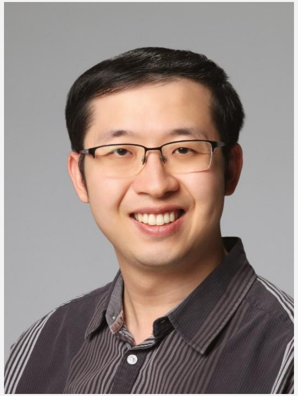
Micro and Macro Level Graph Modeling for Graph Variational Auto-Encoders

Kiarash Zahirnia, Oliver Schulte, Parmis Naddaf, Ke Li

School of Computing Science, Simon Fraser University, Vancouver, Canada

NeurIPS 2022



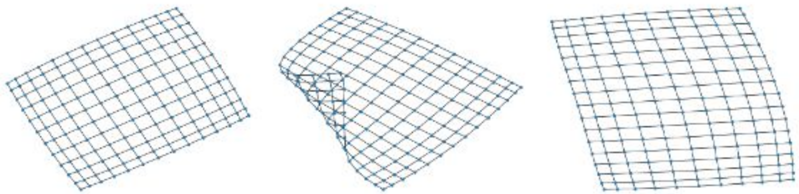
Graph

- $\mathcal{G}$ = (**V**, **E**) is a pair comprising a finite set of |**V**|=N nodes and |**E**| edges.
- A graph can be represented by an adjacency matrix **A**.

Problem Definition

- Given a set of observed graphs **G** = $\{\mathcal{G}_1, \ldots \mathcal{G}_S\}$ sampled from data distribution p(G), the goal of learning generative models for graphs is to learn the distribution of p_θ (G) which is similar to p(G).
- The focus of this paper is on models for generating "realistic-looking" graphs.

Samples from Protein dataset (real data).

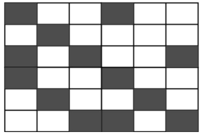


Samples from Grid dataset (synthetic benchmark).

Deep Graph Generative Models (GGMs)

1) All-at-once Models

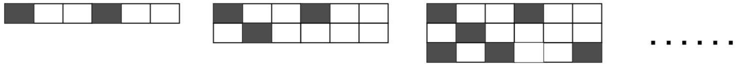
Generate a graph, adjacency matrix, in one-shot.

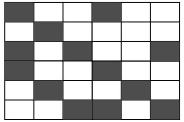


VGAE(Kipf et al 2018)
MolGAN (Cao et al 2018)
GraphVAE (Dai et al 2018)

2) Autoregressive Models

Generate a graph sequentially, an edge, node, or block at a time.



