Graphite:
Iterative Generative Modeling of Graphs

GNN Study Group - Elian

01. INTRODUCTION

- Graph are Ubiquitous
- Learning Deep Latent Variable Models

02. AUTOENCODER

- How to Generate
- AutoEncoder Problem

03. VARIATIONAL INFERENCE

- Evidence Lower Bound (ELBO)
- Variational AutoEncoder (VAE)

04. GRAPHITE

- GCN Encoder
- GCN Decoder

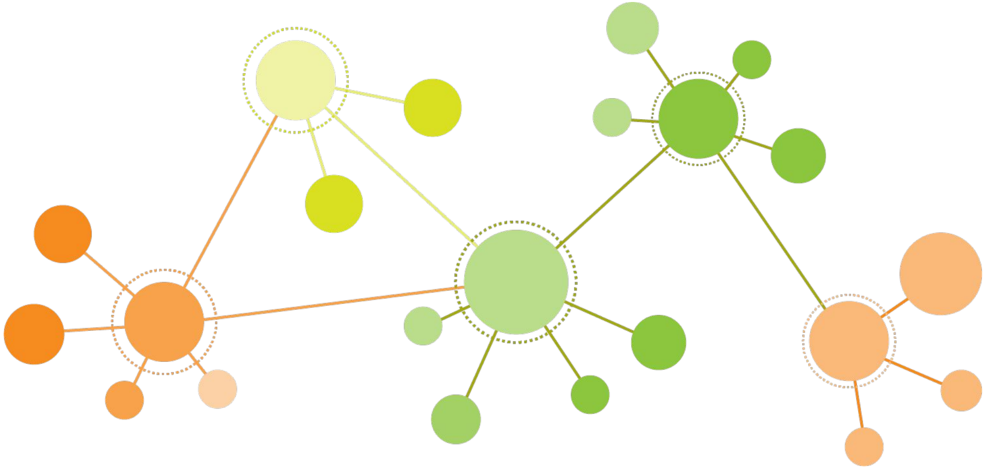
05. EXPERIMENTAL

- Reconstruction Loss
- Link Prediction
- Node Classification

01

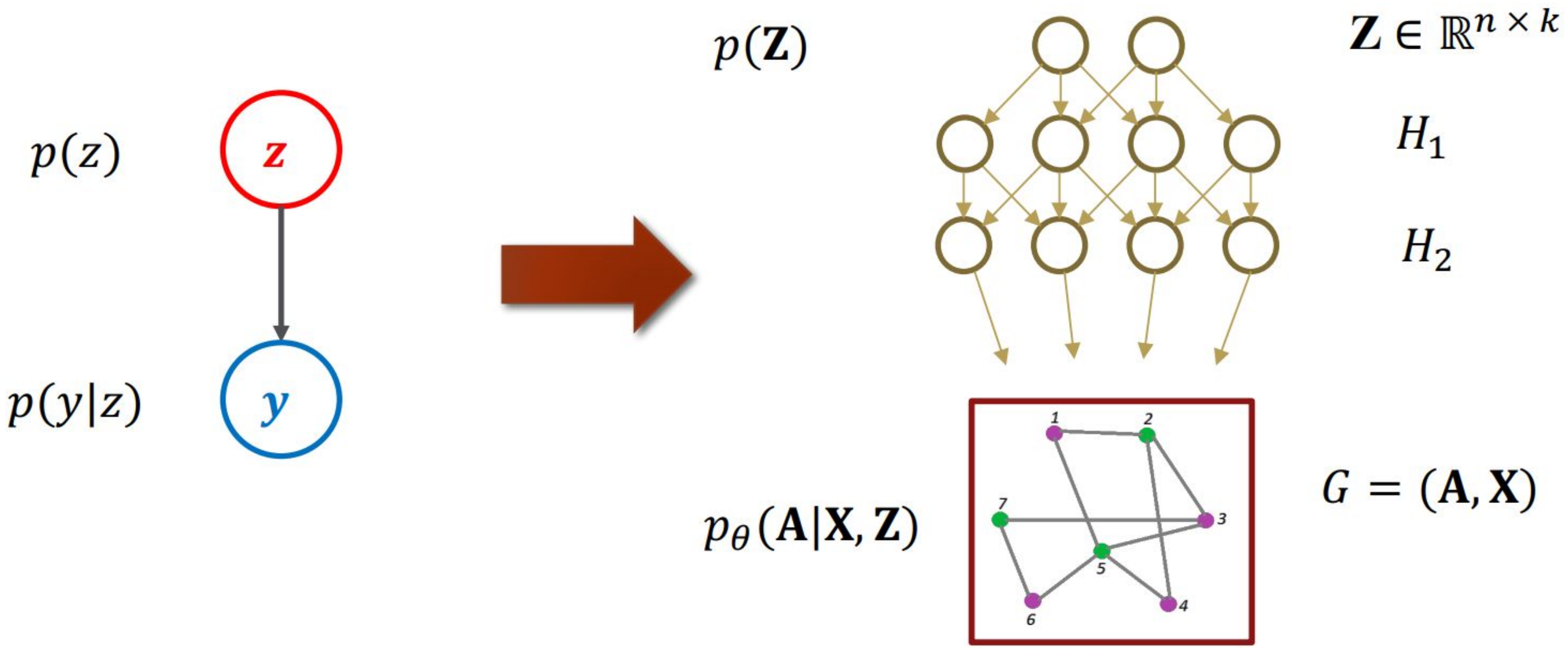
INTRODUCTION

Graph are Ubiquitous



- Ecology:
  - Food web networks
- Biology:
  - Brain networks
  - Protein-protein interaction networks
- Chemistry:
  - Molecules
  - Materials

