

# 1 Stochastic block model

Some thoughts on variational inference for the stochastic block model.

## 1.1 Variational inference

Variational Bayes approximates the posterior distribution  $p(\boldsymbol{\theta}|\mathbf{x})$  using the following identities

$$\begin{aligned}
\log p(\mathbf{x}) &= \int q(\boldsymbol{\theta}) \ln p(\mathbf{x}) d\boldsymbol{\theta} \\
&= \int q(\boldsymbol{\theta}) \log \frac{p(\mathbf{x}, \boldsymbol{\theta}) q(\boldsymbol{\theta})}{p(\boldsymbol{\theta}|\mathbf{x}) q(\boldsymbol{\theta})} d\boldsymbol{\theta} \\
&= \int q(\boldsymbol{\theta}) \log p(\mathbf{x}, \boldsymbol{\theta}) d\boldsymbol{\theta} - \int q(\boldsymbol{\theta}) q(\boldsymbol{\theta}) d\boldsymbol{\theta} + \int q(\boldsymbol{\theta}) \frac{q(\boldsymbol{\theta})}{p(\boldsymbol{\theta}|\mathbf{x})} d\boldsymbol{\theta} \quad (1) \\
&= \mathbb{E}_q[\log p(\mathbf{x}, \boldsymbol{\theta})] + H(q) + D_{KL}(q(\boldsymbol{\theta})||p(\boldsymbol{\theta}|\mathbf{x})) \\
&\geq \mathbb{E}_q[\log p(\mathbf{x}, \boldsymbol{\theta})] + H(q) .
\end{aligned}$$

This last term is famously known as the *evidence lower bound* (ELBO) and commonly optimized wrt  $q(\boldsymbol{\theta})$  in variational Bayes.

### 1.1.1 Stochastic gradient

Using the so called *reparametrization trick* the ELBO is amenable to stochastic gradient ascent. Consider a distribution  $q(\boldsymbol{\theta}; \boldsymbol{\eta})$  with variational parameters  $\boldsymbol{\eta}$ . Further, assume that a sample from  $q$  can be obtained by transforming standard uniform or standard normal random variates  $\mathbf{z}$ , i.e. via some function  $\boldsymbol{\theta} = f_{\boldsymbol{\theta}}(\mathbf{z}; \boldsymbol{\eta})$ . Then, by Leibniz rule and the chain rule the derivative of the first term of the ELBO is given by

$$\begin{aligned}
\frac{\partial}{\partial \boldsymbol{\eta}} \mathbb{E}_q[\log p(\mathbf{x}, \boldsymbol{\theta})] &= \frac{\partial}{\partial \boldsymbol{\eta}} \int q(\boldsymbol{\theta}; \boldsymbol{\eta}) \log p(\mathbf{x}, \boldsymbol{\theta}) d\boldsymbol{\theta} \\
&= \frac{\partial}{\partial \boldsymbol{\eta}} \int p(\mathbf{z}) \log p(\mathbf{x}, f_{\boldsymbol{\theta}}(\mathbf{z}; \boldsymbol{\eta})) d\mathbf{z} \quad (2) \\
&= \mathbb{E}_{\mathbf{z}} \left[ \frac{\partial}{\partial \boldsymbol{\theta}} \log p(\mathbf{x}, \boldsymbol{\theta}) \frac{\partial}{\partial \boldsymbol{\eta}} f_{\boldsymbol{\theta}}(\mathbf{z}; \boldsymbol{\eta}) \right]
\end{aligned}$$

which is then stochastically evaluated using a few Monte-Carlo samples, i.e.

$$\frac{\partial}{\partial \boldsymbol{\eta}} \mathbb{E}_q[\log p(\mathbf{x}, \boldsymbol{\theta})] \approx \frac{1}{M} \sum_i \frac{\partial}{\partial \boldsymbol{\theta}} \log p(\mathbf{x}, \boldsymbol{\theta}) \frac{\partial}{\partial \boldsymbol{\eta}} f_{\boldsymbol{\theta}}(\mathbf{z}_i; \boldsymbol{\eta}) . \quad (3)$$

## 1.2 Stochastic block model

The *stochastic block model* (SBM) is a random graph model able to describe clustering/modularity and other structures in dense graphs. It is a special case

of the more general class of graphin models that can be derived from exchangeability arguments [?].

According to the SBM a graph on  $N$  nodes, actually its adjacency matrix, is produced by the following generative process:

- Randomly assign a type  $c_i \in \{1, \dots, K\}$  to each node  $i$ :

$$c_i \sim \text{Categorical}(\theta_1, \dots, \theta_K)$$

Note: This assignment is unobserved, i.e. the SBM is an example of a *latent/hidden variable model*.

- Draw a link between  $i$  and  $j$  with probability  $p_{c_i c_j}$ , i.e.

$$A_{ij} \sim \text{Bernoulli}(p_{c_i c_j})$$

Note: Links between nodes of the same types are independent (as in the Erdős-Rényi model).

Link probabilities are usually collected in a matrix  $\mathbf{B} \in [0, 1]^{K \times K}$ , i.e.  $(B)_{cc'} = p_{cc'}$  where  $K \ll N$ .

Thus, the parameters of the SBM are  $\boldsymbol{\theta}$  and  $\mathbf{B}$ . In Bayesian statistics, we need to specify their prior distribution  $p(\boldsymbol{\theta}, \mathbf{B})$ . Commonly an independent Dirichlet prior for  $\boldsymbol{\theta}$  and Beta priors for the entries of  $\mathbf{B}$  are assumed. These are also the (conditionally) conjugate priors for the SBM.

Overall, we arrive at the following joint probability:

$$\begin{aligned} p(\mathbf{A}, \mathbf{c}, \boldsymbol{\theta}, \mathbf{B}) &= p(\boldsymbol{\theta})p(\mathbf{B})p(\mathbf{c}|\boldsymbol{\theta})p(\mathbf{A}|\mathbf{c}, \mathbf{B}) \\ &= p(\boldsymbol{\theta}) \prod_{c, c'} p(B_{cc'}) \prod_i p(c_i|\boldsymbol{\theta}) \prod_{ij} p(A_{ij}|B_{c_i c_j}) \\ &= p(\boldsymbol{\theta}) \prod_{c, c'} p(B_{cc'}) \prod_i \theta_{c_i} \prod_{ij} B_{c_i c_j}^{A_{ij}} (1 - B_{c_i c_j})^{1-A_{ij}} \\ &= p(\boldsymbol{\theta}) \prod_{c, c'} p(B_{cc'}) \prod_c \theta_c^{N_c} \prod_{c, c'} B_{cc'}^{n_{cc'}} (1 - B_{cc'})^{N_c N_{c'} - n_{cc'}} \end{aligned} \quad (4)$$

where we have used the sufficient statistics

- $N_c$ , counting number of nodes in each group  $c$
- and  $n_{cc'}$ , counting number of edges between groups  $c$  and  $c'$ .