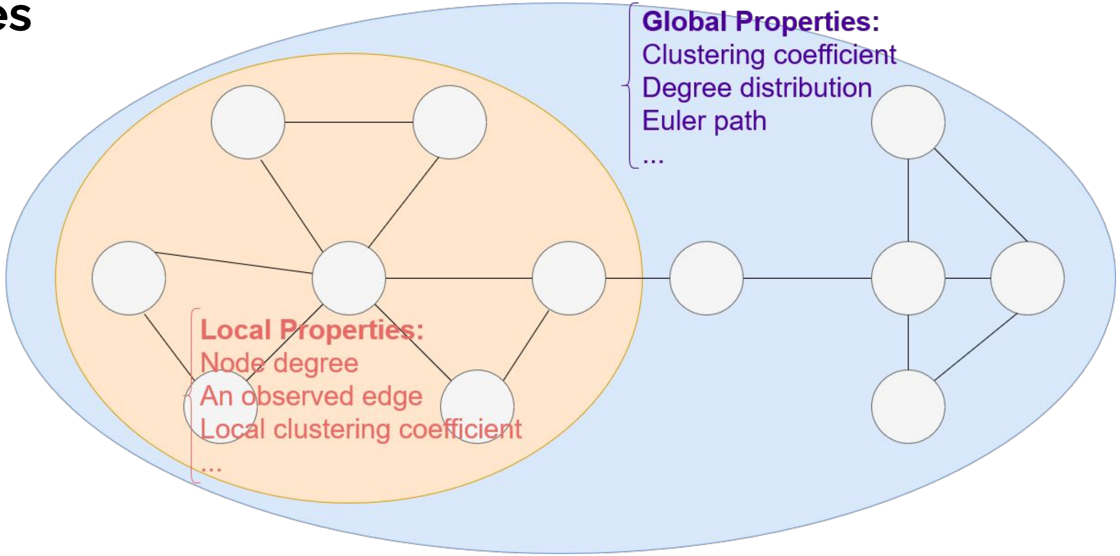


GraphRNN (You et al 2018)
GRAN (Liao et al 2019)
BiGG (Dai et al 2020)

- All-at-once models have fast and tractable sampling and relatively stable training.
- Sequential graph generation allows autoregressive models to capture complex dependencies between new edges/nodes and edges/nodes already generated.

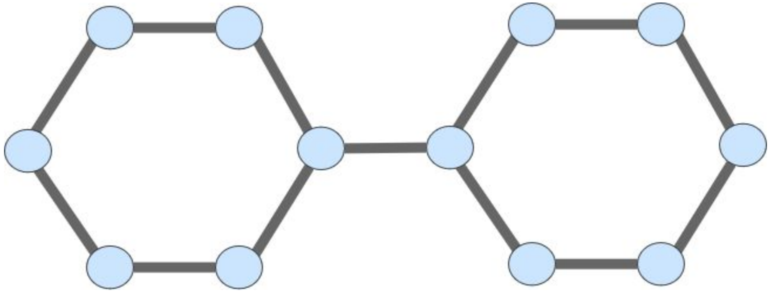
Global and Local Graph Properties

- Two levels of information:
  1) Local node-level properties
  2) Global graph-level properties



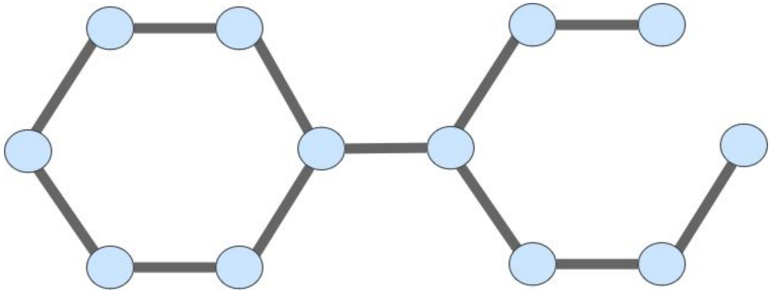
- Most deep GGMs are trained with an objective based on local properties.
  - Local properties does not model different edge roles in the graph global structure.

Global and Local Graph Properties (example)

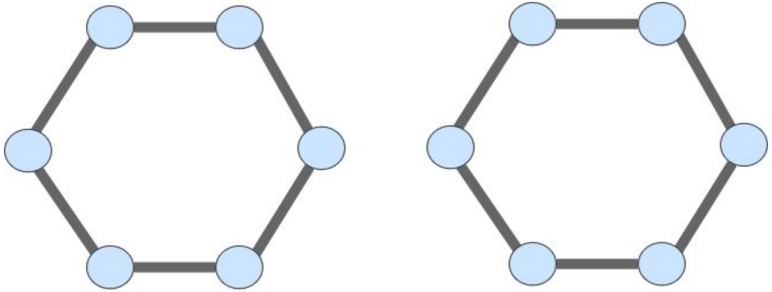


Original Graph

The two right graphs score the same in terms of number of reconstructed edges, however the **Graph 1**, is structurally more similar to the **Original Graph.**

