EECE 571F: Deep Learning with Structures

Lecture 8: Deep Generative Models of Graphs
VAEs and GANs

Renjie Liao

University of British Columbia

Winter, Term 1, 2022

Course Scope

- Brief Intro to Deep Learning

- Geometric Deep Learning

  - Deep Learning Models for Sets and Sequences: Deep Sets & Transformers
  - Deep Learning Models for Graphs: Graph Convolution & Message Passing GNNs
  - Expressiveness & Generalizations of GNNs
  - Unsupervised/Self-supervised Graph Representation Learning

- Probabilistic Deep Learning

  - Deep Generative Models:
    Auto-regressive models, GANs, VAEs, Diffusion/Score based models
  - Discrete/Hybrid Latent Variable Models: RBMs, Latent Graph Models
  - Stochastic Gradient Estimation

Course Scope

- Brief Intro to Deep Learning

- Geometric Deep Learning

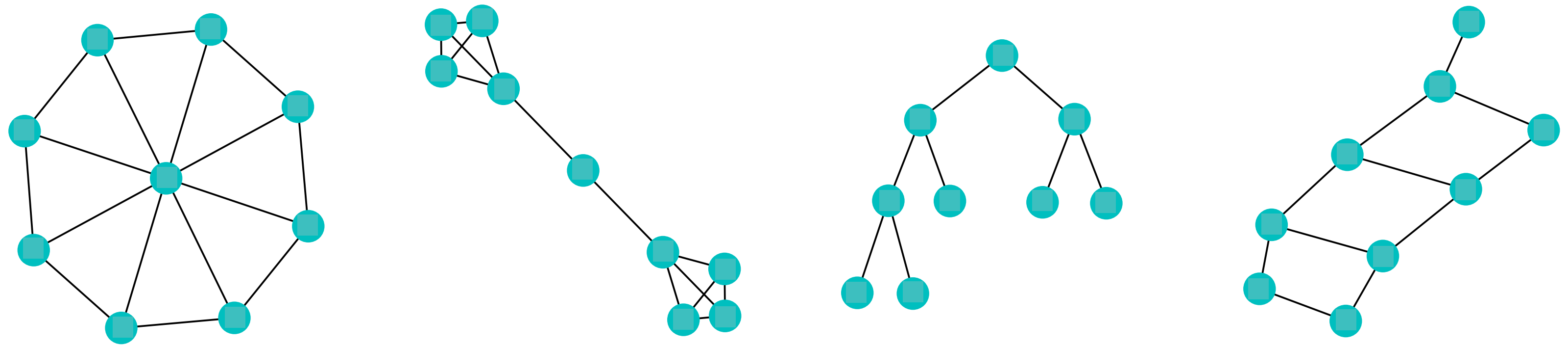
    - Deep Learning Models for Sets and Sequences: Deep Sets & Transformers
    - Deep Learning Models for Graphs: Graph Convolution & Message Passing GNNs
    - Expressiveness & Generalizations of GNNs
    - Unsupervised/Self-supervised Graph Representation Learning

- Probabilistic Deep Learning

    - **Deep Generative Models**:
      Auto-regressive models, **GANs, VAEs**, Diffusion/Score based models
    - Discrete/Hybrid Latent Variable Models: RBMs, Latent Graph Models
    - Stochastic Gradient Estimation

Deep Generative Models of Graphs

G



Wheel

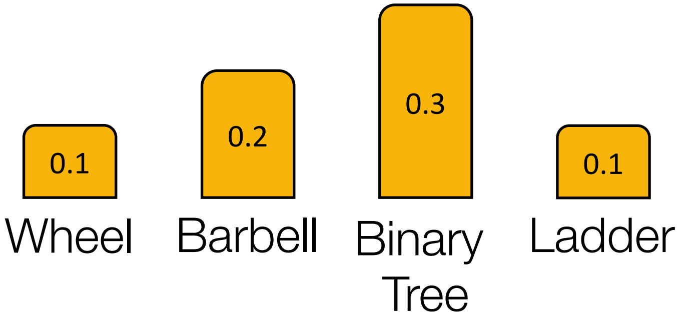
Barbell

Binary Tree

Ladder

......

$P(G)$



......

