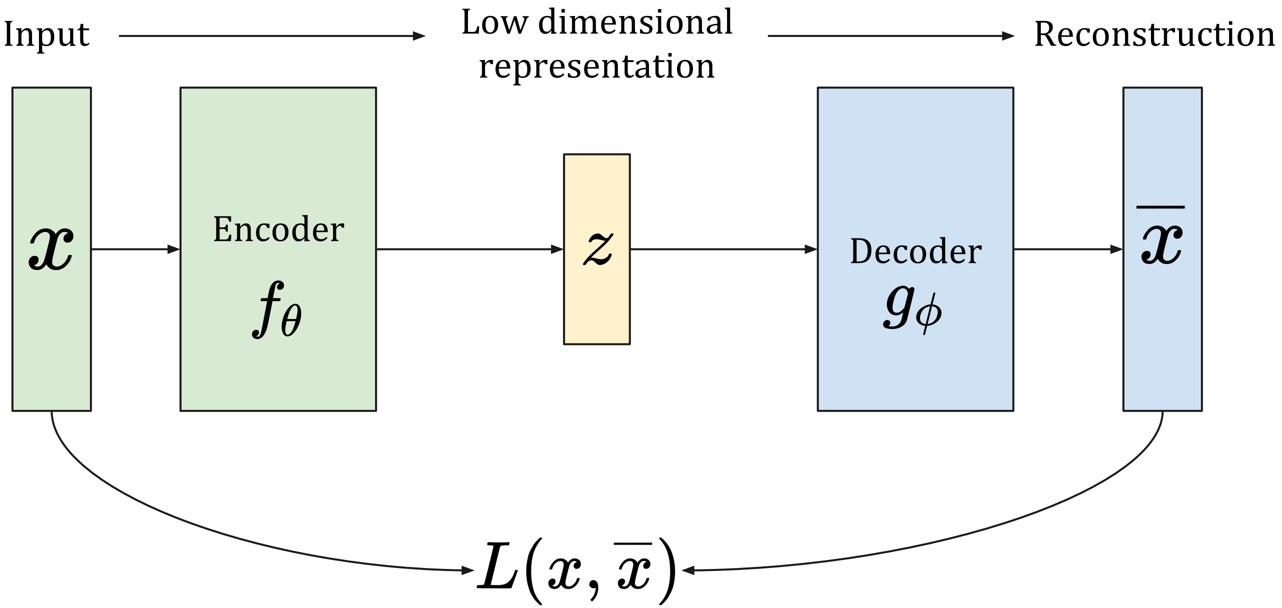
Epitomic Variational Graph Autoencoder

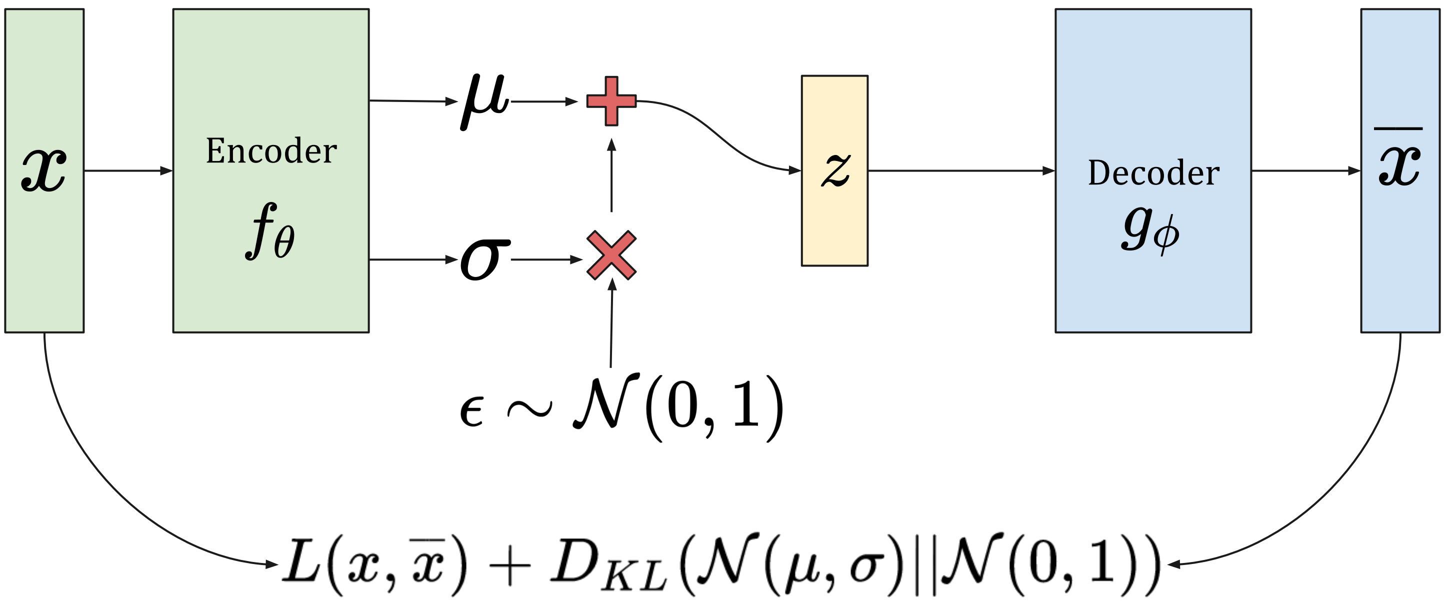
Rayyan Ahmad Khan (Presenter), Muhammad Umer Anwaar and Martin Kleinsteuber