Note that it is not possible to marginalize over the latent class assignment, as the resulting summation

$$\sum_{\mathbf{c}} p(\mathbf{A}, \mathbf{c}, \boldsymbol{\theta}, \mathbf{B}) = \sum_{c_1=1}^K \cdots \sum_{c_N=1}^K p(\mathbf{A}, \mathbf{c}, \boldsymbol{\theta}, \mathbf{B}) \quad (5)$$

involves exponentially many terms and cannot be simplified/separated. Thus, we need to resort to some approximation algorithm ...

### 1.2.1 Variational inference

A tractable approximation for the SBM can be obtained by variational inference. In this case, the full posterior  $p(\mathbf{c}, \boldsymbol{\theta}, \mathbf{B} | \mathbf{A})$  is approximated with simpler distribution that in particular factorizes over all nodes, i.e.

$$q(\mathbf{c}, \boldsymbol{\theta}, \mathbf{B}) = \prod_i q(c_i) q(\boldsymbol{\theta}) \prod_{c, c'} q(B_{cc'}) . \quad (6)$$

Here, the discrete distributions  $q(c_i)$  are obviously categorical, i.e.  $q(c; \boldsymbol{\eta}) = \eta_c$  and by conjugacy the optimal choice for  $q(\boldsymbol{\theta})$  and  $q(B_{cc'})$  can be shown to be the Dirichlet and Beta distribution respectively.

Using that approximation the log marginal likelihood can be approximated by the ELBO and especially the expectation over the approximating distribution for the latent class assignment can be carried out explicitly:

$$\begin{aligned} & \mathbb{E}_{q(\mathbf{c}, \boldsymbol{\theta}, \mathbf{B})} [\log p(\mathbf{A}, \mathbf{c}, \boldsymbol{\theta}, \mathbf{B})] + H(q(\mathbf{c}, \boldsymbol{\theta}, \mathbf{B})) \\ &= \mathbb{E}_{q(\boldsymbol{\theta}, \mathbf{B})} \left[ \mathbb{E}_{q(\mathbf{c})} \left[ \log p(\boldsymbol{\theta}) + \sum_{cc'} \log p(B_{cc'}) + \sum_i \log p(c_i | \boldsymbol{\theta}) + \sum_{ij} \log p(A_{ij} | B_{c_i c_j}) \mid \boldsymbol{\theta}, \mathbf{B} \right] \right. \\ & \quad \left. + H(q(\mathbf{c}, \boldsymbol{\theta}, \mathbf{B})) \right] \\ &= \mathbb{E}_{q(\boldsymbol{\theta}, \mathbf{B})} \left[ \log p(\boldsymbol{\theta}) + \sum_{cc'} \log p(B_{cc'}) + \sum_{c_1=1}^K \cdots \sum_{c_N=1}^K \prod_i q(c_i) \left( \sum_i \log p(c_i | \boldsymbol{\theta}) + \sum_{ij} \log p(A_{ij} | B_{c_i c_j}) \right) \right] \\ & \quad + H(q(\mathbf{c}, \boldsymbol{\theta}, \mathbf{B})) \\ &= \mathbb{E}_{q(\boldsymbol{\theta}, \mathbf{B})} \left[ \log p(\boldsymbol{\theta}) + \sum_{cc'} \log p(B_{cc'}) + \sum_i \sum_{c_i=1}^K q(c_i) \log p(c_i | \boldsymbol{\theta}) \right. \\ & \quad \left. + \sum_{ij} \sum_{c_i=1}^K \sum_{c_j=1}^K q(c_i) q(c_j) \log p(A_{ij} | B_{c_i c_j}) \right] + \sum_i H(q(c_i)) + H(\boldsymbol{\theta}) + \sum_{cc'} H(q(B_{cc'})) \\ &= \mathbb{E}_{q(\boldsymbol{\theta}, \mathbf{B})} \left[ \log p(\boldsymbol{\theta}) + \sum_{cc'} \log p(B_{cc'}) + \sum_i \sum_{c_i=1}^K \eta_{c_i} \log \theta_{c_i} \right. \\ & \quad \left. + \sum_{ij} \sum_{c_i=1}^K \sum_{c_j=1}^K \eta_{c_i} \eta_{c_j} \left( A_{ij} \log B_{c_i c_j} + (1 - A_{ij}) \log(1 - B_{c_i c_j}) \right) \right] \\ & \quad - \sum_i \sum_{c_i=1}^K \eta_{c_i} \log \eta_{c_i} + \mathbb{E}_{q(\boldsymbol{\theta})} [-\log q(\boldsymbol{\theta})] + \sum_{cc'} \mathbb{E}_{q(B_{cc'})} [-\log q(B_{cc'})] \end{aligned} \quad (7)$$

The remaining expectation over  $q(\boldsymbol{\theta}, \mathbf{B})$  can be approximated by Monte-Carlo sampling as explained above. Further, the sum  $\sum_{c_i=1}^K \sum_{c_j=1}^K \eta_{c_i} \eta_{c_j} \log B_{c_i c_j}$  can be conveniently written as a matrix multiplication  $\boldsymbol{\eta}_i^T \log(\mathbf{B}) \boldsymbol{\eta}_j$ .

### 1.2.2 Scalable algorithm

A scalable algorithm is then obtained by maximizing the stochastic ELBO with respect to all variational parameters, i.e. of  $q(\boldsymbol{\theta}, \mathbf{B})$  and  $\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_N$ .

For even larger data sets also minibatch training should be feasible. To this end, the ELBO can be written as a sum over all vertex pairs

$$\begin{aligned} & \sum_{\substack{i,j \\ i \neq j}} \left[ \frac{1}{N(N-1)} \mathbb{E}_{q(\boldsymbol{\theta})} [\log p(\boldsymbol{\theta}) - \log q(\boldsymbol{\theta})] \right. \\ & + \frac{1}{N(N-1)} \sum_{c_i=1}^K \sum_{c_j=1}^K \mathbb{E}_{q(B_{c_i c_j})} [\log p(B_{c_i c_j}) - \log q(B_{c_i c_j})] \\ & + \frac{1}{N-1} \sum_{c_i=1}^K \eta_{c_i} (\mathbb{E}_{q(\boldsymbol{\theta})} [\log \theta_{c_i}] - \log \eta_{c_i}) \\ & \left. + \sum_{c_i=1}^K \sum_{c_j=1}^K \eta_{c_i} \eta_{c_j} \mathbb{E}_{q(B_{c_i c_j})} \left[ A_{ij} \log B_{c_i c_j} + (1 - A_{ij}) \log(1 - B_{c_i c_j}) \right] \right] \end{aligned} \quad (8)$$

where we assume that the diagonal of the adjacency matrix, i.e.  $i = j$ , is excluded from the likelihood. The gradient can then be evaluated with respect to a minibatch containing just a fraction of all  $N(N-1)$  vertex pairs.