$G \sim P(G)$

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:

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

Deep Generative Models of Graphs

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

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

Variational Auto-Encoders (VAEs)

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

Deep Generative Models of Graphs

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

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

Variational Auto-Encoders (VAEs)

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

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

Deep Generative Models of Graphs

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

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

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

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

Deep Generative Models of Graphs

Variational Approximation

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

Integrating from both sides:

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

Deep Generative Models of Graphs

Variational Approximation

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

Integrating from both sides:

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

Deep Generative Models of Graphs

Variational Approximation

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

Integrating from both sides:

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

Why is it a lower bound?
Why is it a variational approximation?

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:

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

Deep Generative Models of Graphs

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

ELBO:

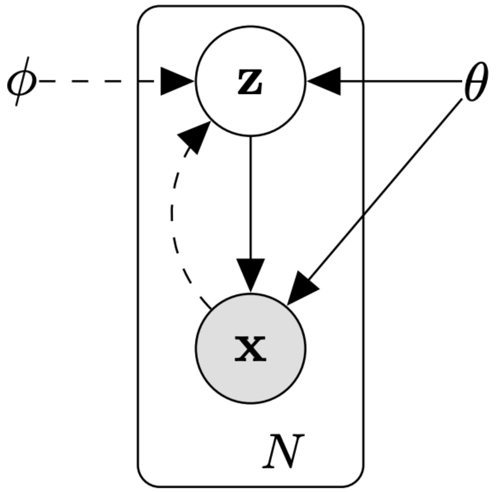
$$\begin{aligned}
\mathbb{E}_{q_\phi(Z|X)}\left[\log\left(\frac{p_\theta(X,Z)}{q_\phi(Z|X)}\right)\right] &= \mathbb{E}_{q_\phi(Z|X)}\left[\log\left(\frac{p_\theta(X|Z)p_\theta(Z)}{q_\phi(Z|X)}\right)\right] \\
&= \mathbb{E}_{q_\phi(Z|X)}\left[\log\left(p_\theta(X|Z)\right)\right] + \mathbb{E}_{q_\phi(Z|X)}\left[\log\left(\frac{p_\theta(Z)}{q_\phi(Z|X)}\right)\r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

Deep Generative Models of Graphs

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

ELBO:

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



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:

$$q_\phi(Z|X,A) = \prod_i q_\phi(Z_i|X,A)$$

$$q_\phi(Z_i|X,A) = \mathcal{N}(Z_i|\mu_i, \sigma_i^2 I)$$

$$H = \mathrm{GNN}_\phi(X,A)$$

$$\mu_i, \log \sigma_i^2 = \mathrm{Readout}_\phi(H)$$

Deep Generative Models of Graphs

Graph Variational Auto-Encoders [4,5]

Encoder:

$$q_\phi(Z|X,A) = \prod_i q_\phi(Z_i|X,A)$$

$$q_\phi(Z_i|X,A) = \mathcal{N}(Z_i|\mu_i, \sigma_i^2 I)$$

$$H = \mathrm{GNN}_\phi(X,A)$$

$$\mu_i, \log\sigma_i^2 = \mathrm{Readout}_\phi(H)$$

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:
$$p_\theta(A|Z) = \prod_i \prod_j p_\theta(A_{ij}|Z)$$
$$H = \mathrm{MLP}(Z)$$
$$p_\theta(A_{ij} = 1|Z_i, Z_j) = \sigma(H_i^\top H_j)$$

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:

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

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

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

Node Feature Decoder:

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

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

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

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

Deep Generative Models of Graphs

Graph Variational Auto-Encoders [4,5]

Learning:

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

Deep Generative Models of Graphs

Graph Variational Auto-Encoders [4,5]

Learning:

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

Are we done?

Deep Generative Models of Graphs

Graph Variational Auto-Encoders [4,5]

Learning:

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

Are we done?

No! We hope ELBO is permutation invariant!

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

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)

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

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]

