Supplementary Material for "Joint Latent Space Model for Social Networks with High-dimensional Multivariate Attributes".

1 Simulation to evaluate the recovery of the true dimension of the network using the cross-validation approach

We perform a small simulation study to evaluate the recovery of the true dimension of the latent space using the cross-validation approach. We generate data with a true dimension between 1 to 4. Then we fit the proposed model to data with the number of latent dimensions ranging between 1 and 5, perform 5-fold cross-validation and compare the out-of-sample AUC values for the different dimensions. If the out-of-sample AUC value is the highest when the fitted model has the number of dimensions the same as the true data generation process, we can conclude that we are able to recover the true number of dimensions using cross-validation. For each number of dimension between 1 to 4, we generate 10 sets of data. To maintain consistency across conditions, we generate data following the model equation with $N = 100$, $M = 100$, $\alpha_0 = \alpha_1 = 0$, and $\lambda_0 = \lambda_1 = 1$. For each data, we perform the cross-validation and report the average out-of-sample AUC values.

The results are shown in Table 1. The highest out-of-sample AUC values for each condition are highlighted in red. The model failed to converge when the number of true dimension is 1 and the dimension specified in the fitted model is 5. The results show that the out-of-sample AUC value is the highest when the fitted model has the number of dimensions the same as the true data generation process. This provides evidence supporting the validity of using cross-validation to recover the true number of dimensions in the proposed model. We have added the simulation to the supplementary material.

Table 1: Out-of-sample AUC values

	Dimension Specified in the Fitted Model				
True Dimension	1	2	3	4	5
1	0.755	0.748	0.734	0.528	
2	0.745	0.829	0.814	0.785	0.509
3	0.717	0.785	0.810	0.794	0.566
4	0.637	0.689	0.698	0.700	0.689

2 Latent Space Model for Bipartite Networks

The development of the latent space model for bipartite networks includes a bipartite version of the latent cluster random effects model in the latentnet package (Krivitsky and Handcock 2008). In addition, the latent space model for a dynamic bipartite network was introduced by Friel et al. (2016) to study the interlocking directorates in Irish companies. In the rest of this section, we introduce a Variational Bayesian EM algorithm for fitting a latent space model to binary bipartite networks (BLSM).

Let $\boldsymbol{Y}$ denote the $N \times M$ bipartite network, whose (i, a) th element y_{ia} is 1 if person i has attribute a, for $i = \{1, 2, \ldots, N\}$ and $a = \{1, 2, \ldots, M\}$. Let $\boldsymbol{V}$ be a $M \times D$ matrix of latent attribute positions, each row of which is a D dimensional vector $\boldsymbol{v_a} = (v_{a1}, v_{a2}, \ldots, v_{aD})$ indicating the latent position of attribute a in the Euclidean space.

The latent distance model for the bipartite network $\boldsymbol{Y}$ can be written as:

$$Y_{ia}|(\boldsymbol{U}, \boldsymbol{V}, \alpha_1) \sim Bernoulli(g(\phi_{ia})), \qquad g(\phi_{ia}) = \frac{\exp(\alpha_1 - |\boldsymbol{u}_i - \boldsymbol{v}_a|^2)}{1 + \exp(\alpha_1 - |\boldsymbol{u}_i - \boldsymbol{v}_a|^2)},$$

We assume $\boldsymbol{u_i} \overset{iid}{\sim} N(0, \lambda_0^2 \mathbf{I}_D)$, $\boldsymbol{v_a} \overset{iid}{\sim} N(0, \lambda_1^2 \mathbf{I}_D)$, and α_1, λ_0 and λ_1 to be unknown parameters. The parameter α_1 accounts for the density of the bipartite network. The probability of a positive response increases as the Euclidean distance between the attribute node and the person node decreases.

2.1 Variational Bayesian EM for the Bipartite Network

We are interested in the posterior inference of the latent variables $\boldsymbol{u_i}$ and $\boldsymbol{v_a}$ following the distance model conditioning on the observed bipartite network. The (conditional) posterior distribution is the ratio of the joint distribution of the observed data and unobserved latent variables to the observed data likelihood

$$P(\boldsymbol{U}, \boldsymbol{V}|\boldsymbol{Y}) = \frac{P(\boldsymbol{Y}|\boldsymbol{U}, \boldsymbol{V})P(\boldsymbol{U}, \boldsymbol{V})}{P(\boldsymbol{Y})}.$$

We can characterize the distribution of latent positions and thus obtain the point and interval estimates by computing this posterior distribution. However, to compute this conditional posterior, we need to evaluate the normalizing constant in the denominator above, which involves integration over the latent variables. This posterior distribution is therefore intractable. To estimate the posterior distribution and obtain both the its mean and variance, we propose a variational inference approach (Blei et al. 2017). The variational inference algorithm is commonly used to estimate latent variables whose posterior distribution is intractable (Beal et al. 2003; Attias 1999; Beal et al. 2006; Blei et al. 2017). In network analysis, the variational approach has been proposed for the stochastic blockmodel (Daudin et al. 2008; Celisse et al. 2012), the mixed-membership stochastic blockmodel (Airoldi et al. 2008), the multi-layer stochastic blockmodel (Xu et al. 2014; Paul and Chen 2016), the dynamic stochastic blockmodel (Matias and Miele 2016), the latent position cluster model (Salter-Townshend and Murphy 2013) and the multiple network latent space model (Gollini and Murphy 2016). Here, we propose a Variational Bayesian Expectation Maximization (VBEM) algorithm to approximate the posterior of the person and the attribute latent positions using the bipartite network. We propose a class of suitable variational posterior distributions for the conditional distribution of $(\boldsymbol{U}, \boldsymbol{V}|\boldsymbol{Y})$ and obtain a distribution from the class that minimizes the Kulback Leibler (KL) divergence from the true but intractable posterior.

We assign the following variational posterior distributions: $q(\boldsymbol{u_i}) = N(\tilde{\mathbf{u}}_i, \tilde{\Lambda}_0)$ and $q(\boldsymbol{v_a}) = N(\tilde{\mathbf{v}}_a, \tilde{\Lambda}_1)$ and set the joint distribution as

$$q(\boldsymbol{U}, \boldsymbol{V}|\boldsymbol{Y}) = \prod_{i=1}^{N} q(\boldsymbol{u_i}) \prod_{a=1}^{M} q(\boldsymbol{v_a}), \tag{2.1}$$

where $\tilde{\mathbf{u}}_i, \tilde{\Lambda}_0, \tilde{\mathbf{v}}_a, \tilde{\Lambda}_1$ are the parameters of the variational distribution, known as variational parameters.

We can estimate the variational parameters by minimizing the Kullback-Leiber (KL) divergence between the variational posterior $q(\boldsymbol{U}, \boldsymbol{V}|\boldsymbol{Y})$ and the true posterior $f(\boldsymbol{U}, \boldsymbol{V}|\boldsymbol{Y})$. Minimizing the KL divergence is equivalent to maximizing the following Evidence Lower Bound (ELBO) function (Blei et al. 2017), (see detailed derivations in the Supplementary Materials)

$$\begin{aligned}
\text{ELBO} &= -\mathbb{E}_{q(\boldsymbol{U},\boldsymbol{V},\alpha_1|\boldsymbol{Y})}\left[\frac{\log q(\boldsymbol{U}, \boldsymbol{V}, \alpha_1|\boldsymbol{Y})}{\log p(\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{Y}|\alpha_1)}\right] \\
&= -\int q(\boldsymbol{U}, \boldsymbol{V}, \alpha_1|\boldsymbol{Y}) \log \frac{q(\boldsymbol{U}, \boldsymbol{V}, \alpha_1|\boldsymbol{Y})}{f(\boldsymbol{U}, \boldsymbol{V}, \alpha_1|\boldsymbol{Y})} d(\boldsymbol{U}, \boldsymbol{V}, \alpha_1) \\
&= -\int \prod_{i=1}^{N} q(\boldsymbol{u_i}) \prod_{a=1}^{M} q(\boldsymbol{v_a}) \log \frac{\prod_{i=1}^{N} q(\boldsymbol{u_i}) \prod_{a=1}^{M} q(\boldsymbol{v_a})}{f(\boldsymbol{Y}|\boldsymbol{U}, \boldsymbol{V}, \alpha_1) \prod_{i=1}^{N} f(\boldsymbol{u_i}) \prod_{a=1}^{M} f(\boldsymbol{v_a})} d(\boldsymbol{U}, \boldsymbol{V}, \alpha_1) \\
&= -\sum_{i=1}^{N} \int q(\boldsymbol{u_i}) \log \frac{q(\boldsymbol{u_i})}{f(\boldsymbol{u_i})} d\boldsymbol{u_i} - \sum_{a=1}^{M} \int q(\boldsymbol{v_a}) \log \frac{q(\boldsymbol{v_a})}{f(\boldsymbol{v_a})} d\boldsymbol{v_a} \\
&\quad + \int q(\boldsymbol{U}, \boldsymbol{V}, \alpha_1|\boldsymbol{Y}) \log f(\boldsymbol{Y}|\boldsymbol{U}, \boldsymbol{V}, \alpha_1) d(\boldsymbol{U}, \boldsymbol{V}, \alpha_1) \\
&= -\sum_{i=1}^{N} \text{KL}[q(\boldsymbol{u_i})|f(\boldsymbol{u_i})] - \sum_{a=1}^{M} \text{KL}[q(\boldsymbol{v_a})|f(\boldsymbol{v_a})] + \mathbb{E}_{q(\boldsymbol{U},\boldsymbol{V},\alpha_1|\boldsymbol{Y})}[\log f(\boldsymbol{Y}|\boldsymbol{U}, \boldsymbol{V}, \alpha_1)] \\
&= -\frac{1}{2}\Big(DN \log(\lambda_0^2) - N \log(\det(\tilde{\Lambda}_0))\Big) - \frac{N \text{tr}(\tilde{\Lambda}_0)}{2\lambda_0^2} - \frac{\sum_{i=1}^{N} \tilde{\boldsymbol{u}}_{\boldsymbol{i}}{}^T \tilde{\boldsymbol{u}}_{\boldsymbol{i}}}{2\lambda_0^2} \\
&\quad - \frac{1}{2}\Big(DM \log(\lambda_1^2) - M \log(\det(\tilde{\Lambda}_1))\Big) - \frac{M \text{tr}(\tilde{\Lambda}_1)}{2\lambda_1^2} - \frac{\sum_{a=1}^{M} \tilde{\boldsymbol{v}}_{\boldsymbol{a}}{}^T \tilde{\boldsymbol{v}}_{\boldsymbol{a}}}{2\lambda_1^2} + \frac{1}{2}(MD + ND) \\
&\quad + \mathbb{E}_{q(\boldsymbol{U},\boldsymbol{V}|\boldsymbol{Y})}[\log f(\boldsymbol{Y}|\boldsymbol{U}, \boldsymbol{V})].
\end{aligned} \tag{2.2}$$