Learning Deep Latent Variable Models of Graphs

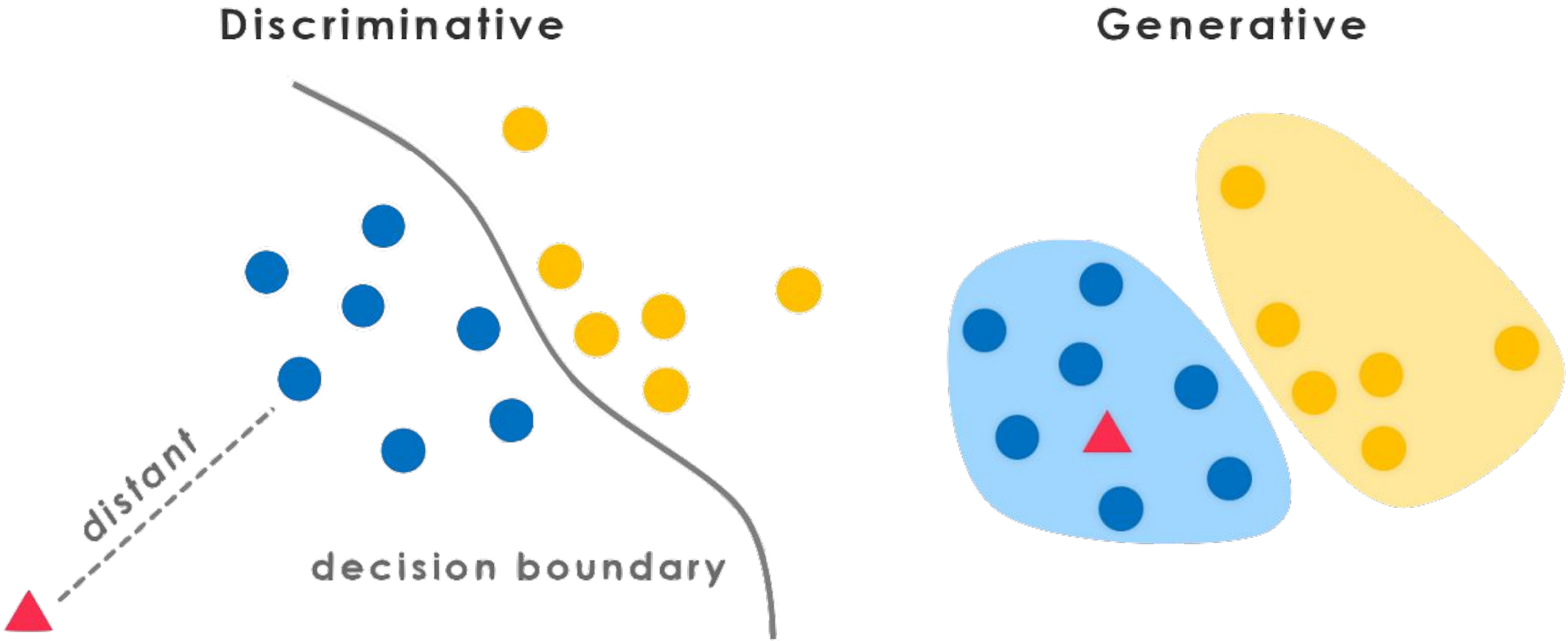


02

AUTOENCODER

What is the Modeling Means?

- Generative vs. Discriminative



What is the Modeling Means?

- Generative vs. Discriminative

What is the Modeling Means?

- Generative vs. Discriminative

Discriminative:

- Skin Color
- Hair Color
- Eye Color
- Facial Different
- Clothing Style
- etc...

Generative:

Asian:

- yellow skin
- small eyes
- black hair
- etc...

African:

- black skin
- black skin
- black skin
- etc...

How to Generate?

- Naive Bayes

Bayes' Theorem:

$$p(Y|\mathbf{X}) = \frac{p(\mathbf{X}, Y)}{p(\mathbf{X})} = \frac{p(Y)p(\mathbf{X}|Y)}{p(\mathbf{X})} \propto p(Y)p(\mathbf{X}|Y)$$

Naive Bayes Generator:

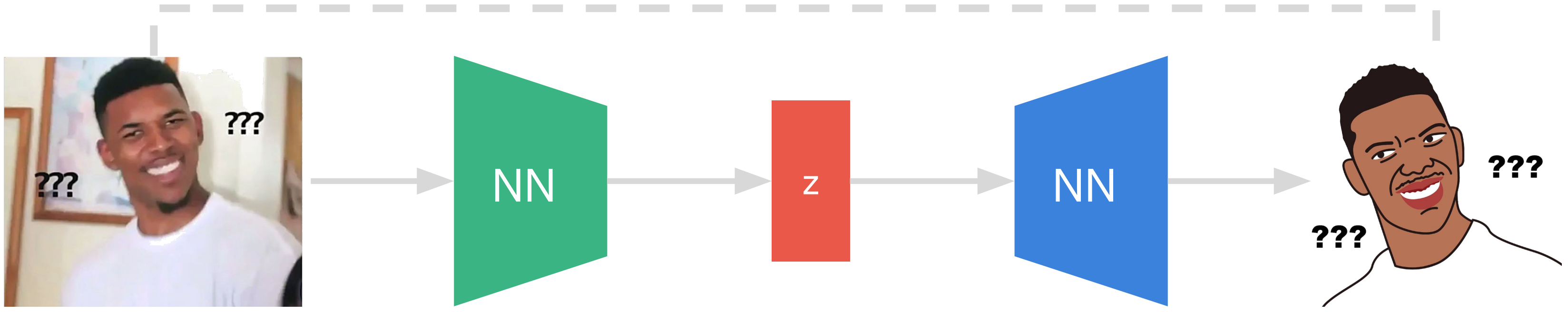
$$p(\mathbf{X}, Y) = p(Y) \times p(X_1, \cdots, X_d|Y) = p(Y) \times \prod_{i=1}^{d} p(X_i|Y)$$

How to Generate?