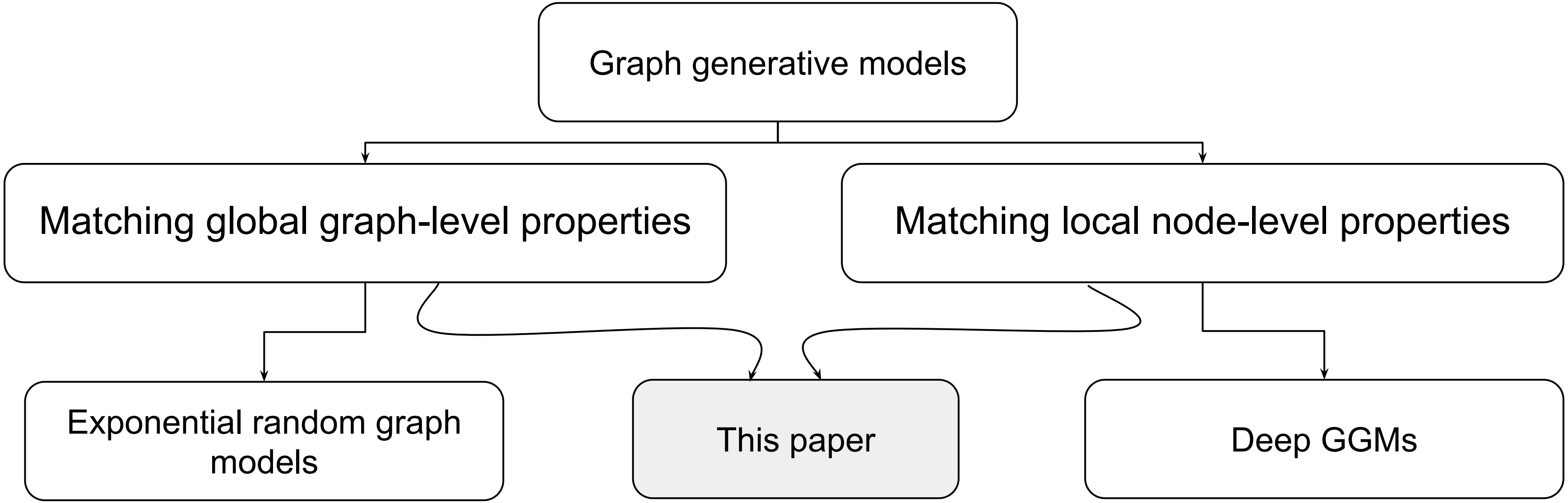


Generated Graph 1



Generated Graph 2

Learning objectives



Approach

- **Micro-macro (MM) Modeling:**
  - A principled probabilistic framework that incorporates both local (Micro) and global (Macro) graph properties.

Approach

- **Micro-macro (MM) Modeling:**
  - A principled probabilistic framework that incorporates both local (Micro) and global (Macro) graph properties.
  - Assuming a predefined finite set of graph global statistics/properties, calculated by $\varphi_1()$, …, $\varphi_m()$ micro-macro loss is of the form:

$$\mathcal{L}_{\boldsymbol{\theta}}(\boldsymbol{A}) = \mathcal{L}^0_{\boldsymbol{\theta}}(\boldsymbol{A}) + \gamma \mathcal{L}^1_{\boldsymbol{\theta}}(\boldsymbol{F}_1, \ldots, \boldsymbol{F}_m)$$

L^0: micro loss.
L^1: macro loss.
A: training graph.
m: number of global properties.
$\mathbf{F}_u$: random variable defined by $\varphi_u(\hat{\mathbf{A}})$.
γ: hyperparameter.

Approach

- **Micro-macro (MM) Modeling:**
  - A principled probabilistic framework that incorporates both local (Micro) and global (Macro) graph properties.
  - Assuming a predefined finite set of graph global statistics/properties, calculated by $\varphi_1()$, …, $\varphi_m()$ micro-macro loss is of the form:

$$\mathcal{L}_{\theta}(A) = \mathcal{L}^0_{\theta}(A) + \gamma \mathcal{L}^1_{\theta}(F_1, \ldots, F_m)$$

- **Advantages**:
  - Realism: Compared to objective functions that are based on predicting local properties, matching graph statistics serves as a regularizer that increases the realism of the generated graph structures
  - User control: the user only needs to specify the target graph statistics and learning will automatically select graph models that match them.

