Learning on **Attribute-Missing Graphs** using **Structure Attribute Transformer**

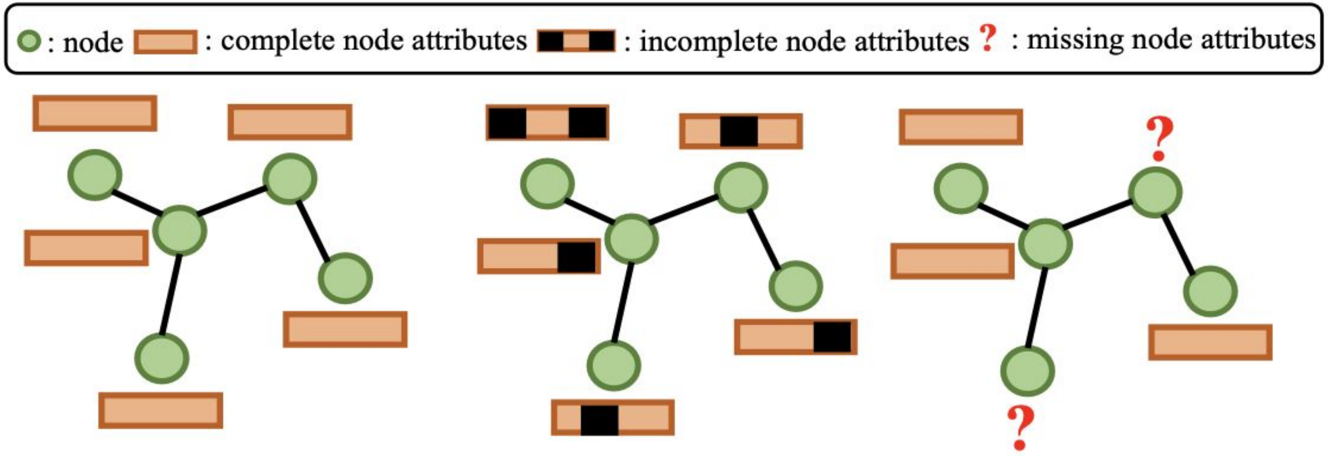
Presentation on the module Analysing Networks
3044093 Sthitadhee Panthadas

Course: Analysing Networks
Prof. Dr. Ulf Brefeld

LEUPHANA
UNIVERSITÄT LÜNEBURG

Table Of Contents

Introduction ~ Motivation



Attribute complete

Attribute incomplete

Attribute missing

Foundation

Autoencoders

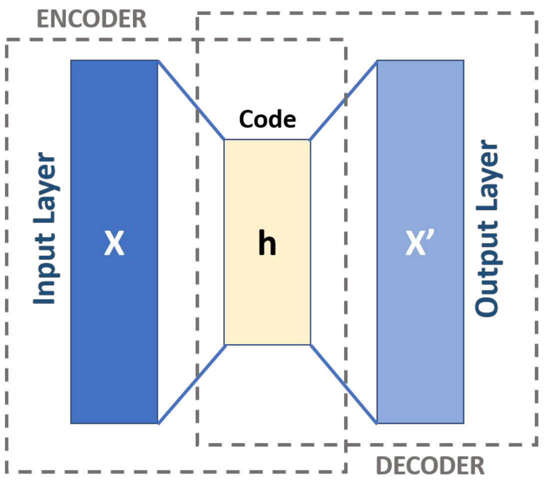


Fig 1: Autoencoder [1]

$$\textbf{Encoder}:\ h = f(x) = \sigma(W_e x + b_e)$$

$$\textbf{Decoder}:\ \hat{x} = g(h) = \sigma(W_d h + b_d)$$

$$\textbf{Loss}:\quad \mathcal{L}(x, \hat{x}) = \frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{x}_i)^2$$

Foundation

Variational Autoencoder (VAE)

