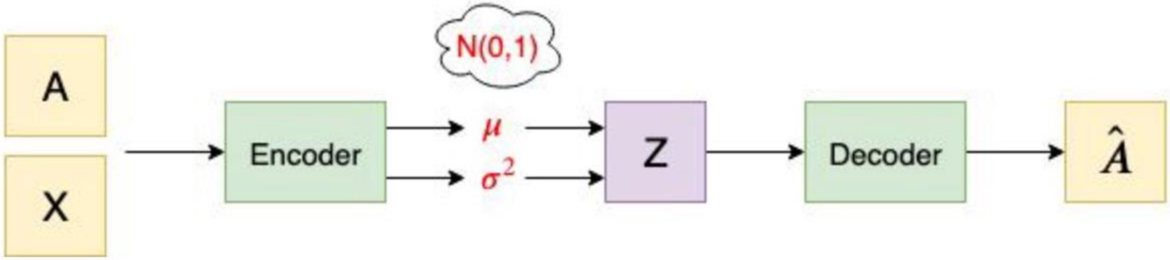
Variational Graph AutoEncoders

The architecture of the Encoder and Decoder



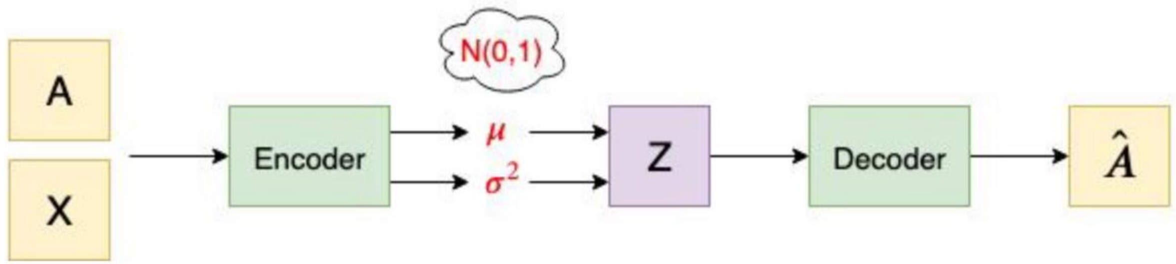
Encoder: two layer GCN

$$\mathrm{GCN}(\mathbf{X}, \mathbf{A}) = \tilde{\mathbf{A}}\,\mathrm{ReLU}\big(\tilde{\mathbf{A}}\mathbf{X}\mathbf{W}_0\big)\mathbf{W}_1$$

$\boldsymbol{\mu} = \mathrm{GCN}_{\boldsymbol{\mu}}(\mathbf{X}, \mathbf{A})$ is the matrix of mean vectors $\boldsymbol{\mu}_i$; similarly $\log \boldsymbol{\sigma} = \mathrm{GCN}_{\boldsymbol{\sigma}}(\mathbf{X}, \mathbf{A})$

$\mathrm{GCN}_{\boldsymbol{\mu}}(\mathbf{X}, \mathbf{A})$ and $\mathrm{GCN}_{\boldsymbol{\sigma}}(\mathbf{X}, \mathbf{A})$ share first-layer parameters $\mathbf{W}_0$

The architecture of the Encoder and Decoder



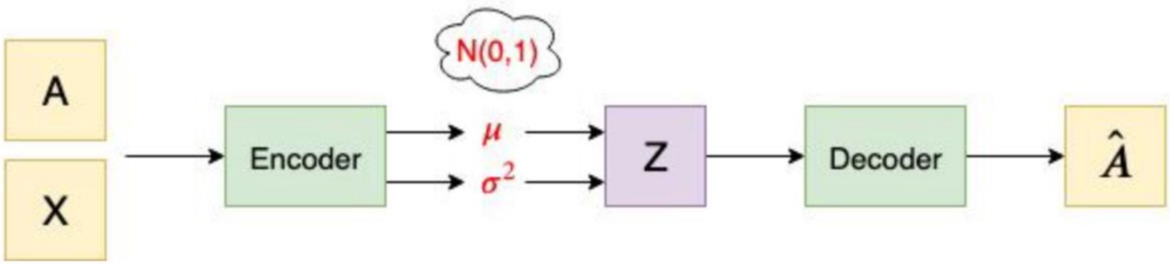
Inference model:

$$q(\mathbf{Z} \,|\, \mathbf{X}, \mathbf{A}) = \prod_{i=1}^{N} q(\mathbf{z}_i \,|\, \mathbf{X}, \mathbf{A})\,, \quad \text{with} \quad q(\mathbf{z}_i \,|\, \mathbf{X}, \mathbf{A}) = \mathcal{N}(\mathbf{z}_i \,|\, \boldsymbol{\mu}_i, \mathrm{diag}(\boldsymbol{\sigma}_i^2))$$

```
tensor([[ 1.3924e+00,  2.3742e-03, -2.0968e-03,  5.6259e-04,  2.5119e-02],
        [ 1.4365e+00,  1.8989e-02, -2.2564e-02,  2.3635e-04,  3.4869e-02],
        [ 1.3924e+00,  2.3742e-03, -2.0968e-03,  5.6259e-04,  2.5119e-02],
        [ 1.5349e+00,  1.7675e-02,  1.1449e-04,  1.0096e-02,  3.4326e-02],
        [ 1.4815e+00,  3.5499e-03, -2.1504e-02,  1.0316e-02,  3.1015e-02],
        [-7.1078e-01, -6.8555e-03, -4.6863e-02, -2.7428e-02,  3.4037e-02],
        [-7.0468e-01, -7.9992e-04, -9.8651e-03,  1.2867e-02, -7.4852e-03],
        [-6.9762e-01,  9.0425e-03, -2.3550e-02, -9.4596e-04,  5.3763e-03],
        [-6.8976e-01,  2.0575e-02,  1.2874e-03,  2.2452e-02, -3.0748e-02],
        [-7.8341e-01,  1.9478e-02, -1.7059e-02, -1.3157e-03,  9.0686e-03]],
       grad_fn=<AddBackward0>),
```

Here, μ and Z are $N \times F$ matrices, in which F is the latent space dimension.

The architecture of the Encoder and Decoder



Generative model:

$$p\left(\mathbf{A} \mid \mathbf{Z}\right) = \prod_{i=1}^{N} \prod_{j=1}^{N} p\left(A_{ij} \mid \mathbf{z}_i, \mathbf{z}_j\right) \,, \quad \text{with} \quad p\left(A_{ij} = 1 \mid \mathbf{z}_i, \mathbf{z}_j\right) = \sigma(\mathbf{z}_i^\top \mathbf{z}_j) \,,$$

where A_{ij} are the elements of $\mathbf{A}$ and $\sigma(\cdot)$ is the logistic sigmoid function.

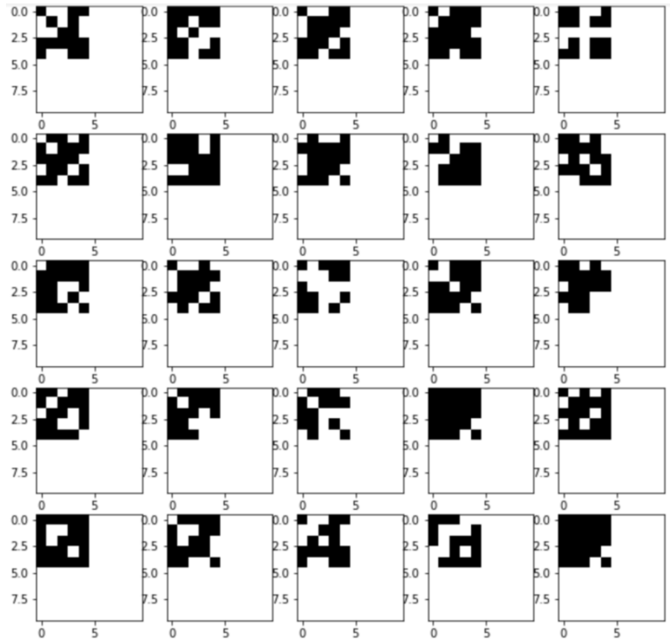
We optimize the variational lower bound $\mathcal{L}$ w.r.t. the variational parameters $\mathbf{W}_i$:

$$\mathcal{L} = \mathbb{E}_{q(\mathbf{Z}|\mathbf{X},\mathbf{A})}\big[\log p\left(\mathbf{A} \mid \mathbf{Z}\right)\big] - \mathrm{KL}\big[q(\mathbf{Z} \mid \mathbf{X}, \mathbf{A}) \,||\, p(\mathbf{Z})\big] \,,$$

in which

Gaussian prior $p(\mathbf{Z}) = \prod_i p(\mathbf{z_i}) = \prod_i \mathcal{N}(\mathbf{z}_i \mid 0, \mathbf{I})$

Input for simulation



Input: $N \times N$ adjacency matrix A and $\mathrm{N} \times \mathrm{D}$ feature matrix X, where N is the number of nodes, D is the dimension of features.