After applying Jensen's inequality (Jensen 1906), a lower-bound on $\mathbb{E}_{q(\boldsymbol{U},\boldsymbol{V}|\boldsymbol{Y})}[\log f(\boldsymbol{Y}|\boldsymbol{U}, \boldsymbol{V})]$

is given by,

$$
\begin{aligned}
&\mathbb{E}_{q(\boldsymbol{U},\boldsymbol{V}|\boldsymbol{Y})}[\log f(\boldsymbol{Y}|\boldsymbol{U},\boldsymbol{V},\alpha_1)] \\
&\geq \sum_{i=1}^{N}\sum_{a=1}^{M} y_{ia}\left[\tilde{\alpha}_1 - \mathrm{tr}(\tilde{\Lambda}_0) - \mathrm{tr}(\tilde{\Lambda}_1) - (\tilde{\boldsymbol{u}}_{\boldsymbol{i}} - \tilde{\boldsymbol{v}}_{\boldsymbol{a}})^T(\tilde{\boldsymbol{u}}_{\boldsymbol{i}} - \tilde{\boldsymbol{v}}_{\boldsymbol{a}})\right] \\
&- \sum_{i=1}^{N}\sum_{a=1}^{M} \log\left(1 + \frac{\exp(\tilde{\alpha}_1)}{\det(\mathbf{I} + 2\tilde{\Lambda}_0 + 2\tilde{\Lambda}_1)^{\frac{1}{2}}}\exp\Big(-(\tilde{\boldsymbol{u}}_{\boldsymbol{i}} - \tilde{\boldsymbol{v}}_{\boldsymbol{a}})^T(\mathbf{I} + 2\tilde{\Lambda}_0 + 2\tilde{\Lambda}_1)^{-1}(\tilde{\boldsymbol{u}}_{\boldsymbol{i}} - \tilde{\boldsymbol{v}}_{\boldsymbol{a}})\Big)\right).
\end{aligned}
$$

We use the Variational Expectation-Maximization algorithm (Jordan et al. 1999; Baum et al. 1970; Dempster et al. 1977) to maximize the ELBO function. Following the variational EM algorithm, we replace the E step of the celebrated EM algorithm, where we compute the expectation of the complete likelihood with respect to the conditional distribution $f(\boldsymbol{U},\boldsymbol{V}|\boldsymbol{Y})$, with a VE step, where we compute the expectation with respect to the best variational distribution (obtained by optimizing the ELBO function) at that iteration.

The detailed procedures are as follows. We start with the initial parameter, $\Theta^{(0)} = \tilde{\alpha}_1^{(0)}$, and $\tilde{\mathbf{u}}_i^{(0)}, \tilde{\Lambda}_0^{(0)}, \tilde{\mathbf{v}}_a^{(0)}, \tilde{\Lambda}_1^{(0)}$, and then we iterate the following VE (Variational expectation) and M (maximization) steps. During the VE step, we maximize the $\mathrm{ELBO}(q(\mathbf{U}), \mathbf{q}(\mathbf{V}), \boldsymbol{\Theta})$ with respect to the variational parameters $\tilde{\boldsymbol{u}}_{\boldsymbol{i}}, \tilde{\boldsymbol{v}}_{\boldsymbol{a}}, \tilde{\lambda}_0$ and $\tilde{\lambda}_1$ given the other model parameters and obtain $\mathrm{ELBO}(q^*(\mathbf{U}), \mathbf{q}^*(\mathbf{V}), \boldsymbol{\Theta})$. During the M step, we fix $\tilde{\boldsymbol{u}}_{\boldsymbol{i}}, \tilde{\boldsymbol{v}}_{\boldsymbol{a}}, \tilde{\Lambda}_0$ and $\tilde{\Lambda}_1$ and maximize the $\mathrm{ELBO}(q(\mathbf{U}), \mathbf{q}(\mathbf{V}), \boldsymbol{\Theta})$ with respect to $\tilde{\alpha}_1$. To do this, we differentiate $\mathrm{ELBO}(q(\mathbf{U}), \mathbf{q}(\mathbf{V}), \boldsymbol{\Theta})$ with respect to each variational parameter. We obtain closed form update rules by setting the partial derivatives to zero while introducing the first- and second-order Taylor series expansion approximation of the log functions in $\mathrm{ELBO}(q(\mathbf{U}), \mathbf{q}(\mathbf{V})\ , \Theta)$ (see detailed derivations in supplementary material). The Taylor series expansions are commonly used in the variational approaches. For example, three first-order Taylor expansions were used by Salter-Townshend and Murphy (2013) to simplify the Euclidean distance in the latent position cluster model, and first- and second-order Taylor expansions were used by Gollini and Murphy (2016) to simplify the squared Euclidean distance in LSM. Following the previous publications using Taylor expansions, we approximate the three log functions in our $\mathrm{ELBO}(q(\mathbf{U}), \mathbf{q}(\mathbf{V}), \boldsymbol{\Theta})$ function to find closed form update rules for the variational parameters. Define the function

$$\boldsymbol{F_A} = \sum_{i=1}^{N}\sum_{a=1}^{M} \log\left(1 + \frac{\exp(\tilde{\alpha}_1)}{\det(\mathbf{I} + 2\tilde{\Lambda}_0 + 2\tilde{\Lambda}_1)^{\frac{1}{2}}} \exp\Big(-(\tilde{\boldsymbol{u}}_{\boldsymbol{i}} - \tilde{\boldsymbol{v}}_{\boldsymbol{a}})^T(\mathbf{I} + 2\tilde{\Lambda}_0 + 2\tilde{\Lambda}_1)^{-1}(\tilde{\boldsymbol{u}}_{\boldsymbol{i}} - \tilde{\boldsymbol{v}}_{\boldsymbol{a}})\Big)\right).$$

The closed form update rules of the $(t+1)$th iteration are as follows

VE-step: Estimate $\tilde{\boldsymbol{u}}_{\boldsymbol{i}}^{(t+1)}$, $\tilde{\boldsymbol{v}}_{\boldsymbol{a}}^{(t+1)}$, $\tilde{\Lambda}_0^{(t+1)}$ and $\tilde{\Lambda}_1^{(t+1)}$ by minimizing $\mathrm{ELBO}(q(\mathbf{U}), \mathbf{q}(\mathbf{V}), \boldsymbol{\Theta})$

$$\begin{aligned}
\tilde{\boldsymbol{u}}_{\boldsymbol{i}}^{(\boldsymbol{t+1})} &= \left[\left(\frac{1}{2\lambda_0^2} + \sum_{a=1}^{M} y_{ia})\boldsymbol{I} + \frac{1}{2}\boldsymbol{H_A}(\tilde{\boldsymbol{u}}_{\boldsymbol{i}}^{(\boldsymbol{t})})\right)\right]^{-1}\left[\sum_{a=1}^{M} y_{ia}\tilde{\boldsymbol{v}}_{\boldsymbol{a}}^{(\boldsymbol{t})} + \frac{1}{2}\boldsymbol{H_A}(\tilde{\boldsymbol{u}}_{\boldsymbol{i}}^{(\boldsymbol{t})})\tilde{\boldsymbol{u}}_{\boldsymbol{i}}^{(\boldsymbol{t})} - \frac{1}{2}\boldsymbol{G_A}(\tilde{\boldsymbol{u}}_{\boldsymbol{i}}^{(\boldsymbol{t})})\right] \\
\tilde{\boldsymbol{v}}_{\boldsymbol{a}}^{(\boldsymbol{t+1})} &= \left[\left(\frac{1}{2\lambda_1^2} + \sum_{i=1}^{N} y_{ia}\right)\boldsymbol{I} - \frac{1}{2}\boldsymbol{H_A}(\tilde{\boldsymbol{v}}_{\boldsymbol{a}}^{(\boldsymbol{t})})\right]^{-1}\left[\sum_{i=1}^{N} y_{ia}\tilde{\boldsymbol{u}}_{\boldsymbol{i}}^{(\boldsymbol{t})} - \frac{1}{2}\boldsymbol{G_A}(\tilde{\boldsymbol{v}}_{\boldsymbol{a}}^{(\boldsymbol{t})})\right] \\
\tilde{\Lambda}_0^{(t+1)} &= \frac{N}{2}\left[\left(\frac{N}{2}\frac{1}{\lambda_0^2} + \sum_{i=1}^{N}\sum_{a=1}^{M} y_{ia}\right)\boldsymbol{I} + \boldsymbol{G_A}(\tilde{\Lambda}_0^{(t)})\right]^{-1} \\
\tilde{\Lambda}_1^{(t+1)} &= \frac{M}{2}\left[\left(\frac{M}{2}\frac{1}{\lambda_1^2} + \sum_{i=1}^{N}\sum_{a=1}^{M} y_{ia}\right)\boldsymbol{I} + \boldsymbol{G_A}(\tilde{\Lambda}_1^{(t)})\right]^{-1},
\end{aligned} \tag{2.3}$$

where $\boldsymbol{G_A}(\tilde{\boldsymbol{u}}_{\boldsymbol{i}}{}^{(t)})$and $\boldsymbol{G_A}(\tilde{\boldsymbol{v}}_{\boldsymbol{a}}{}^{(t)})$ are the partial derivatives (gradients) of $\boldsymbol{F_A}$ with respect to $\tilde{\boldsymbol{u}}_{\boldsymbol{i}}$ and $\tilde{\boldsymbol{v}}_{\boldsymbol{a}}$, evaluated at $\tilde{\boldsymbol{u}}_{\boldsymbol{i}}{}^{(t)}$ and $\tilde{\boldsymbol{v}}_{\boldsymbol{a}}{}^{(t)}$, respectively. In $\boldsymbol{G_A}(\tilde{\boldsymbol{u}}_{\boldsymbol{i}}{}^{(t)})$, the subscript $\boldsymbol{A}$ indicates that the gradient is of function $\boldsymbol{F_A}$, and the subscript i in $\tilde{\boldsymbol{u}}_{\boldsymbol{i}}{}^{(t)}$ indicates that the gradient is with respect to $\tilde{\boldsymbol{u}}_{\boldsymbol{i}}$, evaluated at $\tilde{\boldsymbol{u}}_{\boldsymbol{i}}{}^{(t)}$. Similarly, $\boldsymbol{H_A}(\tilde{\boldsymbol{u}}_{\boldsymbol{i}}{}^{(t)})$ and $\boldsymbol{H_A}(\tilde{\boldsymbol{v}}_{\boldsymbol{a}}{}^{(t)})$ are the second-order partial derivatives of $\boldsymbol{F_A}$ with respect to $\tilde{\boldsymbol{u}}_{\boldsymbol{i}}$ and $\tilde{\boldsymbol{v}}_{\boldsymbol{a}}$, evaluated at $\tilde{\boldsymbol{u}}_{\boldsymbol{i}}{}^{(t)}$ and $\tilde{\boldsymbol{v}}_{\boldsymbol{a}}{}^{(t)}$, respectively.