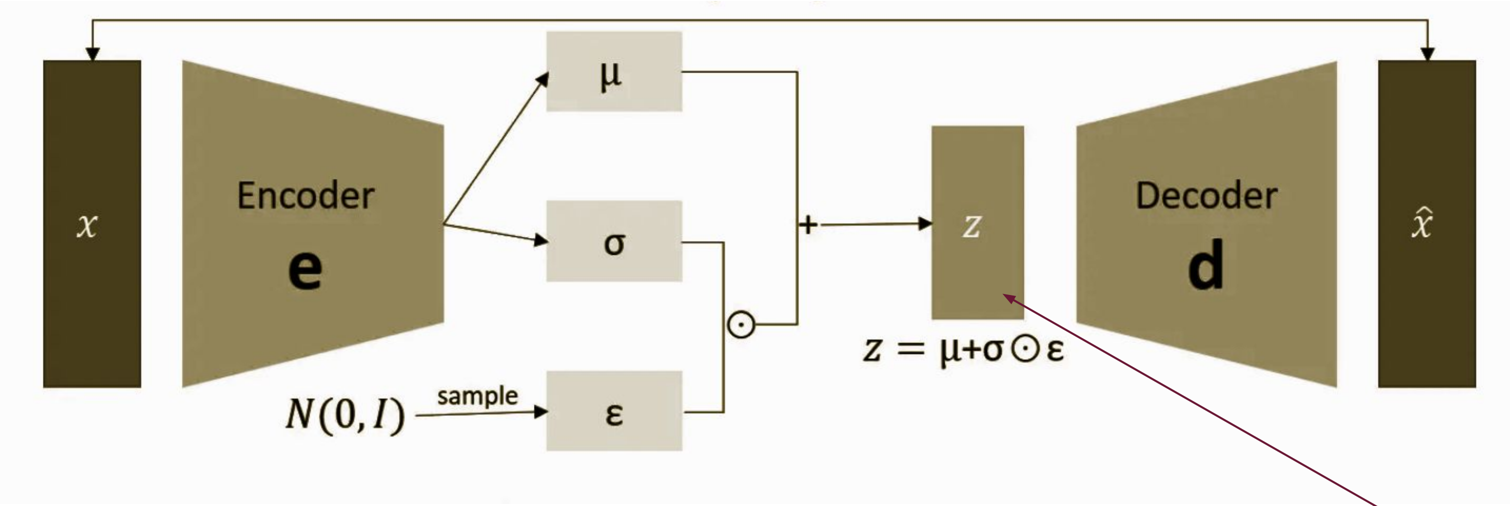


Fig 2: VAE [2]

Latent represent or hidden layer

$$\textbf{Loss}: \quad \mathcal{L}_{vae} = \mathcal{L}(x, \hat{x}) + \mathcal{L}_{regularization}$$

$$\textbf{Prior}: \quad \vec{z} \sim \mathcal{N}(\vec{\mu}, \sigma^2 \mathbf{I})$$

Foundation

KL Divergence

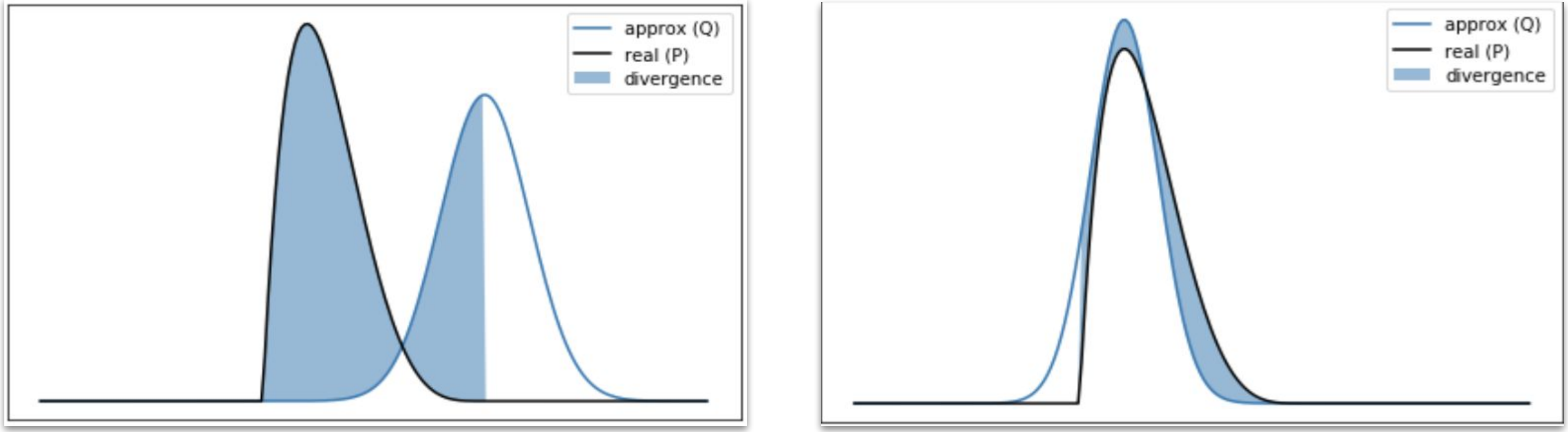


Fig 3: KL divergence [3]