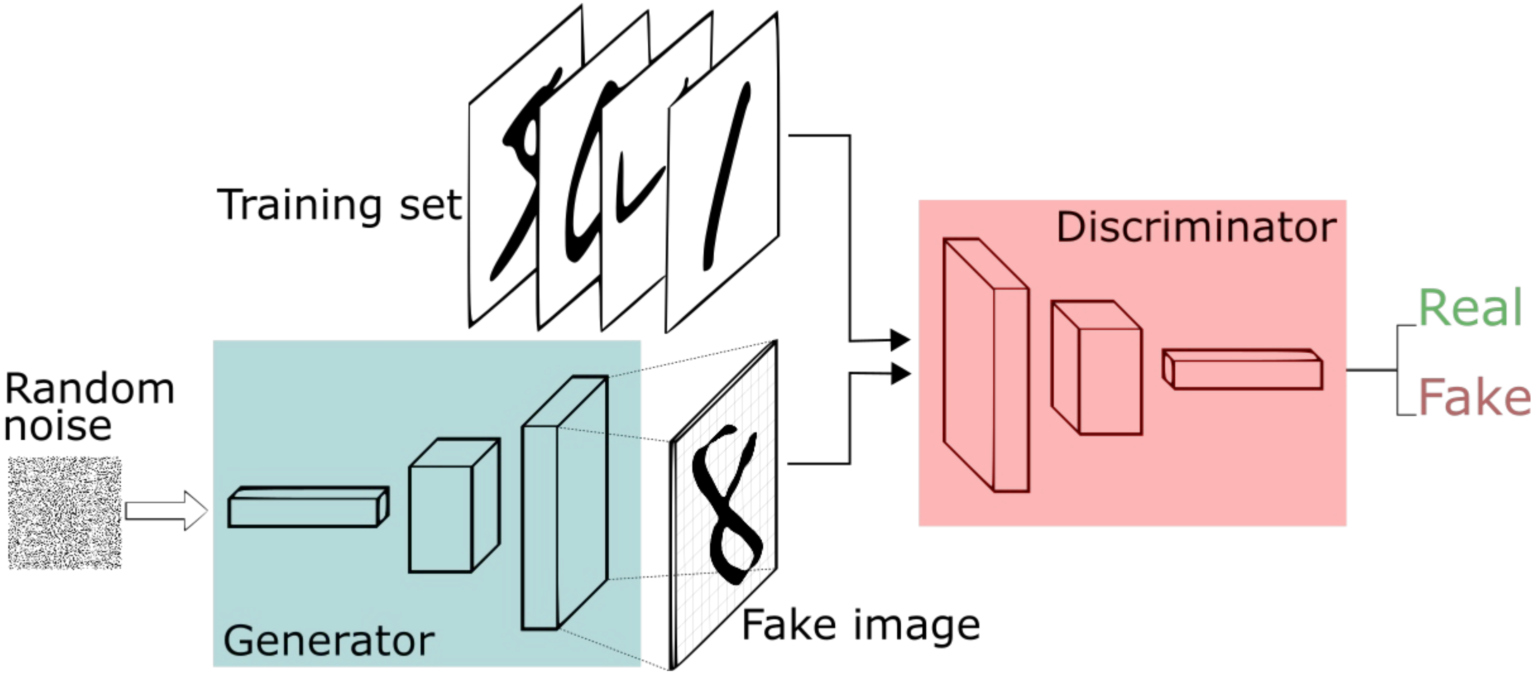
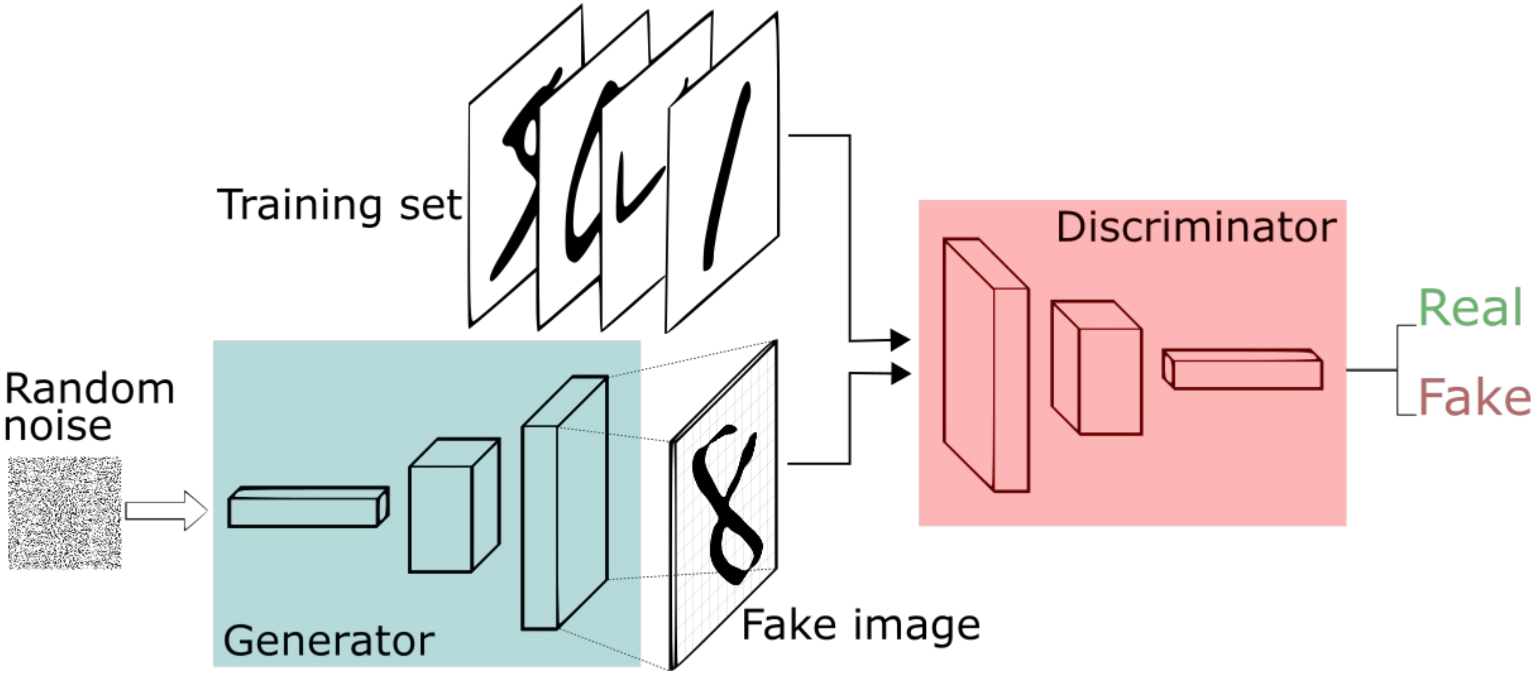