M-step: Estimate $\tilde{\alpha}_1^{(t+1)}$ with the following closed form solution,

$$\tilde{\alpha}_1^{(t+1)} = \frac{\sum_{i=1}^{N}\sum_{a=1}^{M} y_{ia} - g_A(\tilde{\alpha}_1^{(t)}) + \tilde{\alpha}_1^{(t)} h_A(\tilde{\alpha}_1^{(t)})}{h_A(\tilde{\alpha}_1^{(t)})}, \tag{2.4}$$

where $g_A(\tilde{\alpha}_1^{(t)})$ is the partial derivative (gradient) of $\boldsymbol{F_A}$ with respect to $\tilde{\alpha}_1$, evaluated at $\tilde{\alpha}_1^{(t)}$; and $h_A(\tilde{\alpha}_1^{(t)})$ is the second-order partial derivative of $\boldsymbol{F_A}$ with respect to $\tilde{\alpha}_1$, evaluated at $\tilde{\alpha}_1^{(t)}$.

3 Excerpt from 'Instagram Popularity and Topical Interests Study'

Ferrara et al. (2014) collected data for a study of online popularity and topical interests through Instagram during Jan-Feb 2014. The dataset was collected from photographic contests run through Instagram's official blog. Each contest was expressed by a unique (hash)tag prefixed with #whp. This dataset includes a directed social network between the followers' and the followees' Instagram accounts, and the users' participation in the #whp contests. We use the tags adopted by each user to label visual images as attributes and users' follower-followee relationships as friendship links. We focus on an excerpt of 1862 users that participate in at least one of the #whp contests and are also present in the social network. A total of 486 unique #whp tags were identified in the data[1]. A social network and a bipartite network are formed of $128,413$ and $17,024$ edges and have densities of 0.037 and 0.019. Their degree distributions are shown in Figure 3. We note that the degree distribution for tags is extreme skewed indicating that only a few tags are extremely popular.

We fit the JLSM model to Instagram users' social network and their participation in the #whp contest. The estimated $\alpha_0 = -1.7570$ and $\alpha_1 = -1.5459$. The receiver operating characteristic curves (ROCs) and the AUC values for in-sample predictions of social network

[1]This dataset was downloaded from https://people.dimes.unical.it/andreatagarelli/data/

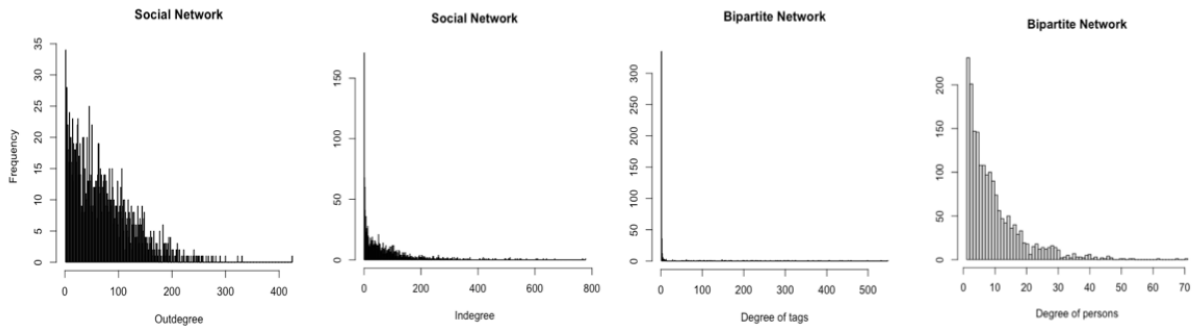

Figure 1: The degree distributions for the Instagram social network and users' participation in #whp contests.

and photo tags are shown in Figure 2, respectively. The average AUCs in the cross-validated samples are 0.67 and 0.79 for social network and photo tags. For the task of predicting missing tag information, we obtain an average AUC value of 0.80 in 10 fold cross-validation.

In comparison, we also fit BLSM to the Instagram users' participation in the #whp contest. The estimated $\alpha_1 = -2.3192$. The estimated latent positions for users $\tilde{\boldsymbol{u}}_{\boldsymbol{i}}$ and tags $\tilde{\boldsymbol{v}}_{\boldsymbol{a}}$ are shown in Figure 2. The in-sample and cross-validated AUCs are 0.87 and 0.78, indicating good fit of the proposed BLSM.

In the bipartite latent space, the popular Instagram tags are found at the center and differentiated from the less popular tags in the peripheral of the space. Along the y axis, the popular tags and the individual nodes follow a vertical line, which shows that they are differentiated along this dimension. Along the x axis, we observe the less popular tags being differentiated at the left and the right latent space. In the joint latent space, the popular tags remain at the center of the space and differentiated from the less popular tags. There seem to be more variations of the person nodes with the added follower-followee network with further distances between nodes observed. It appears that in the right most plot, the latent space captures little structure as the model has already accounted for the attributes. Figure 8 appears to show a core-periphery structure but this is probably down to the lack of accounting for heavily skewed degrees / sociality in the model.

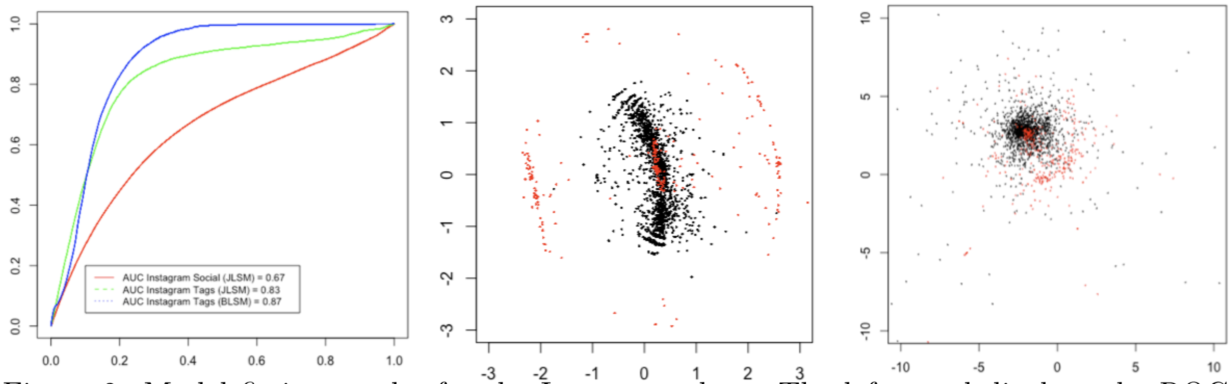

Figure 2: Model fitting results for the Instagram data. The left panel displays the ROCs. The middle and right panel display the estimated latent person and attribute positions, $\tilde{\boldsymbol{u}}_{\boldsymbol{i}}$ (black) and $\tilde{\boldsymbol{v}}_{\boldsymbol{a}}$ (red) based on the BLSM and the JLSM, respectively.

4 French financial elite dataset

We analyse a dataset containing a symmetric social network between 28 French financial elites and the multidimensional attribute covariates. The data were collected through interviews for people who held leading positions in major financial institutions and frequently appeared in financial section press reports. The friendship information was collected by asking the interviewees to name their friends in the social context. Kadushin and de Quillacqi then identified an inner circle of 28 elites from the initial sample based on their influence and their perceived eliteness by other participants. The resulting friendship network is a symmetric adjacency matrix.

The data also contains additional background information, including age, a complex set of post-secondary education experiences, place of birth, political and cabinet careers, political party preference, religion, current residence, and club memberships. Two aspects of the elites' "prestige" include whether the person is named in the social register and whether the person has a *particle* ("de") in front of either his (no woman was in the inner circle), his wife's, his mother's, or his children's names. Having "de" in the name is associated with nobility. Father's occupation is one of the variables used to reflect an elite's social class. Fathers' occupation is considered “high” if the father is in higher management, a professional, an industrialist, or an investor. Unfortunately, upon communications with the original author, we found that the coding procedures regarding some variables have been lost, including Finance Ministry information, religion, etc. Kadushin (1995) found that elites' political affiliations, education, career trajectory and class are interrelated with their social circles. In this paper, we use 13 binary variables including information on education ("Science Po", "Polytechniqu", "University" and "Ecole Nationale d'Administration"), career ("Inspection General de Finance" and "Cabinet"), class ("Social Register", "Father Status", "Particule"), politics ("Socialist", "Capitalist" and "Centrist") and "Age" after excluding the lost or the unrelated information, i.e., mason and location, which are not associated with the social network based on Kadushin (1995) (location is not considered to be related to the social network after adjusting for multiple comparisons). "Age" was converted into a binary variable following Kadushin (1995), where a group of elites was considered of older age with an average birth year of 1938. We will use ENA as an abbreviation for Ecole Nationale d'Administration.

The Science Po or the Institut d'Etudes Politiques de Paris prepares students for the entrance exam of the ENA. An alternative of the Science Po is the (Ecole) Polytechnique, a French military school whose graduates often enter one of the technical ministries. These elites with Polytechnique degrees enter one of the technical ministries. Both the Science Po and the Polytechnique are called Grandes Ecoles. A Grandes Ecoles education is highly respected in France as it leads to membership in the ENA, where the grands corps, which are the French civil service elites, including the Inspection General de Finance, etc-recruit its members (Kadushin 1995).

The authors in Kadushin (1995) first used multidimensional scaling to draw the friendship network's sociogram. Then they applied Quadratic Assignment Procedure regressions and correlations to test each background variable's association with the social network. Based on the social network, two clusters were identified, which the authors called the left and the right moieties. The dependence between the social network and background information was understood through comparisons of the elites between the left and the right moiety. The elites in the right moiety were found to have a higher social class (upper-class parentage with high social standing), to be older (average birth year of 1929), and to have fewer appointments in public offices. The left moiety elites were more likely to be ENA graduates, grand corps members, cabinet members, treasury service members, socialists, and younger (average birth year of 1938).

Using the JLSM, we will construct a joint latent space which will allow us to jointly model elites' friendship connections and their background information. Using the JLSM, we will also replicate Kadushin (1995)'s left and the right moiety, adding simultaneous interpretation for the division in the elite circle. Furthermore, we observe an additional division within the left moiety using JLSM, which provides opportunities for new hypotheses.

