

EECE 571F: Deep Learning with Structures

Lecture 7: Deep Generative Models of Graphs VAEs and GANs

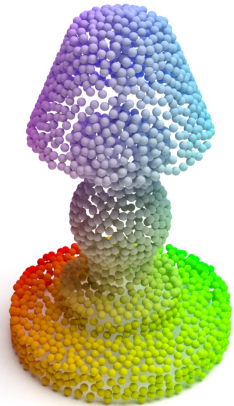
Renjie Liao

University of British Columbia

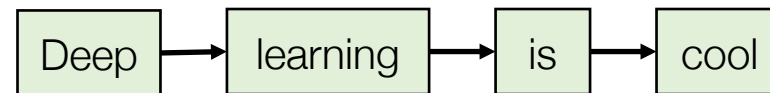
Winter, Term 2, 2021/22

Course Scope

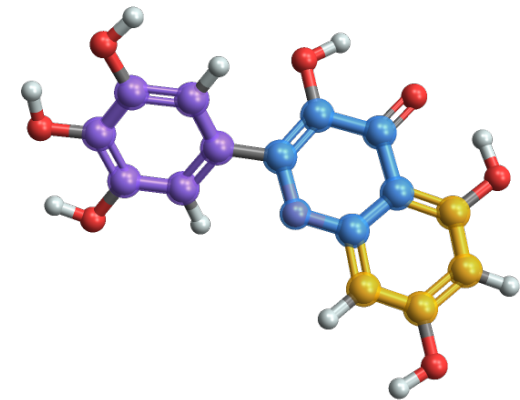
- Supervised Learning with Observable Structures
- **Unsupervised / Self-supervised Learning with Observable Structures**
- Supervised Learning with Latent Structures



Points/Sets

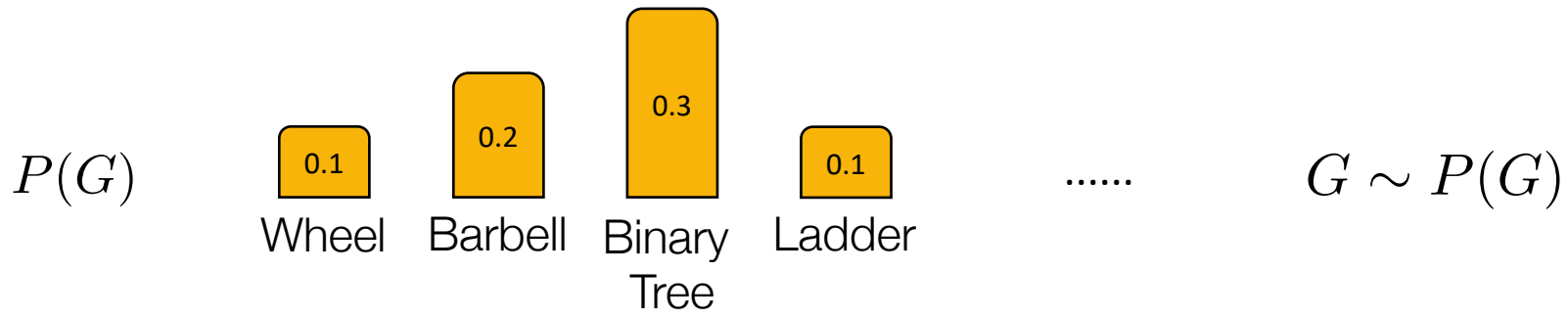
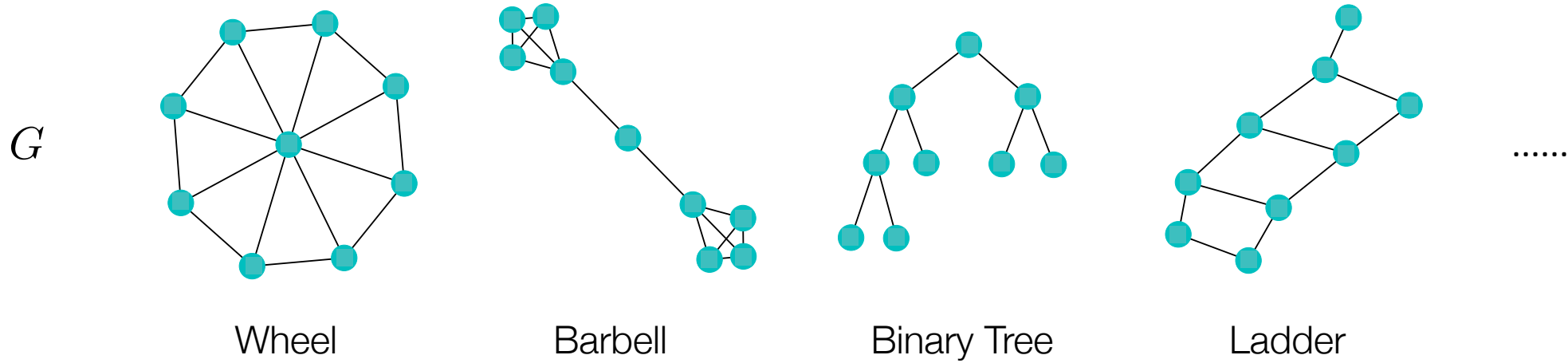


Lists/Sequences



Graphs

Deep Generative Models of Graphs



Deep Generative Models of Graphs

- Variational Auto-Encoders (VAEs), [1,2]
- Generative Adversarial Networks (GANs) [3]

Deep Generative Models of Graphs

- **Variational Auto-Encoders (VAEs)** [1,2]
- **Generative Adversarial Networks (GANs)** [3]

Deep Generative Models of Graphs

Given data $X \in \mathbb{R}^d$, Maximum Likelihood is:

$$\max_{\theta} \log p_{\theta}(X)$$

Deep Generative Models of Graphs

Given data $X \in \mathbb{R}^d$, Maximum Likelihood is:

$$\max_{\theta} \log p_{\theta}(X)$$

Variational Auto-Encoders (VAEs)

We introduce latent variable $Z \in \mathbb{R}^m$

Deep Generative Models of Graphs

Given data $X \in \mathbb{R}^d$, Maximum Likelihood is:

$$\max_{\theta} \log p_{\theta}(X)$$

Variational Auto-Encoders (VAEs)

