistics, degree, clustering coefficient, and orbit. For the datasets with small graphs, we use two-layer graph encoder and decoder and a two-layer convolution for score matching while they are of more layers for modeling large graphs, please refer to Tab.4 for further details.

Implementation Details of Molecular Graph Generation In the experiments of molecular graph generation, each molecule is represented by a graph with the node features $\boldsymbol{X} \in \{0,1\}^{N\times F}$ and the adjacency matrix $\boldsymbol{A} \in \{0,1,2,3\}^{N\times N}$, where N denotes the number of atoms in the molecule, and F denotes the number of atom types. The entries of $\boldsymbol{A}$ represent the bond types, i.e. single, double, or triple bonds. Following the standard process of existing works (Shi et al., 2020; Luo et al., 2021b; Jo et al., 2022), the molecules are kekulized by the RDKit library Landrum et al. (2016) and hydrogen atoms are deleted. The whole training pipeline of molecular graph generation is similar to that of generic graph generation, which has been introduced in last paragraph. For further detailed hyperpameter settings, please refer to Tab.4. The novelty value can influence the FCD and NSPDK MMD values. With respect to the evaluation, we choose the hyperparameters that exhibit the best FCD value among those which show the novelty that exceeds 85%. As demonstrated in section 4, we conduct the experiments for extensibility analysis. The geometrical and torsional diffusion processes that used in experiments are implemented according to the published code of their original works (Xu et al., 2021b; Jing et al., 2022).

Table 4: We provide hyperparameters of SGGM with SLD that used in the generic graph generation tasks and the molecule generation tasks.

	Hyperparameter	Ego-small	Community-small	Enzymes	Grid	QM9	ZINC250k
s_θ	Number of convolutional layers	2	3	5	5	2	2
	Hidden dimension	32	32	32	32	16	16
f_{enc}	Number of graph layers	2	2	3	3	2	2
	Hidden dimension	32	32	32	32	16	16
f_{dec}	Number of graph layers	2	2	3	3	2	2
	Hidden dimension	32	32	32	32	16	16
Hierachical label	Number of hierarchies	2	2	3	3	3	3
	Number of update epochs	50	50	50	50	5	5
	Number of classes	{6, 2}	{6, 2}	{12, 4, 2}	{12, 4, 2}	{12, 4, 2}	{12, 4, 2}
Loss Objective	Weight α	1	1	1	1	1	1
	Weight β	0.5	0.5	0.5	0.5	0.5	0.5
Train	Optimizer	Adam	Adam	Adam	Adam	Adam	Adam
	Learning rate	1×10^{-2}	1×10^{-2}	1×10^{-2}	1×10^{-2}	5×10^{-3}	5×10^{-3}
	Weight decay	1×10^{-4}	1×10^{-4}	1×10^{-4}	1×10^{-4}	1×10^{-4}	1×10^{-4}
	Batch size	128	128	64	8	1024	1024
	Number of epochs	5000	5000	5000	5000	300	500
	Number of sampling steps	600	600	600	600	600	600

B.2 ADDITIONAL EXPERIMENTAL RESULTS

In this subsection, we provide additional experimental results for illustration: generic and molecular graph generation results with the standard deviation; visualization of small, middle, and large graphs generation results.

Table 5: We report the MMD distance between the test datasets and generated graphs with the standard deviation on the Ego-small and the Community-small datasets.

	Ego-small			Community-small		
	Real, $4 \leq \|V\| \leq 18$			Synthetic, $12 \leq \|V\| \leq 20$		
	Deg.	Clus.	Orbit	Deg.	Clus.	Orbit
SGGM	0.025 ± 0.005	0.028 ± 0.004	0.009 ± 0.005	0.041 ± 0.015	0.079 ± 0.134	0.010 ± 0.003
SGGM+SLD	$\mathbf{0.014} \pm 0.006$	$\mathbf{0.019} \pm 0.006$	0.007 ± 0.004	$\mathbf{0.035} \pm 0.017$	$\mathbf{0.071} \pm 0.015$	$\mathbf{0.006} \pm 0.003$

Table 6: We report the MMD distance between the test datasets and generated graphs with the standard deviation on the Enzymes and the Grid datasets.