4.1 Visualization of the French Elite Network

4.2 5-fold Cross-Validation

5 The Estimation Procedure for JLSM

5.1 Derivation of KL Divergence

We set the variational parameter as $\Theta = \tilde{\alpha}_0, \tilde{\alpha}_1$ and $\tilde{\boldsymbol{u}}_{\boldsymbol{i}}, \tilde{\Lambda}_0, \tilde{\boldsymbol{v}}_{\boldsymbol{a}}, \tilde{\Lambda}_1$, where $q(\boldsymbol{u}_{\boldsymbol{i}}) = N(\tilde{\boldsymbol{u}}_{\boldsymbol{i}}, \tilde{\Lambda}_0)$, and $q(\boldsymbol{v}_{\boldsymbol{a}}) = N(\tilde{\boldsymbol{v}}_{\boldsymbol{a}}, \tilde{\Lambda}_1)$. We set the variational posterior as:

$$q(\boldsymbol{U}, \boldsymbol{V}|\boldsymbol{X}, \boldsymbol{Y}) = \prod_{i=1}^{N} q(\boldsymbol{u}_{\boldsymbol{i}}) \prod_{a=1}^{M} q(\boldsymbol{v}_{\boldsymbol{a}})$$

The Kullback-Leiber divergence between the variational posterior and the true posterior is:

$$\begin{aligned}
&\mathrm{KL}[q(\boldsymbol{U}, \boldsymbol{V}, \alpha_0, \alpha_1|\boldsymbol{X}, \boldsymbol{Y})|f(\boldsymbol{U}, \boldsymbol{V}, \alpha_0, \alpha_1|\boldsymbol{X}, \boldsymbol{Y})] \\
&= \int q(\boldsymbol{U}, \boldsymbol{V}, \alpha_0, \alpha_1|\boldsymbol{X}, \boldsymbol{Y}) \log \frac{q(\boldsymbol{U}, \boldsymbol{V}, \alpha_0, \alpha_1|\boldsymbol{X}, \boldsymbol{Y})}{f(\boldsymbol{U}, \boldsymbol{V}, \alpha_0, \alpha_1|\boldsymbol{X}, \boldsymbol{Y})} d(\boldsymbol{U}, \boldsymbol{V}, \alpha_0, \alpha_1)
\end{aligned}$$

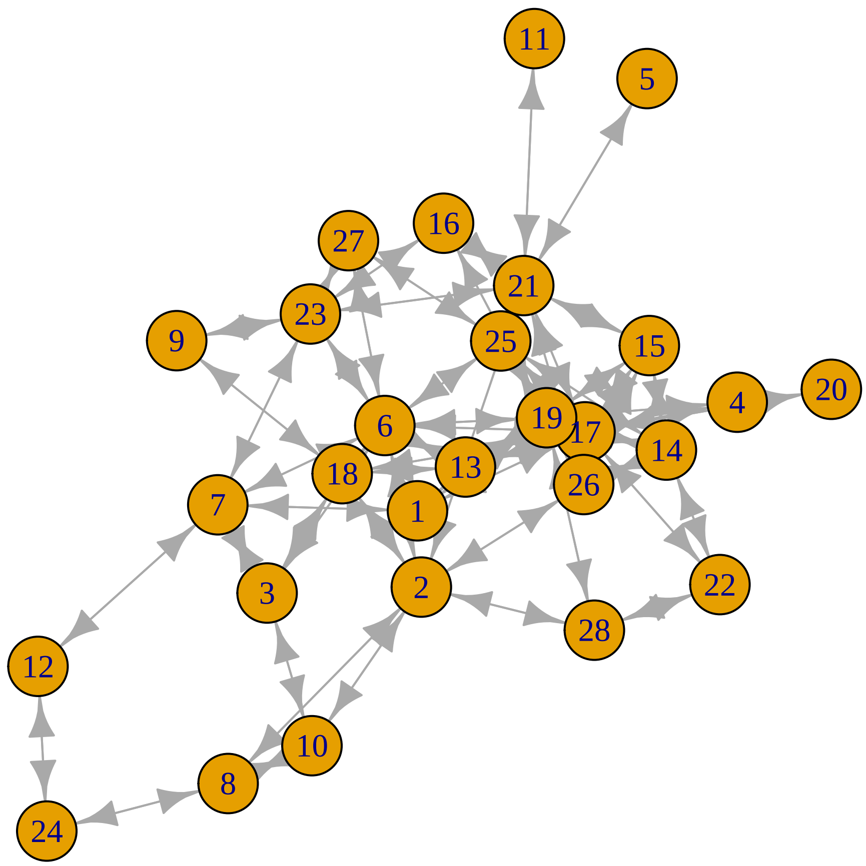

Figure 3: The visualization of the French Elite Network. Each circle represents a french elite, and the edge between two circles represents the presence of a mutual friendship between two elites.

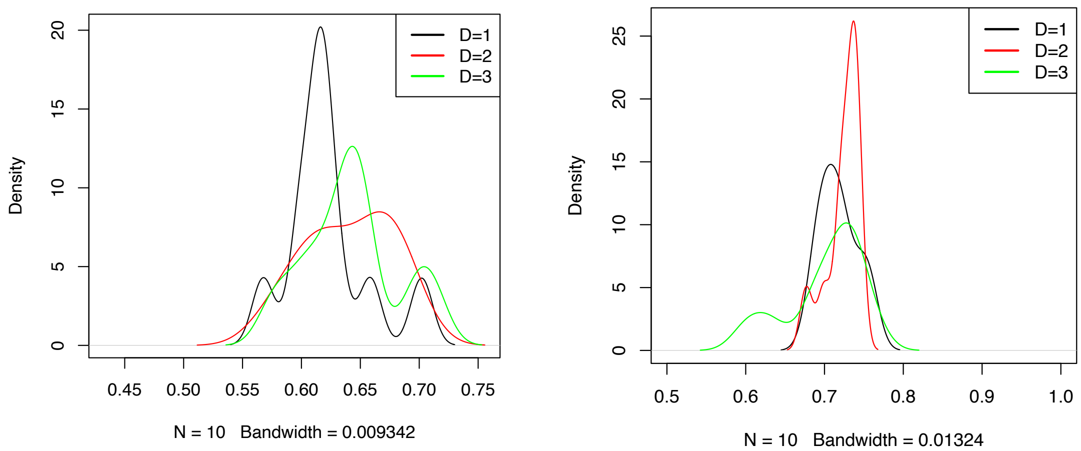

Figure 4: Out-of-sample AUC values for 5-fold cross-validation the French Elite dataset under different numbers of dimensions, $D \in 1, 2, 3$.

$$
\begin{aligned}
&= \int \prod_{i=1}^{N} q(\boldsymbol{u_i}) \prod_{a=1}^{M} q(\boldsymbol{v_a}) \log \frac{\prod_{i=1}^{N} q(\boldsymbol{u_i}) \prod_{a=1}^{M} q(\boldsymbol{v_a})}{f(\boldsymbol{X}, \boldsymbol{Y} | \boldsymbol{U}, \boldsymbol{V}, \alpha_0, \alpha_1) \prod_{i=1}^{N} f(\boldsymbol{u_i}) \prod_{a=1}^{M} f(\boldsymbol{v_a})} d(\boldsymbol{U}, \boldsymbol{V}, \alpha_0, \alpha_1) \\
&= \sum_{i=1}^{N} \int q(\boldsymbol{u_i}) \log \frac{q(\boldsymbol{u_i})}{f(\boldsymbol{u_i})} d\boldsymbol{u_i} + \sum_{a=1}^{M} \int q(\boldsymbol{v_a}) \log \frac{q(\boldsymbol{v_a})}{f(\boldsymbol{v_a})} d\boldsymbol{v_a} \\
&- \int q(\boldsymbol{U}, \boldsymbol{V}, \alpha_0, \alpha_1 | \boldsymbol{X}, \boldsymbol{Y}) \log f(\boldsymbol{X}, \boldsymbol{Y} | \boldsymbol{U}, \boldsymbol{V}, \alpha_0, \alpha_1) d(\boldsymbol{U}, \boldsymbol{V}, \alpha_0, \alpha_1) \\
&= \sum_{i=1}^{N} \mathrm{KL}[q(\boldsymbol{u_i}) | f(\boldsymbol{u_i})] + \sum_{a=1}^{M} \mathrm{KL}[q(\boldsymbol{v_a}) | f(\boldsymbol{v_a})] \\
&- \mathrm{E}_{q(\boldsymbol{U}, \boldsymbol{V}, \alpha_0, \alpha_1 | \boldsymbol{X}, \boldsymbol{Y})} [\log f(\boldsymbol{X}, \boldsymbol{Y} | \boldsymbol{U}, \boldsymbol{V}, \alpha_0, \alpha_1)],
\end{aligned}
$$

where each of the components are calculated as follows:

$$
\begin{aligned}
&\sum_{i=1}^{N} \mathrm{KL}[q(\boldsymbol{u_i}) || f(\boldsymbol{u_i})] \\
&= - \sum_{i=1}^{N} \int q(\boldsymbol{u_i}) \log \frac{f(\boldsymbol{u_i})}{q(\boldsymbol{u_i})} d\boldsymbol{u_i} \\
&= - \sum_{i=1}^{N} \int q(\boldsymbol{u_i}) \Bigg(\frac{1}{2} \Bigg(- D \log(\lambda_0^2) + \log(\det(\tilde{\Lambda}_0)) - \frac{1}{\lambda_0^2} \boldsymbol{u_i}^T \boldsymbol{u_i} + (\boldsymbol{u_i} - \tilde{\boldsymbol{u}}_{\boldsymbol{i}})^T \tilde{\Lambda}_0^{-1} (\boldsymbol{u_i} - \tilde{\boldsymbol{u}}_{\boldsymbol{i}}) \Bigg) \Bigg)
\end{aligned}
$$

$$= \frac{1}{2}\Big(DN\log(\lambda_0^2) - N\log(\det(\tilde{\Lambda}_0))\Big) + \sum_{i=1}^{N}\frac{1}{2}\left(\frac{1}{\lambda_0^2}\mathrm{E}_{q(\boldsymbol{u_i})}[\boldsymbol{u_i}^T\boldsymbol{u_i}] - \mathrm{E}_{q(\boldsymbol{u_i})}[(\boldsymbol{u_i}-\tilde{\boldsymbol{u}}_{\boldsymbol{i}})^T\tilde{\Lambda}_0^{-1}(\boldsymbol{u_i}-\tilde{\boldsymbol{u}}_{\boldsymbol{i}})]\right)$$

$$= \frac{1}{2}\Big(DN\log(\lambda_0^2) - N\log(\det(\tilde{\Lambda}_0))\Big) + \sum_{i=1}^{N}\frac{1}{2\lambda_0^2}\Big(\mathrm{Var}(\boldsymbol{u_i}) + \big(\mathrm{E}_{q(\boldsymbol{u_i})}[\boldsymbol{u_i}]\big)^2\Big) - \frac{1}{2}ND$$

$$= \frac{1}{2}\Big(DN\log(\lambda_0^2) - N\log(\det(\tilde{\Lambda}_0))\Big) + \frac{N\mathrm{tr}(\tilde{\Lambda}_0)}{2\lambda_0^2} + \frac{\sum_{i=1}^{N}\tilde{\boldsymbol{u}}_{\boldsymbol{i}}^T\tilde{\boldsymbol{u}}_{\boldsymbol{i}}}{2\lambda_0^2} - \frac{1}{2}ND$$

$$\sum_{a=1}^{M}\mathrm{KL}[q(\boldsymbol{v_a})||f(\boldsymbol{v_a})]$$

$$= \frac{1}{2}\Big(DM\log(\lambda_1^2) - M\log(\det(\tilde{\Lambda}_1))\Big) + \frac{M\mathrm{tr}(\tilde{\Lambda}_1)}{2\lambda_1^2} + \frac{\sum_{a=1}^{M}\tilde{\boldsymbol{v}}_{\boldsymbol{a}}^T\tilde{\boldsymbol{v}}_{\boldsymbol{a}}}{2\lambda_1^2} - \frac{1}{2}MD$$