For the upper left submatrix of A, let

$$A(i,j) \sim Bernoulli(p = 0.75);\ A(j,i) = A(i,j), \qquad for\ j \leq \frac{N}{2}, i < j.$$

$$A(\mathrm{i}, j) = 0, \text{elsewhere}$$

Let X be the one-hot matrix with respect to the node index.

Hyperparameters

Optimizer: Adams optimizer with learning rate = 0.01
Epoches = 100000

Sampling

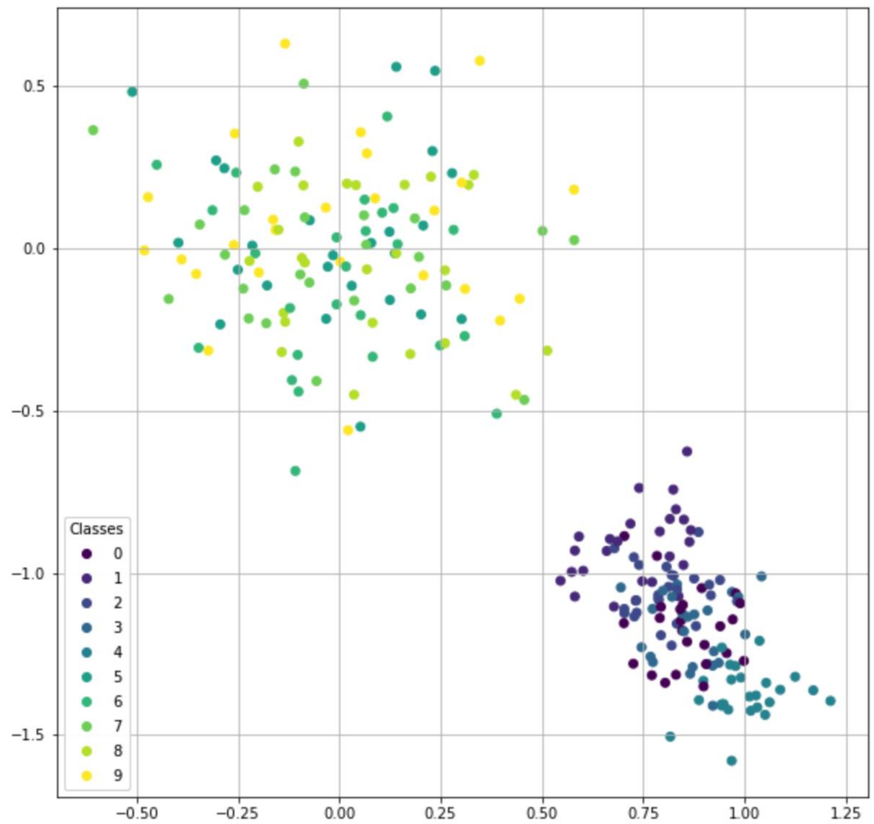
- Flatten the n $N \times F$ μ_i into $1 \times NF$ vectors μ_i'.
- Fit multivariate gaussian distributions $\mathcal{N}_{\mu'}$ from μ_i' with diagonal covariance matrix.
- Sample μ'_{sample} from $\mathcal{N}_{\mu'}$.
- Restore μ'_{sample} to $N \times F$ matrix μ_{sample}.
- Generate a latent variable z from μ_{sample}:

$$q(z) = \mathcal{N}(\mathrm{z}|\mu_{sample}, \mathrm{mean}(\mathrm{diag}(\sigma)))$$

