Multiresolution Equivariant Graph Variational Autoencoder

Truong Son Hy and Risi Kondor

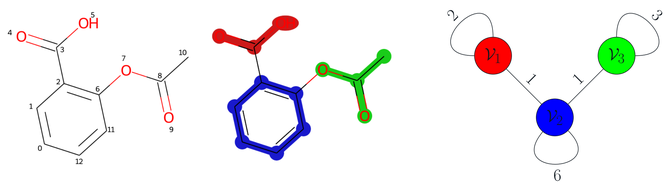
Department of Computer Science, The University of Chicago

Abstract

In this paper, we propose *Multiresolution Equivariant Graph Variational Autoencoders* (MGVAE), the first hierarchical generative model to learn and generate graphs in a multiresolution and equivariant manner. At each resolution level, MGVAE employs higher order message passing to encode the graph while learning to partition it into mutually exclusive clusters and coarsening into a lower resolution that eventually creates a hierarchy of latent distributions. MGVAE then constructs a hierarchical generative model to variationally decode into a hierarchy of coarsened graphs. Importantly, our proposed framework is end-to-end permutation equivariant with respect to node ordering. MGVAE achieves competitive results with several generative tasks including general graph generation, molecular generation, unsupervised molecular representation learning to predict molecular properties, link prediction on citation graphs, and graph-based image generation. Our source code is available at `https://github.com/HyTruongSon/MGVAE`.

Multiresolution Graph Networks

Multiresolution Graph Networks (MGN) constructs multiple resolutions of the input graph via the learning to cluster algorithm in a data-driven manner. The hard clustering can be differentible (for back-propagation) by the Gumbel-softmax trick.



Aspirin $C_9H_8O_4$, its 3-cluster partition and the corresponding coarsen graph.

It is desirable to have a *balanced* K-cluster partition in which clusters $\mathcal{V}_1^{(\ell)}, .., \mathcal{V}_K^{(\ell)}$ (at the ℓ-th resolution level) have similar sizes that are close to $|\mathcal{V}^{(\ell)}|/K$. We enforce the clustering procedure to produce a balanced cut by minimizing the following KL divergence:

$$\mathcal{D}_{KL}(P||Q) = \sum_{k=1}^{K} P(k) \log \frac{P(k)}{Q(k)}, \qquad P = \left(\frac{|\mathcal{V}_1^{(\ell)}|}{|\mathcal{V}^{(\ell)}|}, .., \frac{|\mathcal{V}_K^{(\ell)}|}{|\mathcal{V}^{(\ell)}|} \right), \qquad Q = \left(\frac{1}{K}, .., \frac{1}{K} \right)$$

Multiresolution Generative Models

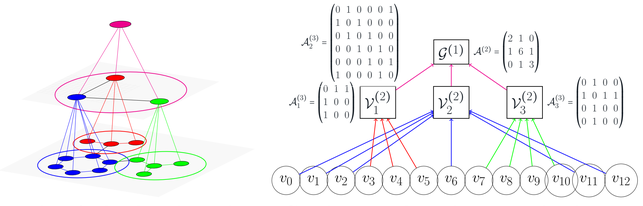
Based on the construction of multiresolution graph network, the latent hierarchy is partitioned into disjoint groups, $\mathcal{Z}_i = \{\mathcal{Z}_i^{(1)}, \mathcal{Z}_i^{(2)}, .., \mathcal{Z}_i^{(L)}\}$ where $\mathcal{Z}_i^{(\ell)}$ is the set of latents at the ℓ-th resolution level. We employ the use of **Hierarchical VAEs**. We write our multiresolution variational lower bound $\mathcal{L}_{\text{MGVAE}}(\phi, \theta)$ on $\log p(\mathcal{G})$ compactly as

$$\mathcal{L}_{\text{MGVAE}}(\phi, \theta) = \sum_i \sum_\ell \left[\mathbb{E}_{q_\phi(\mathcal{Z}_i^{(\ell)}|\mathcal{G}_i^{(\ell)})} [\log p_\theta(\mathcal{G}_i^{(\ell)}|\mathcal{Z}_i^{(\ell)})] - \mathcal{D}_{\text{KL}}(q_\phi(\mathcal{Z}_i^{(\ell)}|\mathcal{G}_i^{(\ell)}) \| p_0(\mathcal{Z}_i^{(\ell)})) \right].$$