$\mathrm{E}_{q(\boldsymbol{U},\boldsymbol{V}|\boldsymbol{X},\boldsymbol{Y})}[\log f(\boldsymbol{X},\boldsymbol{Y}|\boldsymbol{U},\boldsymbol{V})]$ can be expanded into 6 components:

$$\begin{aligned}
&\mathrm{E}_{q(\boldsymbol{U},\boldsymbol{V}|\boldsymbol{X},\boldsymbol{Y})}[\log f(\boldsymbol{X},\boldsymbol{Y}|\boldsymbol{U},\boldsymbol{V})]\\
&= \sum_{i=1}^{N}\sum_{a=1}^{M} y_{ia}\mathrm{E}_{q(\boldsymbol{U},\boldsymbol{V}|\boldsymbol{X},\boldsymbol{Y})}[\alpha_1 - (\boldsymbol{u_i}-\boldsymbol{v_a})^T(\boldsymbol{u_i}-\boldsymbol{v_a})]\\
&+ \sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N} x_{ij}\mathrm{E}_{q(\boldsymbol{U},\boldsymbol{V}|\boldsymbol{X},\boldsymbol{Y})}[\alpha_0 - (\boldsymbol{u_i}-\boldsymbol{u_j})^T(\boldsymbol{u_i}-\boldsymbol{u_j})]\\
&- \sum_{i=1}^{N}\sum_{a=1}^{M}\mathrm{E}_{q(\boldsymbol{U},\boldsymbol{V}|\boldsymbol{X},\boldsymbol{Y})}[\log(1+\exp(\alpha_1 - (\boldsymbol{u_i}-\boldsymbol{v_a})^T(\boldsymbol{u_i}-\boldsymbol{v_a})))]\\
&- \sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\mathrm{E}_{q(\boldsymbol{U},\boldsymbol{V}|\boldsymbol{X},\boldsymbol{Y})}[\log(1+\exp(\alpha_0 - (\boldsymbol{u_i}-\boldsymbol{u_j})^T(\boldsymbol{u_i}-\boldsymbol{u_j})))]
\end{aligned}$$

First 2 components of $\mathrm{E}_{q(\boldsymbol{U},\boldsymbol{V}|\boldsymbol{X},\boldsymbol{Y})}[\log f(\boldsymbol{X},\boldsymbol{Y}|\boldsymbol{U},\boldsymbol{V})]$ are calculated as follows:

$$\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N} x_{ij}\mathrm{E}_{q(\boldsymbol{U},\boldsymbol{V}|\boldsymbol{X},\boldsymbol{Y})}[\alpha_0 - (\boldsymbol{u_i}-\boldsymbol{u_j})(\boldsymbol{u_i}-\boldsymbol{u_j})^T]$$

$$= \sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N} x_{ij}\int\big(\alpha_0 - (\boldsymbol{u_i}-\boldsymbol{u_j})(\boldsymbol{u_i}-\boldsymbol{u_j})^T\big)q(\boldsymbol{u_i})q(\boldsymbol{u_j})d(\boldsymbol{u_i},\boldsymbol{u_j})$$

$$
\begin{aligned}
&= \sum_{i=1}^{N} \sum_{j=1, j\neq i}^{N} x_{ij} \left[\tilde{\alpha}_0 - \int (\boldsymbol{u_i} - \boldsymbol{u_j})(\boldsymbol{u_i} - \boldsymbol{u_j})^T q(\boldsymbol{u_i}) q(\boldsymbol{u_j}) d(\boldsymbol{u_i}, \boldsymbol{u_j}) \right] \\
&= \sum_{i=1}^{N} \sum_{j=1, j\neq i}^{N} x_{ij} \left[\tilde{\alpha}_0 - \int \sum_{d=1}^{D} (u_{id} - u_{jd})^2 q(\boldsymbol{u_i}) q(\boldsymbol{u_j}) d(\boldsymbol{u_i}, \boldsymbol{u_j}) \right] \\
&= \sum_{i=1}^{N} \sum_{j=1, j\neq i}^{N} x_{ij} \left[\tilde{\alpha}_0 - \left[\sum_{d=1}^{D} \left[\int u_{id}^2 q(u_{id}) du_{id} + \int u_{jd}^2 q(u_{jd}) du_{jd} - \int \int 2u_{id} u_{jd} q(u_{id}) q(u_{jd}) du_{id}, du_{jd} \right] \right] \right] \\
&= \sum_{i=1}^{N} \sum_{j=1, j\neq i}^{N} x_{ij} \left[\tilde{\alpha}_0 - \left[\sum_{d=1}^{D} \left[E[u_{id}^2] + E[u_{jd}^2] - 2E[u_{id}] E[u_{jd}] \right] \right] \right] \\
&= \sum_{i=1}^{N} \sum_{j=1, j\neq i}^{N} x_{ij} \left[\tilde{\alpha}_0 - \left[\sum_{d=1}^{D} \left[Var[u_{id}] + E[u_{id}]^2 + Var[u_{jd}] + E[u_{jd}]^2 - 2E[u_{id}] E[u_{jd}] \right] \right] \right] \\
&= \sum_{i=1}^{N} \sum_{j=1, j\neq i}^{N} x_{ij} \left[\tilde{\alpha}_0 - 2\mathrm{tr}(\tilde{\Lambda}_0) - (\tilde{\boldsymbol{u}}_{\boldsymbol{i}} - \tilde{\boldsymbol{u}}_{\boldsymbol{j}})^T (\tilde{\boldsymbol{u}}_{\boldsymbol{i}} - \tilde{\boldsymbol{u}}_{\boldsymbol{j}}) \right]
\end{aligned}
$$

$$
\begin{aligned}
&\sum_{i=1}^{N}\sum_{a=1}^{M} y_{ia}\mathrm{E}_{q(\boldsymbol{U},\boldsymbol{V}|\boldsymbol{X},\boldsymbol{Y})}[\alpha_1 - (\boldsymbol{u_i}-\boldsymbol{v_a})(\boldsymbol{u_i}-\boldsymbol{v_a})^T] \\
&= \sum_{i=1}^{N}\sum_{a=1}^{M} y_{ia} \int \big(\alpha_1 - (\boldsymbol{u_i}-\boldsymbol{v_a})(\boldsymbol{u_i}-\boldsymbol{v_a})^T\big) q(\boldsymbol{u_i})q(\boldsymbol{v_a})d(\boldsymbol{u_i},\boldsymbol{v_a}) \\
&= \sum_{i=1}^{N}\sum_{a=1}^{M} y_{ia}\left[\tilde{\alpha}_1 - \int (\boldsymbol{u_i}-\boldsymbol{v_a})(\boldsymbol{u_i}-\boldsymbol{v_a})^T q(\boldsymbol{u_i})q(\boldsymbol{v_a})d(\boldsymbol{u_i},\boldsymbol{v_a})\right] \\
&= \sum_{i=1}^{N}\sum_{a=1}^{M} y_{ia}\left[\tilde{\alpha}_1 - \int \sum_{d=1}^{D}(u_{id}-v_{ad})^2 q(\boldsymbol{u_i})q(\boldsymbol{v_a})d(\boldsymbol{u_i},\boldsymbol{v_a})\right] \\
&= \sum_{i=1}^{N}\sum_{a=1}^{M} y_{ia}\left[\tilde{\alpha}_1 - \left[\sum_{d=1}^{D}\Big[\int u_{id}^2 q(u_{id})du_{id} + \int v_{ad}^2 q(v_{ad})dv_{ad} - \int\int 2u_{id}v_{ad}q(u_{id})q(v_{ad})du_{id}, dv_{ad}\Big]\right]\right] \\
&= \sum_{i=1}^{N}\sum_{a=1}^{M} y_{ia}\left[\tilde{\alpha}_1 - \left[\sum_{d=1}^{D}\big[E[u_{id}^2] + E[v_{ad}^2] - 2E[u_{id}]E[v_{ad}]\big]\right]\right] \\
&= \sum_{i=1}^{N}\sum_{a=1}^{M} y_{ia}\left[\tilde{\alpha}_1 - \left[\sum_{d=1}^{D}\big[Var[u_{id}] + E[u_{id}]^2 + Var[v_{ad}] + E[v_{ad}]^2 - 2E[u_{id}]E[v_{ad}]\big]\right]\right] \\
&= \sum_{i=1}^{N}\sum_{a=1}^{M} y_{ia}\left[\tilde{\alpha}_1 - \mathrm{tr}(\tilde{\Lambda}_0) - \mathrm{tr}(\tilde{\Lambda}_1) - (\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{v}}_{\boldsymbol{a}})^T(\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{v}}_{\boldsymbol{a}})\right]
\end{aligned}
$$

The last 2 expectations of the log functions can be simplified using Jensen's inequality and $\mathrm{E}_{q(\boldsymbol{U},\boldsymbol{V}|\boldsymbol{X},\boldsymbol{Y})}[\log f(\boldsymbol{X},\boldsymbol{Y}|\boldsymbol{U},\boldsymbol{V})]$ is now:

$$
\begin{aligned}
&\mathrm{E}_{q(\boldsymbol{U},\boldsymbol{V}|\boldsymbol{X},\boldsymbol{Y})}[\log f(\boldsymbol{X},\boldsymbol{Y}|\boldsymbol{U},\boldsymbol{V})] \\
&\leq \sum_{i=1}^{N}\sum_{a=1}^{M} y_{ia}\left[\tilde{\alpha}_1 - \mathrm{tr}(\tilde{\Lambda}_0) - \mathrm{tr}(\tilde{\Lambda}_1) - (\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{v}}_{\boldsymbol{a}})^T(\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{v}}_{\boldsymbol{a}})\right] \\
&+ \sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N} x_{ij}\left[\tilde{\alpha}_0 - 2\mathrm{tr}(\tilde{\Lambda}_0) - (\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{u}}_{\boldsymbol{j}})^T(\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{u}}_{\boldsymbol{j}})\right] \\
&\quad - \sum_{i=1}^{N}\sum_{a=1}^{M} \mathrm{E}_{q(\boldsymbol{U},\boldsymbol{V}|\boldsymbol{X},\boldsymbol{Y})}[\log(1+\exp(\alpha_1 - (\boldsymbol{u_i}-\boldsymbol{v_a})^T(\boldsymbol{u_i}-\boldsymbol{v_a})))] \\
&\quad - \sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N} \log(1+\mathrm{E}_{q(\boldsymbol{U},\boldsymbol{V}|\boldsymbol{X},\boldsymbol{Y})}[\exp(\alpha_0 - (\boldsymbol{u_i}-\boldsymbol{u_j})^T(\boldsymbol{u_i}-\boldsymbol{u_j}))])
\end{aligned}
$$