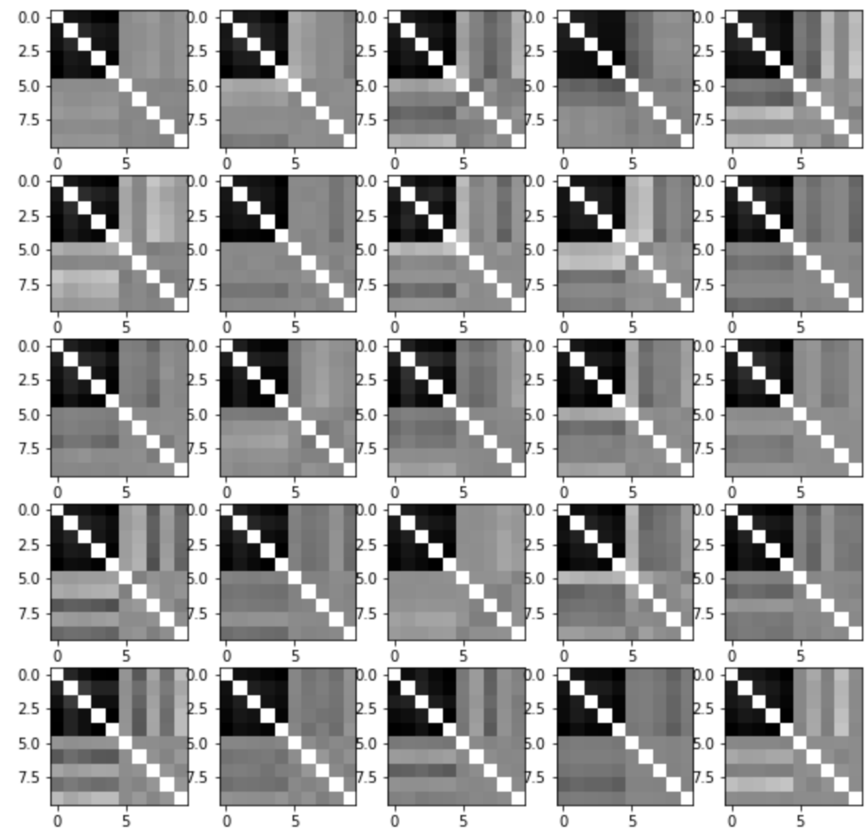
- Decode the latent variable z into a sampling adjacency matrix using the inner product decoder:

$$p(\mathbf{A} \mid \mathbf{Z}) = \prod_{i=1}^{N} \prod_{j=1}^{N} p(A_{ij} \mid \mathbf{z}_i, \mathbf{z}_j) \ , \quad \text{with} \quad p(A_{ij} = 1 \mid \mathbf{z}_i, \mathbf{z}_j) = \sigma(\mathbf{z}_i^\top \mathbf{z}_j)$$

Sampling

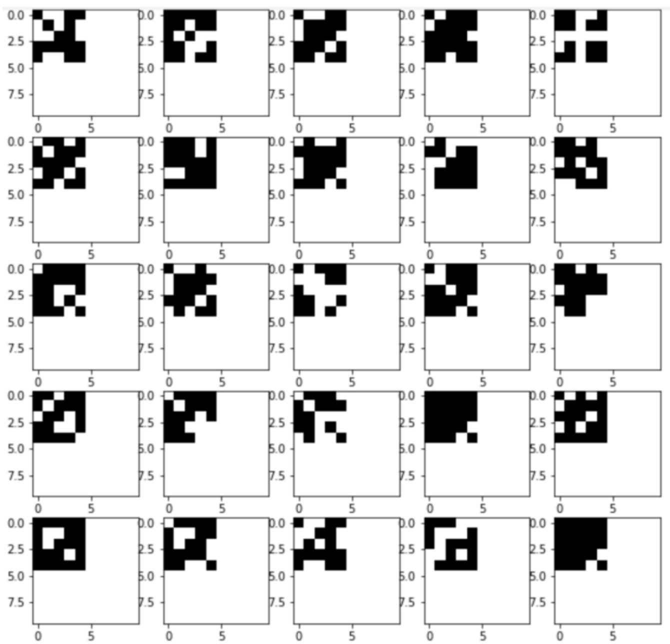


The μ for each nodes for 25 μ sampled from the Gaussian distribution. The blue and purple points are the first five nodes (connected nodes). The yellow and green points are the last five nodes (empty nodes).