Overview

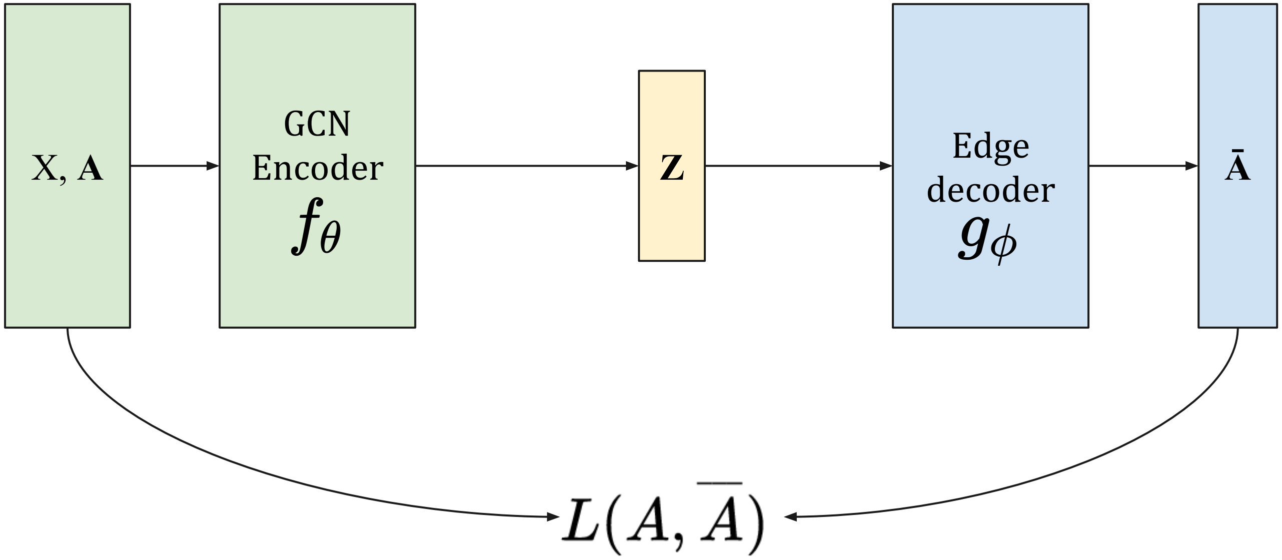
Autoencoder



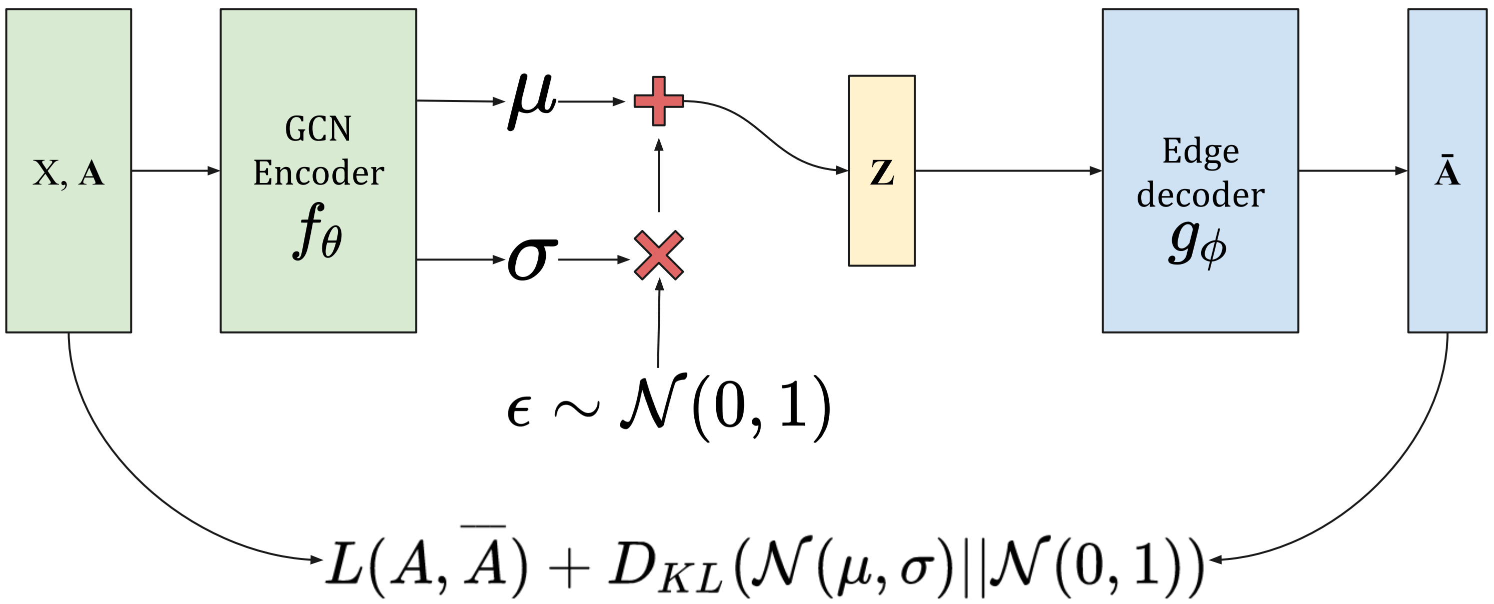
Variational Autoencoder (VAE)



Graph Autoencoder (GAE)



Variational Graph Autoencoder (VGAE)



Over-pruning

Over-pruning in VGAE

VGAE loss function penalizes the latent dimensions/units that **fail to convey enough** information about the input to the decoder block.

Unit Activity

Intuition:

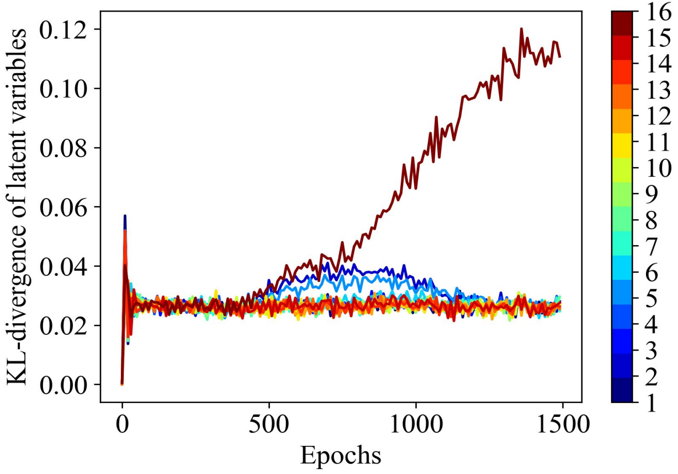
An active unit should have different values for different inputs.