Recall $\boldsymbol{u_i}, \boldsymbol{u_j}$ are $D \times 1$ column vectors. Define $\mathbf{u} = \tilde{\boldsymbol{u}}_{\boldsymbol{i}} - \tilde{\boldsymbol{u}}_{\boldsymbol{j}}$. Then we have, $\boldsymbol{u_i} - \boldsymbol{u_j} \overset{iid}{=} N(\mathbf{u}, 2\tilde{\Lambda}_0)$, where $\mathbf{u}$ is a $D \times 1$ vector and $\tilde{\Lambda}_0$ is an $D \times D$ positive semidefinite matrix. Further define $\mathbf{Z} = (2\tilde{\Lambda}_0)^{-1/2}(\boldsymbol{u_i} - \boldsymbol{u_j} - (\tilde{\boldsymbol{u}}_{\boldsymbol{i}} - \tilde{\boldsymbol{u}}_{\boldsymbol{j}}))$. Then clearly $\mathbf{Z}$ follows D dimensional multivariate standard normal distribution and its density function is given by $f_Z(z) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}\mathbf{z}^T\mathbf{z})$. Consequently, we have $\boldsymbol{u_i} - \boldsymbol{u_j} = 2\tilde{\Lambda}_0^{1/2}\mathbf{Z} + \mathbf{u}$.

Therefore, we can reparameterize

$$
\begin{aligned}
&\mathrm{E}_{q(\boldsymbol{U},\boldsymbol{V}|\boldsymbol{X},\boldsymbol{Y})}[\exp(-(\boldsymbol{u_i} - \boldsymbol{u_j})^T(\boldsymbol{u_i} - \boldsymbol{u_j}))] \\
&= \mathrm{E}_{q(\boldsymbol{U},\boldsymbol{V}|\boldsymbol{X},\boldsymbol{Y})}\left[\exp\left(-\left(\mathbf{Z}^T(2\tilde{\Lambda}_0)^{1/2} + \mathbf{u}^T\right)\left((2\tilde{\Lambda}_0)^{1/2}\mathbf{Z} + \mathbf{u}\right)\right)\right] \\
&= \mathrm{E}_{q(\boldsymbol{U},\boldsymbol{V}|\boldsymbol{X},\boldsymbol{Y})}\left[\exp\left(-\mathbf{Z}^T(2\tilde{\Lambda}_0)\mathbf{Z} - 2\mathbf{Z}^T(2\tilde{\Lambda}_0)^{1/2}\mathbf{u} - \mathbf{u}^T\mathbf{u})\right)\right] \\
&= \frac{1}{\sqrt{2\pi}} \int \exp\left(-\mathbf{Z}^T\big(2\tilde{\Lambda}_0 + \frac{1}{2}\mathbf{I}\big)\mathbf{Z} - 2\mathbf{Z}^T(2\tilde{\Lambda}_0)^{1/2}\mathbf{u} - \mathbf{u}^T\mathbf{u}\right) d\mathbf{Z}
\end{aligned}
$$

Now define $Q = \mathbf{u}(2\tilde{\Lambda}_0 + \frac{1}{2}\mathbf{I})^{-1}(2\tilde{\Lambda}_0)^{1/2}$. Then the above integral becomes

$$
\begin{aligned}
&\frac{1}{\sqrt{2\pi}} \int \exp\left(-(\mathbf{Z} - Q)^T(2\tilde{\Lambda}_0 + \frac{1}{2}\mathbf{I})(\mathbf{Z} - Q) - \mathbf{u}^T\mathbf{u} + \mathbf{u}^T(2\tilde{\Lambda}_0 + \frac{1}{2}\mathbf{I})^{-1}(2\tilde{\Lambda}_0)\mathbf{u}\right) d\mathbf{Z} \\
&= \exp\left(-\mathbf{u}^T\mathbf{u} + \mathbf{u}^T(2\tilde{\Lambda}_0 + \frac{1}{2}\mathbf{I})^{-1}(2\tilde{\Lambda}_0)\mathbf{u}\right) \det(\mathbf{I} + 4\tilde{\Lambda}_0)^{-\frac{1}{2}} \\
&= \exp\left(-\mathbf{u}^T(\mathbf{I} - (2\tilde{\Lambda}_0 + \frac{1}{2}\mathbf{I})^{-1}(2\tilde{\Lambda}_0))\mathbf{u}\right) \det(\mathbf{I} + 4\tilde{\Lambda}_0)^{-\frac{1}{2}} \\
&= \exp\left(-\mathbf{u}^T(4\tilde{\Lambda}_0 + \mathbf{I})^{-1}\mathbf{u}\right) \det(\mathbf{I} + 4\tilde{\Lambda}_0)^{-\frac{1}{2}} \ .
\end{aligned}
$$

The last line follows since for any two invertible matrices A and B, if $A+B$ is also invertible, then by Henderson and Searle (1981)

$$
(A + B)^{-1} = A^{-1} - A^{-1}B(I + A^{-1}B)^{-1}A^{-1}.
$$

Letting $B = 4\tilde{\Lambda}_0$ and $A = I$ gives:

$$\mathrm{E}_{q(\boldsymbol{U},\boldsymbol{V}|\boldsymbol{X},\boldsymbol{Y})}[\exp(-(\boldsymbol{u_i}-\boldsymbol{u_j})^T(\boldsymbol{u_i}-\boldsymbol{u_j}))] = \exp\Big(-(\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{u}}_{\boldsymbol{j}})^T(\mathbf{I}+4\tilde{\Lambda}_0)^{-1}(\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{u}}_{\boldsymbol{j}})\Big)\det(\mathbf{I}+4\tilde{\Lambda}_0)^{-\frac{1}{2}}$$

Recall $\boldsymbol{u_i}, \boldsymbol{v_a}$ are $D \times 1$ column vectors. Define $\mathbf{u} = \tilde{\boldsymbol{u}}_{\boldsymbol{i}} - \tilde{\boldsymbol{v}}_{\boldsymbol{a}}$. Then we have, $\boldsymbol{u_i} - \boldsymbol{v_a} \stackrel{iid}{=} N(\mathbf{u}, \tilde{\Lambda}_0 + \tilde{\Lambda}_1)$, where $\mathbf{u}$ is a $D \times 1$ vector and $\tilde{\Lambda}_0$ is an $D \times D$ positive semidefinite matrix. Further define $\mathbf{Z} = (\tilde{\Lambda}_0 + \tilde{\Lambda}_1)^{-1/2}(\boldsymbol{u_i} - \boldsymbol{v_a} - (\tilde{\boldsymbol{u}}_{\boldsymbol{i}} - \tilde{\boldsymbol{v}}_{\boldsymbol{a}}))$. Then clearly $\mathbf{Z}$ follows D dimensional multivariate standard normal distribution and its density function is given by $f_Z(z) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}\mathbf{z}^T\mathbf{z})$. Consequently, we have $\boldsymbol{u_i} - \boldsymbol{v_a} = (\tilde{\Lambda}_0 + \tilde{\Lambda}_1)^{1/2}\mathbf{Z} + \mathbf{u}$.

Therefore, we can reparameterize

$$\begin{aligned}
&\mathrm{E}_{q(\boldsymbol{U},\boldsymbol{V}|\boldsymbol{X},\boldsymbol{Y})}[\exp(-(\boldsymbol{u_i} - \boldsymbol{v_a})^T(\boldsymbol{u_i} - \boldsymbol{v_a}))] \\
&= \mathrm{E}_{q(\boldsymbol{U},\boldsymbol{V}|\boldsymbol{X},\boldsymbol{Y})}\left[\exp\left(-\Big(\mathbf{Z}^T(\tilde{\Lambda}_0 + \tilde{\Lambda}_1)^{1/2} + \mathbf{u}^T\Big)\Big((\tilde{\Lambda}_0 + \tilde{\Lambda}_1)^{1/2}\mathbf{Z} + \mathbf{u}\Big)\right)\right] \\
&= \mathrm{E}_{q(\boldsymbol{U},\boldsymbol{V}|\boldsymbol{X},\boldsymbol{Y})}\left[\exp\left(-\mathbf{Z}^T(\tilde{\Lambda}_0 + \tilde{\Lambda}_1)\mathbf{Z} - 2\mathbf{Z}^T(\tilde{\Lambda}_0 + \tilde{\Lambda}_1)^{1/2}\mathbf{u} - \mathbf{u}^T\mathbf{u})\right)\right] \\
&= \frac{1}{\sqrt{2\pi}}\int \exp\left(-\mathbf{Z}^T\big(\tilde{\Lambda}_0 + \tilde{\Lambda}_1 + \frac{1}{2}\mathbf{I}\big)\mathbf{Z} - 2\mathbf{Z}^T(\tilde{\Lambda}_0 + \tilde{\Lambda}_1)^{1/2}\mathbf{u} - \mathbf{u}^T\mathbf{u}\right)d\mathbf{Z}
\end{aligned}$$