	Enzymes			Grid		
	Real, $10 \leq \|V\| \leq 125$			Synthetic, $100 \leq \|V\| \leq 400$		
	Deg.	Clus.	Orbit	Deg.	Clus.	Orbit
GraphRNN	$\mathbf{0.017} \pm 0.007$	0.062 ± 0.020	0.046 ± 0.031	$\mathbf{0.064} \pm 0.017$	0.043 ± 0.022	$\mathbf{0.021} \pm 0.007$
GraphAF	1.669 ± 0.024	1.283 ± 0.019	0.266 ± 0.007	-	-	-
GraphDF	1.503 ± 0.011	1.061 ± 0.011	0.202 ± 0.002	-	-	-
GraphVAE	1.369 ± 0.020	0.629 ± 0.005	0.191 ± 0.020	1.619 ± 0.007	0.0 ± 0.000	0.919 ± 0.002
EDP-GNN	0.023 ± 0.012	0.268 ± 0.164	0.082 ± 0.078	0.455 ± 0.319	0.238 ± 0.380	0.328 ± 0.278
GDSS	0.026 ± 0.008	$\mathbf{0.061} \pm 0.010$	0.009 ± 0.005	0.111 ± 0.012	0.005 ± 0.000	0.070 ± 0.044
SGGM	0.030 ± 0.005	0.073 ± 0.008	0.013 ± 0.005	0.114 ± 0.010	$\mathbf{0.0} \pm 0.0$	0.065 ± 0.024
SGGM+SLD	0.022 ± 0.004	$\mathbf{0.062} \pm 0.006$	$\mathbf{0.007} \pm 0.002$	0.103 ± 0.006	$\mathbf{0.0} \pm 0.0$	0.053 ± 0.014

Table 7: Graph generation results are the means and the standard deviations of 3 runs on the QM9 dataset. Validity is the fraction of the generated molecules that do not violate the chemical valency rule. Uniqueness is the fraction of the valid molecules that are unique. Novelty is the fraction of the valid molecules that are not included in the training set.

Method	Validity w/o correction (%) ↑	NSPDK MMD ↓	FCD ↓	Validity (%) ↑	Uniqueness (%) ↑	Novelty (%) ↑
GraphAF	67	0.020±0.003	5.268±0.403	**100.00**	94.51	88.83
GraphDF	82.67	0.063±0.001	10.816±0.020	**100.00**	97.62	**98.10**
MoFlow	91.36±1.23	0.017±0.003	4.467±0.595	**100.00**±0.00	98.65±0.57	94.72±0.77
EDP-GNN	47.52±3.60	0.005±0.001	2.680±0.221	**100.00**±0.00	99.25±0.05	86.58±1.85
GraphEBM	8.22±2.24	0.030±0.004	6.143±0.411	**100.00**±0.00	97.90±0.14	97.01±0.17
GDSS	95.72±1.94	**0.003**±0.000	2.900±0.282	**100.00**±0.00	98.46±0.61	86.27±2.29
SGGM	95.91±1.73	0.006±0.001	2.745±0.264	**100.00**±0.00	98.52±0.15	96.61±1.75
SGGM+SLD	**97.35**±1.21	**0.004**±0.000	**2.593**±0.191	**100.00**±0.00	**99.41**±0.11	97.49±1.32

Table 8: Graph generation results are the means and the standard deviations of 3 runs on the ZINC250k dataset. Validity is the fraction of the generated molecules that do not violate the chemical valency rule. Uniqueness is the fraction of the valid molecules that are unique. Novelty is the fraction of the valid molecules that are not included in the training set.

Method	Validity w/o correction (%) ↑	NSPDK MMD ↓	FCD ↓	Validity (%) ↑	Uniqueness (%) ↑	Novelty (%) ↑
GraphAF	68	0.044±0.006	16.289±0.482	**100.00**	99.10	**100.00**
GraphAF+FC	68.47±0.99	0.044±0.005	16.023±0.451	**100.00**±0.00	98.64±0.69	99.99±0.01
GraphDF	89.03	0.176±0.001	34.202±0.160	**100.00**	99.16	**100.00**
GraphDF+FC	90.61±4.30	0.177±0.001	33.546±0.150	**100.00**±0.00	99.63±0.01	**100.00**±0.00
MoFlow	63.11±5.17	0.046±0.002	20.931±0.184	**100.00**±0.00	**99.99**±0.01	**100.00**±0.00
EDP-GNN	82.97±2.73	0.049±0.006	16.737±1.300	**100.00**±0.00	99.79±0.08	**100.00**±0.00
GraphEBM	5.29±3.83	0.212±0.075	35.471±5.331	99.96±0.02	98.79±0.15	**100.00**±0.00
GDSS	97.01±0.77	0.019±0.001	14.656±0.680	**100.00**±0.00	99.64±0.13	**100.00**±0.00
SGGM	97.28±0.78	0.018±0.001	13.931±0.625	**100.00**±0.00	98.87±0.14	**100.00**±0.00
SGGM+SLD	**98.32**±0.71	**0.014**±0.001	**11.379**±0.597	**100.00**±0.00	99.83±0.11	**100.00**±0.00

C GENERATION OF SMALL, MIDDLE AND LARGE GRAPHS

We visualize the graphs of small, middle, and large scales from the training datasets and the generated graphs of SGGM with SLD for each dataset in Fig.3-5. We randomly choose samples from the training datasets and the generated graph set for visualization, with e denotes edges number and n denotes nodes number.