- AutoEncoder

$$loss = \frac{1}{n}\sum_{i}^{n}(x_i - G(x_i))^2$$

reconstruction loss

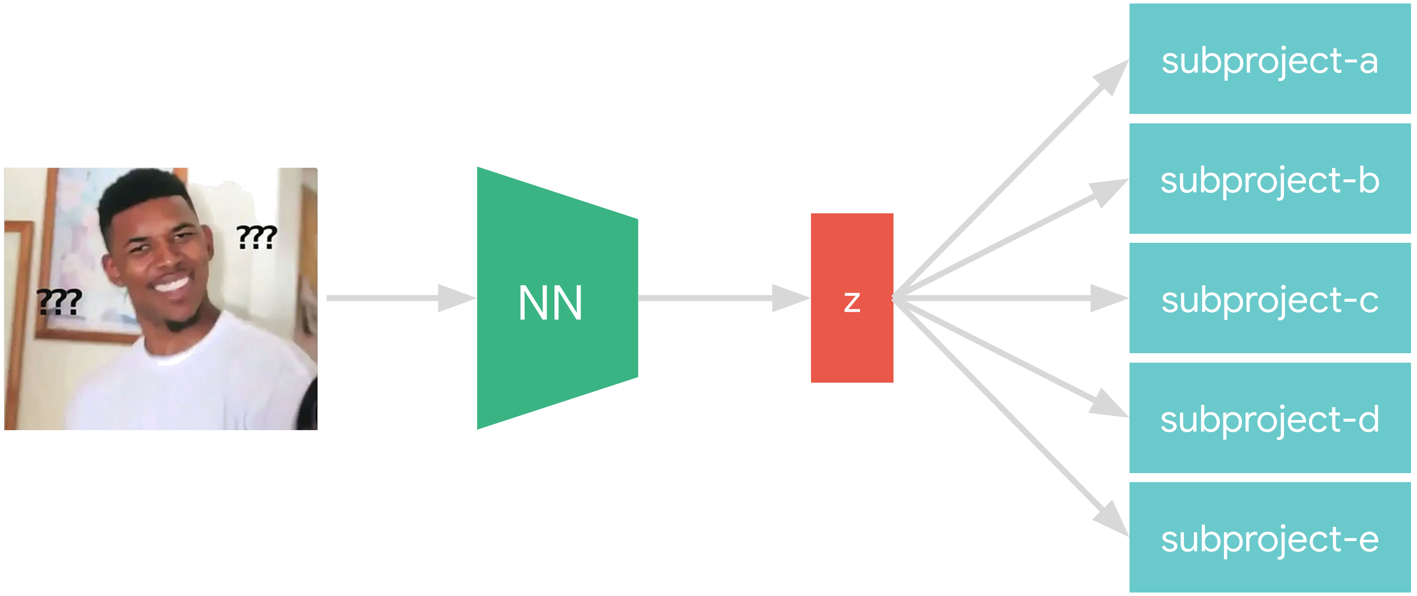


Encoder

Decoder

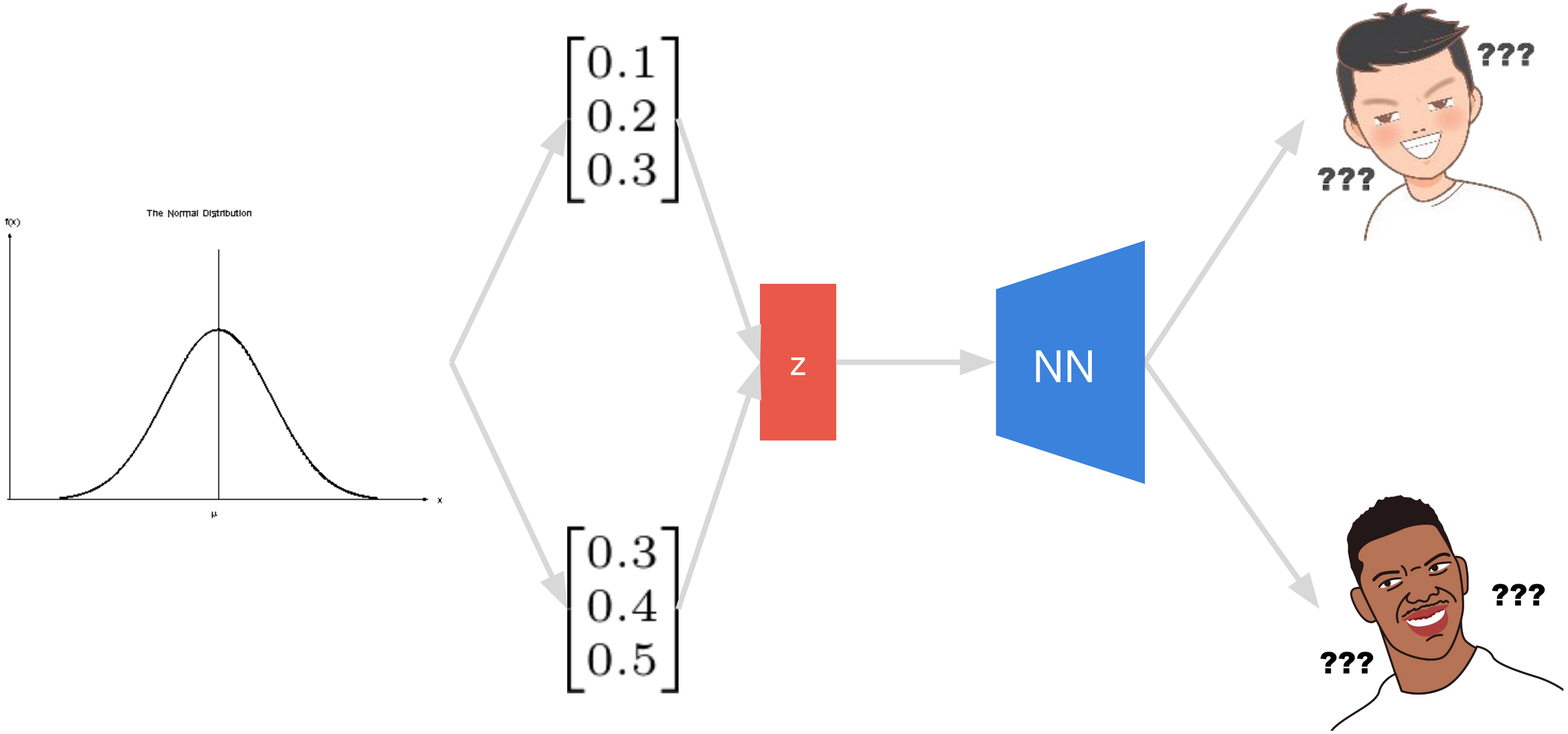
How to Generate?

- Image Compression

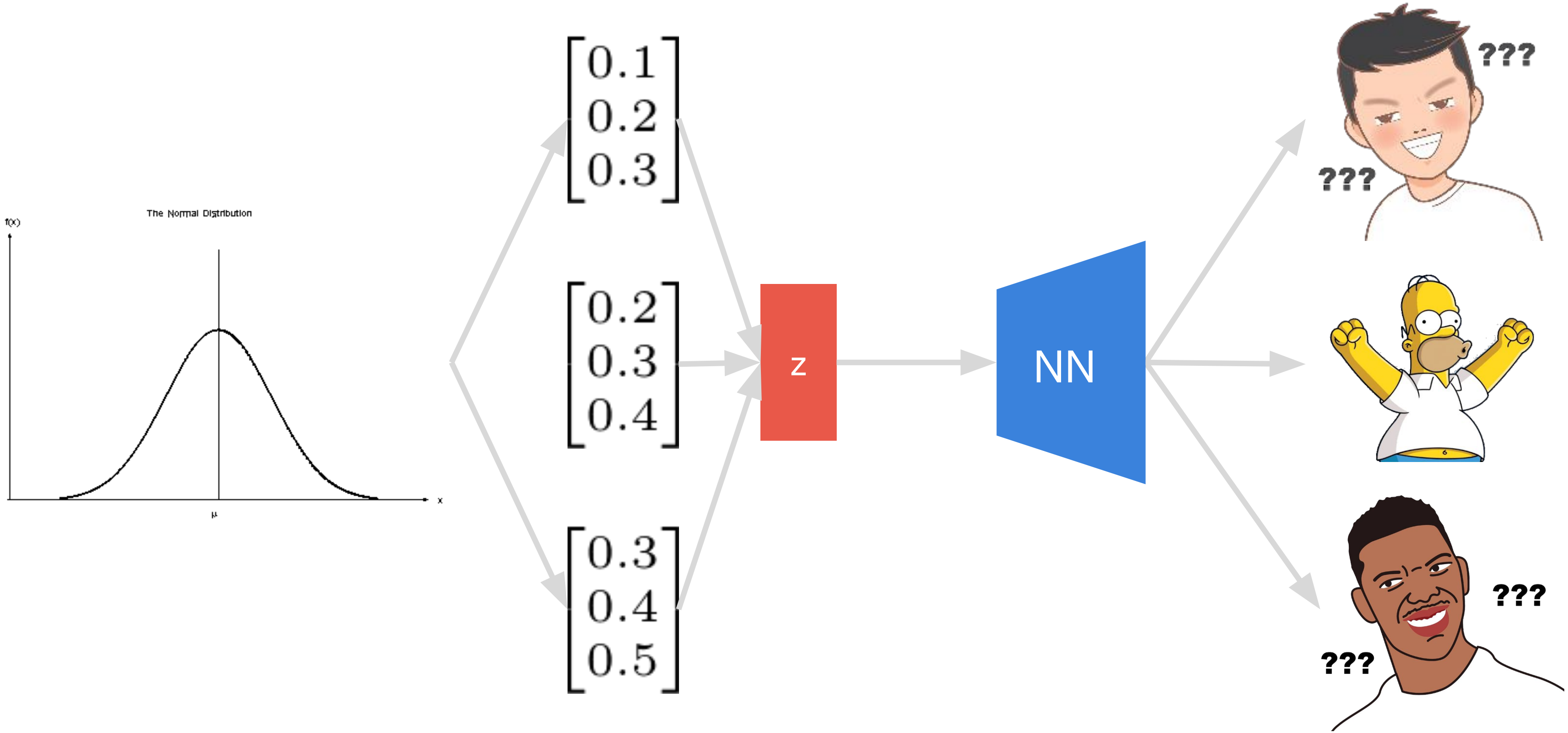


How to Generate?

