Applied to our setting

- Settings
  - Node number: N
  - Feature dimension: d
  - First hidden dimension: h
  - Number of out of channel: f
  - Sample size: n
  - Sample size of $\boldsymbol{z}$ for compute $E[log \prod_{k=1}^{n} p(\boldsymbol{A}_k|\boldsymbol{Z})]$: L

- Assumptions
  - Each node is independent
  - Each graph is independent
  - Each out of channel is independent

- Notations:
  - **A**: observation of adjacency matrix (with diagonal elements 1), representing graphs, $\boldsymbol{A}_k, k = 1, \dots, n$
  - **D**: degree matrix, $\boldsymbol{D}_k, k = 1, \dots, n$
  - $\widetilde{\boldsymbol{A}}$: normalized adjacency matrix, $\widetilde{\boldsymbol{A}} = \boldsymbol{D}^{-1/2}\boldsymbol{A}\boldsymbol{D}^{-1/2}$, $\widetilde{\boldsymbol{A}}_k, k = 1, \dots, n$
  - $\bar{\bar{\boldsymbol{A}}}$: adjacency matrix without diagonal elements
  - **Z**: latent variable, $\boldsymbol{Z}_i, i = 1, \dots, N$
  - **X**: feature matrix, $\boldsymbol{X}_k, k = 1, \dots, n$

- Inference model
  - $q_\phi(\boldsymbol{Z}|\boldsymbol{X},\boldsymbol{A}) = \prod_{i=1}^{N} q(\boldsymbol{Z}_i|\boldsymbol{X},\boldsymbol{A})$ with $q(\boldsymbol{Z}_i|\boldsymbol{X},\boldsymbol{A}) = N(\boldsymbol{\mu}_i, \boldsymbol{\sigma}_i^2)$
  - $\boldsymbol{Z}_i$ and $\boldsymbol{\mu}_i$ are of dimension f, $\boldsymbol{\sigma}_i^2$ is a $f \times f$ diagonal matrix

- Prior
  - $p_\theta(\boldsymbol{Z}) = \prod_{i=1}^{N} p(\boldsymbol{Z}_i)$ with $p(\boldsymbol{Z}_i) = N(\boldsymbol{0}, \boldsymbol{I}_f)$

- Generative model
  - $p(\boldsymbol{A}|\boldsymbol{Z}) = \prod_{k=1}^{n} p(\boldsymbol{A}_k|\boldsymbol{Z})$
  - $p(\boldsymbol{A}_k|\boldsymbol{Z}) = \prod_{i=1}^{N}\prod_{j=1}^{N} p(\boldsymbol{A}_{kij}|\boldsymbol{Z}_i,\boldsymbol{Z}_j)$ with $p\left(\boldsymbol{A}_{kij} = 1\middle|\boldsymbol{Z}_i,\boldsymbol{Z}_j\right) = sigmoid(\boldsymbol{Z}_i^T\boldsymbol{Z}_j)$

- Objective function

$$ELBO(q) = E[logp(\boldsymbol{A}|\boldsymbol{Z})] - KL\left(q_\phi(\boldsymbol{Z})||p_\theta(\boldsymbol{Z})\right)$$

$$= E\{log \prod_{k=1}^{n} p(\boldsymbol{A}_k|\boldsymbol{Z})\} + E\left\{log \prod_{i=1}^{N} p_\theta(\boldsymbol{Z}_i)\right\} - E\left\{log \prod_{j=1}^{N} q_\phi\left(\boldsymbol{Z}_j\right)\right\}$$

$$= \sum_{k=1}^{n} Elogp(\boldsymbol{A}_k|\boldsymbol{Z}) + \sum_{i=1}^{N} Elogp_\theta(\boldsymbol{Z}_i) - \sum_{j=1}^{N} Elogq_\phi(\boldsymbol{Z}_j)$$

- Input
  - $(\widetilde{\boldsymbol{A}}_k, \boldsymbol{X}_k), k = 1, \ldots, n$
  - $\widetilde{\boldsymbol{A}}_k, \boldsymbol{X}_k$ are of dimension $N \times N$