We introduce latent variable $Z \in \mathbb{R}^m$

$$\begin{aligned} p_{\theta}(X) &= \int_{\mathcal{Z}} p_{\theta}(X, Z) dZ \\ &= \int_{\mathcal{Z}} p_{\theta}(X|Z) p_{\theta}(Z) dZ \end{aligned}$$

Deep Generative Models of Graphs

Given data $X \in \mathbb{R}^d$, Maximum Likelihood is:

$$\max_{\theta} \log p_{\theta}(X)$$

Variational Auto-Encoders (VAEs)

We introduce latent variable $Z \in \mathbb{R}^m$

$$\begin{aligned} p_{\theta}(X) &= \int_Z p_{\theta}(X, Z) dZ \\ &= \int_Z p_{\theta}(X|Z) p_{\theta}(Z) dZ \end{aligned}$$

Intractable Integration!

Deep Generative Models of Graphs

Variational Approximation

$$\begin{aligned}\log p_{\theta}(X) &= \log \left(\frac{p_{\theta}(X, Z)}{p_{\theta}(Z|X)} \right) \\ &= \log \left(\frac{p_{\theta}(X, Z) q_{\phi}(Z|X)}{q_{\phi}(Z|X) p_{\theta}(Z|X)} \right)\end{aligned}$$

Deep Generative Models of Graphs

Variational Approximation

$$\begin{aligned}\log p_\theta(X) &= \log \left(\frac{p_\theta(X, Z)}{p_\theta(Z|X)} \right) \\ &= \log \left(\frac{p_\theta(X, Z) q_\phi(Z|X)}{q_\phi(Z|X) p_\theta(Z|X)} \right)\end{aligned}$$

Integrating from both sides:

$$\begin{aligned}\log p_\theta(X) &= \int q_\phi(Z|X) \log p_\theta(X) dZ \\ &= \int q_\phi(Z|X) \log \left(\frac{p_\theta(X, Z) q_\phi(Z|X)}{q_\phi(Z|X) p_\theta(Z|X)} \right) dZ \\ &= \int q_\phi(Z|X) \log \left(\frac{p_\theta(X, Z)}{q_\phi(Z|X)} \right) dZ + \int q_\phi(Z|X) \log \left(\frac{q_\phi(Z|X)}{p_\theta(Z|X)} \right) dZ \\ &= \mathbb{E}_{q_\phi(Z|X)} \left[\log \left(\frac{p_\theta(X, Z)}{q_\phi(Z|X)} \right) \right] + \text{KL} (q_\phi(Z|X) || p_\theta(Z|X))\end{aligned}$$

Deep Generative Models of Graphs

Variational Approximation

$$\begin{aligned}\log p_{\theta}(X) &= \log \left(\frac{p_{\theta}(X, Z)}{p_{\theta}(Z|X)} \right) \\ &= \log \left(\frac{p_{\theta}(X, Z) q_{\phi}(Z|X)}{q_{\phi}(Z|X) p_{\theta}(Z|X)} \right)\end{aligned}$$

Integrating from both sides:

$$\begin{aligned}\log p_{\theta}(X) &= \int q_{\phi}(Z|X) \log p_{\theta}(X) dZ \\ &= \int q_{\phi}(Z|X) \log \left(\frac{p_{\theta}(X, Z) q_{\phi}(Z|X)}{q_{\phi}(Z|X) p_{\theta}(Z|X)} \right) dZ \\ &= \int q_{\phi}(Z|X) \log \left(\frac{p_{\theta}(X, Z)}{q_{\phi}(Z|X)} \right) dZ + \int q_{\phi}(Z|X) \log \left(\frac{q_{\phi}(Z|X)}{p_{\theta}(Z|X)} \right) dZ \\ &= \underbrace{\mathbb{E}_{q_{\phi}(Z|X)} \left[\log \left(\frac{p_{\theta}(X, Z)}{q_{\phi}(Z|X)} \right) \right]}_{\text{Evidence Lower Bound (ELBO)}} + \underbrace{\text{KL} (q_{\phi}(Z|X) || p_{\theta}(Z|X))}_{\text{Kullback-Leibler Divergence}}\end{aligned}$$

Evidence Lower Bound (ELBO) Kullback-Leibler Divergence

Deep Generative Models of Graphs

Variational Approximation

$$\begin{aligned}\log p_\theta(X) &= \log \left(\frac{p_\theta(X, Z)}{p_\theta(Z|X)} \right) \\ &= \log \left(\frac{p_\theta(X, Z) q_\phi(Z|X)}{q_\phi(Z|X) p_\theta(Z|X)} \right)\end{aligned}$$

Integrating from both sides:

$$\begin{aligned}\log p_\theta(X) &= \int q_\phi(Z|X) \log p_\theta(X) dZ \\ &= \int q_\phi(Z|X) \log \left(\frac{p_\theta(X, Z) q_\phi(Z|X)}{q_\phi(Z|X) p_\theta(Z|X)} \right) dZ \\ &= \int q_\phi(Z|X) \log \left(\frac{p_\theta(X, Z)}{q_\phi(Z|X)} \right) dZ + \int q_\phi(Z|X) \log \left(\frac{q_\phi(Z|X)}{p_\theta(Z|X)} \right) dZ \\ &= \underbrace{\mathbb{E}_{q_\phi(Z|X)} \left[\log \left(\frac{p_\theta(X, Z)}{q_\phi(Z|X)} \right) \right]}_{\text{Evidence Lower Bound (ELBO)}} + \underbrace{\text{KL} (q_\phi(Z|X) || p_\theta(Z|X))}_{\text{Kullback-Leibler Divergence}}\end{aligned}$$

Why is it a lower bound?

Why is it a variational approximation?