$$KL_D\,(P\,||\,Q) = \sum P \log P/Q$$

Foundation

Generative Adversarial Learning

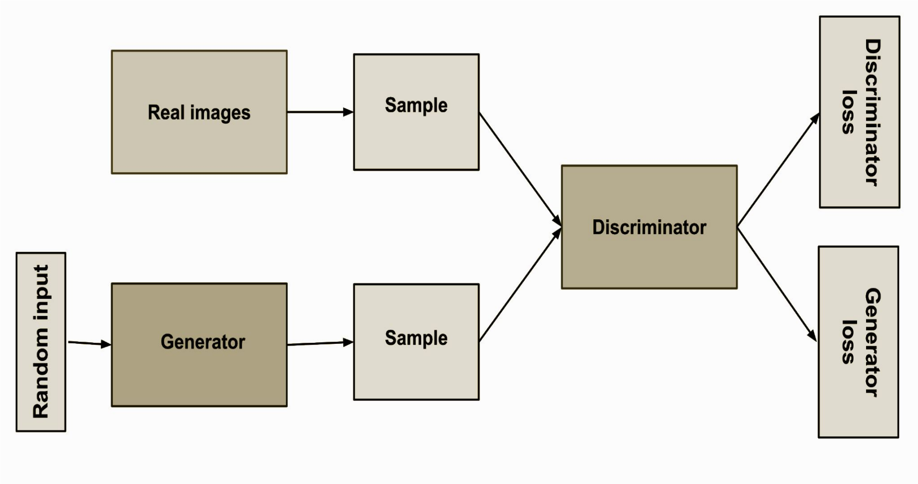
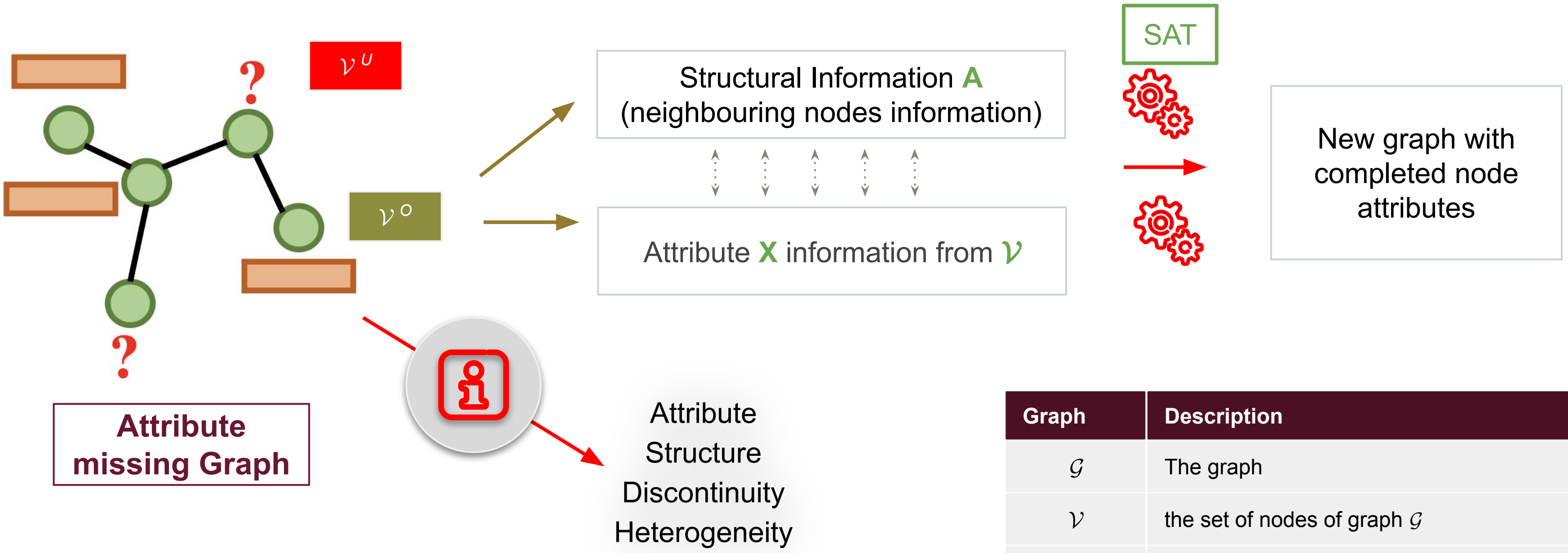


Fig 4: GAN [4]