In general, the overall optimization is given as follows:

$$\min_{\phi, \theta, \{\hat{\boldsymbol{\mu}}^{(\ell)}, \hat{\boldsymbol{\Sigma}}^{(\ell)}\}_\ell} \mathcal{L}_{\text{MGVAE}}(\phi, \theta; \{\hat{\boldsymbol{\mu}}^{(\ell)}, \hat{\boldsymbol{\Sigma}}^{(\ell)}\}_\ell) + \sum_{i,\ell} \lambda^{(\ell)} \mathcal{D}_{\text{KL}}(P_i^{(\ell)} \| Q_i^{(\ell)}),$$

where ϕ denotes all learnable parameters of the encoders, θ denotes all learnable parameters of the decoders, $\mathcal{D}_{\text{KL}}(P_i^{(\ell)} \| Q_i^{(\ell)})$ is the balanced-cut loss for graph $\mathcal{G}_i$ at level ℓ, $\hat{\boldsymbol{\mu}}^{(\ell)}$ and $\hat{\boldsymbol{\Sigma}}^{(\ell)}$ are learnable parameters of the prior in an equivariant manner.

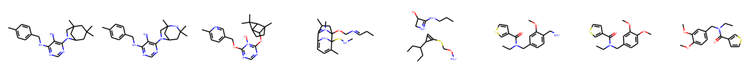


Hierarchy of 3-level Multiresolution Graph Network on Aspirin molecular graph.

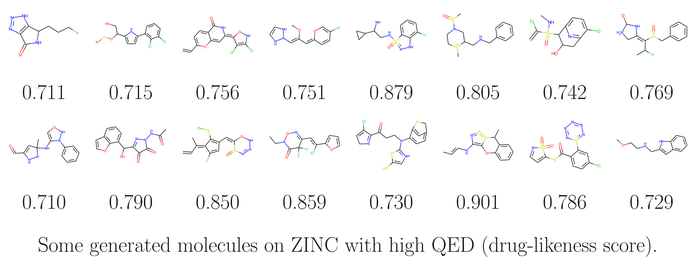
Molecule generation

We examine the generative power of MGVAE in the challenging task of molecule generation, in which the graphs are highly structured. We demonstrate that MGVAE is the first generative model generating graphs in a permutation-equivariant manner that is competitive against autoregressive results. We train on two standard datasets: QM9 & ZINC.

Dataset	Method	Training size	Input features	Validity	Novelty	Uniqueness
QM9	GraphVAE	~ 100K	Graph	61.00%	85.00%	40.90%
	CGVAE			100%	94.35%	98.57%
	MolGAN			98.1%	94.2%	10.4%
	Autoregressive MGN	10K		100%	95.01%	97.44%
	All-at-once MGVAE			100%	100%	95.16%
ZINC	GraphVAE	~ 200K	Graph	14.00%	100%	31.60%
	CGVAE			100%	100%	99.82%
	JT-VAE			100%	-	-
	Autoregressive MGN	1K		100%	99.89%	99.69%
	All-at-once MGVAE	10K	Chemical	99.92%	100%	99.34%



Interpolation on the latent space: we randomly select two molecules from ZINC and we reconstruct the corresponding molecular graphs on the interpolation line.



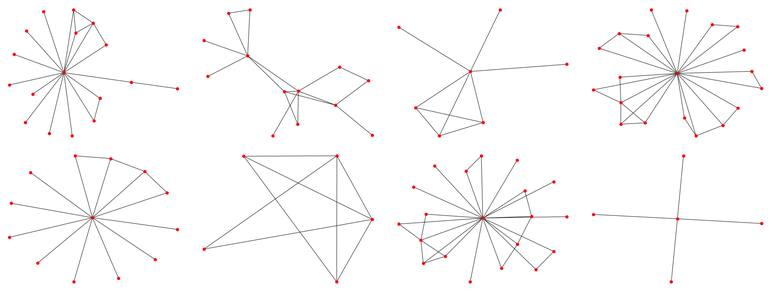
Some generated molecules on ZINC with high QED (drug-likeness score).

General graph generation

We further examine the expressive power of hierarchical latent structure of MGVAE in the task of general graph generation. We choose two datasets from GraphRNN paper (You et al., 2018). MGVAE outperforms all competing methods.