Evidence Lower Bound (ELBO) Kullback-Leibler Divergence

Deep Generative Models of Graphs

Since true posterior $p_{\theta}(Z|X)$ is often unknown, KL term is intractable

Deep Generative Models of Graphs

Since true posterior $p_\theta(Z|X)$ is often unknown, KL term is intractable

ELBO:

$$\begin{aligned}\mathbb{E}_{q_\phi(Z|X)} \left[\log \left(\frac{p_\theta(X, Z)}{q_\phi(Z|X)} \right) \right] &= \mathbb{E}_{q_\phi(Z|X)} \left[\log \left(\frac{p_\theta(X|Z)p_\theta(Z)}{q_\phi(Z|X)} \right) \right] \\ &= \mathbb{E}_{q_\phi(Z|X)} [\log(p_\theta(X|Z))] + \mathbb{E}_{q_\phi(Z|X)} \left[\log \left(\frac{p_\theta(Z)}{q_\phi(Z|X)} \right) \right] \\ &= -\mathbb{E}_{q_\phi(Z|X)} [-\log(p_\theta(X|Z))] - \text{KL}(q_\phi(Z|X) \| p_\theta(Z))\end{aligned}$$

Deep Generative Models of Graphs

Since true posterior $p_\theta(Z|X)$ is often unknown, KL term is intractable

ELBO:

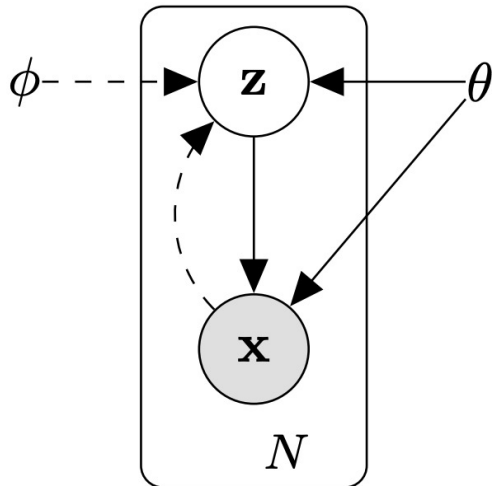
$$\begin{aligned}\mathbb{E}_{q_\phi(Z|X)} \left[\log \left(\frac{p_\theta(X, Z)}{q_\phi(Z|X)} \right) \right] &= \mathbb{E}_{q_\phi(Z|X)} \left[\log \left(\frac{p_\theta(X|Z)p_\theta(Z)}{q_\phi(Z|X)} \right) \right] \\ &= \mathbb{E}_{q_\phi(Z|X)} [\log(p_\theta(X|Z))] + \mathbb{E}_{q_\phi(Z|X)} \left[\log \left(\frac{p_\theta(Z)}{q_\phi(Z|X)} \right) \right] \\ &= \underbrace{-\mathbb{E}_{q_\phi(Z|X)} [-\log(p_\theta(X|Z))]}_{\text{Reconstruction Error/Loss}} - \underbrace{\text{KL}(q_\phi(Z|X) \| p_\theta(Z))}_{\text{Regularizer}}\end{aligned}$$

Deep Generative Models of Graphs

Since true posterior $p_\theta(Z|X)$ is often unknown, KL term is intractable

ELBO:

$$\begin{aligned}\mathbb{E}_{q_\phi(Z|X)} \left[\log \left(\frac{p_\theta(X, Z)}{q_\phi(Z|X)} \right) \right] &= \mathbb{E}_{q_\phi(Z|X)} \left[\log \left(\frac{p_\theta(X|Z)p_\theta(Z)}{q_\phi(Z|X)} \right) \right] \\ &= \mathbb{E}_{q_\phi(Z|X)} [\log(p_\theta(X|Z))] + \mathbb{E}_{q_\phi(Z|X)} \left[\log \left(\frac{p_\theta(Z)}{q_\phi(Z|X)} \right) \right] \\ &= \underbrace{-\mathbb{E}_{q_\phi(Z|X)} [-\log(p_\theta(X|Z))]}_{\text{Reconstruction Error/Loss}} - \underbrace{\text{KL}(q_\phi(Z|X) \| p_\theta(Z))}_{\text{Regularizer}}\end{aligned}$$



Encoder: $q_\phi(Z|X)$

Decoder: $p_\theta(X|Z)$

Prior: $p_\theta(Z)$

Deep Generative Models of Graphs

Graph Variational Auto-Encoders [4,5]

Node feature: $X \in \mathbb{R}^{n \times d}$

Node latent variables: $Z \in \mathbb{R}^{n \times m}$

Adjacency matrix: $A \in \mathbb{R}^{n \times n}$

Deep Generative Models of Graphs

Graph Variational Auto-Encoders [4,5]

Encoder:

$$q_\phi(Z|X, A) = \prod_i q_\phi(Z_i|X, A)$$
$$q_\phi(Z_i|X, A) = \mathcal{N}(Z_i|\mu_i, \sigma_i^2 I)$$
$$H = \text{GNN}_\phi(X, A)$$
$$\mu_i, \log \sigma_i^2 = \text{Readout}_\phi(H)$$

Deep Generative Models of Graphs

Graph Variational Auto-Encoders [4,5]

Encoder:

$$q_\phi(Z|X, A) = \prod_i q_\phi(Z_i|X, A)$$
$$q_\phi(Z_i|X, A) = \mathcal{N}(Z_i|\mu_i, \sigma_i^2 I)$$
$$H = \text{GNN}_\phi(X, A)$$
$$\mu_i, \log \sigma_i^2 = \text{Readout}_\phi(H)$$

Prior:

$$p(Z) = \prod_i p(Z_i) = \prod_i \mathcal{N}(Z_i|0, I)$$

Deep Generative Models of Graphs

Graph Variational Auto-Encoders [4,5]

Decoder:

$$p_{\theta}(X, A|Z) = p_{\theta}(A|Z)p_{\theta}(X|A, Z)$$

Deep Generative Models of Graphs

Graph Variational Auto-Encoders [4,5]

Decoder:

$$p_{\theta}(X, A|Z) = p_{\theta}(A|Z)p_{\theta}(X|A, Z)$$

Adjacency Matrix Decoder:

$$p_{\theta}(A|Z) = \prod_i \prod_j p_{\theta}(A_{ij}|Z)$$

$$H = \text{MLP}(Z)$$

$$p_{\theta}(A_{ij} = 1|Z_i, Z_j) = \sigma(H_i^{\top} H_j)$$

Deep Generative Models of Graphs

Graph Variational Auto-Encoders [4,5]