Now define $Q = \mathbf{u}(\tilde{\Lambda}_0 + \tilde{\Lambda}_1 + \frac{1}{2}\mathbf{I})^{-1}(\tilde{\Lambda}_0 + \tilde{\Lambda}_1)^{1/2}$. Then the above integral becomes

$$\begin{aligned}
&\frac{1}{\sqrt{2\pi}}\int \exp\left(-(\mathbf{Z} - Q)^T(\tilde{\Lambda}_0 + \tilde{\Lambda}_1 + \frac{1}{2}\mathbf{I})(\mathbf{Z} - Q) - \mathbf{u}^T\mathbf{u} + \mathbf{u}^T(\tilde{\Lambda}_0 + \tilde{\Lambda}_1 + \frac{1}{2}\mathbf{I})^{-1}(\tilde{\Lambda}_0 + \tilde{\Lambda}_1)\mathbf{u}\right)d\mathbf{Z} \\
&= \exp\Big(-\mathbf{u}^T\mathbf{u} + \mathbf{u}^T(\tilde{\Lambda}_0 + \tilde{\Lambda}_1 + \frac{1}{2}\mathbf{I})^{-1}(\tilde{\Lambda}_0 + \tilde{\Lambda}_1)\mathbf{u}\Big)\det(\mathbf{I} + 2\tilde{\Lambda}_0 + 2\tilde{\Lambda}_1)^{-\frac{1}{2}} \\
&= \exp\Big(-\mathbf{u}^T(\mathbf{I} - (\tilde{\Lambda}_0 + \tilde{\Lambda}_1 + \frac{1}{2}\mathbf{I})^{-1}(\tilde{\Lambda}_0 + \tilde{\Lambda}_1))\mathbf{u}\Big)\det(\mathbf{I} + 2\tilde{\Lambda}_0 + 2\tilde{\Lambda}_1)^{-\frac{1}{2}} \\
&= \exp\Big(-\mathbf{u}^T(\mathbf{I} + 2\tilde{\Lambda}_0 + 2\tilde{\Lambda}_1)^{-1}\mathbf{u}\Big)\det(\mathbf{I} + 2\tilde{\Lambda}_0 + 2\tilde{\Lambda}_1)^{-\frac{1}{2}}\ .
\end{aligned}$$

The last line follows since for any two invertible matrices A and B, if $A+B$ is also invertible, then by Henderson and Searle (1981)

$$(A+B)^{-1} = A^{-1} - A^{-1}B(I+A^{-1}B)^{-1}A^{-1}.$$

Letting $A = I$ and $B = 2\tilde{\Lambda}_0 + 2\tilde{\Lambda}_1$ gives:

$$\mathrm{E}_{q(\boldsymbol{U},\boldsymbol{V}|\boldsymbol{X},\boldsymbol{Y})}[\exp(-(\boldsymbol{u_i}-\boldsymbol{v_a})^T(\boldsymbol{u_i}-\boldsymbol{v_a}))] = \exp\Big(-(\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{v}}_{\boldsymbol{a}})^T(\mathbf{I}+2\tilde{\Lambda}_0+2\tilde{\Lambda}_1)^{-1}(\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{v}}_{\boldsymbol{a}})\Big)\det(\mathbf{I}+2\tilde{\Lambda}_0+2\tilde{\Lambda}_1)^{-\frac{1}{2}}$$

Finally, the Kullback-Leiber divergence between the variational posterior and the true posterior is

$$\begin{aligned}
&\mathrm{KL}[q(\boldsymbol{U},\boldsymbol{V}|\boldsymbol{X},\boldsymbol{Y})||f(\boldsymbol{U},\boldsymbol{V}|\boldsymbol{X},\boldsymbol{Y})]\\
\geq&\frac{1}{2}\Big(DN\log(\lambda_0^2) - N\log(\det(\tilde{\Lambda}_0))\Big) + \frac{N\mathrm{tr}(\tilde{\Lambda}_0)}{2\lambda_0^2} + \frac{\sum_{i=1}^N \tilde{\boldsymbol{u}}_{\boldsymbol{i}}^T\tilde{\boldsymbol{u}}_{\boldsymbol{i}}}{2\lambda_0^2} - \frac{1}{2}ND\\
+&\frac{1}{2}\Big(DM\log(\lambda_1^2) - M\log(\det(\tilde{\Lambda}_1))\Big) + \frac{M\mathrm{tr}(\tilde{\Lambda}_1)}{2\lambda_1^2} + \frac{\sum_{a=1}^M \tilde{\boldsymbol{v}}_{\boldsymbol{a}}^T\tilde{\boldsymbol{v}}_{\boldsymbol{a}}}{2\lambda_1^2} - \frac{1}{2}MD\\
-&\sum_{i=1}^N\sum_{a=1}^M y_{ia}\left[\tilde{\alpha}_1 - \mathrm{tr}(\tilde{\Lambda}_0) - \mathrm{tr}(\tilde{\Lambda}_1) - (\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{v}}_{\boldsymbol{a}})^T(\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{v}}_{\boldsymbol{a}})\right]\\
-&\sum_{i=1}^N\sum_{j=1,j\neq i}^N x_{ij}\left[\tilde{\alpha}_0 - 2\mathrm{tr}(\tilde{\Lambda}_0) - (\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{u}}_{\boldsymbol{j}})^T(\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{u}}_{\boldsymbol{j}})\right]\\
+&\sum_{i=1}^N\sum_{a=1}^M \log\left(1+\frac{\exp(\tilde{\alpha}_1)}{\det(\mathbf{I}+2\tilde{\Lambda}_0+2\tilde{\Lambda}_1)^{\frac{1}{2}}}\exp\Big(-(\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{v}}_{\boldsymbol{a}})^T(\mathbf{I}+2\tilde{\Lambda}_0+2\tilde{\Lambda}_1)^{-1}(\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{v}}_{\boldsymbol{a}})\Big)\right)\\
+&\sum_{i=1}^N\sum_{j=1,j\neq i}^N \log\left(1+\frac{\exp(\tilde{\alpha}_0)}{\det(\mathbf{I}+4\tilde{\Lambda}_0)^{1/2}}\exp\Big(-(\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{u}}_{\boldsymbol{j}})^T(\mathbf{I}+4\tilde{\Lambda}_0)^{-1}(\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{u}}_{\boldsymbol{j}})\Big)\right) + \mathrm{Const}\tilde{\boldsymbol{u}}_{\boldsymbol{i}}
\end{aligned}$$

5.2 Derivations of EM algorithms

E-step: Estimate $\tilde{\boldsymbol{u}}_{\boldsymbol{i}}$, $\tilde{\boldsymbol{v}}_{\boldsymbol{a}}$, $\tilde{\Lambda}_0$ and $\tilde{\Lambda}_1$ by minimizing the KL divergence.

$$\mathrm{KL}_{\tilde{\boldsymbol{u}}_{\boldsymbol{i}}}[q(\boldsymbol{U},\boldsymbol{V}|\boldsymbol{X},\boldsymbol{Y})||f(\boldsymbol{U},\boldsymbol{V}|\boldsymbol{X},\boldsymbol{Y})]$$

$$
\begin{aligned}
&\geq \frac{\sum_{i=1}^{N} \boldsymbol{\tilde{u}_i}^T \boldsymbol{\tilde{u}_i}}{2\lambda_0^2} \\
&- \sum_{i=1}^{N}\sum_{a=1}^{M} y_{ia}\left[\tilde{\alpha}_1 - \mathrm{tr}(\tilde{\Lambda}_0) - \mathrm{tr}(\tilde{\Lambda}_1) - (\boldsymbol{\tilde{u}_i} - \boldsymbol{\tilde{v}_a})^T(\boldsymbol{\tilde{u}_i} - \boldsymbol{\tilde{v}_a})\right] \\
&- \sum_{i=1}^{N}\sum_{j=1, j\neq i}^{N} x_{ij}\left[\tilde{\alpha}_0 - 2\mathrm{tr}(\tilde{\Lambda}_0) - (\boldsymbol{\tilde{u}_i} - \boldsymbol{\tilde{u}_j})^T(\boldsymbol{\tilde{u}_i} - \boldsymbol{\tilde{u}_j})\right] \\
&+ \sum_{i=1}^{N}\sum_{a=1}^{M} \log\left(1 + \frac{\exp(\tilde{\alpha}_1)}{\det(\mathbf{I} + 2\tilde{\Lambda}_0 + 2\tilde{\Lambda}_1)^{\frac{1}{2}}}\exp\Big(-(\boldsymbol{\tilde{u}_i} - \boldsymbol{\tilde{v}_a})^T(\mathbf{I} + 2\tilde{\Lambda}_0 + 2\tilde{\Lambda}_1)^{-1}(\boldsymbol{\tilde{u}_i} - \boldsymbol{\tilde{v}_a})\Big)\right) \\
&+ \sum_{i=1}^{N}\sum_{j=1, j\neq i}^{N} \log\left(1 + \frac{\exp(\tilde{\alpha}_0)}{\det(\mathbf{I} + 4\tilde{\Lambda}_0)^{1/2}}\exp\Big(-(\boldsymbol{\tilde{u}_i} - \boldsymbol{\tilde{u}_j})^T(\mathbf{I} + 4\tilde{\Lambda}_0)^{-1}(\boldsymbol{\tilde{u}_i} - \boldsymbol{\tilde{u}_j})\Big)\right) + \mathrm{Const}\boldsymbol{\tilde{u}_i}
\end{aligned}
$$

To find the closed form updates of $\boldsymbol{\tilde{u}_i}$, we use second-order Taylor-expansions of

$$
\boldsymbol{F_A} = \sum_{i=1}^{N}\sum_{a=1}^{M} \log\left(1 + \frac{\exp(\tilde{\alpha}_1)}{\det(\mathbf{I} + 2\tilde{\Lambda}_0 + 2\tilde{\Lambda}_1)^{\frac{1}{2}}}\exp\Big(-(\boldsymbol{\tilde{u}_i} - \boldsymbol{\tilde{v}_a})^T(\mathbf{I} + 2\tilde{\Lambda}_0 + 2\tilde{\Lambda}_1)^{-1}(\boldsymbol{\tilde{u}_i} - \boldsymbol{\tilde{v}_a})\Big)\right) \tag{5.1}
$$

$$
\boldsymbol{F_I} = \sum_{i=1}^{N}\sum_{j=1, j\neq i}^{N} \log\left(1 + \frac{\exp(\tilde{\alpha}_0)}{\det(\mathbf{I} + 4\tilde{\Lambda}_0)^{1/2}}\exp\Big(-(\boldsymbol{\tilde{u}_i} - \boldsymbol{\tilde{u}_j})^T(\mathbf{I} + 4\tilde{\Lambda}_0)^{-1}(\boldsymbol{\tilde{u}_i} - \boldsymbol{\tilde{u}_j})\Big)\right) \tag{5.2}
$$

