Semi-Implicit Graph Variational Auto-Encoders

Arman Hasanzadeh* Ehsan Hajiramezanali* Nick Duffield* Krishna Narayanan* Mingyuan Zhou** Xiaoning Qian*

*Department of Electrical and Computer Engineering, Texas A&M University
**McCombs School of Business, University of Texas at Austin

Introduction

Variational graph auto-encoder (VGAE) embeds each node to a latent space as a random vector. Despite its popularity,

1) The Gaussian assumption imposed on the variational distribution restricts its variational inference expressive power.
2) The adopted inner-product (IP) decoder restricts its generative model flexibility.

We address the issues by, first, combining two non-Gaussian variational inference methods with VGAE, to introduce:

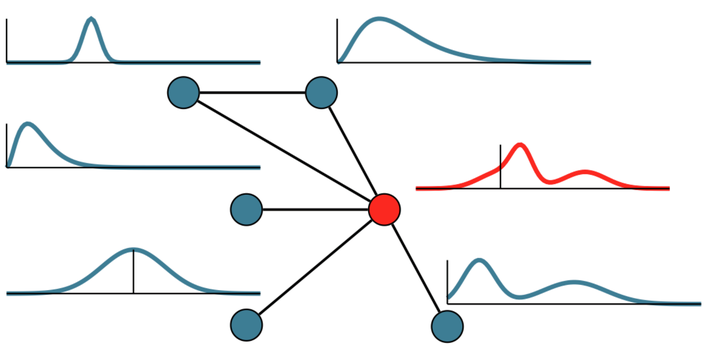
- Semi-implicit graph variational auto-encoder (SIG-VAE)
- Normalizing flow variational graph auto-encoder (NF-VGAE).

In addition, we use Bernoulli-Poisson link decoder to further increase the expressiveness of the generative model.

Compared to VGAE, the derived graph latent representations by our proposed methods are more interpretable, due to more expressive generative model, especially for sparse graphs.

Methods

SIG-VAE:



- The inference model with semi-implicit hierarchical construction:

$$\mathbf{Z} \sim q(\mathbf{Z} \mid \psi), \qquad \psi \sim q_{\boldsymbol{\phi}}(\psi \mid \mathbf{X}, \mathbf{A})$$

$$\mathbf{h}_u = \mathrm{GNN}_u(\mathbf{A}, \mathrm{CONCAT}(\mathbf{X}, \boldsymbol{\epsilon}_u, \mathbf{h}_{u-1})),$$
$$\text{where } \boldsymbol{\epsilon}_u \sim q_u(\boldsymbol{\epsilon}) \text{ for } u = 1, \ldots, L, \ \mathbf{h}_0 = \mathbf{0}$$
$$\boldsymbol{\mu}(\mathbf{A}, \mathbf{X}) = \mathrm{GNN}_\mu(\mathbf{A}, \mathrm{CONCAT}(\mathbf{X}, \mathbf{h}_L))$$
$$\boldsymbol{\Sigma}(\mathbf{A}, \mathbf{X}) = \mathrm{GNN}_\Sigma(\mathbf{A}, \mathrm{CONCAT}(\mathbf{X}, \mathbf{h}_L))$$

$$q(\mathbf{Z} \mid \mathbf{A}, \mathbf{X}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \prod_{i=1}^{N} q(\mathbf{z}_i \mid \mathbf{A}, \mathbf{X}, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$$
$$q(\mathbf{z}_i \mid \mathbf{A}, \mathbf{X}, \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) = \mathcal{N}(\boldsymbol{\mu}_i(\mathbf{A}, \mathbf{X}), \boldsymbol{\Sigma}_i(\mathbf{A}, \mathbf{X}))$$