- 1_{st} layer:
  - $\boldsymbol{X}_k^1 = ReLU(\widetilde{\boldsymbol{A}}_k \boldsymbol{X}_k \boldsymbol{W}^1), \quad ReLU = \max(0, \cdot)$
  - $\boldsymbol{W}^1, \boldsymbol{X}_k^1$ are of dimension $N \times h$

- 2_{nd} layer:
  - $\boldsymbol{X}_k^{\mu} = \widetilde{\boldsymbol{A}}_k \boldsymbol{X}_k^1 \boldsymbol{W}^{\mu}$
  - $\boldsymbol{X}_k^{log\sigma} = \widetilde{\boldsymbol{A}}_k \boldsymbol{X}_k^1 \boldsymbol{W}^{log\sigma}$
  - $\boldsymbol{W}^{\mu}, \boldsymbol{W}^{log\sigma}$ are of dimension $h \times f$
  - $\boldsymbol{X}_k^{\mu}, \boldsymbol{X}_k^{log\sigma}$ are of dimension $N \times f$

- Sample mean layer
  - $\boldsymbol{X}^{\mu} = \frac{1}{n} \sum_{k=1}^{n} \boldsymbol{X}_k^{\mu}$
  - $\boldsymbol{X}^{log\sigma} = \frac{1}{n} \sum_{k=1}^{n} \boldsymbol{X}_k^{log\sigma}$
  - $\boldsymbol{X}^{\mu}, \boldsymbol{X}^{log\sigma}$ are of dimension $N \times f$

- Generate Z
  - $\boldsymbol{Z} = \boldsymbol{X}^{\mu} + exp\{\boldsymbol{X}^{log\sigma}\} \times \boldsymbol{\varepsilon}$
  - $\boldsymbol{Z}$ is of dimension $N \times f$

Objective function

$$ELBO(q) = E[logp(\boldsymbol{A}|\boldsymbol{Z})] - KL(q(\boldsymbol{Z})||p(\boldsymbol{Z}))$$

$$E[logp(A|Z)]$$

$$= E\left[log \prod_{k=1}^{n} p(\boldsymbol{A}_k|\boldsymbol{Z})\right]$$ (each sample is independent)

$$= \sum_{k=1}^{n} E[\log p(\boldsymbol{A}_k|Z_{kl})] = \sum_{k=1}^{n} \frac{1}{L} \sum_{l=1}^{L} logp(\boldsymbol{A}_k|\boldsymbol{Z}_{kl})$$

$$= \sum_{k=1}^{n} \frac{1}{L} \sum_{l=1}^{L} log \prod_{i=1}^{N} \prod_{j=1}^{N} p(A_{ijk}|\boldsymbol{Z}_{\boldsymbol{kl}})$$ (each element of adjacency matrix is independent)

$$= \sum_{k=1}^{n} \frac{1}{L} \sum_{l=1}^{L} log \prod_{i=1}^{N} \prod_{j=1}^{N} \left\{\sigma\left(\boldsymbol{Z}_{ikl}^{T} \boldsymbol{Z}_{jkl}\right)^{A_{ijk}} \left[1 - \sigma\left(\boldsymbol{Z}_{ikl}^{T} \boldsymbol{Z}_{jkl}\right)\right]^{1-A_{ijk}}\right\}$$ ($\sigma(x) = \frac{1}{1+e^{-x}}$)

$$= \frac{1}{L} \sum_{l=1}^{L} \sum_{k=1}^{n} \sum_{i=1}^{N} \sum_{j=1}^{N} A_{ijk} log\sigma\left(\boldsymbol{Z}_{ikl}^{T} \boldsymbol{Z}_{jkl}\right) + \left(1 - A_{ijk}\right) log[1 - \sigma(\boldsymbol{Z}_{ikl}^{T} \boldsymbol{Z}_{jkl})]$$