L^0: micro loss.
L^1: macro loss.
A: training graph.
m: number of global properties.
F$_u$: random variable defined by $\varphi_u(\hat{\mathbf{A}})$.
γ: hyperparameter.

GraphVAE-MM

- This paper works with negative log-likelihood losses:

$$\mathcal{L}^0_{\boldsymbol{\psi}}(\boldsymbol{A}) = -\ln p^0_{\boldsymbol{\psi}}(\boldsymbol{A}) = -\ln \int P(\boldsymbol{A}|\tilde{\boldsymbol{A}}_{\boldsymbol{z}})p(\boldsymbol{z})d\boldsymbol{z}$$

$$\mathcal{L}^1_{\boldsymbol{\psi},\boldsymbol{\sigma}}(\boldsymbol{F}_1,\ldots,\boldsymbol{F}_m) = -\sum_{u=1}^{m} \frac{1}{|\boldsymbol{F}_u|} \ln p^1_{\boldsymbol{\psi},\boldsymbol{\sigma}}(\boldsymbol{F}_u)$$

L^0: micro loss.
L^1: macro loss.
A: training graph.
Ã: underlying probabilistic adjacency matrix.
$\mathbf{F}_u$: random variable defined by $\varphi_u(\hat{\mathbf{A}})$.
γ: hyperparameter.
z: graph embedding.
$\tilde{\mathbf{A}}_z$: probabilistic adjacency matrix computed as a function of graph embedding z.
m: number of target statistic.
$|\mathbf{F}_u|$: dimensionality of target statistic $\mathbf{F}_u$.

GraphVAE-MM

- This paper works with negative log-likelihood losses:

$$\mathcal{L}^0_{\psi}(A) = -\ln p^0_{\psi}(A) = -\ln \int P(A|\tilde{A}_z)p(z)dz$$

$$\mathcal{L}^1_{\psi,\sigma}(F_1, \ldots, F_m) = -\sum_{u=1}^{m} \frac{1}{|F_u|} \ln p^1_{\psi,\sigma}(F_u)$$

- By approximating with variational Lower bound we have:

$$\mathcal{L}_{\theta}(A) \leq E_{z \sim q_{\varphi}(z|A)}\left[-\ln p^0_{\psi}(A|\tilde{A}_z) - \sum_{u=1}^{m} \frac{1}{|F_u|} \ln p^1_{\psi,\sigma}(F_u)\right] + (1+\gamma m)KL(q_{\varphi}(z|A)||p(z))$$

L^0: micro loss.
L^1: macro loss.
A: training graph.
Ã: underlying probabilistic adjacency matrix.
$\mathbf{F}_u$: random variable defined by $\varphi_u(\hat{\mathbf{A}})$.
γ: hyperparameter.
z: graph embedding.
$\tilde{\mathbf{A}}_z$: probabilistic adjacency matrix computed as a function of graph embedding z.
m: number of global properties.
$|\mathbf{F}_u|$: dimensionality of target statistic $\mathbf{F}_u$.