Note that using conjugate priors and matches variational distributions all expectations can be computed analytically. The minibatch ELBO is thus a function of the variational parameters and can be optimized with standard optimizers, e.g. minibatch stochastic gradient ascent or BFGS over all data points. Obviously, using the reparameterization explained above also a doubly stochastic variant with additional Monte-Carlo sampling of the expectations is feasible.

### 1.2.3 Algorithm implementation

Implementing this version of the algorithm, we separated different parts of ELBO as building blocks responsible for different model characteristics. Later on, adding new blocks we can modify ELBO for more advanced models, those utilize degree correction and non-zero-one weights of edges.

As we said before, we assume

$$\begin{aligned} \boldsymbol{\theta} & \sim \text{Dirichlet}(\theta_1, \dots, \theta_K) \\ \mathbf{B} & \sim \text{Beta}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \end{aligned}$$

This allows us to calculate the whole ELBO analytically. The first and the second terms can be presented through Kullback-Leibler divergence between two Dirichlet distributions [?]:

$$\mathbb{E}_{q(\boldsymbol{\theta})} [\log p(\boldsymbol{\theta}) - \log q(\boldsymbol{\theta})] = -D_{KL}(q(\boldsymbol{\theta}) || p(\boldsymbol{\theta})) \quad (9)$$

$$\begin{aligned}
\sum_{c_i=1}^K \sum_{c_j=1}^K \mathbb{E}_{q(B_{c_i c_j})} [\log p(B_{c_i c_j}) - \log q(B_{c_i c_j})] &= \mathbb{E}_{q(\mathbf{B})} [\log p(\mathbf{B}) - \log q(\mathbf{B})] \\
&= -D_{KL}(q(\mathbf{B})||p(\mathbf{B}))
\end{aligned} \tag{10}$$

Let us denote the third term as  $\Phi(i)$  and the last one as  $\Omega(i, j)$ . In both cases, the expectation of the logarithm can be calculated with digamma function  $\psi()$ , i.e.

$$\begin{aligned}
\Phi(i) &= \sum_{c_i=1}^K \eta_{c_i} \left( \mathbb{E}_{q(\boldsymbol{\theta})} [\log \theta_{c_i}] - \log \eta_{c_i} \right) \\
&= \sum_{c_i=1}^K \eta_{c_i} \left( \psi(\theta_{c_i}) - \psi \left( \sum_{c_i} \theta_{c_i} \right) - \log \eta_{c_i} \right)
\end{aligned} \tag{11}$$

$$\begin{aligned}
\Omega(i, j) &= \sum_{c_i=1}^K \sum_{c_j=1}^K \eta_{c_i} \eta_{c_j} \mathbb{E}_{q(B_{c_i c_j})} \left[ A_{ij} \log B_{c_i c_j} + (1 - A_{ij}) \log(1 - B_{c_i c_j}) \right] \\
&= \sum_{c_i=1}^K \sum_{c_j=1}^K \eta_{c_i} \eta_{c_j} \left( A_{ij} \left( \psi(\alpha_{c_i c_j}) - \psi(\alpha_{c_i c_j} + \beta_{c_i c_j}) \right) + (1 - A_{ij}) \left( \psi(\beta_{c_i c_j}) - \psi(\alpha_{c_i c_j} + \beta_{c_i c_j}) \right) \right)
\end{aligned} \tag{12}$$

Combining Eq. 8-12, ELBO for a minibatch  $\Lambda$  of length  $L$  which contains edges  $(i, j)$  can be written as:

$$\begin{aligned}
ELBO_{SBM}(\Lambda) &= \sum_{\substack{i, j \\ i \neq j}} \left( -\frac{L}{N(N-1)} D_{KL}(q(\boldsymbol{\theta})||p(\boldsymbol{\theta})) \right. \\
&\quad - \frac{L}{N(N-1)} D_{KL}(q(\mathbf{B})||p(\mathbf{B})) \\
&\quad + \frac{L}{N-1} \Phi(i) \\
&\quad \left. + \Omega(i, j) \right)
\end{aligned} \tag{13}$$

### 1.3 Stochastic block model with degree correction

If network's nodes have significantly different degree, SBM can prefer to separate nodes not based on their latent class but on their degree. In order to overcome this difficulties, the degree-corrected SBM (DCSBM) was proposed. The original model [?] uses Poisson distribution to define connection's probability between two nodes, which makes it more suitable for multigraphs. We decided to introduce degree correction in the standard SBM within Bernoulli distribution:

- A link between  $i$  and  $j$  is drawn with probability  $p_{i,j}$ , i.e.

$$A_{ij} \sim \text{Bernoulli}(p_{i,j})$$

where

$$p_{i,j} = \frac{e^{\delta_i + \delta_j}}{1 + e^{\delta_i + \delta_j}} \cdot B_{c_i, c_j} = \sigma(\delta_i + \delta_j) \cdot B_{c_i, c_j}$$

Here  $\sigma()$  is a sigmoid function,  $\delta_i$  and  $\delta_j$  correspond to expected degree of  $i$  and  $j$ .

Further, we arrive at the following joint probability:

$$\begin{aligned} p(\mathbf{A}, \mathbf{c}, \boldsymbol{\theta}, \mathbf{B}, \boldsymbol{\delta}) &= p(\boldsymbol{\theta})p(\mathbf{B})p(\boldsymbol{\delta})p(\mathbf{c}|\boldsymbol{\theta})p(\mathbf{A}|\mathbf{c}, \mathbf{B}, \boldsymbol{\delta}) \\ &= p(\boldsymbol{\theta}) \prod_{c, c'} p(B_{cc'}) \prod_i p(\delta_i) \prod_i p(c_i|\boldsymbol{\theta}) \prod_{ij} p(A_{ij}|B_{c_i c_j}, \delta_i, \delta_j) \\ &= p(\boldsymbol{\theta}) \prod_{c, c'} p(B_{cc'}) \prod_i p(\delta_i) \prod_i \theta_{c_i} \\ &\quad \cdot \prod_{ij} (\sigma(\delta_i + \delta_j) \cdot B_{c_i, c_j})^{A_{ij}} (1 - (\sigma(\delta_i + \delta_j) \cdot B_{c_i, c_j}))^{1-A_{ij}} \end{aligned} \tag{14}$$