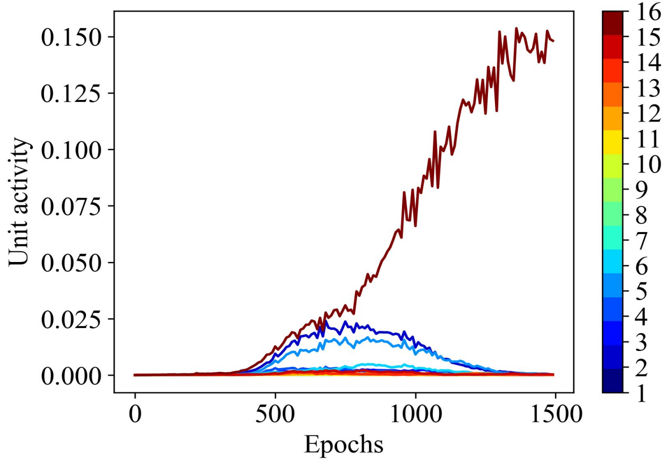
Definition:

$$A_u = \mathrm{Cov}_x\left(\mathbb{E}_{u\sim q(u|x)}[u]\right)$$

A unit u is considered *active* if $A_u > 0.02$

KLD & Unit Activity in *Pure* VGAE - Cora Dataset





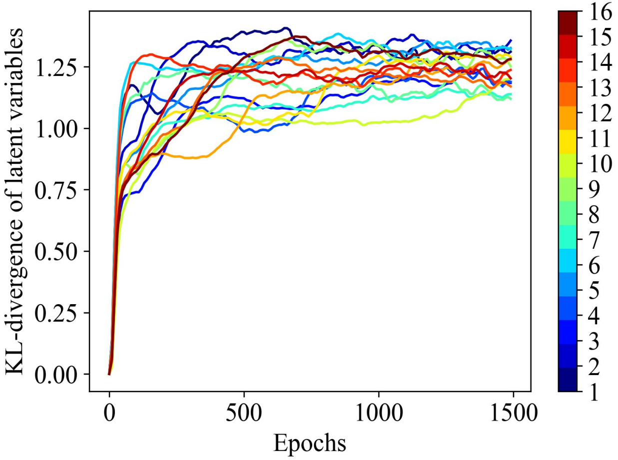
Only one out of 16 hidden units is actively encoding input information required for the reconstruction.

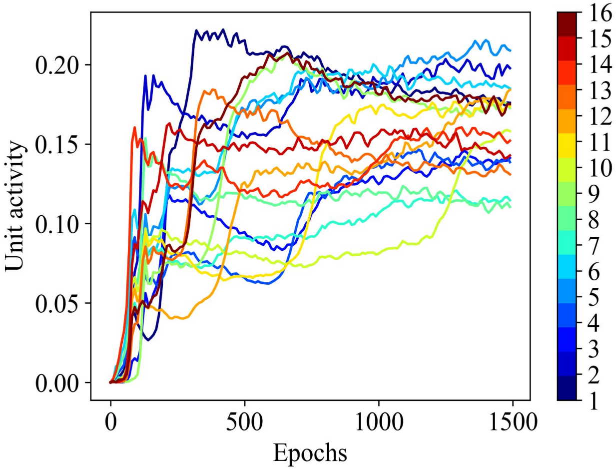
VGAE Approach to Tackle Over-pruning

$$L_{\text{VGAE}} = L(A, \overline{A}) + \beta D_{KL}(\mathcal{N}(\mu, \sigma)||\mathcal{N}(0, 1))$$

- VGAE applies $\beta = \frac{1}{N}$
- Less pruning compared to *pure* VGAE
- Poor distribution matching $\rightarrow$ poor generativeness
- VGAE $\rightarrow$ GAE as $\beta \rightarrow 0$

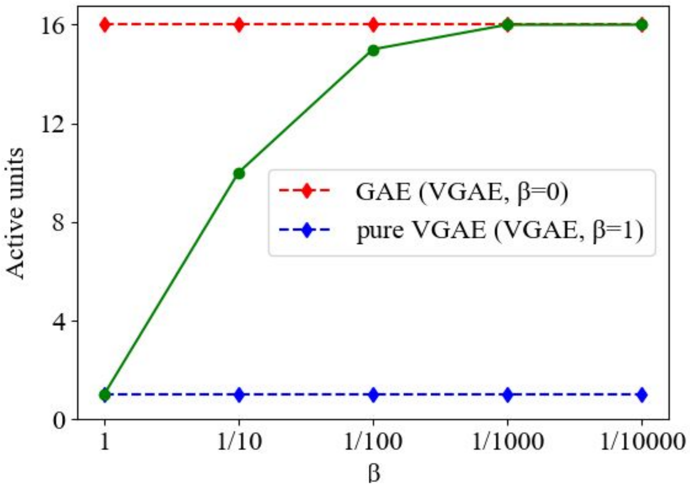
KLD & Unit Activity in VGAE - Cora Dataset