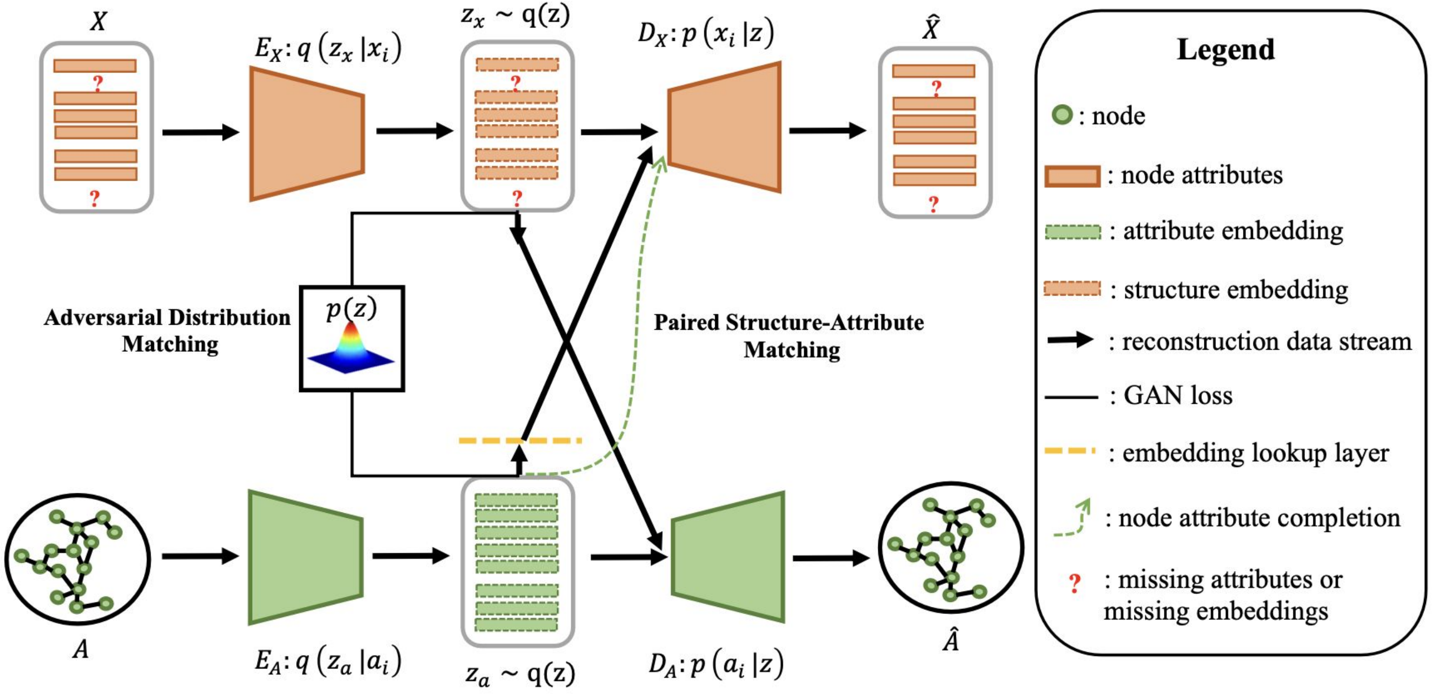
$$\mathcal{L}_{adv} = \mathbb{E}_{x_{real}}[\log D(x_{real})] + \mathbb{E}_z[1 - \log D(z)]$$

Model formulation



Graph	Description
$\mathcal{G}$	The graph
$\mathcal{V}$	the set of nodes of graph $\mathcal{G}$
A	the adjacent matrix of graph $\mathcal{G}$
X	the attribute matrix of graph $\mathcal{G}$
$\mathcal{V}^O$	the set of attribute-observed nodes $\mathcal{G}$
$\mathcal{V}^U$	the set of attribute-unobserved nodes $\mathcal{G}$

Structure Attribute Transformer



Likelihood And Loss Function

Symbol	Description
q	approximate posterior
p	true posterior
z_x	latent Factor with respect to attribute
z_a	latent Factor with respect to structure
D_{KL}	KL divergence
L	Loss function
E	Expectation with respect to a distribution

$$\mathbb{E}_{x\sim p(x)}[f(x)] = \sum p(x)f(x)$$

Joint Probability

$$\log p_\theta(x_i, a_i) = D_{KL}[q_\phi(z_x, z_a|x_i, a_i)||p_\theta(z_x, z_a|x_i, a_i)] + \mathcal{L}(\theta, \phi; x_i, a_i)$$

$$\mathcal{L}(\theta, \phi; x_i, a_i) = \underbrace{\mathbb{E}_{q_\phi(z_x, z_a|x_i, a_i)}[\log p_\theta(x_i, a_i|z_x, z_a)]}_{\text{Reconstruction Loss}} - \underbrace{D_{KL}[q_\phi(z_x, z_a|x_i, a_i)||p(z_x, z_a)]}_{\text{Prior regularization loss}}$$