### 1.3.1 Variational inference

Analog to simple SBM, the posterior is approximated with

$$q(\mathbf{c}, \boldsymbol{\theta}, \mathbf{B}, \boldsymbol{\delta}) = \prod_i q(c_i) q(\boldsymbol{\theta}) \prod_{c, c'} q(B_{cc'}) \prod_i q(\delta_i). \tag{15}$$

and the ELBO is

$$\begin{aligned}
& \mathbb{E}_{q(\mathbf{c}, \boldsymbol{\theta}, \mathbf{B}, \boldsymbol{\delta})} [\log p(\mathbf{A}, \mathbf{c}, \boldsymbol{\theta}, \mathbf{B}, \boldsymbol{\delta})] + H(q(\mathbf{c}, \boldsymbol{\theta}, \mathbf{B}, \boldsymbol{\delta})) \\
&= \mathbb{E}_{q(\boldsymbol{\theta}, \mathbf{B}, \boldsymbol{\delta})} \left[ \mathbb{E}_{q(\mathbf{c})} \left[ \log p(\boldsymbol{\theta}) + \sum_{cc'} \log p(B_{cc'}) + \sum_i \log p(\delta_i) \right. \right. \\
&\quad \left. \left. + \sum_i \log p(c_i | \boldsymbol{\theta}) + \sum_{ij} \log p(A_{ij} | B_{c_i c_j}, \delta_i, \delta_j) \mid \boldsymbol{\theta}, \mathbf{B}, \boldsymbol{\delta} \right] \right] \\
&\quad + H(q(\mathbf{c}, \boldsymbol{\theta}, \mathbf{B}, \boldsymbol{\delta})) \\
&= \mathbb{E}_{q(\boldsymbol{\theta}, \mathbf{B}, \boldsymbol{\delta})} \left[ \log p(\boldsymbol{\theta}) + \sum_{cc'} \log p(B_{cc'}) + \sum_i \log p(\delta_i) \right. \\
&\quad \left. + \sum_{c_1=1}^K \cdots \sum_{c_N=1}^K \prod_i q(c_i) \left( \sum_i \log p(c_i | \boldsymbol{\theta}) + \sum_{ij} \log p(A_{ij} | B_{c_i c_j}, \delta_i, \delta_j) \right) \right] \\
&\quad + H(q(\mathbf{c}, \boldsymbol{\theta}, \mathbf{B}, \boldsymbol{\delta})) \\
&= \mathbb{E}_{q(\boldsymbol{\theta}, \mathbf{B}, \boldsymbol{\delta})} \left[ \log p(\boldsymbol{\theta}) + \sum_{cc'} \log p(B_{cc'}) + \sum_i \log p(\delta_i) \right. \\
&\quad \left. + \sum_i \sum_{c_i=1}^K q(c_i) \log p(c_i | \boldsymbol{\theta}) + \sum_{ij} \sum_{c_i=1}^K \sum_{c_j=1}^K q(c_i) q(c_j) \log p(A_{ij} | B_{c_i c_j}, \delta_i, \delta_j) \right] \\
&\quad + \sum_i H(q(c_i)) + H(\boldsymbol{\theta}) + \sum_{cc'} H(q(B_{cc'})) + \sum_i H(q(\delta_i)) \\
&= \mathbb{E}_{q(\boldsymbol{\theta}, \mathbf{B}, \boldsymbol{\delta})} \left[ \log p(\boldsymbol{\theta}) + \sum_{cc'} \log p(B_{cc'}) + \sum_i \log p(\delta_i) + \sum_i \sum_{c_i=1}^K \eta_{c_i} \log \theta_{c_i} \right. \\
&\quad \left. + \sum_{ij} \sum_{c_i=1}^K \sum_{c_j=1}^K \eta_{c_i} \eta_{c_j} \left( A_{ij} \log(\sigma(\delta_i + \delta_j) \cdot B_{c_i c_j}) + (1 - A_{ij}) \log(1 - (\sigma(\delta_i + \delta_j) \cdot B_{c_i c_j})) \right) \right] \\
&\quad - \sum_i \sum_{c_i=1}^K \eta_{c_i} \log \eta_{c_i} + \mathbb{E}_{q(\boldsymbol{\theta})} [-\log q(\boldsymbol{\theta})] + \sum_{cc'} \mathbb{E}_{q(B_{cc'})} [-\log q(B_{cc'})] + \sum_i \mathbb{E}_{q(\delta_i)} [-\log q(\delta_i)]
\end{aligned} \tag{16}$$

### 1.3.2 Scalable algorithm

As in the previous section, ELBO for a minibatch can be written as a sum over all vertex pairs

$$\begin{aligned}
& \sum_{\substack{ij \\ i \neq j}} \left[ \frac{1}{N(N-1)} \mathbb{E}_{q(\boldsymbol{\theta})} [\log p(\boldsymbol{\theta}) - \log q(\boldsymbol{\theta})] \right. \\
& + \frac{1}{N(N-1)} \sum_{c_i=1}^K \sum_{c_j=1}^K \mathbb{E}_{q(B_{c_i c_j})} [\log p(B_{c_i c_j}) - \log q(B_{c_i c_j})] \\
& + \frac{1}{N-1} \sum_{c_i=1}^K \eta_{c_i} (\mathbb{E}_{q(\boldsymbol{\theta})} [\log \theta_{c_i}] - \log \eta_{c_i}) \\
& + \frac{1}{N-1} \sum_i \mathbb{E}_{q(\boldsymbol{\delta})} [\log p(\delta_i) - \log q(\delta_i)] \\
& \left. + \sum_{c_i=1}^K \sum_{c_j=1}^K \eta_{c_i} \eta_{c_j} \mathbb{E}_{q(B_{c_i c_j})} \left[ A_{ij} \log(\sigma(\delta_i + \delta_j) \cdot B_{c_i c_j}) + (1 - A_{ij}) \log(1 - (\sigma(\delta_i + \delta_j) \cdot B_{c_i c_j})) \right] \right] \\
\end{aligned} \tag{17}$$

### 1.3.3 Algorithm implementation

Comparing the algorithm for DCSBM and SBM, we see that the first three terms stay the same. The next assumption

$$\boldsymbol{\delta} \sim \text{Normal}(\mu_{\delta}, \sigma_{\delta}^2)$$

allows us to calculate also the third term analytically, i.e.

$$\sum_i \mathbb{E}_{q(\boldsymbol{\delta})} [\log p(\delta_i) - \log q(\delta_i)] = -D_{KL}(q(\boldsymbol{\delta}) || p(\boldsymbol{\delta})) \tag{18}$$