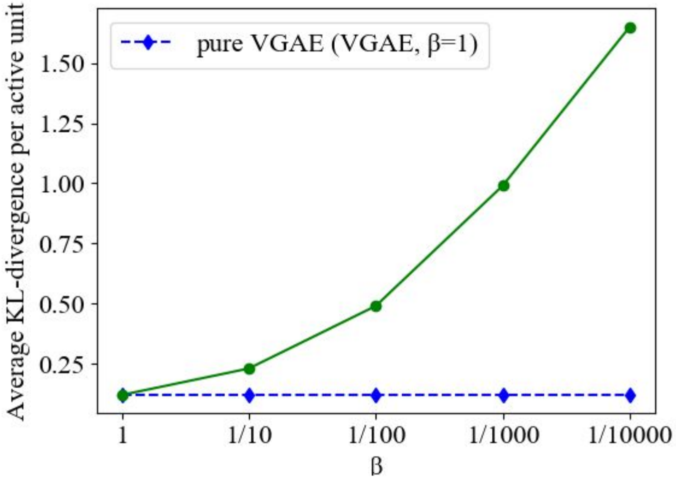




All the hidden units are active but KL-divergence is quite high, indicating poor matching of learnt distribution with prior, consequently affecting generative ability of the model

Effect of β on VGAE - Cora Dataset



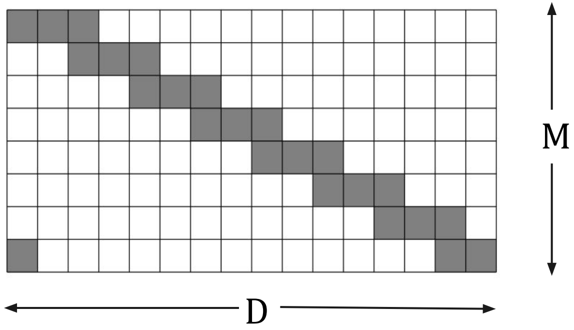


As **β** decreases, both number of active units and average KLD of active units increases.

Epitomic Variational Graph Autoencoder

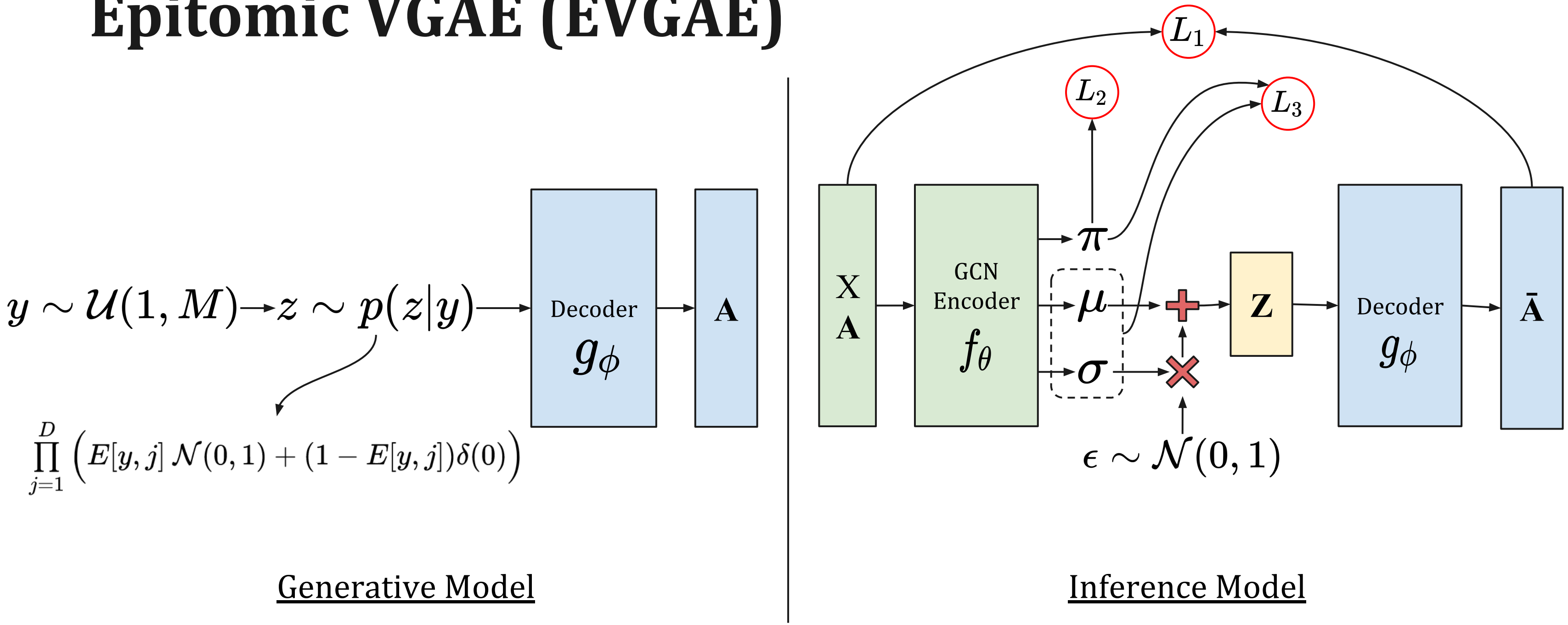
Epitomic VGAE (EVGAE)