Objective function

$$ELBO(q) = E[logp(\boldsymbol{A}|\boldsymbol{Z})] - KL(q(\boldsymbol{Z})||p(\boldsymbol{Z}))$$

$$\frac{1}{N^2}ELBO(q) = \frac{1}{N^2}E[logp(\boldsymbol{A}|\boldsymbol{Z})] - \frac{1}{N^2}KL(q(\boldsymbol{Z})||p(\boldsymbol{Z}))$$

$$\frac{1}{N^2}E[logp(A|Z)]$$

$$= \frac{1}{L}\frac{1}{N^2}\sum_{l=1}^{L}\sum_{k=1}^{n}\sum_{i=1}^{N}\sum_{j=1}^{N} A_{ijk}log\sigma\left(\boldsymbol{Z}_{il}^T\boldsymbol{Z}_{jl}\right) + \left(1 - A_{ijk}\right)log[1 - \sigma(\boldsymbol{Z}_{il}^T\boldsymbol{Z}_{jl})]$$

$$= \frac{1}{L}\frac{1}{N^2}\sum_{l=1}^{L}\sum_{k=1}^{n}\sum_{i=1}^{N}\sum_{j=1}^{N} A_{ijk}log\sigma\left(\boldsymbol{Z}_{il}^T\boldsymbol{Z}_{jl}\right) + \frac{1}{L}\frac{1}{N^2}\sum_{l=1}^{L}\sum_{k=1}^{n}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(1 - A_{ijk}\right)log[1 - \sigma(\boldsymbol{Z}_{il}^T\boldsymbol{Z}_{jl})]$$

For sparse A, it can be beneficial to re-weight terms with Aij = 1 in L or alternatively sub-sample terms with Aij = 0 to balance the weight of positive edge (Aij = 1) information and negative edge (Aij = 0) information.

In our model, we reweight $\frac{1}{N^2}\ E[logp(A|Z)] \approx$

$$= \frac{1}{L}\sum_{l=1}^{L}\sum_{k=1}^{n}\frac{1}{\#\ (A_{ij}=1)}\sum_{i=1}^{N}\sum_{j=1}^{N} A_{ijk}log\sigma\left(\boldsymbol{Z}_{il}^T\boldsymbol{Z}_{jl}\right) +$$

$$\frac{1}{L}\sum_{l=1}^{L}\sum_{k=1}^{n}\frac{1}{\#\ (A_{ij}=0)}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(1 - A_{ijk}\right)log[1 - \sigma(\boldsymbol{Z}_{il}^T\boldsymbol{Z}_{jl})]$$

= positive edge reconstruction likelihood +
negative edge reconstruction likelihood

For node i, j

- $E_\phi log p_\theta(\boldsymbol{Z}_i) = \int N(\boldsymbol{Z}_i;\mu,\sigma^2) logN(\boldsymbol{Z}_i;0,I)d\boldsymbol{Z}_i = \int \log\left\{\frac{1}{\sqrt{(2\pi)^f}}\exp\left[-\frac{1}{2}\boldsymbol{Z}_i^T\boldsymbol{Z}_i\right]\right\} q_\phi(\boldsymbol{Z}_i)d(\boldsymbol{Z}_i)$

$= \int\left\{-\frac{f}{2}log(2\pi) - \frac{1}{2}\sum_{t=1}^{f} Z_{it}^2\right\} q_\phi(\boldsymbol{Z}_i)d(\boldsymbol{Z}_i) = -\frac{f}{2}lo\,g(2\pi) - \frac{1}{2}\sum_{t=1}^{f} E_\phi(Z_{it}) = -\frac{f}{2}lo\,g(2\pi) - \frac{1}{2}(\sigma_{it}^2+\mu_{it}^2)$

- $E_\phi log q_\phi(\boldsymbol{Z}_j) = E_\phi lo\,g\left\{\frac{1}{\sqrt{(2\pi)^f\prod_{t=1}^{f}\sigma_{jt}^2}}\exp\left[-\frac{1}{2}\sum_{t=1}^{f}\frac{(Z_{jt}-\mu_{jt})^2}{\sigma_{jt}^2}\right]\right\}$