- Image Generator



AutoEncoder Problem?

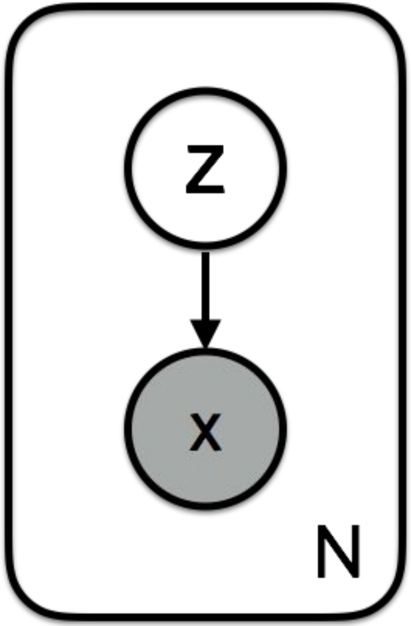


03

VARIANCE

Evidence Lower Bound (ELBO)

- Modeling



Latent Variables: $z_i \sim p(z)$

Data Points: $x_i \sim p_\theta(x \mid z)$

Joint Distribution: $p_\theta(x, z) = p_\theta(x \mid z)p(z)$

Model: $p_\theta(x) = \int p_\theta(x, z)dz$

$$p_\theta(x) = \frac{p_\theta(x, z)}{p(z \mid x)}$$

$$p(z \mid \theta) \to p(z \mid x)$$

Evidence Lower Bound (ELBO)

- KL Divergence

KL Divergence: $\min_{\theta} D_{KL}\left[q(z \mid \theta) \parallel p(z \mid x)\right]$

$$
\begin{aligned}
D_{KL}\left[q(z \mid \theta) \parallel p(z \mid x)\right] &= \int q(z \mid \theta) \log \frac{q(z \mid \theta)}{p(z \mid x)} dz \\
&= \int q(z \mid \theta) \log \frac{q(z \mid \theta) p(x)}{p(x, z)} dz \\
&= \int q(z \mid \theta) \log \frac{q(z \mid \theta)}{p(x, z)} dz + \int q(z \mid \theta) \log \frac{p(x)}{p(x, z)} dz \\
&= \int q(z \mid \theta) \left(\log q(z \mid \theta) - \log p(z, x)\right) + \log p(x) \\
&= -\left(E_{q(z \mid \theta)}\left[\log \frac{p(x, z)}{q(z \mid \theta)}\right]\right) + \log p(x)
\end{aligned}
$$

Evidence Lower Bound (ELBO)

- KL Divergence

KL Divergence: $\min_{\theta} D_{KL}\left[q(z \mid \theta) \parallel p(z \mid x)\right]$

$$D_{KL}\left[q(z \mid \theta) \parallel \boxed{p(z \mid x)}\right] = -\boxed{\left(E_{q(z|\theta)}\left[\log \frac{p(x,z)}{q(z \mid \theta)}\right]\right)} + \boxed{\log p(x)}$$

$$\log p(x) = E_{q(z|\theta)}\left[\log \frac{p(x,z)}{q(z \mid \theta)}\right] + D_{KL}\left[q(z \mid \theta) \parallel p(z \mid x)\right]$$

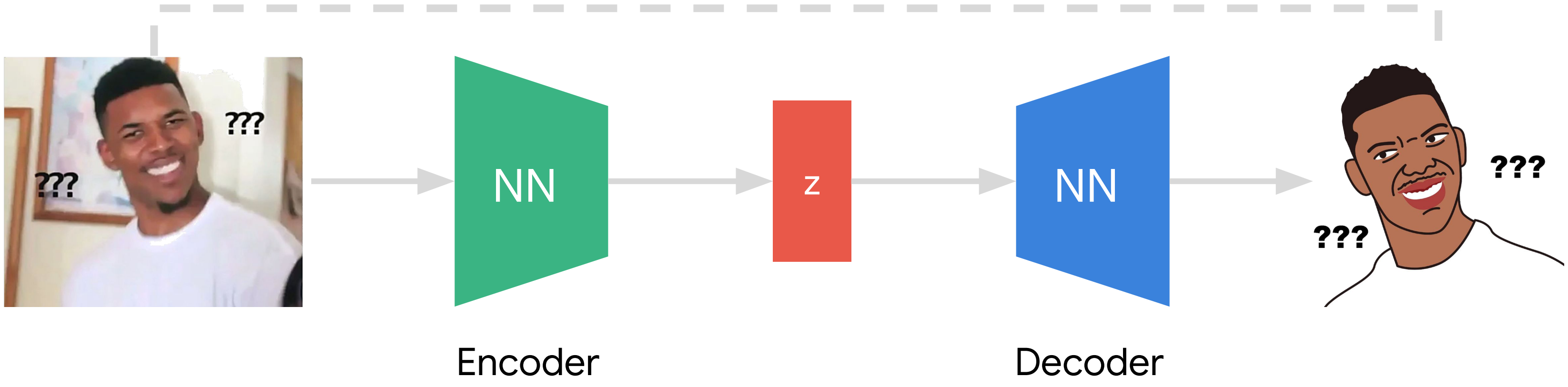
$$\log p(x) \geq E_{q(z|\theta)}\left[\log \frac{p(x,z)}{q(z \mid \theta)}\right]$$