Decoder:

$$p_{\theta}(X, A|Z) = p_{\theta}(A|Z)p_{\theta}(X|A, Z)$$

Adjacency Matrix Decoder:

$$p_{\theta}(A|Z) = \prod_i \prod_j p_{\theta}(A_{ij}|Z)$$

$$H = \text{MLP}(Z)$$

$$p_{\theta}(A_{ij} = 1|Z_i, Z_j) = \sigma(H_i^{\top} H_j)$$

Node Feature Decoder:

$$p_{\theta}(X|A, Z) = \prod_i p_{\theta}(X_i|A, Z)$$

$$p_{\theta}(X_i|A, Z) = \mathcal{N}(X_i|\tilde{\mu}_i, \tilde{\sigma}_i^2 I)$$

$$\tilde{H} = \text{GNN}_{\theta}(Z, A)$$

$$\tilde{\mu}_i, \log \tilde{\sigma}_i^2 = \text{Readout}_{\theta}(\tilde{H})$$

Deep Generative Models of Graphs

Graph Variational Auto-Encoders [4,5]

Learning:

$$\begin{aligned}\log p_{\theta}(X, A) &\geq \text{ELBO} \\ &= -\mathbb{E}_{q_{\phi}(Z|A, X)} [-\log (p_{\theta}(X, A|Z))] - \text{KL} (q_{\phi}(Z|X, A) \| p_{\theta}(Z))\end{aligned}$$

Deep Generative Models of Graphs

Graph Variational Auto-Encoders [4,5]

Learning:

$$\begin{aligned}\log p_{\theta}(X, A) &\geq \text{ELBO} \\ &= -\mathbb{E}_{q_{\phi}(Z|A, X)} [-\log (p_{\theta}(X, A|Z))] - \text{KL} (q_{\phi}(Z|X, A) \| p_{\theta}(Z))\end{aligned}$$

Are we done?

Deep Generative Models of Graphs

Graph Variational Auto-Encoders [4,5]

Learning:

$$\begin{aligned}\log p_{\theta}(X, A) &\geq \text{ELBO} \\ &= -\mathbb{E}_{q_{\phi}(Z|A, X)} [-\log(p_{\theta}(X, A|Z))] - \text{KL}(q_{\phi}(Z|X, A) \| p_{\theta}(Z))\end{aligned}$$

Are we done?

No! We hope ELBO is permutation invariant!

$$\max_{\theta} \log \left(\sum_{P \in \Pi} p_{\theta}(PX, PAP^{\top}) \right)$$

Deep Generative Models of Graphs

Graph Variational Auto-Encoders [4,5]

How to approximately achieve permutation-invariant ELBO?

- Sample a few random permutations
(e.g., importance sampling, special permutations from domain knowledge)

Deep Generative Models of Graphs

Graph Variational Auto-Encoders [4,5]

How to approximately achieve permutation-invariant ELBO?

- Sample a few random permutations
(e.g., importance sampling, special permutations from domain knowledge)

$$\begin{aligned} \log \left(\sum_{P \in \Pi} p_{\theta}(PX, PAP^{\top}) \right) &\geq \log \left(\sum_{P \in S} p_{\theta}(PX, PAP^{\top}) \right) \\ &= \log \left(\sum_{P \in S} \exp(\log p_{\theta}(PX, PAP^{\top})) \right) \\ &\geq \log \left(\sum_{P \in S} \exp(\text{ELBO}) \right) \end{aligned}$$

Deep Generative Models of Graphs

Graph Variational Auto-Encoders [4,5]

More generalizations:

- Hierarchical encoder and decoder [6]
- Normalizing Flow based prior [7]

Deep Generative Models of Graphs

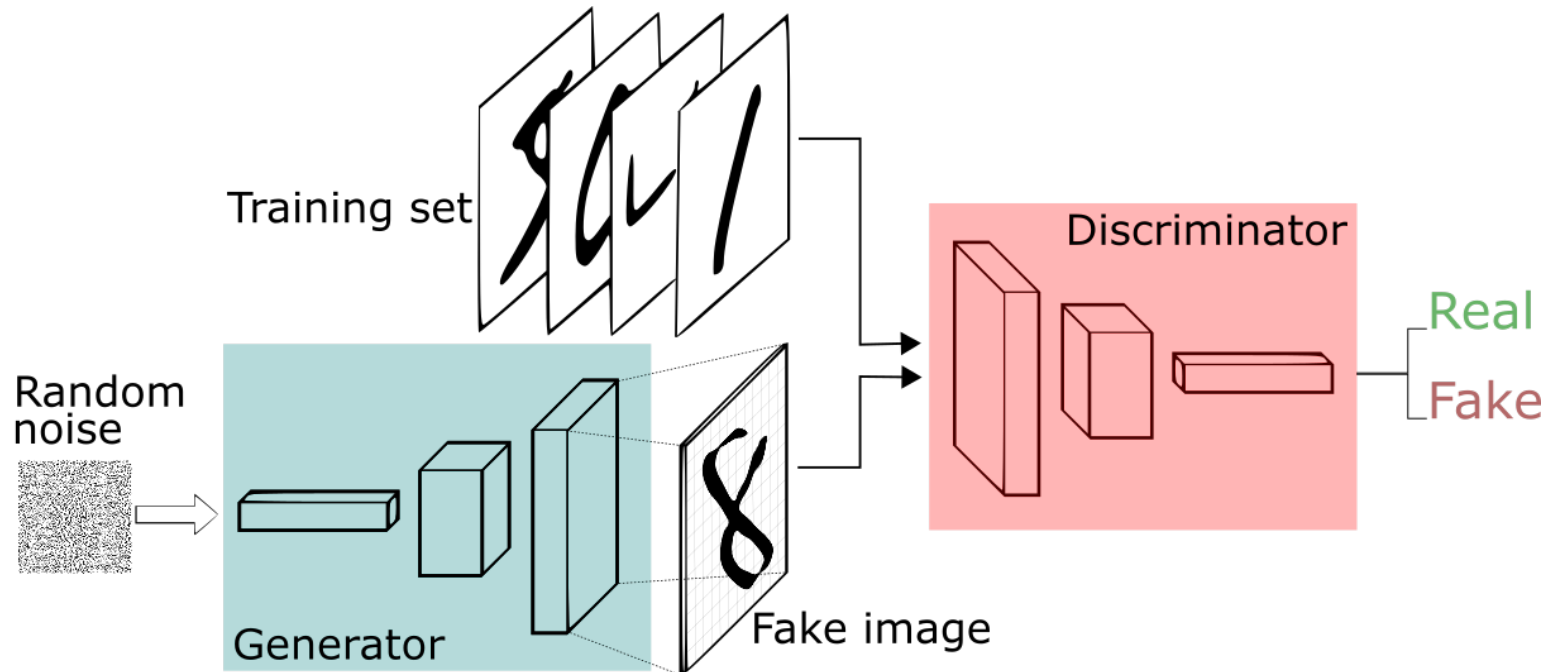
- Variational Auto-Encoders (VAEs) [1,2]
- Generative Adversarial Networks (GANs) [3]