	COMMUNITY-SMALL			EGO-SMALL		
Model	Degree	Cluster	Orbit	Degree	Cluster	Orbit
GraphVAE	0.35	0.98	0.54	0.13	0.17	0.05
DeepGMG	0.22	0.95	0.4	0.04	0.10	0.02
GraphRNN	0.08	0.12	0.04	0.09	0.22	0.003
GNF	0.20	0.20	0.11	0.03	0.10	0.001
GraphAF	0.06	0.10	0.015	0.04	0.04	0.008
MGVAE	**0.002**	**0.01**	**0.01**	**1.74e-05**	**0.0006**	**6.53e-05**

Graph generation results depicting Maximum Mean Discrepancy (MMD) for various graph statistics between the test set and generated graphs.



The top row and the bottom row show some generated and training examples on the **EGO-SMALL** dataset, respectively.

Link prediction on citation graphs

We demonstrate the ability of the MGVAE models to learn meaningful latent embeddings on a link prediction task on popular citation network datasets Cora and Citeseer (Sen et al., 2008). MGVAE outperforms all other methods.

Dataset	**Cora**		**Citeseer**	
Method	AUC (ROC)	AP	AUC (ROC)	AP
SC	84.6 ± 0.01	88.5 ± 0.00	80.5 ± 0.01	85.0 ± 0.01
DW	83.1 ± 0.01	85.0 ± 0.00	80.5 ± 0.02	83.6 ± 0.01
VGAE	90.97 ± 0.77	91.88 ± 0.83	89.63 ± 1.04	91.10 ± 1.02
MGVAE (Spectral)	91.19 ± 0.76	92.27 ± 0.73	90.55 ± 1.17	91.89 ± 1.27
MGVAE (K-Means)	93.07 ± 5.61	92.49 ± 5.77	90.81 ± 1.19	91.98 ± 1.02
MGVAE	**95.67 ± 3.11**	**95.02 ± 3.36**	**93.93 ± 5.87**	**93.06 ± 6.33**

Citation graph link prediction results (AUC & AP).

Method	Min	Max	STD	KL divergence
Spectral	1	3320	292.21	4.51
K-Means	1	326	41.69	0.74
Learn to cluster	11	38	4.93	**0.01**

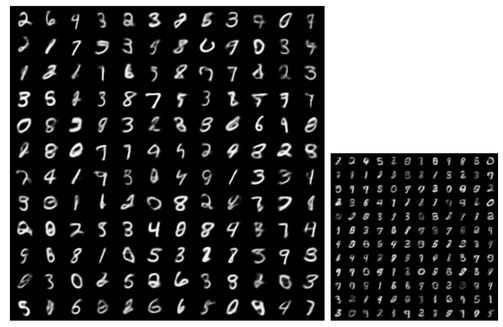
Learning to cluster algorithm returns balanced cuts on Citeseer.

Graph-based image generation

We apply MGVAE into the task of image generation by representing an image as a grid graph. MGVAE outperforms all the baselines for the highest resolution generation on MNIST. Importantly, this is the only model able to generate on multiple resolutions.

Method	FID$_\downarrow$ (32×32)	FID$_\downarrow$ (16×16)	FID$_\downarrow$ (8×8)
DCGAN	113.129	N/A	N/A
VEEGAN	68.749		
PACGAN	58.535		
PresGAN	42.019		
MGVAE	**39.474**	64.289	39.038

Quantitative evaluation of the generated set by FID metric for each resolution



Generated examples at 32×32 and 16×16 resolutions

Supervised molecular properties prediction

Furthermore, to demonstrate the comprehensiveness of MGN, we apply our model in a supervised regression task to predict the solubility (LogP) on the ZINC dataset. The baselines include Graph Attention Networks (GAT) (Velikovi et al., 2018), MoNet (Monti et al., 2017), Disentangled GCNs (Ma et al., 2019), Factorizable GCNs (Yang et al., 2020), and GatedGCN$_E$ (Dwivedi et al., 2020). Our supervised result shows that MGN outperforms the state-of-the-art models in the field with a margin of **20%**.

Method	MLP	GCN	GAT	MoNet	DiscenGCN	FactorGCN	GatedGCN$_E$	**MGN**
MAE	0.667	0.503	0.479	0.407	0.538	0.366	0.363	**0.290**