Unfortunately, there is no analytical solution the last term  $\widehat{\Omega}(i, j)$  and it requires approximation of the expectation through sampling. The final formula for the ELBO is written as

$$\begin{aligned}
ELBO_{DCSBM}(\Lambda) = \sum_{\substack{ij \\ i \neq j}} & \left( -\frac{L}{N(N-1)} D_{KL}(q(\boldsymbol{\theta}) || p(\boldsymbol{\theta})) \right. \\
& - \frac{L}{N(N-1)} D_{KL}(q(\mathbf{B}) || p(\mathbf{B})) \\
& + \frac{L}{N-1} \Phi(i) \\
& - \frac{L}{N-1} D_{KL}(q(\boldsymbol{\delta}) || p(\boldsymbol{\delta})) \\
& \left. + \widehat{\Omega}(i, j) \right) \\
\end{aligned} \tag{19}$$



## 1.4 Stochastic block models for weighted graphs

Two of the presented models and algorithms can be easily modified to work with weighted graphs. In weighted SBM and DCSBM (WSBM and WDCSBM) after a link between nodes is drawn, its weight is sampled from the distribution

$$w_{ij} \sim \text{Log} - \text{Normal}(\mu_{c_i c_j}, 1/\tau_{c_i c_j}^2)$$

which depends on the latent class assignments of the nodes. Here  $\tau_{c_i c_j}$  is the precision that is inversely proportional to the scale. Note, that in this case all weights are non-negative. This allows us to interpret all edges with zero weights as not existing.

The join probability of WSBM will have the following form:

$$\begin{aligned} p(\mathbf{A}, \mathbf{c}, \boldsymbol{\theta}, \mathbf{B}, \mathbf{w}) &= p(\boldsymbol{\theta})p(\mathbf{B})p(\mathbf{w})p(\mathbf{c}|\boldsymbol{\theta})p(\mathbf{A}|\mathbf{c}, \mathbf{B}) \\ &= p(\boldsymbol{\theta}) \prod_{c, c'} p(B_{cc'}) \prod_{c, c'} p(\mu_{cc'}) \prod_{c, c'} p(\tau_{cc'}) \prod_i p(c_i|\boldsymbol{\theta}) \\ &\quad \cdot \prod_{ij} p(A_{ij}|B_{c_i c_j}) \prod_{ij} p(w_{ij}|\mu_{c_i c_j}, \tau_{c_i c_j}) \\ &= p(\boldsymbol{\theta}) \prod_{c, c'} p(B_{cc'}) \prod_{c, c'} p(\mu_{cc'}) \prod_{c, c'} p(\tau_{cc'}) \prod_i \theta_{c_i} \\ &\quad \cdot \prod_{ij} B_{c_i c_j}^{A_{ij}} (1 - B_{c_i c_j})^{1-A_{ij}} \prod_{ij} p(w_{ij}|\mu_{c_i c_j}, \tau_{c_i c_j}), \end{aligned} \tag{20}$$

its variational distribution will be

$$q(\mathbf{c}, \boldsymbol{\theta}, \mathbf{B}, \boldsymbol{\mu}, \boldsymbol{\tau}) = \prod_i q(c_i)q(\boldsymbol{\theta}) \prod_{c, c'} q(B_{cc'}) \prod_{c, c'} q(\mu_{cc'}) \prod_{c, c'} q(\tau_{cc'}) , \tag{21}$$

and the ELBO

$$\begin{aligned}
& \mathbb{E}_{q(\mathbf{c}, \boldsymbol{\theta}, \mathbf{B}, \boldsymbol{\mu}, \boldsymbol{\tau})} [\log p(\mathbf{A}, \mathbf{c}, \boldsymbol{\theta}, \mathbf{B}, \boldsymbol{\mu}, \boldsymbol{\tau})] + H(q(\mathbf{c}, \boldsymbol{\theta}, \mathbf{B}, \boldsymbol{\mu}, \boldsymbol{\tau})) \\
&= \mathbb{E}_{q(\boldsymbol{\theta}, \mathbf{B}, \boldsymbol{\mu}, \boldsymbol{\tau})} \left[ \mathbb{E}_{q(\mathbf{c})} \left[ \log p(\boldsymbol{\theta}) + \sum_{cc'} \log p(B_{cc'}) + \sum_{cc'} \log p(\mu_{cc'}) + \sum_{cc'} \log p(\tau_{cc'}) \right. \right. \\
&\quad \left. \left. + \sum_i \log p(c_i | \boldsymbol{\theta}) + \sum_{ij} \log p(A_{ij} | B_{c_i c_j}) + \sum_{ij} \log p(w_{ij} | \mu_{c_i c_j}, \tau_{c_i c_j}) \mid \boldsymbol{\theta}, \mathbf{B}, \boldsymbol{\mu}, \boldsymbol{\tau} \right] \right] \\
&\quad + H(q(\mathbf{c}, \boldsymbol{\theta}, \mathbf{B}, \boldsymbol{\mu}, \boldsymbol{\tau})) \\
&= \mathbb{E}_{q(\boldsymbol{\theta}, \mathbf{B}, \boldsymbol{\mu}, \boldsymbol{\tau})} \left[ \log p(\boldsymbol{\theta}) + \sum_{cc'} \log p(B_{cc'}) + \sum_{cc'} \log p(\mu_{cc'}) + \sum_{cc'} \log p(\tau_{cc'}) \right. \\
&\quad \left. + \sum_{c_1=1}^K \cdots \sum_{c_N=1}^K \prod_i q(c_i) \left( \sum_i \log p(c_i | \boldsymbol{\theta}) + \sum_{ij} \log p(A_{ij} | B_{c_i c_j}) + \sum_{ij} \log p(w_{ij} | \mu_{c_i c_j}, \tau_{c_i c_j}) \right) \right] \\
&\quad + H(q(\mathbf{c}, \boldsymbol{\theta}, \mathbf{B}, \boldsymbol{\mu}, \boldsymbol{\tau})) \\
&= \mathbb{E}_{q(\boldsymbol{\theta}, \mathbf{B}, \boldsymbol{\mu}, \boldsymbol{\tau})} \left[ \log p(\boldsymbol{\theta}) + \sum_{cc'} \log p(B_{cc'}) + \sum_{cc'} \log p(\mu_{cc'}) + \sum_{cc'} \log p(\tau_{cc'}) \right. \\
&\quad + \sum_i \sum_{c_i=1}^K q(c_i) \log p(c_i | \boldsymbol{\theta}) + \sum_{ij} \sum_{c_i=1}^K \sum_{c_j=1}^K q(c_i) q(c_j) \log p(A_{ij} | B_{c_i c_j}) \\
&\quad \left. + \sum_{ij} \sum_{c_i=1}^K \sum_{c_j=1}^K q(c_i) q(c_j) \log p(w_{ij} | \mu_{c_i c_j}, \tau_{c_i c_j}) \right] \\
&\quad + \sum_i H(q(c_i)) + H(\boldsymbol{\theta}) + \sum_{cc'} H(q(B_{cc'})) + \sum_{cc'} H(q(\mu_{cc'})) + \sum_{cc'} H(q(\tau_{cc'})) \\
&= \mathbb{E}_{q(\boldsymbol{\theta}, \mathbf{B}, \boldsymbol{\mu}, \boldsymbol{\tau})} \left[ \log p(\boldsymbol{\theta}) + \sum_{cc'} \log p(B_{cc'}) + \sum_{cc'} \log p(\mu_{cc'}) + \sum_{cc'} \log p(\tau_{cc'}) + \sum_i \sum_{c_i=1}^K \eta_{c_i} \log \theta_{c_i} \right. \\
&\quad + \sum_{ij} \sum_{c_i=1}^K \sum_{c_j=1}^K \eta_{c_i} \eta_{c_j} \left( A_{ij} \log B_{c_i c_j} + (1 - A_{ij}) \log(1 - B_{c_i c_j}) \right) \\
&\quad + \sum_{ij} \sum_{c_i=1}^K \sum_{c_j=1}^K \eta_{c_i} \eta_{c_j} \log p(w_{ij} | \mu_{c_i c_j}, \tau_{c_i c_j}) \left. \right] - \sum_i \sum_{c_i=1}^K \eta_{c_i} \log \eta_{c_i} + \mathbb{E}_{q(\boldsymbol{\theta})} [-\log q(\boldsymbol{\theta})] \\
&\quad + \sum_{cc'} \mathbb{E}_{q(B_{cc'})} [-\log q(B_{cc'})] + \sum_{cc'} \mathbb{E}_{q(\mu_{cc'})} [-\log q(\mu_{cc'})] + \sum_{cc'} \mathbb{E}_{q(\tau_{cc'})} [-\log q(\tau_{cc'})] \\
&\hspace{15em} (22)
\end{aligned}$$