AutoEncoder Review

- AutoEncoder

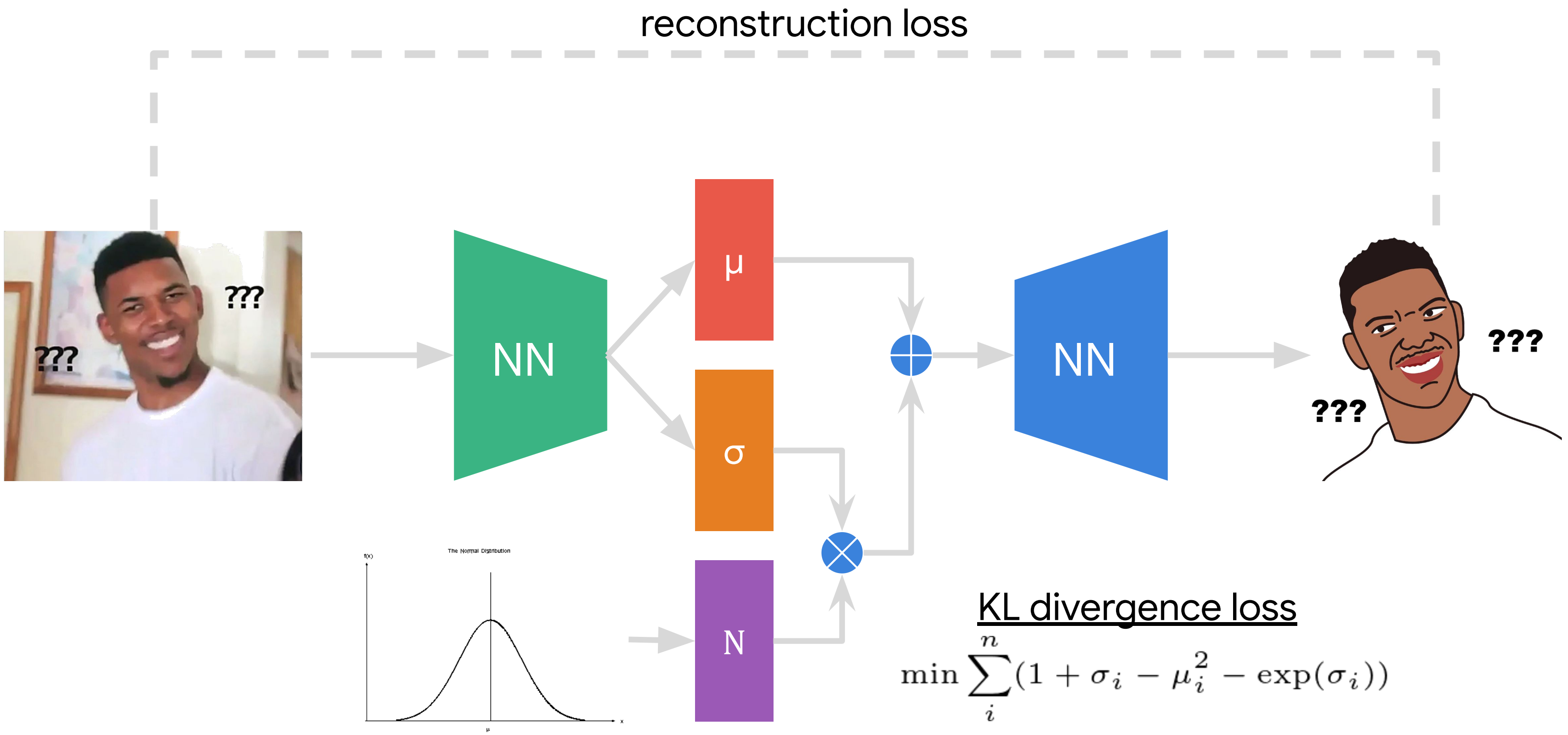
$$loss = \frac{1}{n}\sum_{i}^{n}(x_i - G(x_i))^2$$

reconstruction loss



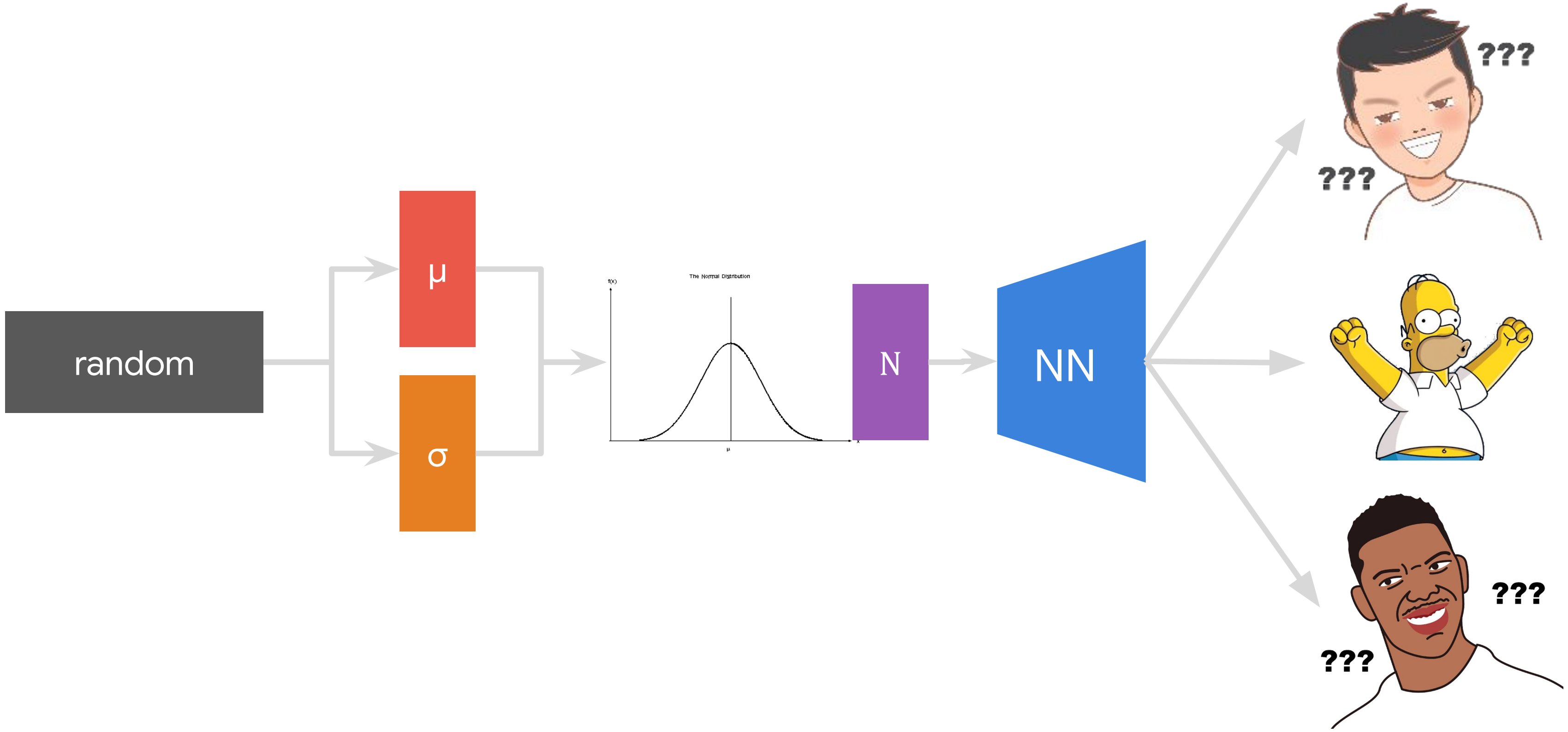
How to Generate?

- Variational AutoEncoder



How to Generate?