Deep Generative Models of Graphs

- Variational Auto-Encoders (VAEs) [1,2]
- **Generative Adversarial Networks (GANs) [3]**

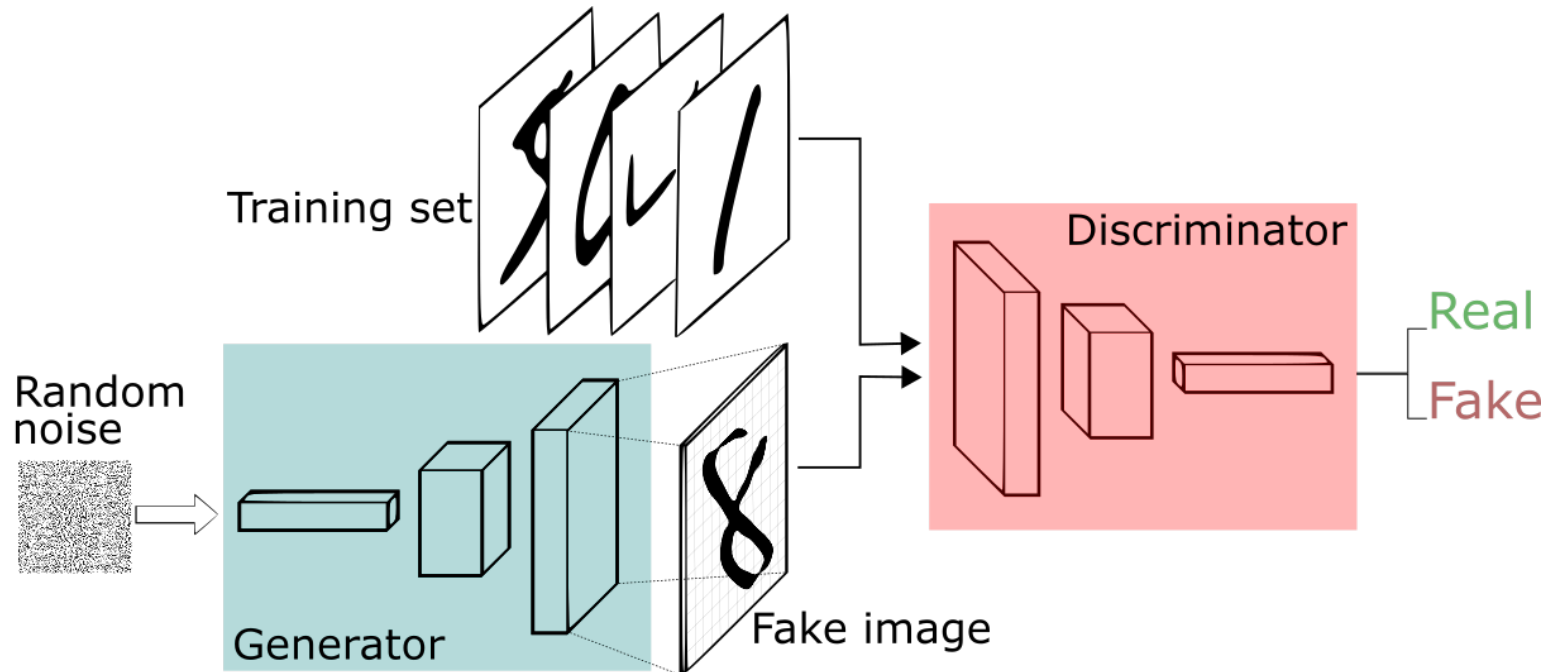
Deep Generative Models of Graphs

Generative Adversarial Networks (GANs) [3]



Deep Generative Models of Graphs

Generative Adversarial Networks (GANs) [3]



$$\text{Learning: } \min_{\theta} \max_{\phi} \mathbb{E}_{X \sim p_{\text{data}}(X)} [\log D_{\phi}(X)] + \mathbb{E}_{Z \sim p(Z)} [\log(1 - D_{\phi}(G_{\theta}(Z)))]$$

Deep Generative Models of Graphs

Generative Adversarial Networks (GANs) [3]

1. Fix generator, the optimal discriminator is

$$D_{\phi}^*(X) = \frac{p_{\text{data}}(X)}{p_{\text{data}}(X) + p_{G_{\theta}}(X)}$$

Deep Generative Models of Graphs

Generative Adversarial Networks (GANs) [3]

1. Fix generator, the optimal discriminator is

$$D_{\phi}^*(X) = \frac{p_{\text{data}}(X)}{p_{\text{data}}(X) + p_{G_{\theta}}(X)}$$

Why?