Graph Statistics

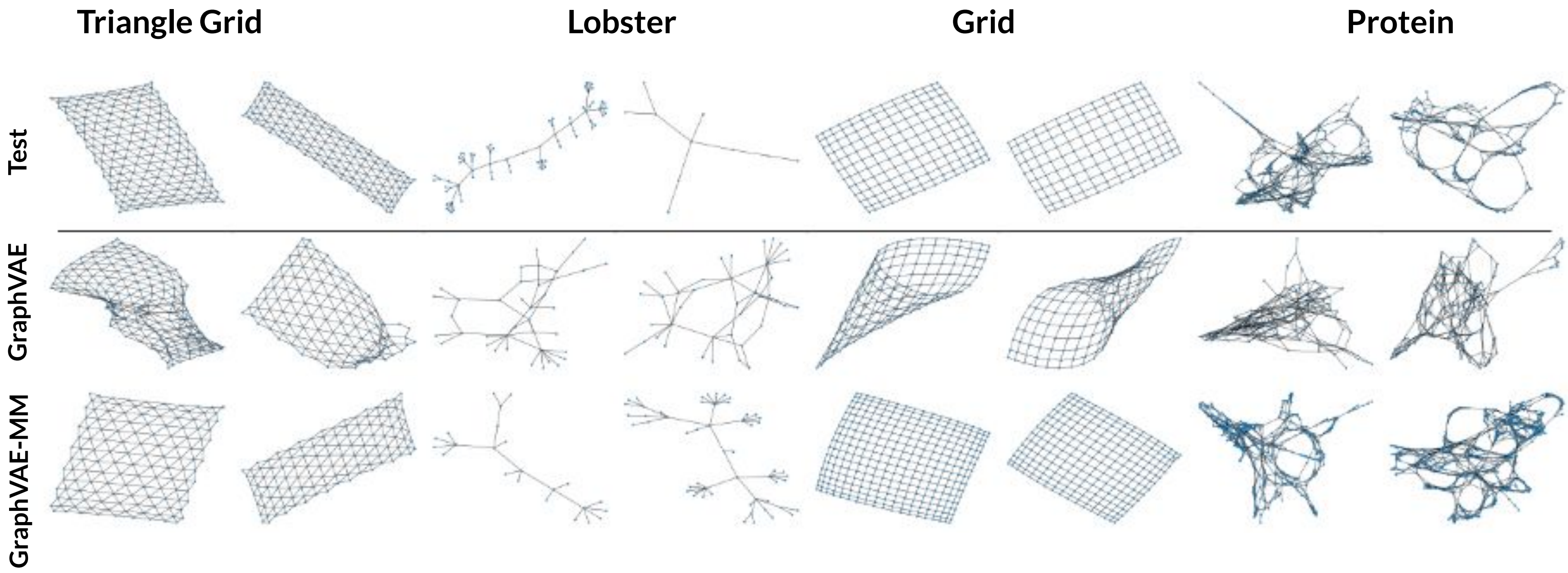
- **GraphVAE-MM:** We utilize the micro-macro objective to improve graph generation with a **GraphVAE** (Dai et al 2018) architecture.

Graph Statistics

- **GraphVAE-MM:** We utilize the micro-macro objective to improve graph generation with a **GraphVAE** (Dai et al 2018) architecture.

- In our experiments, we utilize 3 default graph global properties:
    - Degree histogram
    - Number of triangles
    - S-Step transition probability for S=2,...,5

Qualitative Evaluation



- GraphVAE–MM achieves much better visual match than GraphVAE.

Quantitative Evaluation (GNN-based evaluation metrics Thompson et al 2022)

Method	Triangle Grid		Lobster		Grid		ogbg-molbbbp		Protein	
	MMD RBF	F1 PR	MMD RBF	F1 PR	MMD RBF	F1 PR	MMD RBF	F1 PR	MMD RBF	F1 PR
50/50 split	0.03 ± 0.00	98.58 ± 0.00	0.04 ± 0.00	98.58 ± 0.00	0.009 ± 0.00	98.70 ± 0.00	0.002 ± 0.00	98.07 ± 0.00	0.04 ± 0.00	98.67 ± 1.11
GraphVAE	0.23 ± 0.01	75.92 ± 8.96	0.36 ± 0.11	78.48 ± 24.13	0.17 ± 0.01	75.52 ± 2.53	0.20 ± 0.07	54.53 ± 6.15	0.10 ± 0.05	84.11 ± 9.56
GraphVAE-MM	**0.17 ± 0.01**	**83.58 ± 5.50**	**0.10 ± 0.00**	**100.00 ± 0.00**	**0.13 ± 0.01**	**97.09 ± 6.33**	**0.02 ± 0.01**	93.78 ± 1.33	**0.03 ± 0.01**	90.78 ± 3.76
GraphRNN-S (You et al. 2018)	0.72 ± 0.17	33.68 ± 19.44	0.98 ± 0.13	58.72 ± 7.55	0.79 ± 0.08	71.18 ± 2.36	0.48 ± 0.02	81.41 ± 0.71	0.28 ± 0.26	72.36 ± 27.63
GraphRNN (You et al. 2018)	0.64 ± 0.11	25.80 ± 11.75	0.87 ± 0.04	61.97 ± 0.00	0.99 ± 0.03	13.22 ± 0.05	1.45 ± 0.19	**98.94 ± 0.56**	0.32 ± 0.14	93.94 ± 0.56
GRAN (Liao et al. 2019b)	0.88 ± 0.09	23.71 ± 9.72	0.24 ± 0.04	50.53 ± 12.12	0.40 ± 0.00	78.73 ± 0.02	0.39 ± 0.07	94.06 ± 2.60	0.07 ± 0.00	98.05 ± 0.76
BiGG (Dai et al. 2020)	0.41 ± 0.13	62.08 ± 0.14	0.12 ± 0.00	99.74 ± 0.76	0.35 ± 0.00	92.43 ± 0.00	0.04 ± 0.00	96.16 ± 0.31	0.15 ± 0.00	**98.11 ± 0.62**