- EVGAE consists of multiple sparse VGAE models, called epitomes.
- One epitome is active for each training sample.
- Latent space is shared between the epitomes.



Example of 8 epitomes in 16 dimensions. Gray and white cells refer to 1 and 0 respectively.

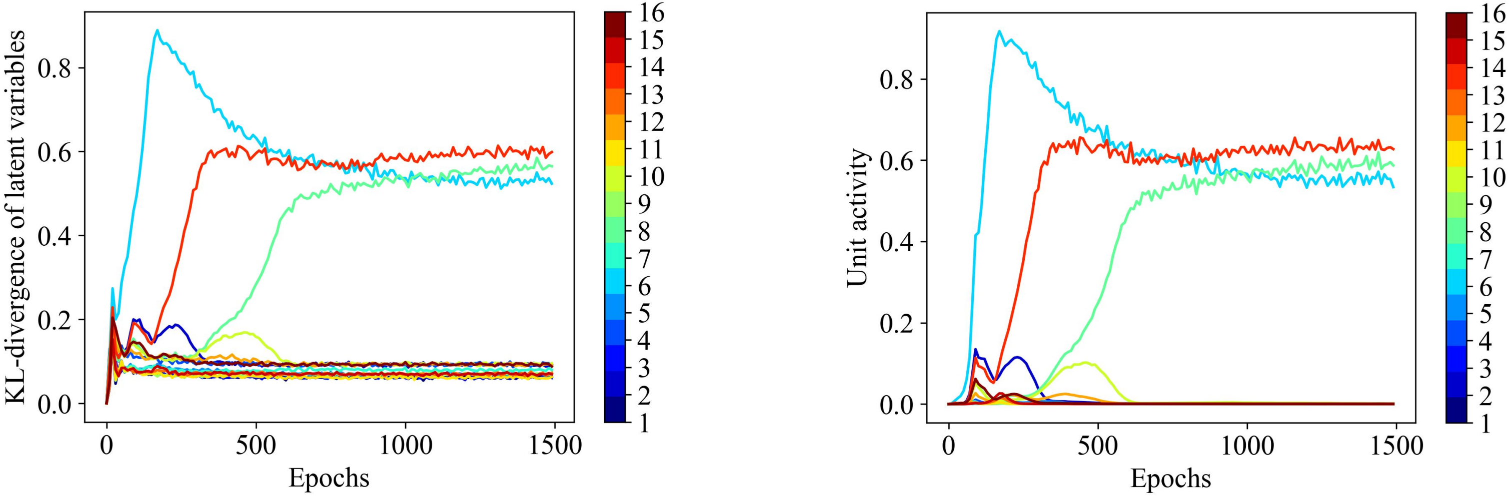
Epitomic VGAE (EVGAE)