- Evidence lower bound (ELBO):

$$\begin{aligned}
\mathcal{L} &= \mathbb{E}_{\psi \sim q_{\boldsymbol{\phi}}(\psi \mid \mathbf{X}, \mathbf{A})}\left[\mathbb{E}_{\mathbf{Z} \sim q(\mathbf{Z} \mid \psi)}\left[\log\left(\frac{p(\mathbf{A}|\mathbf{Z})p(\mathbf{Z})}{q(\mathbf{Z} \mid \psi)}\right)\right]\right] \\
&= -\mathbb{E}_{\psi \sim q_{\boldsymbol{\phi}}(\psi \mid \mathbf{X}, \mathbf{A})}[\mathbf{KL}(q(\mathbf{Z} \mid \psi) \,||\, p(\mathbf{Z}))] \\
&\quad + \mathbb{E}_{\psi \sim q_{\boldsymbol{\phi}}(\psi \mid \mathbf{X}, \mathbf{A})}\left[\mathbb{E}_{\mathbf{Z} \sim q(\mathbf{Z} \mid \psi)}[\log p(\mathbf{A} \mid \mathbf{Z})]\right] \\
&= -\mathbf{KL}(\mathbb{E}_{\psi \sim q_{\boldsymbol{\phi}}(\psi \mid \mathbf{X}, \mathbf{A})}[q(\mathbf{Z} \mid \psi)] \,||\, p(\mathbf{Z})) \\
&\quad + \mathbb{E}_{\psi \sim q_{\boldsymbol{\phi}}(\psi \mid \mathbf{X}, \mathbf{A})}[\mathbb{E}_{\mathbf{Z} \sim q(\mathbf{Z} \mid \psi)}[\log p(\mathbf{A} \mid \mathbf{Z})]] = \mathcal{L}
\end{aligned}$$

NF-VGAE:

- The inference model with normalizing flow construction:

$$\mathbf{h}_u = \mathrm{GNN}_u(\mathbf{A}, \mathrm{CONCAT}(\mathbf{X}, \mathbf{h}_{u-1})),$$
$$\text{for } u = 1, \ldots, L, \ \mathbf{h}_0 = \mathbf{0}$$
$$\boldsymbol{\mu}(\mathbf{A}, \mathbf{X}) = \mathrm{GNN}_\mu(\mathbf{A}, \mathrm{CONCAT}(\mathbf{X}, \mathbf{h}_L))$$
$$\boldsymbol{\Sigma}(\mathbf{A}, \mathbf{X}) = \mathrm{GNN}_\Sigma(\mathbf{A}, \mathrm{CONCAT}(\mathbf{X}, \mathbf{h}_L))$$

$$q_0(\mathbf{Z}^{(0)} \mid \mathbf{A}, \mathbf{X}) = \prod_{i=1}^{N} q_0(\mathbf{z}_i^{(0)} \mid \mathbf{A}, \mathbf{X}),$$
$$\text{with} \quad q_0(\mathbf{z}_i^{(0)} \mid \mathbf{A}, \mathbf{X}) = \mathcal{N}(\boldsymbol{\mu}_i, \mathrm{diag}(\boldsymbol{\sigma}_i^2)),$$
$$q_K(\mathbf{Z}^{(K)} \mid \mathbf{A}, \mathbf{X}) = \prod_{i=1}^{N} q_0(\mathbf{z}_i^{(K)} | \mathbf{A}, \mathbf{X}),$$
$$\ln(q_K(\mathbf{z}_i^{(K)} \mid -)) = \ln(q_0(\mathbf{z}_i^{(0)})) - \sum_k \ln|\det \frac{\partial f_k}{\partial \mathbf{z}_i^{(k)}}|,$$

Bernoulli-Poisson Link Decoder:

$$A_{i,j} = \delta(m_{ij} > 0)$$
$$m_{ij} \sim \mathrm{Poisson}\big(\exp(\sum_{k=1}^{l} r_k z_{ik}\, z_{jk})\big)$$
$$p(\mathbf{A} \mid \mathbf{Z}, \mathbf{R}) = \prod_{i=1}^{N}\prod_{j=1}^{N} p(A_{i,j} \mid \mathbf{z}_i, \mathbf{z}_j, \mathbf{R})$$
$$p(A_{i,j} = 1 \mid \mathbf{z}_i, \mathbf{z}_j, \mathbf{R}) = 1 - e^{-\exp(\sum_{k=1}^{l} r_k z_{ik}\, z_{jk})}$$

Results

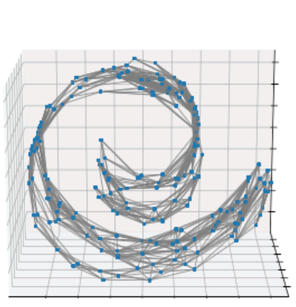
Link Prediction:

- We measure the performance by average precision (AP) and area under the ROC curve (AUC) based on 10 runs on a test set of previously removed links (10% of all links) in these graphs.