Deep Generative Models of Graphs

Generative Adversarial Networks (GANs) [3]

1. Fix generator, the optimal discriminator is

$$D_{\phi}^*(X) = \frac{p_{\text{data}}(X)}{p_{\text{data}}(X) + p_{G_{\theta}}(X)}$$

Why?

$$\begin{aligned} \ell(G_{\theta}, D_{\phi}) &= \mathbb{E}_{X \sim p_{\text{data}}(X)} [\log D_{\phi}(X)] + \mathbb{E}_{Z \sim p(Z)} [\log(1 - D_{\phi}(G_{\theta}(Z)))] \\ &= \mathbb{E}_{X \sim p_{\text{data}}(X)} [\log D_{\phi}(X)] + \mathbb{E}_{X \sim p_{G_{\theta}}(X)} [\log(1 - D_{\phi}(X))] \\ &= \int p_{\text{data}}(X) \log D_{\phi}(X) + p_{G_{\theta}}(X) \log(1 - D_{\phi}(X)) dX \end{aligned}$$

Deep Generative Models of Graphs

Generative Adversarial Networks (GANs) [3]

1. Fix generator, the optimal discriminator is

$$D_{\phi}^*(X) = \frac{p_{\text{data}}(X)}{p_{\text{data}}(X) + p_{G_{\theta}}(X)}$$

Why?

$$\begin{aligned} \ell(G_{\theta}, D_{\phi}) &= \mathbb{E}_{X \sim p_{\text{data}}(X)} [\log D_{\phi}(X)] + \mathbb{E}_{Z \sim p(Z)} [\log(1 - D_{\phi}(G_{\theta}(Z)))] \\ &= \mathbb{E}_{X \sim p_{\text{data}}(X)} [\log D_{\phi}(X)] + \mathbb{E}_{X \sim p_{G_{\theta}}(X)} [\log(1 - D_{\phi}(X))] \\ &= \int p_{\text{data}}(X) \log D_{\phi}(X) + p_{G_{\theta}}(X) \log(1 - D_{\phi}(X)) dX \end{aligned}$$

Law Of The Unconscious
Statistician (LOTUS)

Set the gradient of loss w.r.t. D to be zero

Deep Generative Models of Graphs

Generative Adversarial Networks (GANs) [3]

$$\begin{aligned} C(G_\theta) &= \max_{D_\phi} \ell(G_\theta, D_\phi) \\ &= \mathbb{E}_{X \sim p_{\text{data}}(X)} [\log D_\phi^*(X)] + \mathbb{E}_{X \sim p_{G_\theta}(X)} [\log(1 - D_\phi^*(X))] \\ &= \mathbb{E}_{X \sim p_{\text{data}}(X)} \left[\log \left(\frac{p_{\text{data}}(X)}{p_{\text{data}}(X) + p_{G_\theta}(X)} \right) \right] + \mathbb{E}_{X \sim p_{G_\theta}(X)} \left[\log \left(\frac{p_{G_\theta}(X)}{p_{\text{data}}(X) + p_{G_\theta}(X)} \right) \right] \end{aligned}$$

Deep Generative Models of Graphs

Generative Adversarial Networks (GANs) [3]

$$\begin{aligned} C(G_\theta) &= \max_{D_\phi} \ell(G_\theta, D_\phi) \\ &= \mathbb{E}_{X \sim p_{\text{data}}(X)} [\log D_\phi^*(X)] + \mathbb{E}_{X \sim p_{G_\theta}(X)} [\log(1 - D_\phi^*(X))] \\ &= \mathbb{E}_{X \sim p_{\text{data}}(X)} \left[\log \left(\frac{p_{\text{data}}(X)}{p_{\text{data}}(X) + p_{G_\theta}(X)} \right) \right] + \mathbb{E}_{X \sim p_{G_\theta}(X)} \left[\log \left(\frac{p_{G_\theta}(X)}{p_{\text{data}}(X) + p_{G_\theta}(X)} \right) \right] \end{aligned}$$

2. The global minimum of $C(G_\theta)$ is achieved iff. $p_{\text{data}}(X) = p_{G_\theta}(X)$

Why?

Deep Generative Models of Graphs

Generative Adversarial Networks (GANs) [3]

$$\begin{aligned} C(G_\theta) &= \max_{D_\phi} \ell(G_\theta, D_\phi) \\ &= \mathbb{E}_{X \sim p_{\text{data}}(X)} [\log D_\phi^*(X)] + \mathbb{E}_{X \sim p_{G_\theta}(X)} [\log(1 - D_\phi^*(X))] \\ &= \mathbb{E}_{X \sim p_{\text{data}}(X)} \left[\log \left(\frac{p_{\text{data}}(X)}{p_{\text{data}}(X) + p_{G_\theta}(X)} \right) \right] + \mathbb{E}_{X \sim p_{G_\theta}(X)} \left[\log \left(\frac{p_{G_\theta}(X)}{p_{\text{data}}(X) + p_{G_\theta}(X)} \right) \right] \end{aligned}$$

2. The global minimum of $C(G_\theta)$ is achieved iff. $p_{\text{data}}(X) = p_{G_\theta}(X)$

Why?

$$\begin{aligned} C(G_\theta) &= \mathbb{E}_{X \sim p_{\text{data}}(X)} \left[\log \left(\frac{p_{\text{data}}(X)}{(p_{\text{data}}(X) + p_{G_\theta}(X))/2} \right) \right] \\ &\quad + \mathbb{E}_{X \sim p_{G_\theta}(X)} \left[\log \left(\frac{p_{G_\theta}(X)}{(p_{\text{data}}(X) + p_{G_\theta}(X))/2} \right) \right] + 2 \log\left(\frac{1}{2}\right) \\ &= \text{JSD}(p_{\text{data}}(X) \| p_{G_\theta}(X)) - \log(4) \end{aligned}$$

Deep Generative Models of Graphs

Generative Adversarial Networks (GANs) [3]

2. The global minimum of $C(G_\theta)$ is achieved iff. $p_{\text{data}}(X) = p_{G_\theta}(X)$

Why?