- Variational AutoEncoder

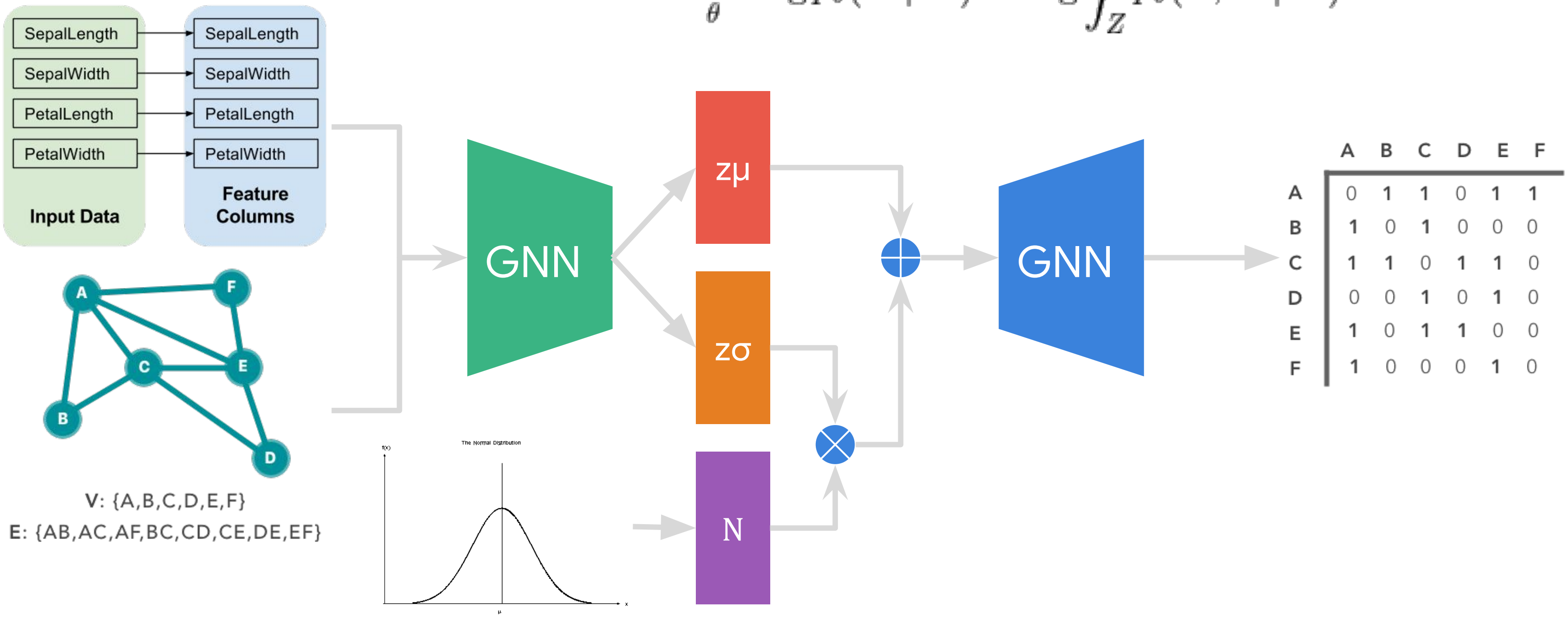


04

Graphite

Graphite Objective

$$\max_{\theta} \log p_{\theta}(A \mid X) = \log \int_{Z} p_{\theta}(A, Z \mid X) dz$$



Graphite Objective

KL Divergence $= E_{q(z|\theta)}\left[\log \frac{p(x,z)}{q(z \mid \theta)}\right] + \log p(x)$

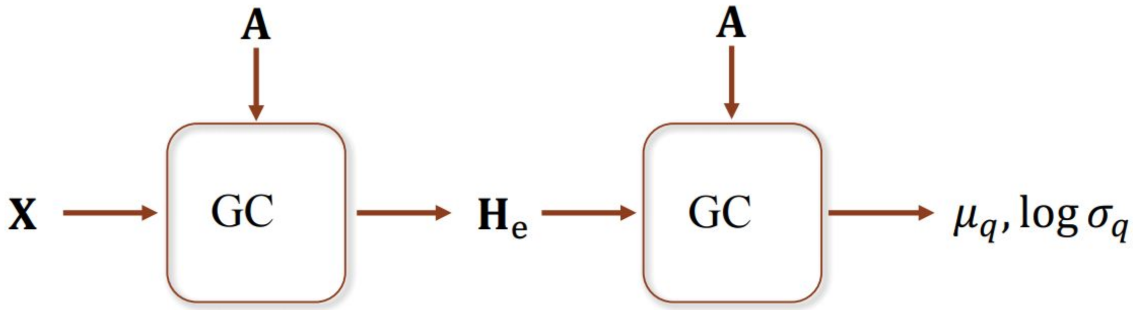
$$\log p(x) = E_{q(z|\theta)}\left[\log \frac{p(x,z)}{q(z \mid \theta)}\right] + D_{KL}\left[q(z \mid \theta) \parallel p(z \mid x)\right]$$

$$\log p(x) \geq E_{q(z|\theta)}\left[\log \frac{p(x,z)}{q(z \mid \theta)}\right]$$

$$\log p(A \mid X) \geq E_{q_\phi}(Z \mid A, X)\left[\log \frac{p_\theta(A, Z \mid X)}{q_\phi(Z \mid A, X)}\right]$$

Graphite Encoder

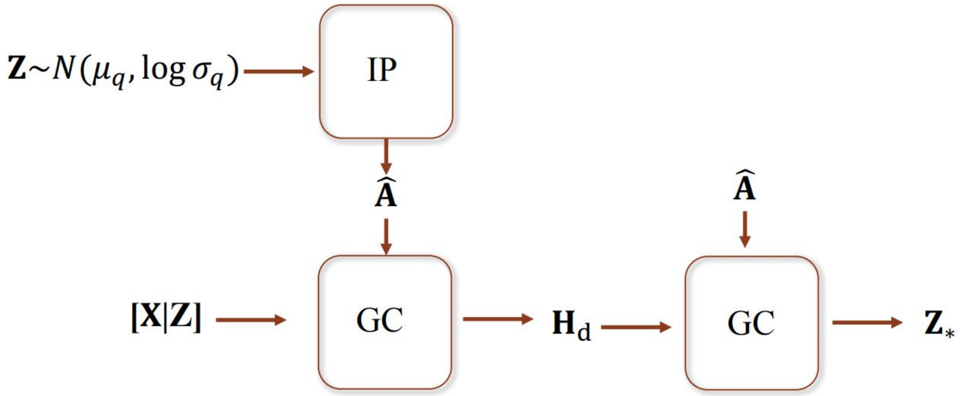
- Variational posterior $q_\theta(Z \mid A, X)$ is a multivariate Gaussian with diagonal covariance
- Encoder parameterized by a graph convolutional network



Forward pass of a two layer encoding GCN

Graphite Decoder

- Decoder is a hybrid that iterates between:
  - intermediate graph construction using an inner product decoder
  - message passing on the intermediate graph using graph convolutions



Graphite Decoder

- The final latent feature matrix is specified as a convex combination of the latent layers

$$Z' = \lambda Z + (1 - \lambda) Z_*$$

where λ ∈ [0,1] is a tunable hyperparameter.