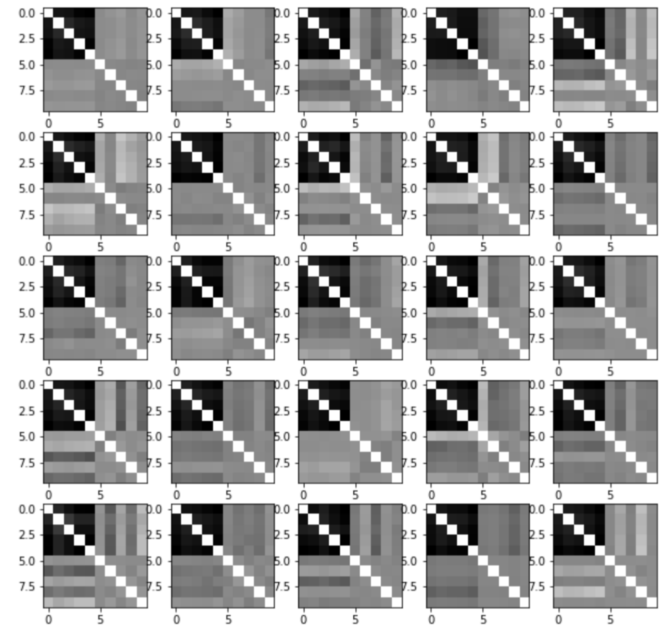


25 generated adjacency matrices.

Comparision

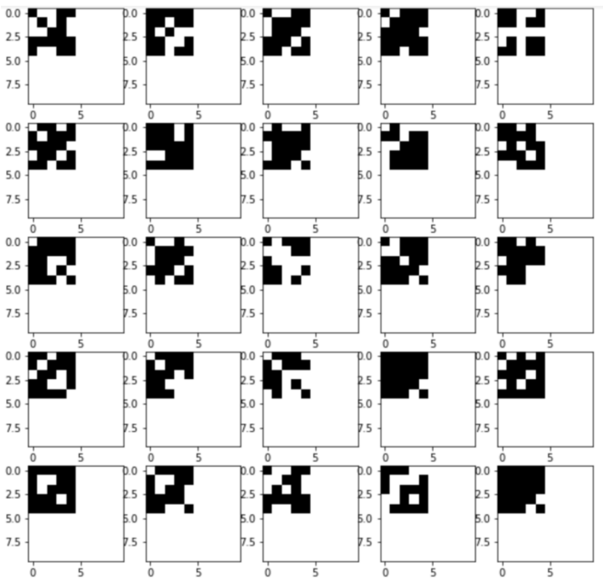


original

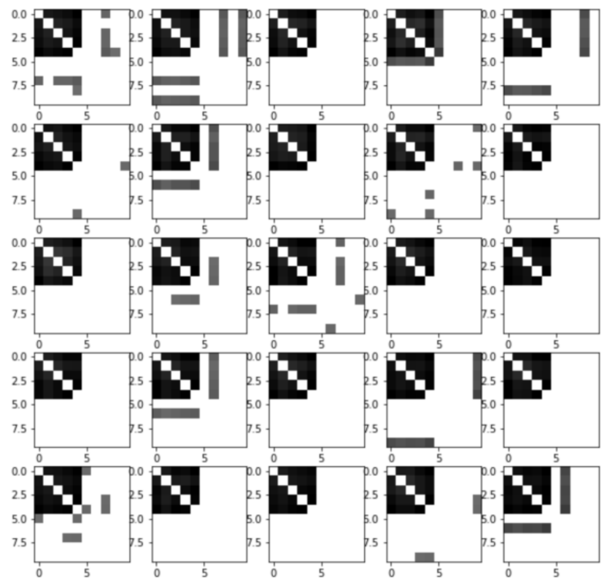


generated

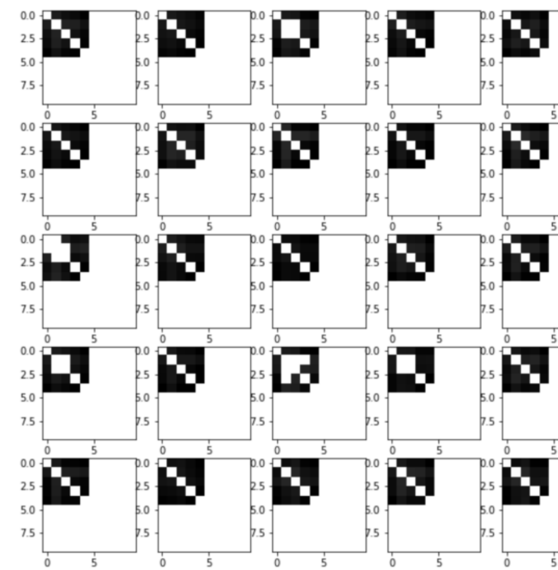
Comparision



original



Generated with filter value=0.3



Generated with filter value=0.5