- **MMD RBF and F1 PR** capture the reality and diversity of generated graphs, respectively.
- **Impact on GraphVAE.** MM modeling provides a large improvement in the realism and diversity of graphs generated by a GraphVAE architecture.
- **GraphVAE-MM vs. Benchmark GGMs.** Micro-macro (MM) modeling greatly improved the GraphVAE, to match or exceed that of benchmark models.

Quantitative Evaluation (statistic-based evaluation metrics O'Bray et al 2022)

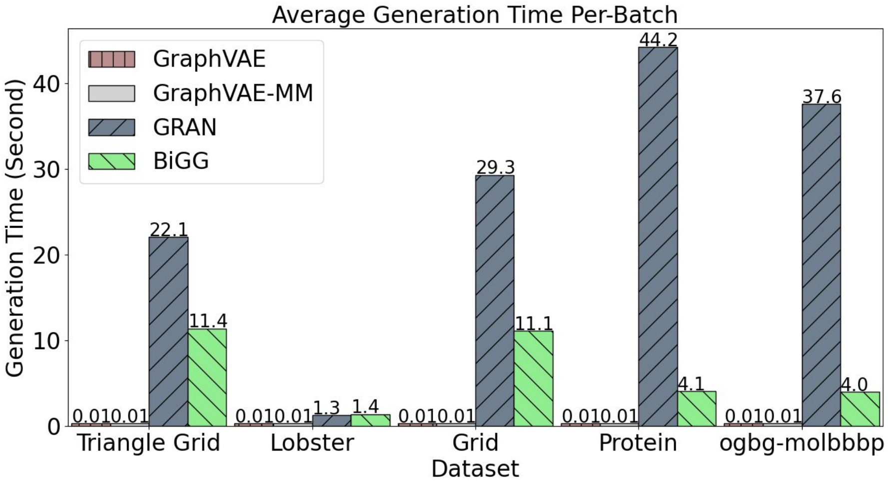
(a) Synthetic Graphs

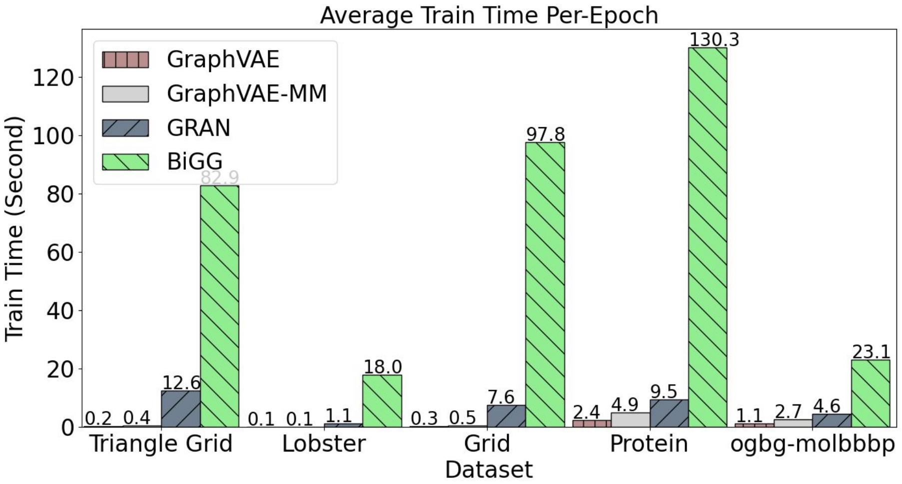
Method	Triangle Grid					Lobster					Grid				
	Deg.	Clus.	Orbit	Spect	Diam.	Deg.	Clus.	Orbit	Spect	Diam.	Deg.	Clus.	Orbit	Spect	Diam.
50/50 split	$3e^{-5}$	0.002	$8e^{-5}$	0.004	0.014	0.002	0	0.002	0.005	0.032	$1e^{-5}$	0	$2e^{-5}$	0.004	0.014
GraphVAE	0.0821	0.442	0.421	0.020	0.152	0.081	0.739	0.372	0.056	0.129	0.062	0.055	0.515	0.018	0.143
GraphVAE-MM	**0.001**	**0.093**	**0.001**	**0.013**	**0.133**	**$2e^{-4}$**	**0**	0.008	0.017	0.187	**$5e^{-4}$**	**0**	**0.001**	0.014	**0.065**
GraphRNN-S (You et al. [47])	0.053	1.094	0.121	0.033	1.124	0.016	0.319	0.285	0.045	0.242	0.014	0.003	0.090	0.112	0.128
GraphRNN (You et al. [47])	0.033	1.167	0.107	0.030	1.121	0.004	**0**	0.033	0.035	0.384	0.013	0.166	0.019	0.111	0.460
GRAN (Liao et al. [32])	0.134	0.678	0.673	0.184	1.133	0.005	0.304	0.331	0.043	0.446	0.003	$1e^{-4}$	0.007	**0.012**	0.281
BiGG (Dai et al. [11])	**0.001**	0.107	0.004	0.020	1.123	0.001	**0**	**$6e^{-4}$**	**0.012**	**0.101**	0.002	$3e^{-5}$	0.003	0.018	0.328