ELBO for a minibatch is

$$\begin{aligned}
& \sum_{\substack{ij \\ i \neq j}} \left[ \frac{1}{N(N-1)} \mathbb{E}_{q(\boldsymbol{\theta})} \left[ \log p(\boldsymbol{\theta}) - \log q(\boldsymbol{\theta}) \right] \right. \\
& + \frac{1}{N(N-1)} \sum_{c_i=1}^K \sum_{c_j=1}^K \mathbb{E}_{q(B_{c_i c_j})} \left[ \log p(B_{c_i c_j}) - \log q(B_{c_i c_j}) \right] \\
& + \frac{1}{N-1} \sum_{c_i=1}^K \eta_{c_i} \left( \mathbb{E}_{q(\boldsymbol{\theta})} [\log \theta_{c_i}] - \log \eta_{c_i} \right) \\
& + \sum_{c_i=1}^K \sum_{c_j=1}^K \eta_{c_i} \eta_{c_j} \mathbb{E}_{q(B_{c_i c_j})} \left[ A_{ij} \log B_{c_i c_j} + (1 - A_{ij}) \log(1 - B_{c_i c_j}) \right] \quad (23) \\
& + \frac{1}{N(N-1)} \sum_{c_i=1}^K \sum_{c_j=1}^K \mathbb{E}_{q(\mu_{c_i c_j})} \left[ \log p(\mu_{c_i c_j}) - \log q(\mu_{c_i c_j}) \right] \\
& + \frac{1}{N(N-1)} \sum_{c_i=1}^K \sum_{c_j=1}^K \mathbb{E}_{q(\tau_{c_i c_j})} \left[ \log p(\tau_{c_i c_j}) - \log q(\tau_{c_i c_j}) \right] \\
& + \left. \sum_{c_i=1}^K \sum_{c_j=1}^K \eta_{c_i} \eta_{c_j} \mathbb{E}_{q(\mu_{c_i c_j}, \tau_{c_i c_j})} \left[ \log p(w_{ij} | \mu_{c_i c_j}, \tau_{c_i c_j}) \right] \right]
\end{aligned}$$

#### 1.4.1 Algorithm implementation

Assume

$$\begin{aligned}
\boldsymbol{\mu} & \sim \text{Normal}(\mathbf{m}, \mathbf{s}^2) \\
\boldsymbol{\tau} & \sim \text{Gamma}(\mathbf{k}, t)
\end{aligned}$$

where  $k$  is a shape and  $t$  is a scale.

First four terms of ELBO are similar to the algorithm for SBM. The fifth and sixth terms can be computed through Kullback-Leibler divergence between two Normal and two Gamma distributions:

$$\begin{aligned}
\sum_{c_i=1}^K \sum_{c_j=1}^K \mathbb{E}_{q(\mu_{c_i c_j})} [\log p(\mu_{c_i c_j}) - \log q(\mu_{c_i c_j})] &= \mathbb{E}_{q(\boldsymbol{\mu})} [\log p(\boldsymbol{\mu}) - \log q(\boldsymbol{\mu})] \quad (24) \\
&= -D_{KL}(q(\boldsymbol{\mu}) || p(\boldsymbol{\mu}))
\end{aligned}$$

$$\begin{aligned}
\sum_{c_i=1}^K \sum_{c_j=1}^K \mathbb{E}_{q(\tau_{c_i c_j})} [\log p(\tau_{c_i c_j}) - \log q(\tau_{c_i c_j})] &= \mathbb{E}_{q(\boldsymbol{\tau})} [\log p(\boldsymbol{\tau}) - \log q(\boldsymbol{\tau})] \quad (25) \\
&= -D_{KL}(q(\boldsymbol{\tau}) || p(\boldsymbol{\tau}))
\end{aligned}$$

The last term  $\Psi(i, j)$  can be also calculated as the logarithm of PDF of a Normal distribution:

$$\begin{aligned}
\Psi(i, j) &= \mathbb{E}_{q(\mu_{c_i c_j}, \tau_{c_i c_j})} \left[ \log p(w_{ij} | \mu_{c_i c_j}, \tau_{c_i c_j}) \right] \\
&= -\frac{1}{2} \log(2\pi) + \frac{1}{2} \mathbb{E}_{q(\tau_{c_i c_j})} [\log(\tau_{c_i c_j})] - \log(w_{ij}) \\
&\quad - \frac{1}{2} \mathbb{E}_{q(\tau_{c_i c_j})} [\tau_{c_i c_j}] \mathbb{E}_{q(\mu_{c_i c_j})} [\log^2(w_{ij}) - 2\mu_{c_i c_j} \log(w_{ij}) + \mu_{c_i c_j}^2] \quad (26) \\
&= -\frac{1}{2} \log(2\pi) + \frac{1}{2} \left( \psi(k_{c_i c_j}) + \log(t_{c_i c_j}) \right) - \log(w_{ij}) \\
&\quad - \frac{1}{2} k_{c_i c_j} t_{c_i c_j} \left( \log^2(w_{ij}) - 2m_{c_i c_j} \log(w_{ij}) + m_{c_i c_j}^2 + s_{c_i c_j}^2 \right)
\end{aligned}$$