Image Credit: https://sthalles.github.io/intro-to-gans/

Deep Generative Models of Graphs

Generative Adversarial Networks (GANs) [3]



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

Image Credit: https://sthalles.github.io/intro-to-gans/

Deep Generative Models of Graphs

Generative Adversarial Networks (GANs) [3]

1. Fix generator, the optimal discriminator is $$D_\phi^*(X) = \frac{p_{\mathrm{data}}(X)}{p_{\mathrm{data}}(X) + p_{G_\theta}(X)}$$

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]

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

Deep Generative Models of Graphs

Generative Adversarial Networks (GANs) [3]

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

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]

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

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

Why?

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

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?

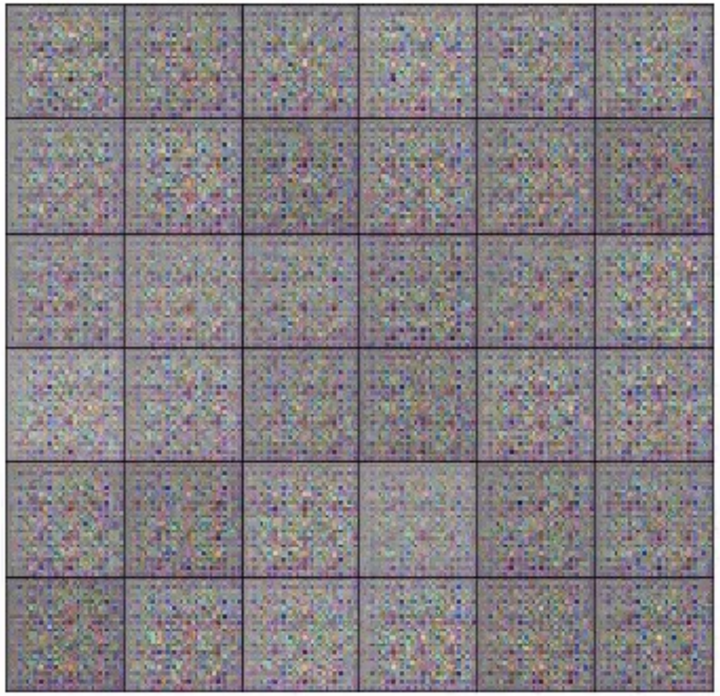
$$\begin{aligned} C(G_\theta) &= \mathbb{E}_{X \sim p_{\text{data}}(X)}\left[\log\left(\frac{p_{\text{data}}(X)}{(p_{\text{data}}(X) + p_{G_\theta}(X))/2}\right)\right] \\ &+ \mathbb{E}_{X \sim p_{G_\theta}(X)}\left[\log\left(\frac{p_{G_\theta}(X)}{(p_{\text{data}}(X) + p_{G_\theta}(X))/2}\right)\right]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