$= -\frac{f}{2}lo\,g(2\pi) - \frac{1}{2}\sum_{t=1}^{f} log\sigma_{jt}^2 - \frac{1}{2}\sum_{t=1}^{f} E_\phi\left(\frac{Z_{jt}-\mu_{jt}}{\sigma_{jt}}\right)^2 = -\frac{f}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{f}(1+log\sigma_{jt}^2)$

- $-KL(q_\phi(\boldsymbol{Z})||p_\theta(\boldsymbol{Z})) = E_\phi log p_\theta(\boldsymbol{Z}) - E_\phi log q_\phi(\boldsymbol{Z})$

$= E_\phi\left[log\prod_{i=1}^{N} p_\theta(\boldsymbol{Z}_i)\right] - E_\phi\left[log\prod_{j=1}^{N} q_\phi(\boldsymbol{Z}_j)\right]$ (each node is independent)

$= \sum_{i=1}^{N} E_\phi[log p_\theta(\boldsymbol{Z}_i)] - \sum_{j=1}^{N} E_\phi[log q_\phi(\boldsymbol{Z}_j)] = \frac{1}{2}\sum_{i=1}^{N}\sum_{t=1}^{f}(1+2log\sigma_{it}-\mu_{ij}^2-\sigma_{ij}^2)$

$-\frac{1}{N^2}KL(q_\phi(\boldsymbol{Z})||p_\theta(\boldsymbol{Z})) = \frac{1}{2}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{t=1}^{f}(1+2log\sigma_{it}-\mu_{ij}^2-\sigma_{ij}^2)$

Loss function

$$Loss = -\frac{1}{N^2} ELBO(q)$$

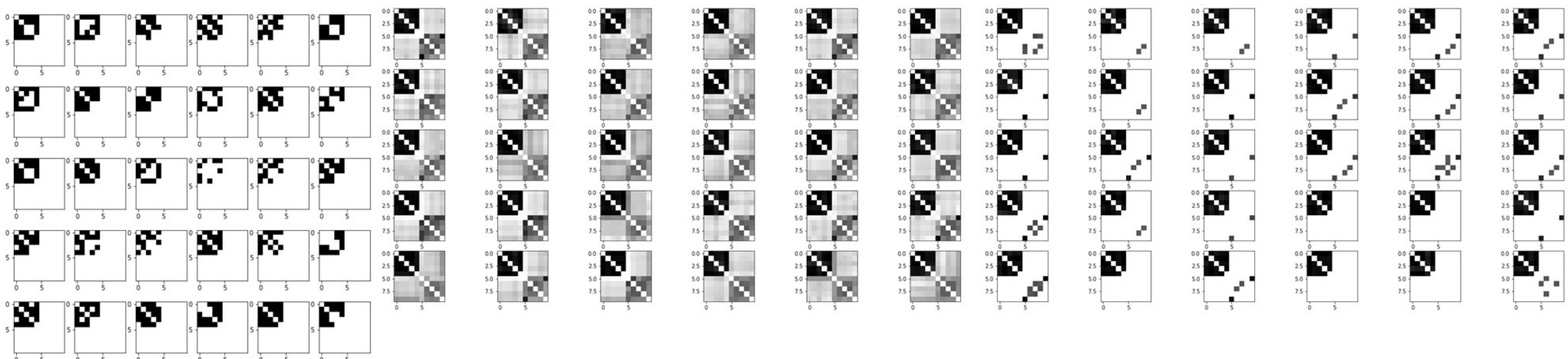
$$= -\frac{1}{N^2} E[logp(\boldsymbol{A}|\boldsymbol{Z})] + \frac{1}{N^2} KL(q_\phi(\boldsymbol{Z})||p_\theta(\boldsymbol{Z}))$$

$$= -\frac{1}{L}\sum_{l=1}^{L}\sum_{k=1}^{n}\frac{1}{\#\,(A_{ij}=1)}\sum_{i=1}^{N}\sum_{j=1}^{N} A_{ijk} log\sigma\left(\boldsymbol{Z}_{il}^{T}\boldsymbol{Z}_{jl}\right)$$ (postive edge recon loss)

$$-\frac{1}{L}\sum_{l=1}^{L}\sum_{k=1}^{n}\frac{1}{\#\,(A_{ij}=0)}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(1 - A_{ijk}\right) log[1-\sigma(\boldsymbol{Z}_{il}^{T}\boldsymbol{Z}_{jl})]$$ (negative edge recon loss)

$$-\frac{1}{2}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{t=1}^{f}(1 + 2log\sigma_{it} - \mu_{ij}^2 - \sigma_{ij}^2)$$ (KL loss)