Reconstruction Loss

Prior regularization loss

Paired Structure-Attribute Matching

Problem

Coupling theory states that there exists infinite joint distribution formulations that can reach the given marginal distributions

Heterogeneity and Discontinuity

Shared Latent space assumption

z_x is **independent** of z_a **given** x_i and a_i

x_i is **independent** of a_i z_x and z_a

$$
\begin{aligned}
q_\phi(z_x, z_a | x_i, a_i) &= q_\phi(z_x | x_i, a_i) q_\phi(z_a | \boxed{z_x}, x_i, a_i) \\
&= q_{\phi_x}(z_x | x_i, \boxed{a_i}) q_{\phi_a}(z_a | \boxed{x_i}, a_i) \\
&= q_{\phi_x}(z_x | x_i) q_{\phi_a}(z_a | a_i)
\end{aligned}
$$

1

$$
\begin{aligned}
p_\theta(x_i, a_i | z_x, z_a) &= p_\theta(x_i | z_x, z_a) p_\theta(a_i | \boxed{x_i}, z_x, z_a) \\
&= p_{\theta_x}(x_i | z_x, z_a) p_{\theta_a}(a_i | z_x, z_a) \\
&= p_{\theta_x}(x_i | z_x) p_{\theta_x}(x_i | z_a) p_{\theta_a}(a_i | z_a) p_{\theta_a}(a_i | z_x)
\end{aligned}
$$

2

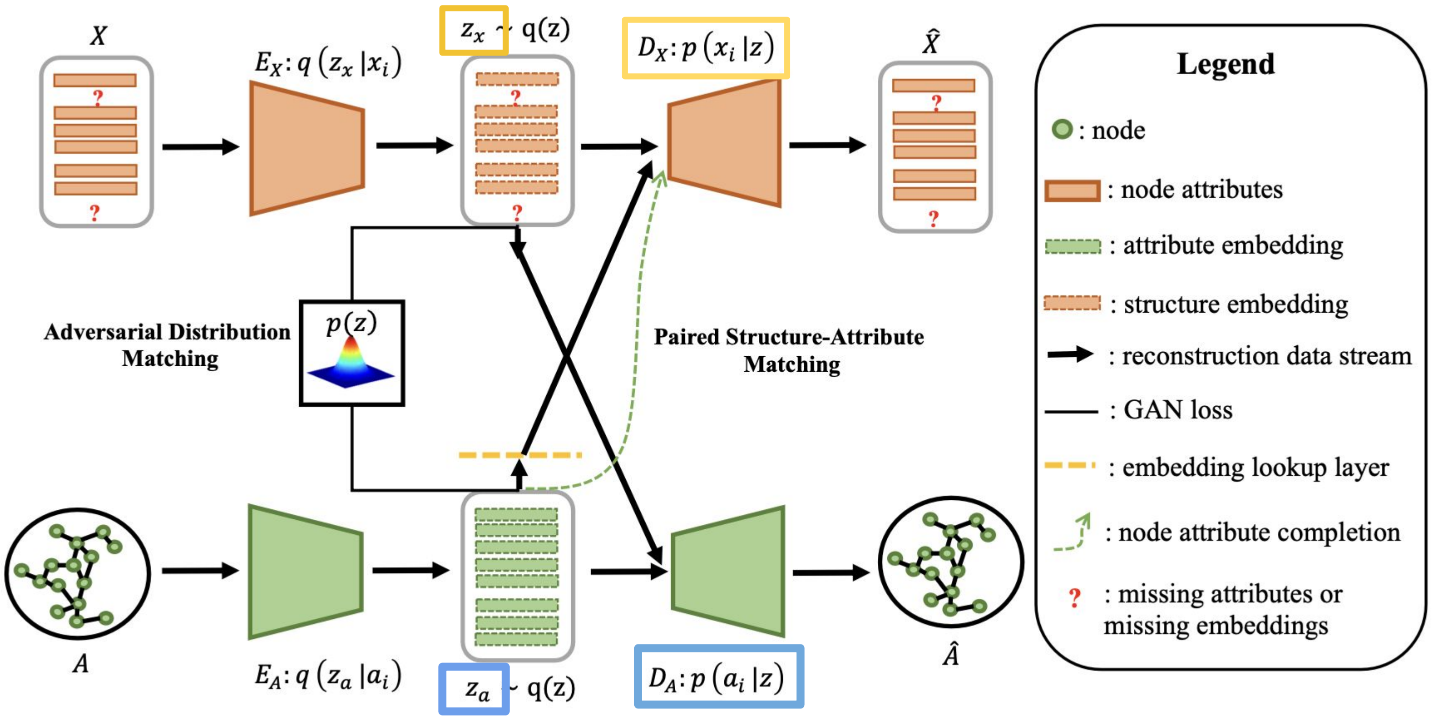
Paired Structure-Attribute Matching

Symbol	Description
q	approximate posterior
p	true posterior
z_x	latent Factor with respect to attribute
z_a	latent Factor with respect to structure
$\boldsymbol{\lambda}_c$	Cross construction hyperparameter
L_r	Joint Construction Loss
E	Expectation with respect to a distribution