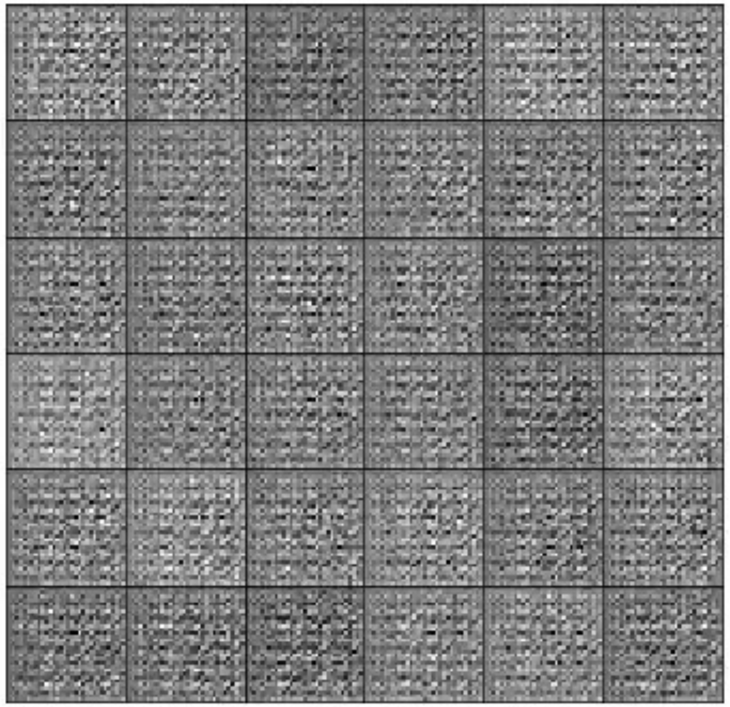
Jensen–Shannon divergence is non-negative and is zero iff. P = Q

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

Deep Generative Models of Graphs

Generative Adversarial Networks (GANs) [3]





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

Image Credit: https://sthalles.github.io/intro-to-gans/

Deep Generative Models of Graphs

MolGAN [8]