A possible explanation of simulation



Input adjacency matrices (p=0.8)

Raw generated adjacency matrices (filter_value=0)

Generated adjacency matrices (filter_value=0.5)

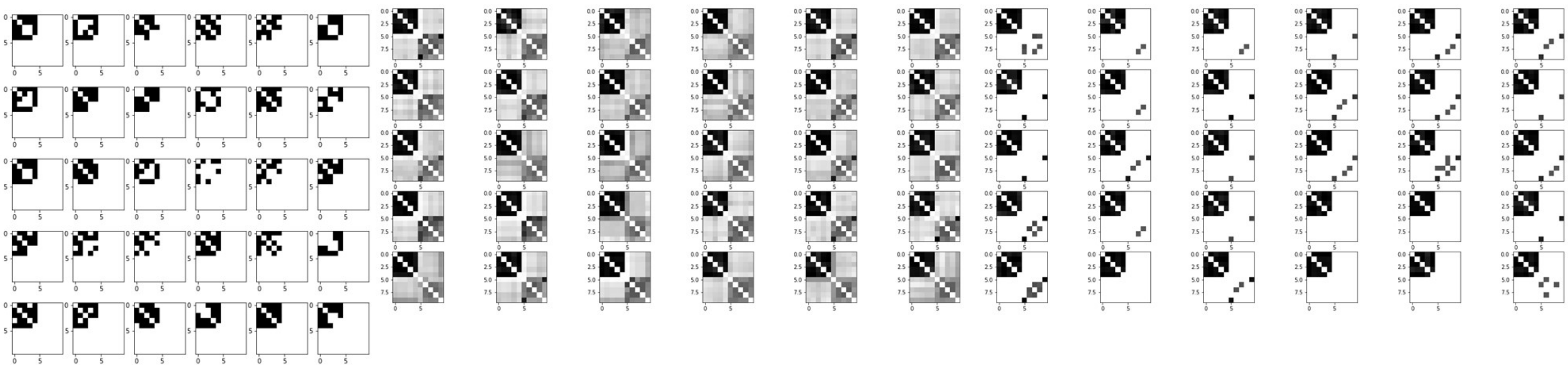
In our model, we reweight $\frac{1}{N^2}\,E[logp(A|Z)] \approx$

$$= \frac{1}{L}\sum_{l=1}^{L}\sum_{k=1}^{n}\frac{1}{\#\,(A_{ij}=1)}\sum_{i=1}^{N}\sum_{j=1}^{N} A_{ijk}\,log\sigma\left(\boldsymbol{Z}_{il}^{T}\boldsymbol{Z}_{jl}\right) +$$

$$\frac{1}{L}\sum_{l=1}^{L}\sum_{k=1}^{n}\frac{1}{\#\,(A_{ij}=0)}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(1 - A_{ijk}\right)log[1-\sigma(\boldsymbol{Z}_{il}^{T}\boldsymbol{Z}_{jl})]$$

= positive edge reconstruction likelihood +
negative edge reconstruction likelihood

A possible explanation of simulation



Input adjacency matrices (p=0.8)

Raw generated adjacency matrices (filter_value=0)

Generated adjacency matrices (filter_value=0.5)

The weights of positive edge information and negative edge information are equal. Hence, though the number of positive edges is less than the number of negative edges, the reconstruction loss considers them equally and they may generates more positive edges than the input.

With proper weights between positive edge loss and negative edge loss, we may not need a filter value.