- Observation model $p_\theta(A \mid X, Z)$ is a factorized multivariate Bernoulli

$$p_\theta(A \mid X, Z) = \prod_{i=1}^{n} \prod_{j=1}^{n} p_\theta(A_{ij} \mid X, Z)$$

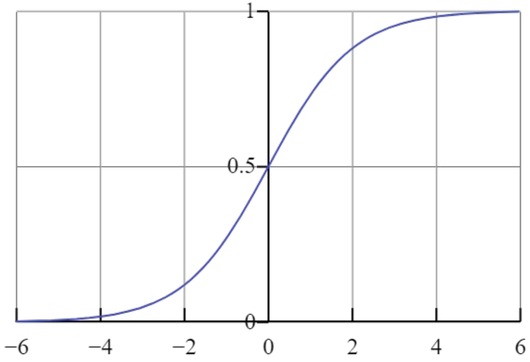
$$p_\theta(A_{ij} \mid X, Z) = \sigma(Z_i' Z_j')$$

05

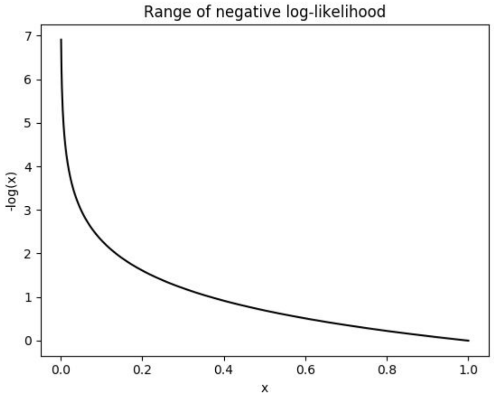
EXPERIMENTAL

Reconstruction Loss

- SCE and NLL




```
  z * -log(sigmoid(x)) + (1 - z) * -log(1 - sigmoid(x))
= z * -log(1 / (1 + exp(-x))) + (1 - z) * -log(exp(-x) / (1 + exp(-x)))
= z * log(1 + exp(-x)) + (1 - z) * (-log(exp(-x)) + log(1 + exp(-x)))
= z * log(1 + exp(-x)) + (1 - z) * (x + log(1 + exp(-x))
= (1 - z) * x + log(1 + exp(-x))
= x - x * z + log(1 + exp(-x))
```



$$
\begin{aligned}
L(b_0, b_1, s^2) &= \log \prod_{i=1}^{n} p(y_i | x_i; b_0, b_1, s^2) & (1)\\
&= \sum_{i=1}^{n} \log p(y_i | x_i; b_0, b_1, s^2) & (2)\\
&= -\frac{n}{2} \log 2\pi - n \log s - \frac{1}{2s^2} \sum_{i=1}^{n} (y_i - (b_0 + b_1 x_i))^2 & (3)
\end{aligned}
$$

Reconstruction Loss

- Sigmoid Cross Entropy

Table 1. Mean reconstruction errors and negative log-likelihood estimates (in nats) for autoencoders and variational autoencoders respectively on test instances from six different generative families. Lower is better.

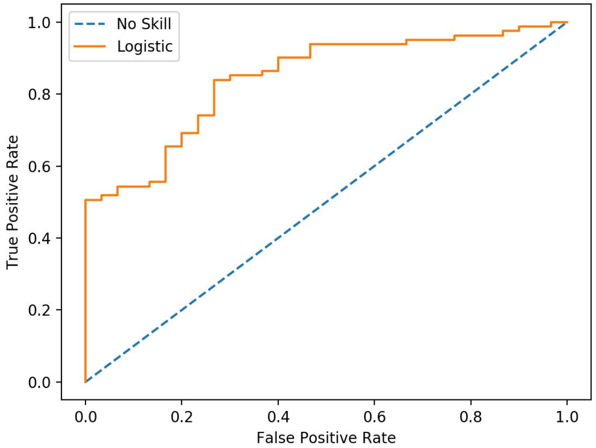
	Erdos-Renyi	Ego	Regular	Geometric	Power Law	Barabasi-Albert
GAE	221.79 ± 7.58	197.3 ± 1.99	198.5 ± 4.78	514.26 ± 41.58	519.44 ± 36.30	236.29 ± 15.13
Graphite-AE	$\mathbf{195.56} \pm 1.49$	$\mathbf{182.79} \pm 1.45$	$\mathbf{191.41} \pm 1.99$	$\mathbf{181.14} \pm 4.48$	$\mathbf{201.22} \pm 2.42$	$\mathbf{192.38} \pm 1.61$
VGAE	273.82 ± 0.07	273.76 ± 0.06	275.29 ± 0.08	274.09 ± 0.06	278.86 ± 0.12	274.4 ± 0.08
Graphite-VAE	$\mathbf{270.22} \pm 0.15$	$\mathbf{270.70} \pm 0.32$	$\mathbf{266.54} \pm 0.12$	$\mathbf{269.71} \pm 0.08$	$\mathbf{263.92} \pm 0.14$	$\mathbf{268.73} \pm 0.09$

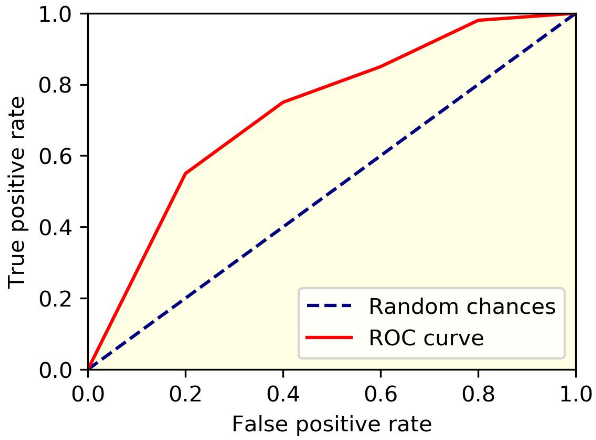
Link Prediction

- **Given two nodes in a graph, does an edge exist between the nodes?**
- **Baselines**:
  - Spectral Clustering (SC)
  - DeepWalk (DW): random walks and skip-gram objective
  - (Variational) Graph Autoencoder (VGAE, GAE):
    - GCN encoder but a single-step inner product decoder
- For Graphite, the task can be formulated as denoising.
- **Datasets**: Protein-protein Interaction, Cora, Citeseer, Pubmed
- **Evaluation metrics**:
  - Area Under the ROC Curve
  - Average Precision

Link Prediction

- AUC of ROC Curve





Link Prediction

- AUC of ROC curve

Table 3. Area Under the ROC Curve (AUC) for link prediction (* denotes dataset with features). Higher is better.

	Cora	Citeseer	Pubmed	Cora*	Citeseer*	Pubmed*
SC	89.9 ± 0.20	91.5 ± 0.17	$\mathbf{94.9} \pm 0.04$	-	-	-
DeepWalk	85.0 ± 0.17	88.6 ± 0.15	91.5 ± 0.04	-	-	-
node2vec	85.6 ± 0.15	89.4 ± 0.14	91.9 ± 0.04	-	-	-
GAE	90.2 ± 0.16	92.0 ± 0.14	92.5 ± 0.06	93.9 ± 0.11	94.9 ± 0.13	96.8 ± 0.04
VGAE	90.1 ± 0.15	92.0 ± 0.17	92.3 ± 0.06	94.1 ± 0.11	96.7 ± 0.08	95.5 ± 0.13
Graphite-AE	91.0 ± 0.15	92.6 ± 0.16	94.5 ± 0.05	94.2 ± 0.13	96.2 ± 0.10	$\mathbf{97.8} \pm 0.03$
Graphite-VAE	$\mathbf{91.5} \pm 0.15$	$\mathbf{93.5} \pm 0.13$	94.6 ± 0.04	$\mathbf{94.7} \pm 0.11$	$\mathbf{97.3} \pm 0.06$	97.4 ± 0.04

Link Prediction

- Average Precision

Table 4. Average Precision (AP) scores for link prediction (* denotes dataset with features). Higher is better.