EVGAE - Loss Function

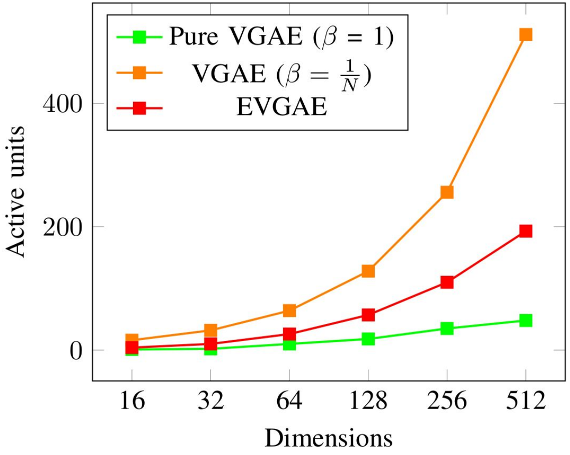
$$L = \overbrace{\mathrm{BCE}}^{L_1} + \overbrace{\sum_{i=1}^{N} D_{KL}\Big(\mathrm{Cat}(\pi_i(\mathcal{G}))||\,\mathcal{U}(1, M)\Big)}^{L_2}$$
$$+ \underbrace{\sum_{i=1}^{N} \sum_{y_i} \pi_i(\mathcal{G}) \sum_{j=1}^{D} E[y_i, j] D_{KL}\Bigg(\mathcal{N}\Big(\mu_i^j(\mathcal{G}), (\sigma_i^2)^j(\mathcal{G})\Big)||\mathcal{N}(0, 1)\Bigg)}_{L_3}.$$

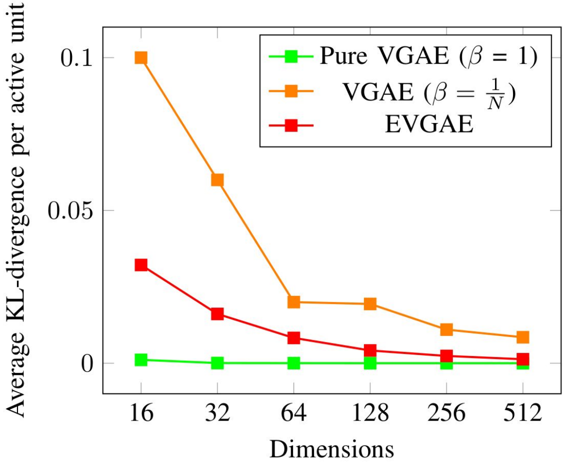
KLD & Unit Activity in EVGAE - Cora Dataset



EVGAE achieves better distribution matching compared to VGAE, while simultaneously getting more units active.

EVGAE vs VGAE vs *Pure* VGAE - Cora Dataset





EVGAE achieves better distribution matching compared to VGAE, while simultaneously getting more units active.

Results on Link Prediction

Method	Cora		Citeseer		PubMed	
	AUC	AP	AUC	AP	AUC	AP
DeepWalk	83.1 ± 0.01	85.0 ± 0.00	80.5 ± 0.02	83.6 ± 0.01	84.4 ± 0.00	84.1 ± 0.0
Spectral Clustering	84.6 ± 0.01	88.5 ± 0.00	80.5 ± 0.01	85.0 ± 0.01	84.2 ± 0.02	87.8 ± 0.01
GAE (VGAE with β = 0)	91.0 ± 0.02	92.0 ± 0.03	89.5 ± 0.04	89.9 ± 0.05	96.4 ± 0.00	96.5 ± 0.0
VGAE ($\beta \sim 10^{-4} - 10^{-5}$)	91.4 ± 0.01	92.6 ± 0.01	90.8 ± 0.02	92.0 ± 0.02	94.4 ± 0.02	94.7 ± 0.0
pure VGAE (β = 1)	79.44 ± 0.03	80.51 ± 0.02	77.08 ± 0.03	79.07 ± 0.02	82.79 ± 0.01	83.88 ± 0.01
EVGAE (β = 1)	$\mathbf{92.96 \pm 0.02}$	$\mathbf{93.58 \pm 0.03}$	$\mathbf{91.55 \pm 0.03}$	$\mathbf{93.24 \pm 0.02}$	$\mathbf{96.80 \pm 0.01}$	$\mathbf{96.91 \pm 0.02}$

Thanks!