So the whole ELBO is implemented as

$$\begin{aligned}
ELBO_{WSBM}(\Lambda) &= \sum_{\substack{ij \\ i \neq j}} \left( -\frac{L}{N(N-1)} D_{KL}(q(\boldsymbol{\theta}) || p(\boldsymbol{\theta})) \right. \\
&\quad - \frac{L}{N(N-1)} D_{KL}(q(\mathbf{B}) || p(\mathbf{B})) \\
&\quad - \frac{L}{N(N-1)} D_{KL}(q(\boldsymbol{\mu}) || p(\boldsymbol{\mu})) \\
&\quad - \frac{L}{N(N-1)} D_{KL}(q(\boldsymbol{\tau}) || p(\boldsymbol{\tau})) \\
&\quad + \frac{L}{N-1} \Phi(i) \\
&\quad + \Psi(i) \\
&\quad \left. + \Omega(i, j) \right) \quad (27)
\end{aligned}$$

## 1.5 Degree-corrected stochastic block models for weighted graphs

This model combines part those were already seen in DCSBM and WSBM.

The join probability of WDCSBM will have the following form:

$$\begin{aligned}
p(\mathbf{A}, \mathbf{c}, \boldsymbol{\theta}, \mathbf{B}, \boldsymbol{\delta}, \mathbf{w}) &= p(\boldsymbol{\theta})p(\mathbf{B})p(\boldsymbol{\delta})p(\mathbf{w})p(\mathbf{c}|\boldsymbol{\theta})p(\mathbf{A}|\mathbf{c}, \mathbf{B}, \boldsymbol{\delta}) \\
&= p(\boldsymbol{\theta}) \prod_{c, c'} p(B_{cc'}) \prod_{c, c'} p(\mu_{cc'}) \prod_{c, c'} p(\tau_{cc'}) \prod_i p(\delta_i) \prod_i p(c_i|\boldsymbol{\theta}) \\
&\quad \cdot \prod_{ij} p(A_{ij}|B_{c_i c_j}, \delta_i, \delta_j) \prod_{ij} p(w_{ij}|\mu_{c_i c_j}, \tau_{c_i c_j}) \\
&= p(\boldsymbol{\theta}) \prod_{c, c'} p(B_{cc'}) \prod_{c, c'} p(\mu_{cc'}) \prod_{c, c'} p(\tau_{cc'}) \prod_i p(\delta_i) \prod_i \theta_{c_i} \\
&\quad \cdot \prod_{ij} (\sigma(\delta_i + \delta_j) \cdot B_{c_i, c_j})^{A_{ij}} (1 - (\sigma(\delta_i + \delta_j) \cdot B_{c_i, c_j}))^{1-A_{ij}} \\
&\quad \cdot \prod_{ij} p(w_{ij}|\mu_{c_i c_j}, \tau_{c_i c_j}),
\end{aligned} \tag{28}$$

Its variational distribution will be

$$q(\mathbf{c}, \boldsymbol{\theta}, \mathbf{B}, \boldsymbol{\delta}, \boldsymbol{\mu}, \boldsymbol{\tau}) = \prod_i q(c_i)q(\boldsymbol{\theta}) \prod_{c, c'} q(B_{cc'}) \prod_i q(\delta_i) \prod_{c, c'} q(\mu_{cc'}) \prod_{c, c'} q(\tau_{cc'}) , \tag{29}$$