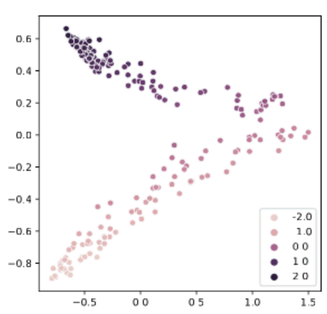
Method	Cora		Citeseer		Pubmed	
	AUC	AP	AUC	AP	AUC	AP
SC	84.6 ± 0.01	88.5 ± 0.00	80.5 ± 0.01	85.0 ± 0.01	84.2 ± 0.02	87.8 ± 0.01
DW	83.1 ± 0.01	85.0 ± 0.00	80.5 ± 0.02	83.6 ± 0.01	84.4 ± 0.00	84.1 ± 0.00
GAE	91.0 ± 0.02	92.0 ± 0.03	89.5 ± 0.04	89.9 ± 0.05	96.4 ± 0.00	96.5 ± 0.00
VGAE	91.4 ± 0.01	92.6 ± 0.01	90.8 ± 0.02	92.0 ± 0.02	94.4 ± 0.02	94.7 ± 0.02
$\mathcal{S}$-VGAE	94.10 ± 0.1	94.10 ± 0.3	94.70 ± 0.2	95.20 ± 0.2	96.00 ± 0.1	96.00 ± 0.1
SEAL	90.09 ± 0.1	83.01 ± 0.3	83.56 ± 0.2	77.58 ± 0.2	96.71 ± 0.1	90.10 ± 0.1
G2G	92.10 ± 0.9	92.58 ± 0.8	95.32 ± 0.7	95.57 ± 0.7	94.28 ± 0.3	93.38 ± 0.5
NF-VGAE	92.42 ± 0.6	93.08 ± 0.5	91.76 ± 0.3	93.04 ± 0.8	96.59 ± 0.3	96.68 ± 0.4
SIG-VAE (IP)	$\mathbf{94.37} \pm 0.1$	$\mathbf{94.41} \pm 0.1$	$\mathbf{95.90} \pm 0.1$	$\mathbf{95.46} \pm 0.1$	$\mathbf{96.73} \pm 0.1$	$\mathbf{96.67} \pm 0.1$
SIG-VAE	$\mathbf{96.04} \pm 0.04$	$\mathbf{95.82} \pm 0.06$	$\mathbf{96.43} \pm 0.02$	$\mathbf{96.32} \pm 0.02$	$\mathbf{97.01} \pm 0.07$	$\mathbf{97.15} \pm 0.04$

Interpretable Latent Representation:

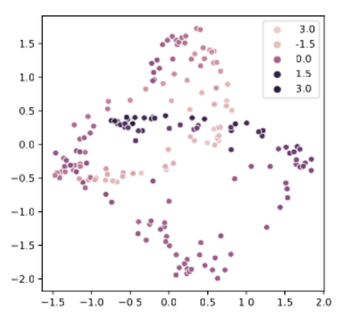
- We show that SIG-VAE captures the graph structure in the latent space better and has a more interpretable embedding than VGAE on a generated Swiss roll graph with 200 nodes and 1244 edges.



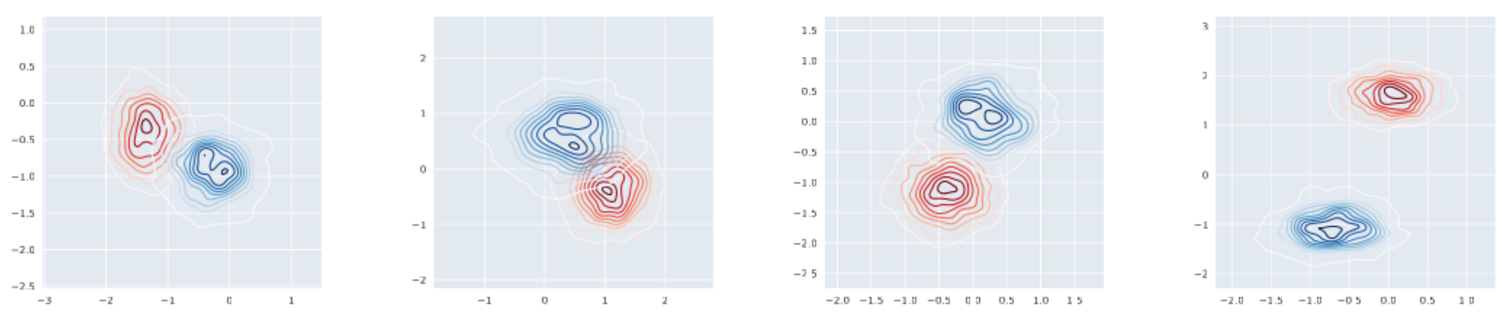
a) Swiss roll graph.