	Cora	Citeseer	Pubmed	Cora*	Citeseer*	Pubmed*
SC	92.8 ± 0.12	94.4 ± 0.11	$\mathbf{96.0} \pm 0.03$	-	-	-
DeepWalk	86.6 ± 0.17	90.3 ± 0.12	91.9 ± 0.05	-	-	-
node2vec	87.5 ± 0.14	91.3 ± 0.13	92.3 ± 0.05	-	-	-
GAE	92.4 ± 0.12	94.0 ± 0.12	94.3 ± 0.5	94.3 ± 0.12	94.8 ± 0.15	96.8 ± 0.04
VGAE	92.3 ± 0.12	94.2 ± 0.12	94.2 ± 0.04	94.6 ± 0.11	97.0 ± 0.08	95.5 ± 0.12
Graphite-AE	92.8 ± 0.13	94.1 ± 0.14	95.7 ± 0.06	94.5 ± 0.14	96.1 ± 0.12	$\mathbf{97.7} \pm 0.03$
Graphite-VAE	$\mathbf{93.2} \pm 0.13$	$\mathbf{95.0} \pm 0.10$	$\mathbf{96.0} \pm 0.03$	$\mathbf{94.9} \pm 0.13$	$\mathbf{97.4} \pm 0.06$	97.4 ± 0.04

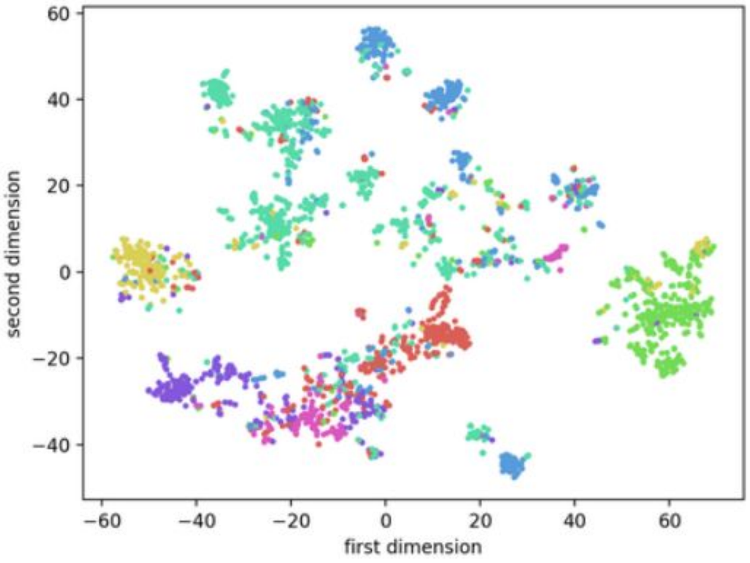
Node Classification

- Accuracy

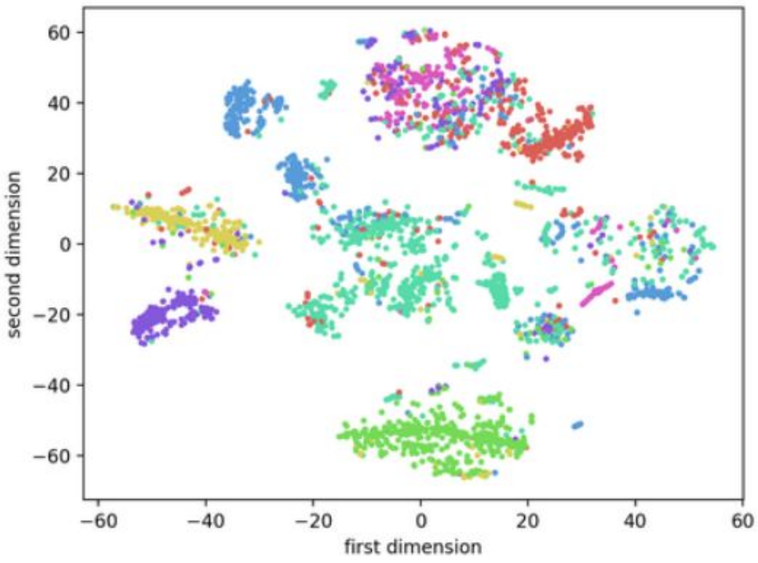
Table 5. Classification accuracies (* denotes dataset with features). Baseline numbers from Kipf & Welling (2017).

	Cora*	Citeseer*	Pubmed*
SemiEmb	59.0	59.6	71.1
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
Graphite	**82.1** $\pm$ 0.06	**71.0** $\pm$ 0.07	**79.3** $\pm$ 0.03

Visualization of Latent Space



(a) Graphite-AE



(b) Graphite-VAE

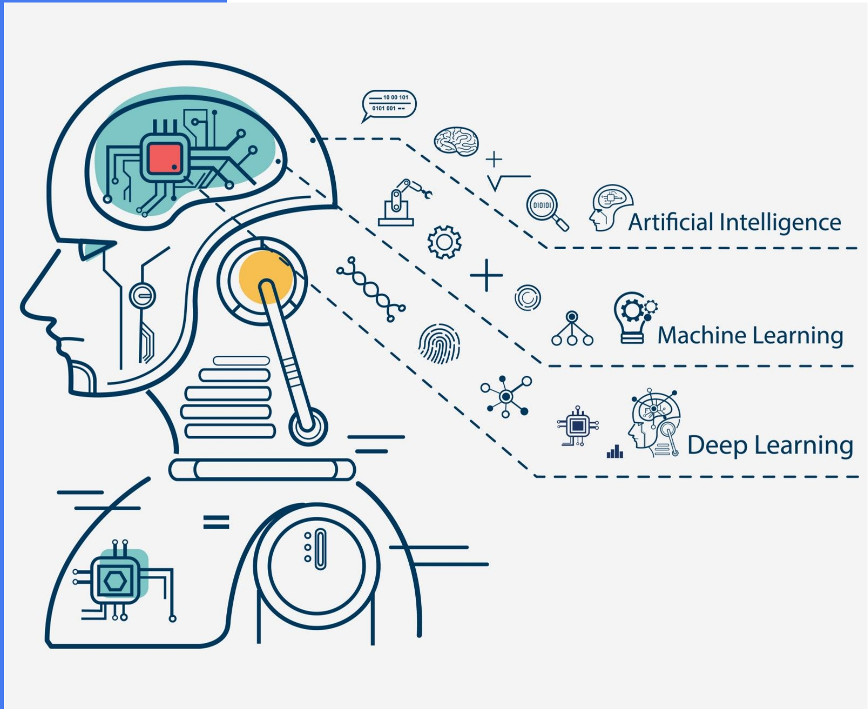
DISCUSSION
CONCLUSION

Proposed Graphite, an algorithmic for generative modeling of graphs.

Future and ongoing work entail applications of Graphite to other inference tasks.

THANKS!

Do you have any questions?
stevenzx123@gmail.com
elian.lu@ikala.tv
ikala.tv



Elian Lu

Please feel free to contact me

ALTERNATIVE RESOURCES

- Study group GitHub link
- The paper link
- The code of this paper imple
- Variational graph auto-encoders(GAE)
- The code of GAE
- What is AutoEncoder?
- Variational AutoEncoder
- Variational Inference