and the ELBO

$$\begin{aligned}
& \mathbb{E}_{q(\mathbf{c}, \boldsymbol{\theta}, \mathbf{B}, \boldsymbol{\delta}, \boldsymbol{\mu}, \boldsymbol{\tau})} [\log p(\mathbf{A}, \mathbf{c}, \boldsymbol{\theta}, \mathbf{B}, \boldsymbol{\delta}, \boldsymbol{\mu}, \boldsymbol{\tau})] + H(q(\mathbf{c}, \boldsymbol{\theta}, \mathbf{B}, \boldsymbol{\delta}, \boldsymbol{\mu}, \boldsymbol{\tau})) \\
&= \mathbb{E}_{q(\boldsymbol{\theta}, \mathbf{B}, \boldsymbol{\delta}, \boldsymbol{\mu}, \boldsymbol{\tau})} \left[ \mathbb{E}_{q(\mathbf{c})} \left[ \log p(\boldsymbol{\theta}) + \sum_{cc'} \log p(B_{cc'}) + \sum_i \log p(\delta_i) + \sum_{cc'} \log p(\mu_{cc'}) + \sum_{cc'} \log p(\tau_{cc'}) \right. \right. \\
&\quad \left. \left. + \sum_i \log p(c_i | \boldsymbol{\theta}) + \sum_{ij} \log p(A_{ij} | B_{c_i c_j}, \delta_i, \delta_j) + \sum_{ij} \log p(w_{ij} | \mu_{c_i c_j}, \tau_{c_i c_j}) \mid \boldsymbol{\theta}, \mathbf{B}, \boldsymbol{\delta}, \boldsymbol{\mu}, \boldsymbol{\tau} \right] \right] \\
&\quad + H(q(\mathbf{c}, \boldsymbol{\theta}, \mathbf{B}, \boldsymbol{\delta}, \boldsymbol{\mu}, \boldsymbol{\tau})) \\
&= \mathbb{E}_{q(\boldsymbol{\theta}, \mathbf{B}, \boldsymbol{\delta}, \boldsymbol{\mu}, \boldsymbol{\tau})} \left[ \log p(\boldsymbol{\theta}) + \sum_{cc'} \log p(B_{cc'}) + \sum_i \log p(\delta_i) + \sum_{cc'} \log p(\mu_{cc'}) + \sum_{cc'} \log p(\tau_{cc'}) \right. \\
&\quad \left. + \sum_{c_1=1}^K \cdots \sum_{c_N=1}^K \prod_i q(c_i) \left( \sum_i \log p(c_i | \boldsymbol{\theta}) + \sum_{ij} \log p(A_{ij} | B_{c_i c_j}, \delta_i, \delta_j) + \sum_{ij} \log p(w_{ij} | \mu_{c_i c_j}, \tau_{c_i c_j}) \right) \right] \\
&\quad + H(q(\mathbf{c}, \boldsymbol{\theta}, \mathbf{B}, \boldsymbol{\delta}, \boldsymbol{\mu}, \boldsymbol{\tau})) \\
&= \mathbb{E}_{q(\boldsymbol{\theta}, \mathbf{B}, \boldsymbol{\delta}, \boldsymbol{\mu}, \boldsymbol{\tau})} \left[ \log p(\boldsymbol{\theta}) + \sum_{cc'} \log p(B_{cc'}) + \sum_i \log p(\delta_i) + \sum_{cc'} \log p(\mu_{cc'}) + \sum_{cc'} \log p(\tau_{cc'}) \right. \\
&\quad + \sum_i \sum_{c_i=1}^K q(c_i) \log p(c_i | \boldsymbol{\theta}) + \sum_{ij} \sum_{c_i=1}^K \sum_{c_j=1}^K q(c_i) q(c_j) \log p(A_{ij} | B_{c_i c_j}, \delta_i, \delta_j) \\
&\quad \left. + \sum_{ij} \sum_{c_i=1}^K \sum_{c_j=1}^K q(c_i) q(c_j) \log p(w_{ij} | \mu_{c_i c_j}, \tau_{c_i c_j}) \right] \\
&\quad + \sum_i H(q(c_i)) + H(\boldsymbol{\theta}) + \sum_{cc'} H(q(B_{cc'})) + \sum_i H(q(\delta_i)) + \sum_{cc'} H(q(\mu_{cc'})) + \sum_{cc'} H(q(\tau_{cc'})) \\
&= \mathbb{E}_{q(\boldsymbol{\theta}, \mathbf{B}, \boldsymbol{\delta}, \boldsymbol{\mu}, \boldsymbol{\tau})} \left[ \log p(\boldsymbol{\theta}) + \sum_{cc'} \log p(B_{cc'}) + \sum_i \log p(\delta_i) + \sum_{cc'} \log p(\mu_{cc'}) + \sum_{cc'} \log p(\tau_{cc'}) \right. \\
&\quad + \sum_{ij} \sum_{c_i=1}^K \sum_{c_j=1}^K \eta_{c_i} \eta_{c_j} \left( A_{ij} \log(\sigma(\delta_i + \delta_j) \cdot B_{c_i c_j}) + (1 - A_{ij}) \log(1 - (\sigma(\delta_i + \delta_j) \cdot B_{c_i c_j})) \right) \\
&\quad + \sum_i \sum_{c_i=1}^K \eta_{c_i} \log \theta_{c_i} + \sum_{ij} \sum_{c_i=1}^K \sum_{c_j=1}^K \eta_{c_i} \eta_{c_j} \log p(w_{ij} | \mu_{c_i c_j}, \tau_{c_i c_j}) \left. \right] \\
&\quad - \sum_i \sum_{c_i=1}^K \eta_{c_i} \log \eta_{c_i} + \mathbb{E}_{q(\boldsymbol{\theta})} [-\log q(\boldsymbol{\theta})] + \sum_{cc'} \mathbb{E}_{q(B_{cc'})} [-\log q(B_{cc'})] + \sum_i \mathbb{E}_{q(\delta_i)} [-\log q(\delta_i)] \\
&\quad + \sum_{cc'} \mathbb{E}_{q(\mu_{cc'})} [-\log q(\mu_{cc'})] + \sum_{cc'} \mathbb{E}_{q(\tau_{cc'})} [-\log q(\tau_{cc'})]
\end{aligned} \tag{30}$$

ELBO for a minibatch is

$$\begin{aligned}
& \sum_{\substack{ij \\ i \neq j}} \left[ \frac{1}{N(N-1)} \mathbb{E}_{q(\boldsymbol{\theta})} \left[ \log p(\boldsymbol{\theta}) - \log q(\boldsymbol{\theta}) \right] \right. \\
& + \frac{1}{N(N-1)} \sum_{c_i=1}^K \sum_{c_j=1}^K \mathbb{E}_{q(B_{c_i c_j})} \left[ \log p(B_{c_i c_j}) - \log q(B_{c_i c_j}) \right] \\
& + \frac{1}{N-1} \sum_i \mathbb{E}_{q(\boldsymbol{\delta})} \left[ \log p(\delta_i) - \log q(\delta_i) \right] \\
& + \frac{1}{N-1} \sum_{c_i=1}^K \eta_{c_i} \left( \mathbb{E}_{q(\boldsymbol{\theta})} [\log \theta_{c_i}] - \log \eta_{c_i} \right) \\
& + \sum_{c_i=1}^K \sum_{c_j=1}^K \eta_{c_i} \eta_{c_j} \mathbb{E}_{q(B_{c_i c_j})} \left[ A_{ij} \log(\sigma(\delta_i + \delta_j) \cdot B_{c_i c_j}) + (1 - A_{ij}) \log(1 - (\sigma(\delta_i + \delta_j) \cdot B_{c_i c_j})) \right] \\
& + \frac{1}{N(N-1)} \sum_{c_i=1}^K \sum_{c_j=1}^K \mathbb{E}_{q(\mu_{c_i c_j})} \left[ \log p(\mu_{c_i c_j}) - \log q(\mu_{c_i c_j}) \right] \\
& + \frac{1}{N(N-1)} \sum_{c_i=1}^K \sum_{c_j=1}^K \mathbb{E}_{q(\tau_{c_i c_j})} \left[ \log p(\tau_{c_i c_j}) - \log q(\tau_{c_i c_j}) \right] \\
& + \left. \sum_{c_i=1}^K \sum_{c_j=1}^K \eta_{c_i} \eta_{c_j} \mathbb{E}_{q(\mu_{c_i c_j}, \tau_{c_i c_j})} \left[ \log p(w_{ij} | \mu_{c_i c_j}, \tau_{c_i c_j}) \right] \right]
\end{aligned} \tag{31}$$

and its implementation

$$\begin{aligned}
ELBO_{WDCSBM}(\Lambda) = & \sum_{\substack{ij \\ i \neq j}} \left( - \frac{L}{N(N-1)} D_{KL}(q(\boldsymbol{\theta}) || p(\boldsymbol{\theta})) \right. \\
& - \frac{L}{N(N-1)} D_{KL}(q(\mathbf{B}) || p(\mathbf{B})) \\
& - \frac{L}{N(N-1)} D_{KL}(q(\boldsymbol{\mu}) || p(\boldsymbol{\mu})) \\
& - \frac{L}{N(N-1)} D_{KL}(q(\boldsymbol{\tau}) || p(\boldsymbol{\tau})) \\
& - \frac{L}{N-1} D_{KL}(q(\boldsymbol{\delta}) || p(\boldsymbol{\delta})) \\
& + \frac{L}{N-1} \Phi(i) \\
& + \Psi(i) \\
& \left. + \hat{\Omega}(i, j) \right)
\end{aligned} \tag{32}$$