(b) Real Graphs

Method	Protein					ogbg-molbbbp				
	Deg.	Clus.	Orbit	Spect	Diam.	Deg.	Clus.	Orbit	Spect	Diam.
50/50 split	$4e^{-5}$	0.004	$5e^{-4}$	$4e^{-4}$	0.003	$2e^{-4}$	$2e^{-5}$	$9e^{-5}$	$5e^{-4}$	0.002
GraphVAE	0.022	0.108	0.577	0.016	0.080	0.028	0.442	0.047	0.015	0.055
GraphVAE-MM	0.006	**0.059**	0.152	0.007	0.091	**0.001**	0.005	$8e^{-4}$	**0.005**	**0.018**
GraphRNN-S (You et al. [47])	0.046	0.324	0.316	0.028	0.302	0.016	0.572	0.006	0.045	0.199
GraphRNN ((You et al. [47])	0.012	0.123	0.264	0.018	0.342	0.002	**$9e^{-4}$**	$4e^{-4}$	0.135	0.495
GRAN (Liao et al. [32])	**0.003**	**0.059**	**0.053**	**0.004**	**0.009**	0.008	0.353	0.013	0.056	0.317
BiGG (Dai et al. [11])	0.007	0.099	0.316	0.012	0.181	0.003	0.001	**$5e^{-5}$**	0.007	0.033

Generation and Train Time





- **Generation time.** The autoregressive methods require substantially more generation time.
- **Training time overhead.** The training time is still less than for the autoregressive methods.

Conclusion

- This paper proposes a new multi-level framework that jointly models node-level properties and graph-level properties, as mutually reinforcing sources of information.

- We derive a joint ELBO as a new micro-macro objective function for training graph encoder-decoder models.

- Our experiments show that adding micro-macro modeling to the GraphVAE model improves graph quality scores up to 2 orders of magnitude while maintaining the GraphVAE generation speed advantage.