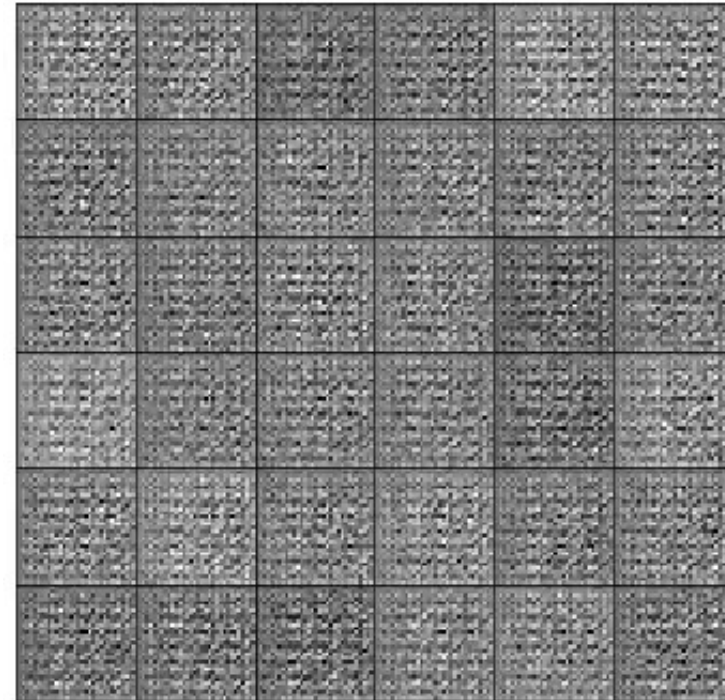
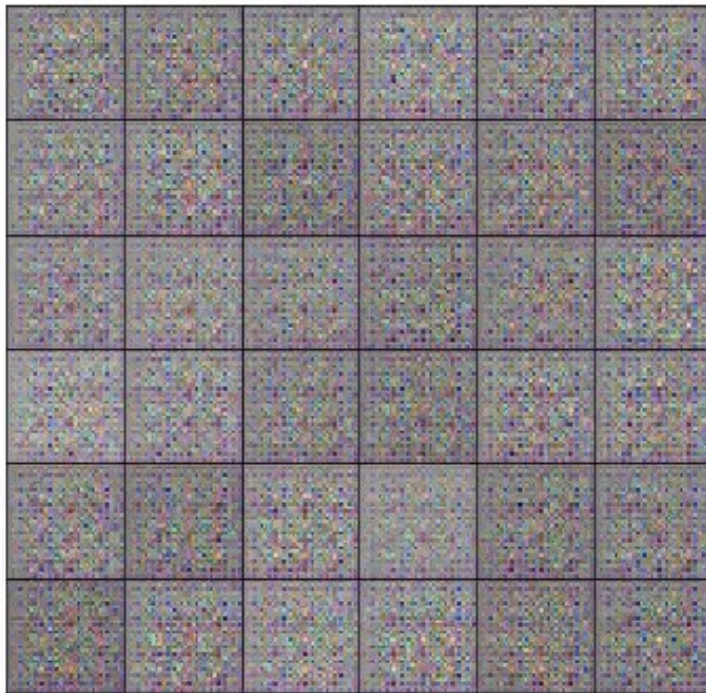
$$\begin{aligned} C(G_\theta) &= \mathbb{E}_{X \sim p_{\text{data}}(X)} \left[\log \left(\frac{p_{\text{data}}(X)}{(p_{\text{data}}(X) + p_{G_\theta}(X))/2} \right) \right] \\ &\quad + \mathbb{E}_{X \sim p_{G_\theta}(X)} \left[\log \left(\frac{p_{G_\theta}(X)}{(p_{\text{data}}(X) + p_{G_\theta}(X))/2} \right) \right] + 2 \log\left(\frac{1}{2}\right) \\ &= \text{JSD}(p_{\text{data}}(X) \| p_{G_\theta}(X)) - \log(4) \end{aligned}$$

Jensen–Shannon divergence is non-negative and is zero iff. $P = Q$

$$\text{JSD}(P \| Q) = \frac{1}{2} \text{KL}\left(P \left\| \frac{P + Q}{2}\right.\right) + \frac{1}{2} \text{KL}\left(Q \left\| \frac{P + Q}{2}\right.\right)$$

Deep Generative Models of Graphs

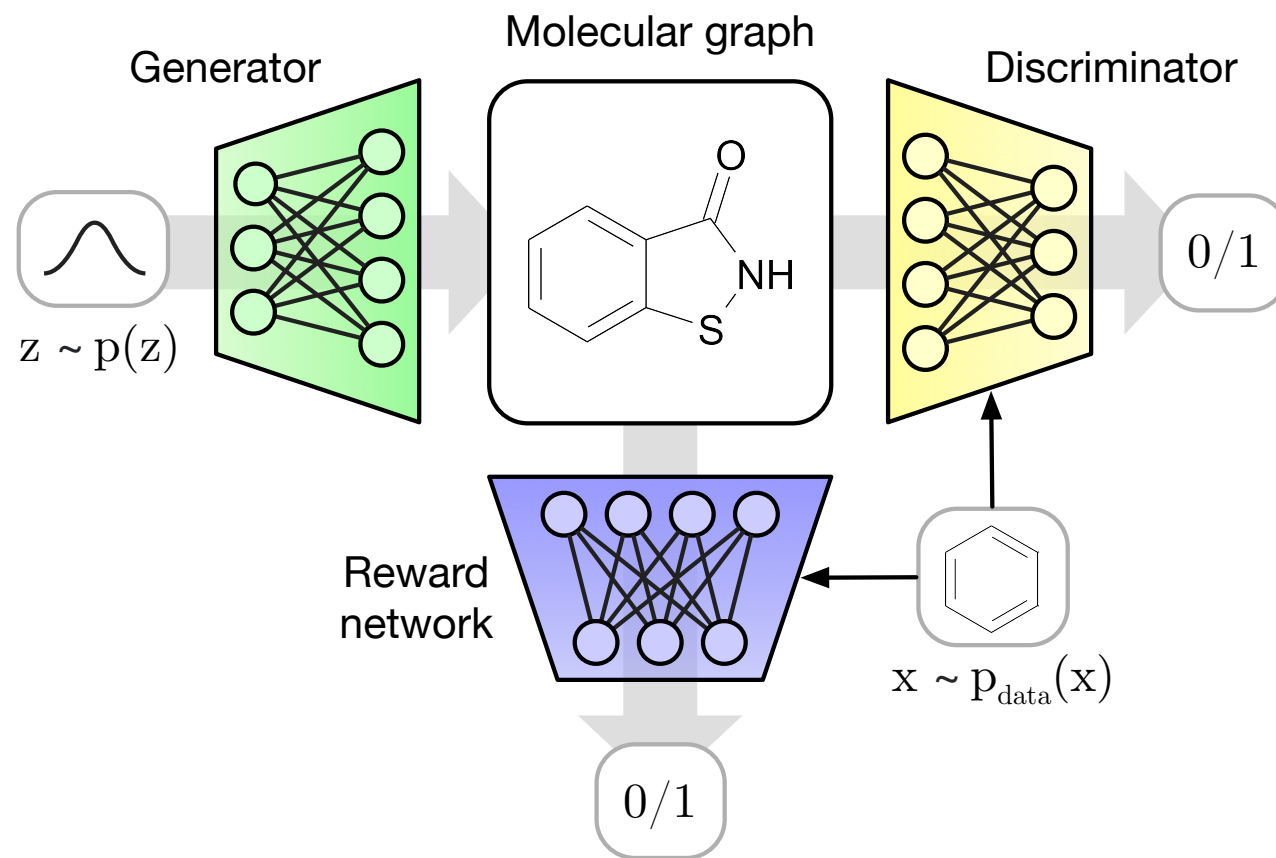
Generative Adversarial Networks (GANs) [3]