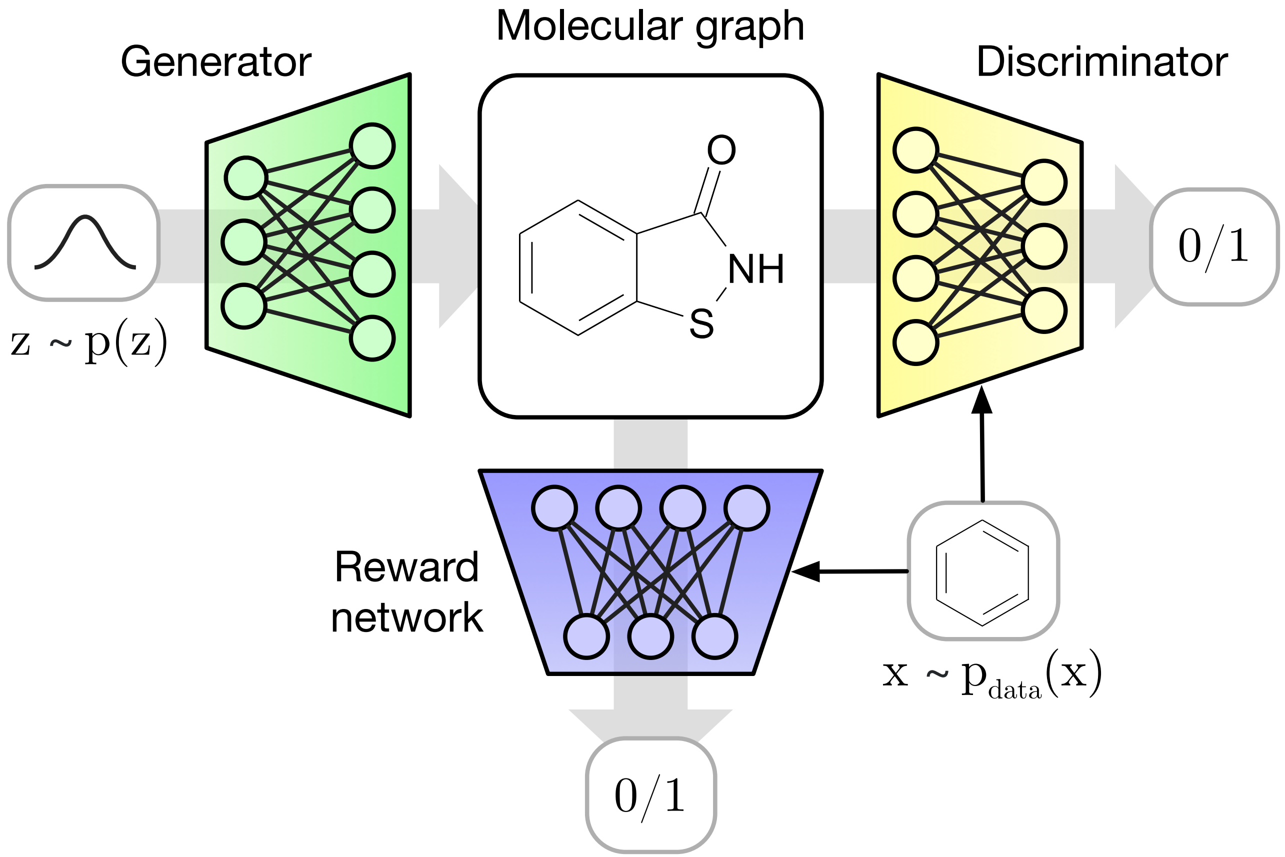


Image Credit: [8]

Deep Generative Models of Graphs

MolGAN [8]

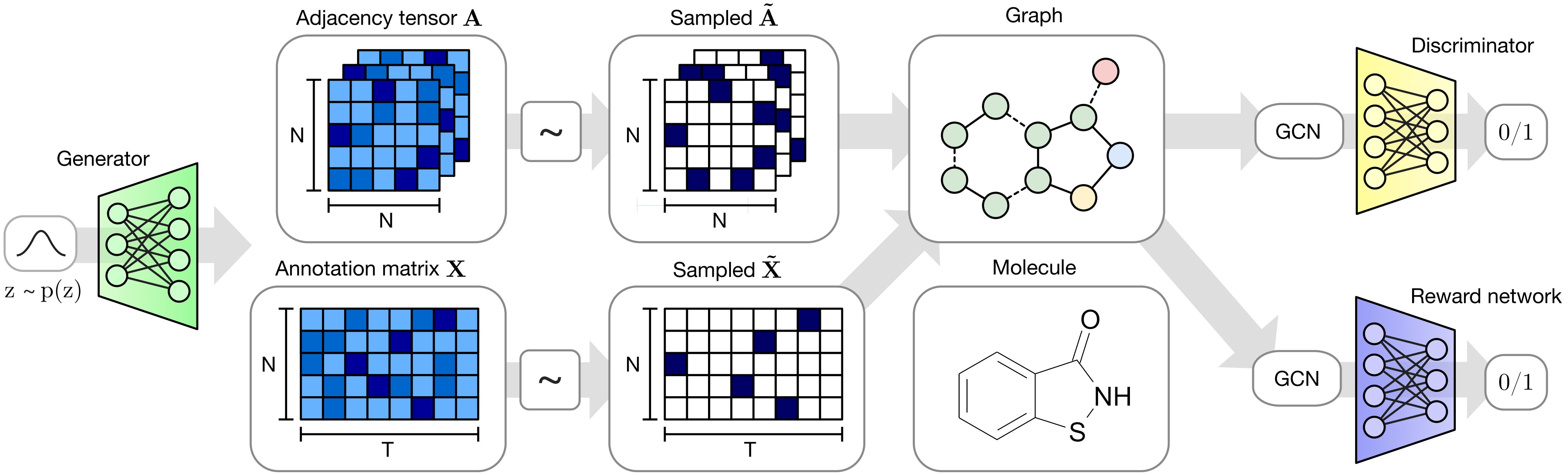


Image Credit: [8]

Deep Generative Models of Graphs

MolGAN [8]

Learning objective

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

Deep Generative Models of Graphs

MolGAN [8]

Learning objective

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

Wasserstein distance (using Kantorovich-Rubinstein duality)

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

Wasserstein-GAN [9]:

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

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?