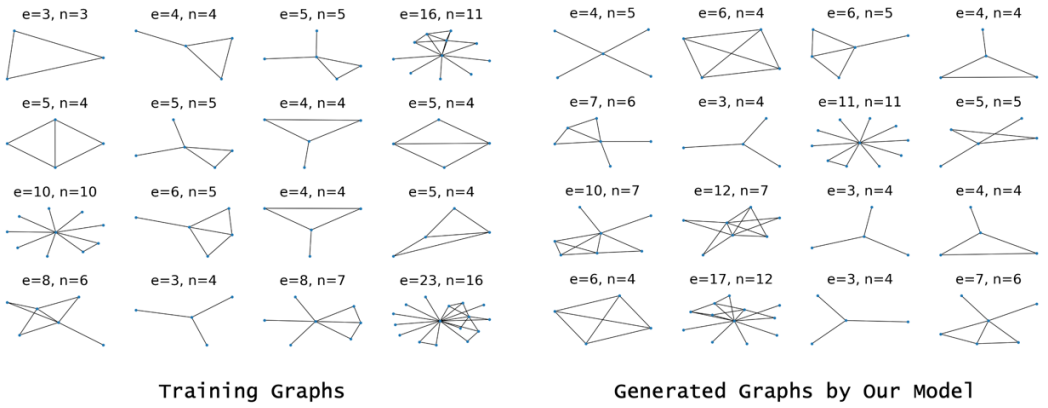

Figure 3: Visualization of the graphs sampled from the Ego-small dataset and the generated graphs of SGGM with SLD.

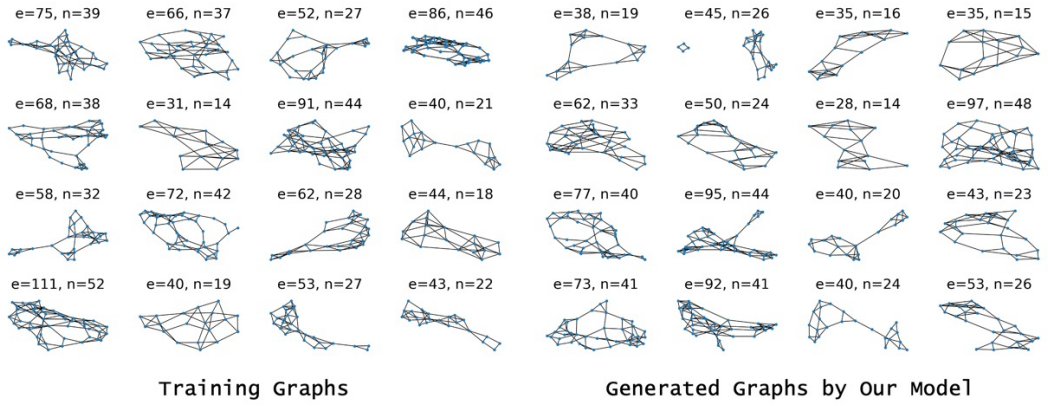

Figure 4: Visualization of the graphs sampled from the ENZYMES dataset and the generated graphs of SGGM with SLD.

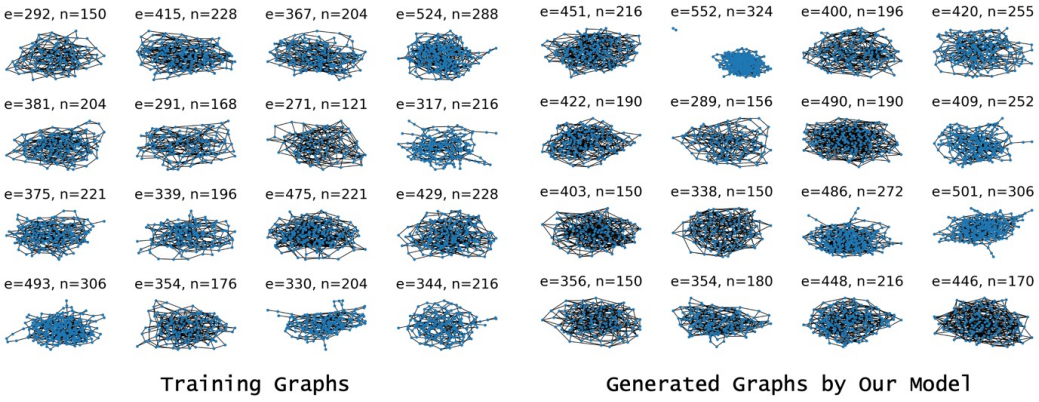

Figure 5: Visualization of the graphs sampled from the Grid dataset and the generated graphs of SGGM with SLD.