b) SIG-VAE (IP) representation in R^3.



c) VGAE representation in R^3.



Latent representation distributions of four example nodes from the Swiss roll graph using SIG-VAE (blue) and VGAE (red).

Graph Generation:

- We used the learned embedding to generate new graphs and compared to original graph based on density and clustering coefficients.

Detasets	Orignial Graph		VGAE		**SIG-VAE** (IP)		**SIG-VAE**	
	Dens.	Clus.	Dens.	Clus.	Dens.	Clus.	Dens.	Clus.
Cora	0.00143	0.24	0.1178	0.49	0.1178	0.49	**0.00147**	**0.25**
Citeseer	0.0008	0.14	0.09	0.45	0.26	0.42	**0.0008**	**0.16**
USAir	0.038	0.62	0.18	0.40	0.21	**0.56**	**0.043**	0.45
NS	0.002	0.63	0.36	0.47	0.26	0.42	**0.02**	**0.49**
Router	0.0004	0.01	0.16	0.49	0.16	0.49	**0.0010**	**0.09**

Node Classification:

- We applied SIG-VAE for node classification on citation graphs with labels by modifying the loss function to include graph reconstruction and semi-supervised classification terms.

Method	Cora	Citeseer	Pubmed
ManiReg	59.5	60.1	70.7
SemiEmb	59.0	59.6	71.1
LP	68.0	45.3	63.0
DeepWalk	67.2	43.2	65.3
ICA	75.1	69.1	73.9
Planetoid	75.7	64.7	77.2
GCN	**81.5**	70.3	79.0
SIG-VAE	79.7	**70.4**	**79.3**

Graph Clustering:

- We first tried SIG-VAE for getting low-dimensional feature space and then applied Gaussian mixture clustering (GMC) on citation graphs with labels including Cora and Citeseer and compared its results with VGAE. We report the normalized mutual information (NMI) and unsupervised clustering accuracy (ACC) of 10 runs.

Method	Cora		Citeseer	
	NMI	ACC	NMI	ACC
VGAE	0.43	59.2	0.20	51.5
SIG-VAE	**0.58**	**68.8**	**0.34**	**57.4**

Analysis of the complexity: For the analysis of the real-world graph dataset Cora on a single GeForce GTX 1080 GPU node, it took 24.5, 11.7 , and 9.5 seconds for SIG-VAE, NF-VGAE, and VGAE methods with 100 epochs, respectively. For the analysis of small real-word graph dataset NS on the same GPU node, it took 7.23, 7.84, and 7.09 seconds for SIG-VAE, NF-VGAE, and VGAE methods.

Conclusion

- Combining the advantages of normalizing flow and semi-implicit hierarchical variational distribution and VGAE with a Bernoulli-Poisson link decoder, NF-VGAE and SIG-VAE are developed to enrich the representation power of the posterior distribution of node embedding given graphs so that both the graph structural and node attribute information can be best captured in the latent space.
- Our experiments with different graph datasets have shown the promising capability of SIG-VAE (and NF-VGAE) in a range of graph analysis applications with interpretable latent representations, thanks to the hierarchical construction that diffuses the distributions of neighborhood nodes in given graphs.

Acknowledgment

We thank Texas A&M High Performance Research Computing for providing computational resources to perform experiments in this work.

Reference

A. Hasanzadeh, E. Hajiramezanali, K. Narayanan, N. Duffield, M. Zhou, and X. Qian, "Semi-implicit graph variational auto-encoders," Neural Information Processing Systems (NeurIPS2019), Vancouver, Canada, Dec. 2019.