Reconstruction Loss

$$
\begin{aligned}
\min_{\theta_x,\theta_a,\phi_x,\phi_a} \mathcal{L}_r = & -\mathbb{E}_{x_i\sim p_X}[\mathbb{E}_{q_{\phi_x}(z_x|x_i)}[\log p_{\theta_x}(x_i|z_x)]] && \text{Self construction stream for attributes}\\
& -\mathbb{E}_{a_i\sim p_A}[\mathbb{E}_{q_{\phi_a}(z_a|a_i)}[\log p_{\theta_a}(a_i|z_a)]] && \text{Self construction stream for structure}\\
& -\lambda_c\cdot\mathbb{E}_{a_i\sim p_A}[\mathbb{E}_{q_{\phi_a}(z_a|a_i)}[\log p_{\theta_x}(x_i|z_a)]] && \text{Cross construction stream for attributes}\\
& -\lambda_c\cdot\mathbb{E}_{x_i\sim p_X}[\mathbb{E}_{q_{\phi_x}(z_x|x_i)}[\log p_{\theta_a}(a_i|z_x)]]. && \text{Cross construction stream for structure}
\end{aligned}
$$

Structure Attribute Transformer PSAM



Handle on Prior Regularization Loss

$$\mathcal{L}(\theta, \phi; x_i, a_i) = \mathbb{E}_{q_\phi(z_x, z_a | x_i, a_i)}[\log p_\theta(x_i, a_i | z_x, z_a)] - D_{KL}[q_\phi(z_x, z_a | x_i, a_i) || p(z_x, z_a)]$$

Reconstruction Loss

Prior regularization loss

Paired Structure Attribute Matching

Adversarial Distribution Matching

Adversarial Distribution Matching

$$D_{KL}[q_\phi(z_x, z_a|x_i, a_i)||\underline{p(z_x, z_a)}]$$

Simplifying by $p(z_x, z_y) = p(z)p(z)$

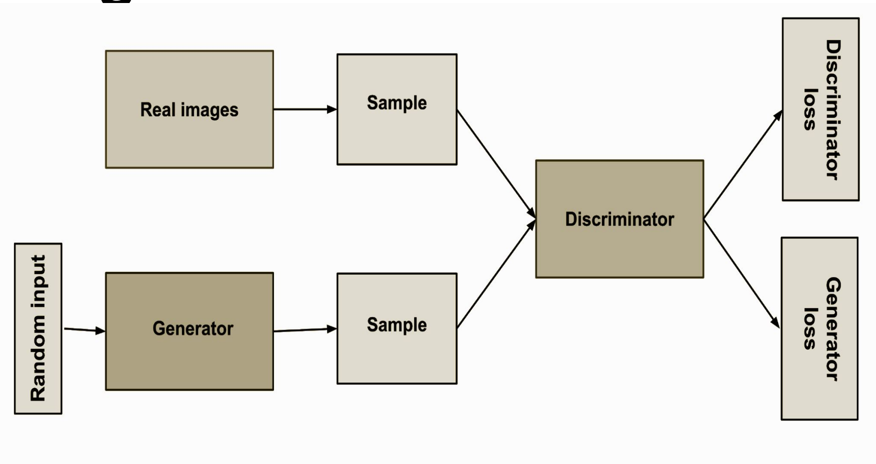
$$D_{KL}[q_{\phi_x}(z_x|x_i)||p(z)] + D_{KL}[q_{\phi_a}(z_a|a_i)||p(z)]$$

Adversarial Distribution matching

New Prior regularization loss

Symbol	Description
q	approximate posterior
p	true posterior
z_x	latent Factor with respect to attribute
z_a	latent Factor with respect to structure
$p(z)$	prior of the our approximate posterior q
x_i	attribute of node i
a_i	neighbours of node i