Samples from generator during training on SVHNs (left) and MNIST (right)

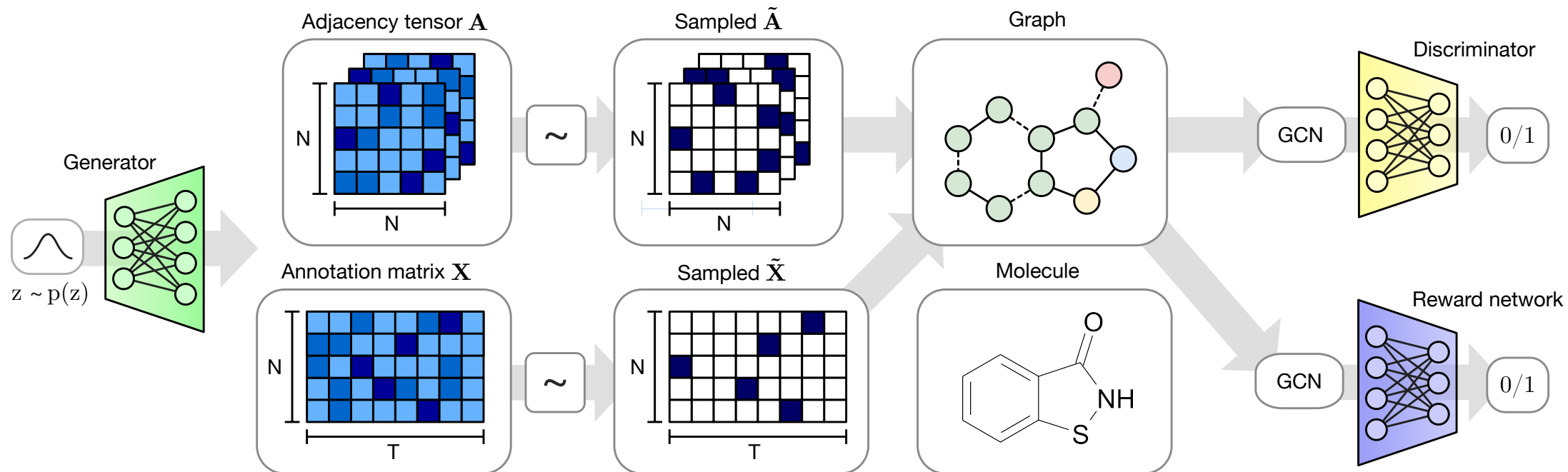
Deep Generative Models of Graphs

MolGAN [8]



Deep Generative Models of Graphs

MolGAN [8]



Deep Generative Models of Graphs

MolGAN [8]

Learning objective

$$\text{GAN:} \quad \min_{\theta} \max_{\phi} \mathbb{E}_{X \sim p_{\text{data}}(X)} [\log D_{\phi}(X)] + \mathbb{E}_{Z \sim p(Z)} [\log(1 - D_{\phi}(G_{\theta}(Z)))]$$

Deep Generative Models of Graphs

MolGAN [8]

Learning objective

$$\text{GAN: } \min_{\theta} \max_{\phi} \mathbb{E}_{X \sim p_{\text{data}}(X)} [\log D_{\phi}(X)] + \mathbb{E}_{Z \sim p(Z)} [\log(1 - D_{\phi}(G_{\theta}(Z)))]$$

Wasserstein distance (using Kantorovich-Rubinstein duality)

$$D(p||q) = \sup_{\|f\|_L \leq K} \mathbb{E}_{X \sim p}[f(X)] - \mathbb{E}_{X \sim q}[f(X)]$$

Wasserstein-GAN [9]:

$$\min_{\theta} \max_{\phi} \mathbb{E}_{X \sim p_{\text{data}}(X)} [D_{\phi}(X)] - \mathbb{E}_{X \sim p_{G_{\theta}}(X)} [D_{\phi}(X)]$$

References

- [1] Kingma, D.P. and Welling, M., 2013. Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114.
- [2] Rezende, D.J., Mohamed, S. and Wierstra, D., 2014, June. Stochastic backpropagation and approximate inference in deep generative models. In International conference on machine learning (pp. 1278-1286). PMLR.
- [3] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y., 2014. Generative adversarial nets. Advances in neural information processing systems, 27.
- [4] Kipf, T.N. and Welling, M., 2016. Variational graph auto-encoders. arXiv preprint arXiv:1611.07308.
- [5] Simonovsky, M. and Komodakis, N., 2018, October. Graphvae: Towards generation of small graphs using variational autoencoders. In International conference on artificial neural networks (pp. 412-422). Springer, Cham.
- [6] Jin, W., Barzilay, R. and Jaakkola, T., 2018, July. Junction tree variational autoencoder for molecular graph generation. In International conference on machine learning (pp. 2323-2332). PMLR.
- [7] Liu, J., Kumar, A., Ba, J., Kiros, J. and Swersky, K., 2019. Graph normalizing flows. Advances in Neural Information Processing Systems, 32.
- [8] De Cao, N. and Kipf, T., 2018. MolGAN: An implicit generative model for small molecular graphs. arXiv preprint arXiv:1805.11973.
- [9] Arjovsky, M., Chintala, S. and Bottou, L., 2017, July. Wasserstein generative adversarial networks. In International conference on machine learning (pp. 214-223). PMLR.

Questions?