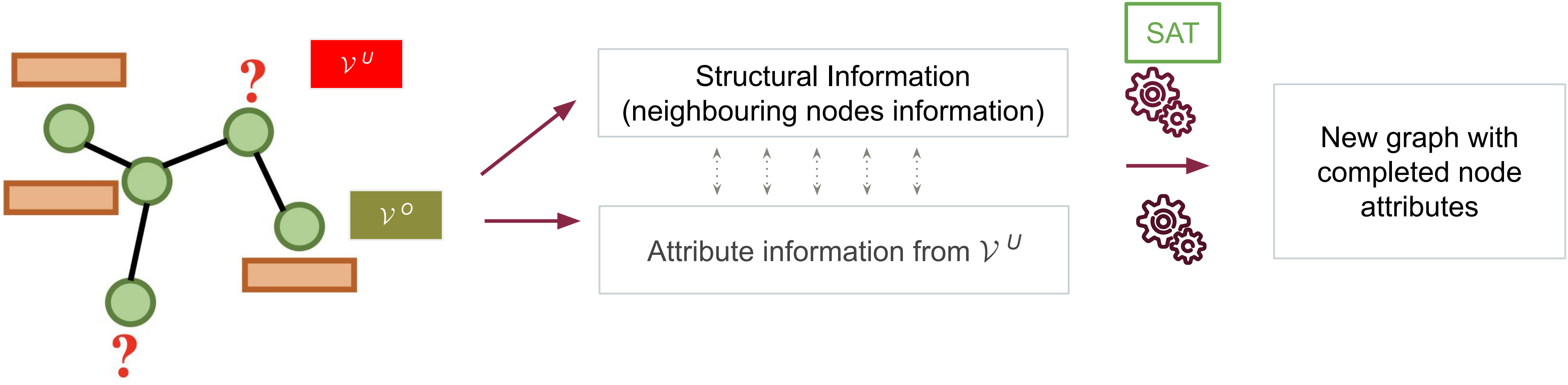
Adversarial Distribution Matching



$$
\begin{aligned}
\min_{\psi}\max_{\phi_x,\phi_a}\mathcal{L}_{adv} = &-\mathbb{E}_{z_p\sim p(z)}[\log\mathcal{D}(z_p)] \\
&-\mathbb{E}_{z_x\sim q_{\phi_x}(z_x|x_i)}[\log(1-\mathcal{D}(z_x))] \\
&-\mathbb{E}_{z_p\sim p(z)}[\log\mathcal{D}(z_p)] \\
&-\mathbb{E}_{z_a\sim q_{\phi_a}(z_a|a_i)}[\log(1-\mathcal{D}(z_a))]
\end{aligned}
$$

Symbol	*Description*
q	approximate posterior
$p(z)$	prior of the our approximate posterior q
z_x	latent Factor with respect to attribute
z_a	latent Factor with respect to structure
z_p	prior of the our approximate posterior q
$D(.)$	Discrimination operation or classification operation
ψ	parameter of shared Discriminator function

MODEL REVISIT



Attribute missing Graph

Graph	Description
$\mathcal{G}$	The graph
$\mathcal{V}$	the set of nodes of graph $\mathcal{G}$
A	the adjacent matrix of graph $\mathcal{G}$
X	the attribute matrix of graph $\mathcal{G}$
$\mathcal{V}^O$	the set of attribute-observed nodes $\mathcal{G}$
$\mathcal{V}^U$	the set of attribute-unobserved nodes $\mathcal{G}$

Final Objective Function And Implementation

Inputs

$A \in R^{N \times N}$ and $X^o \in R^{N \times F}$

Output

$X^U \in R^{N \times F}$

Initialize And Learn

All parameters and hyperparameter

Cost function

$$\min_{\Theta} \max_{\Phi} \mathcal{L} = \mathcal{L}_r + \mathcal{L}_{adv}$$

$$\Theta = \{\theta_x, \theta_a, \phi_x, \phi_a, \psi\}$$

$$\Phi = \{\phi_x, \phi_a\}$$

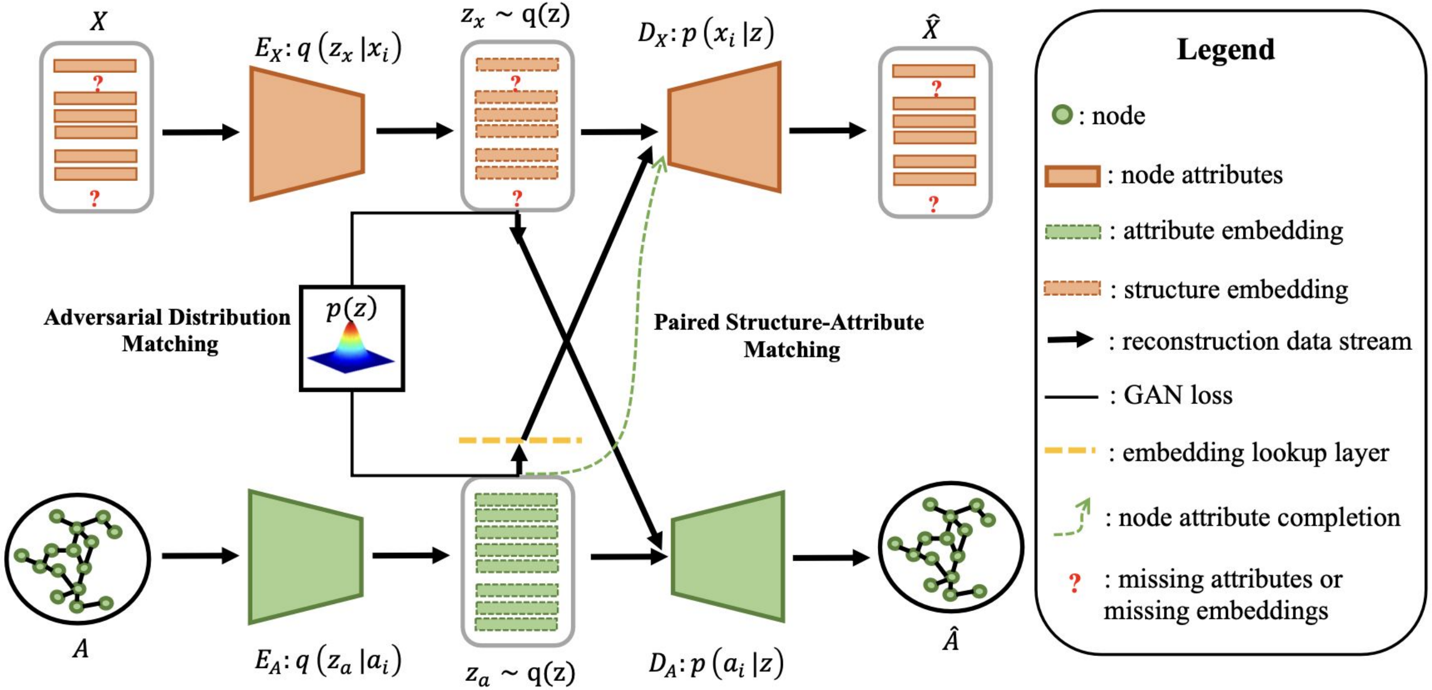
Step 1	Calculate the latent code Z_x and Z_a by encoder for X^o and A^o resp.
Step 2	Make embedding look up operation to get Z_A^o from all Z_A
Step 3	Get samples for adversarial distribution matching from prior $z_p \sim N(0,1)$
Step 4	Conduct Pair Structure-Attribute Matching and Adversarial matching
Step 5	Calculate cost function
Step 6	Gradient descent by adam/SGD to update parameters
Step 7	Repeat until convergence

After Training

Sample from Z_A^U from Z_A to restore node attributes

Predict missing links

Putting it all together



What to take home with you

Things we talked about	Things we could not talk about in detail

Things we talked about

- **Autoencoders**
- Shared **latent space** assumption
- Variational Autoencoders
- KL divergence
- Expectation of probabilities
- **Generators and discriminators**
- Generative **Adversarial Learning**
- SAT
- Paired Structure-Attribute Matching
- **Attribute missing problem** solution
- …

Things we could not talk about in detail

- **Variational Inference**
- **Mode collapse problem**
- KL divergence types
- **GAN** in detail
- **GNN**
- **Time complexity** of SAT
- Experimental results
- GAT - graph attention networks

Conclusion

Use cases

- Recommendation systems
- Fraud detection and network security
- Description Generation in networks
- Molecular graph learning
- Community detection

Further research

- Variational Inference
- GNN
- Different Generative modelling techniques and their connections

REFERENCES

[01] https://de.wikipedia.org/wiki/Autoencoder

[02] https://www.researchgate.net/figure/Basic-schematic-illustration-of-a-variational-autoencoder-VAE_fig1_346355664

[03] https://jessicastringham.net/2018/12/27/KL-Divergence/

[04] https://developers.google.com/machine-learning/gan/gan_structure

[05] Xu Chen, Siheng Chen, Jiangchao Yao, Huangjie Zheng, Ya Zhang, and Ivor W. Tsang. Learning on attribute-missing graphs. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(2):740–757, February 2022

[06] https://www.pinterest.com/pin/32088216083886948/

THANK YOU!

Any Questions?