The gradients of $\boldsymbol{F_i}$ and $\boldsymbol{F_{ia}}$ with respect to $\boldsymbol{\tilde{u}_i}$ are

$$
\boldsymbol{G_I}(\boldsymbol{\tilde{u}_i}) = -2(\mathbf{I} + 4\tilde{\Lambda}_0)^{-1}\sum_{j=1, j\neq i}^{N}(\boldsymbol{\tilde{u}_i} - \boldsymbol{\tilde{u}_j})\left[1 + \frac{\det(\mathbf{I} + 4\tilde{\Lambda}_0)^{1/2}}{\exp(\tilde{\alpha}_0)}\exp\Big((\boldsymbol{\tilde{u}_i} - \boldsymbol{\tilde{u}_j})^T(\mathbf{I} + 4\tilde{\Lambda}_0)^{-1}(\boldsymbol{\tilde{u}_i} - \boldsymbol{\tilde{u}_j})\Big)\right]^{-1}
$$

$$
\begin{aligned}
&\boldsymbol{G_A}(\boldsymbol{\tilde{u}_i}) = -2(\mathbf{I} + 2\tilde{\Lambda}_0 + 2\tilde{\Lambda}_1)^{-1}\sum_{a=1}^{M}(\boldsymbol{\tilde{u}_i} - \boldsymbol{\tilde{v}_a}) \\
&\left[1 + \frac{\det(\mathbf{I} + \tilde{\Lambda}_0 + \tilde{\Lambda}_1)^{1/2}}{\exp(\tilde{\alpha}_1)}\exp\Big((\boldsymbol{\tilde{u}_i} - \boldsymbol{\tilde{v}_a})^T(\mathbf{I} + 2\tilde{\Lambda}_0 + 2\tilde{\Lambda}_1)^{-1}(\boldsymbol{\tilde{u}_i} - \boldsymbol{\tilde{v}_a})\Big)\right]^{-1}
\end{aligned}
$$

The second-order partial derivatives (Hessian matrices) of $\boldsymbol{F_I}, \boldsymbol{F_A}$ with respect to $\boldsymbol{\tilde{u}_i}$ are

$$\boldsymbol{H_I}(\tilde{\boldsymbol{u}}_{\boldsymbol{i}}) = -2(\mathbf{I}+4\tilde{\Lambda}_0)^{-1}\sum_{j=1,j\neq i}^{N}\left[1+\frac{\det(\mathbf{I}+4\tilde{\Lambda}_0)^{1/2}}{\exp(\tilde{\alpha}_0)}\exp\Big((\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{u}}_{\boldsymbol{j}})^T(\mathbf{I}+4\tilde{\Lambda}_0)^{-1}(\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{u}}_{\boldsymbol{j}})\Big)\right]^{-1}$$
$$\left[\mathbf{I}-\frac{2(\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{u}}_{\boldsymbol{j}})(\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{u}}_{\boldsymbol{j}})^T(\mathbf{I}+4\tilde{\Lambda}_0)^{-1}}{1+\frac{\exp(\tilde{\alpha}_0)}{\det(\mathbf{I}+4\tilde{\Lambda}_0)^{1/2}}\exp\Big(-(\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{u}}_{\boldsymbol{j}})^T(\mathbf{I}+4\tilde{\Lambda}_0)^{-1}(\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{u}}_{\boldsymbol{j}})\Big)}\right]$$
$$\boldsymbol{H_A}(\tilde{\boldsymbol{u}}_{\boldsymbol{i}}) = -2(\mathbf{I}+2\tilde{\Lambda}_0+2\tilde{\Lambda}_1)^{-1}$$
$$\sum_{a=1}^{M}\left[1+\frac{\det(\mathbf{I}+2\tilde{\Lambda}_0+2\tilde{\Lambda}_1)^{1/2}}{\exp(\tilde{\alpha}_1)}\exp\Big((\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{v}}_{\boldsymbol{a}})^T(\mathbf{I}+2\tilde{\Lambda}_0+2\tilde{\Lambda}_1)^{-1}(\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{v}}_{\boldsymbol{a}})\Big)\right]^{-1}$$
$$\left[\mathbf{I}-\frac{2(\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{v}}_{\boldsymbol{a}})(\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{v}}_{\boldsymbol{a}})^T(\mathbf{I}+2\tilde{\Lambda}_0+2\tilde{\Lambda}_1)^{-1}}{1+\frac{\exp(\tilde{\alpha}_1)}{\det(\mathbf{I}+2\tilde{\Lambda}_0+2\tilde{\Lambda}_1)^{1/2}}\exp\Big(-(\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{v}}_{\boldsymbol{a}})^T(\mathbf{I}+2\tilde{\Lambda}_0+2\tilde{\Lambda}_1)^{-1}(\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{v}}_{\boldsymbol{a}})\Big)}\right]$$

$$\mathrm{KL}_{\tilde{\boldsymbol{u}}_{\boldsymbol{i}}} = \tilde{\boldsymbol{u}}_{\boldsymbol{i}}^T\big(\frac{1}{2\lambda_0^2}+\sum_{i=a}^{M}y_{ia}+\sum_{j=1,j\neq i}^{N}(x_{ij}+y_{ji})+\boldsymbol{H_I}(\tilde{\boldsymbol{u}}_{\boldsymbol{i}})+\frac{1}{2}\boldsymbol{H_A}(\tilde{\boldsymbol{u}}_{\boldsymbol{i}})\big)\tilde{\boldsymbol{u}}_{\boldsymbol{i}}$$
$$-2\tilde{\boldsymbol{u}}_{\boldsymbol{i}}\big(\sum_{i=a}^{M}y_{ia}\tilde{\boldsymbol{v}}_{\boldsymbol{a}}+\sum_{j=1,j\neq i}^{N}(x_{ij}+y_{ji})\tilde{\boldsymbol{u}}_{\boldsymbol{j}}-\boldsymbol{G_I}(\tilde{\boldsymbol{u}}_{\boldsymbol{i}})-\frac{1}{2}\boldsymbol{G_A}(\tilde{\boldsymbol{u}}_{\boldsymbol{i}})+(\boldsymbol{H_I}(\tilde{\boldsymbol{u}}_{\boldsymbol{i}})+\frac{1}{2}\boldsymbol{H_A}(\tilde{\boldsymbol{u}}_{\boldsymbol{i}}))\tilde{\boldsymbol{u}}_{\boldsymbol{i}}\big).$$

With the Taylor-expansions of the log functions, we can obtain the closed form update rule of $\tilde{\boldsymbol{u}}_{\boldsymbol{i}}$ by setting the partial derivative of KL equal to 0. Finally, we have

$$\tilde{\boldsymbol{u}}_{\boldsymbol{i}} = \left[\left(\frac{1}{2\lambda_0^2}+\sum_{j=1,j\neq i}^{N}(x_{ij}+y_{ji})+\sum_{a=1}^{M}y_{ia}\right)\boldsymbol{I}+\boldsymbol{H_I}(\tilde{\boldsymbol{u}}_{\boldsymbol{i}})+\frac{1}{2}\boldsymbol{H_A}(\tilde{\boldsymbol{u}}_{\boldsymbol{i}})\right]^{-1}$$
$$\left[\sum_{j=1,j\neq i}^{N}(x_{ij}+y_{ji})\tilde{\boldsymbol{u}}_{\boldsymbol{j}}+\sum_{a=1}^{M}y_{ia}\tilde{\boldsymbol{v}}_{\boldsymbol{a}}-\boldsymbol{G_I}(\tilde{\boldsymbol{u}}_{\boldsymbol{i}})+\Big(\boldsymbol{H_I}(\tilde{\boldsymbol{u}}_{\boldsymbol{i}})+\frac{1}{2}\boldsymbol{H_A}(\tilde{\boldsymbol{u}}_{\boldsymbol{i}})\Big)\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\frac{1}{2}\boldsymbol{G_A}(\tilde{\boldsymbol{u}}_{\boldsymbol{i}})\right]$$

Similarly, we can obtain the closed form update rule for $\tilde{\boldsymbol{v}}_{\boldsymbol{a}}$ by taking the second order Taylor-expansion of $\boldsymbol{F_{ia}}$ (see Equation 5.1) The gradient and Hessian matrix of $\boldsymbol{F_{ia}}$ with respect to $\tilde{\boldsymbol{v}}_{\boldsymbol{a}}$ are

$$\boldsymbol{G_A}(\tilde{\boldsymbol{v}}_{\boldsymbol{a}}) = -2(\mathbf{I} + 2\tilde{\Lambda}_0 + 2\tilde{\Lambda}_1)^{-1} \sum_{i=1}^{N} (\tilde{\boldsymbol{v}}_{\boldsymbol{a}} - \tilde{\boldsymbol{u}}_{\boldsymbol{i}})$$

$$\left[1 + \frac{\det(\mathbf{I} + \tilde{\Lambda}_0 + \tilde{\Lambda}_1)^{1/2}}{\exp(\tilde{\alpha}_1)} \exp\Big((\tilde{\boldsymbol{u}}_{\boldsymbol{i}} - \tilde{\boldsymbol{v}}_{\boldsymbol{a}})^T (\mathbf{I} + 2\tilde{\Lambda}_0 + 2\tilde{\Lambda}_1)^{-1} (\tilde{\boldsymbol{u}}_{\boldsymbol{i}} - \tilde{\boldsymbol{v}}_{\boldsymbol{a}})\Big)\right]^{-1}$$

$$\boldsymbol{H_A}(\tilde{\boldsymbol{v}}_{\boldsymbol{a}}) = -2(\mathbf{I} + 2\tilde{\Lambda}_0 + 2\tilde{\Lambda}_1)^{-1}$$

$$\sum_{i=1}^{N} \left[1 + \frac{\det(\mathbf{I} + 2\tilde{\Lambda}_0 + 2\tilde{\Lambda}_1)^{1/2}}{\exp(\tilde{\alpha}_1)} \exp\Big((\tilde{\boldsymbol{u}}_{\boldsymbol{i}} - \tilde{\boldsymbol{v}}_{\boldsymbol{a}})^T (\mathbf{I} + 2\tilde{\Lambda}_0 + 2\tilde{\Lambda}_1)^{-1} (\tilde{\boldsymbol{u}}_{\boldsymbol{i}} - \tilde{\boldsymbol{v}}_{\boldsymbol{a}})\Big)\right]^{-1}$$

$$\left[\mathbf{I} - \frac{2(\tilde{\boldsymbol{v}}_{\boldsymbol{a}} - \tilde{\boldsymbol{u}}_{\boldsymbol{i}})(\tilde{\boldsymbol{v}}_{\boldsymbol{a}} - \tilde{\boldsymbol{u}}_{\boldsymbol{i}})^T (\mathbf{I} + 2\tilde{\Lambda}_0 + 2\tilde{\Lambda}_1)^{-1}}{1 + \frac{\exp(\tilde{\alpha}_1)}{\det(\mathbf{I} + 2\tilde{\Lambda}_0 + 2\tilde{\Lambda}_1)^{1/2}} \exp\Big(-(\tilde{\boldsymbol{u}}_{\boldsymbol{i}} - \tilde{\boldsymbol{v}}_{\boldsymbol{a}})^T (\mathbf{I} + 2\tilde{\Lambda}_0 + 2\tilde{\Lambda}_1)^{-1} (\tilde{\boldsymbol{u}}_{\boldsymbol{i}} - \tilde{\boldsymbol{v}}_{\boldsymbol{a}})\Big)}\right]$$

$$\mathrm{KL}_{\tilde{\boldsymbol{v}}_{\boldsymbol{a}}} = \tilde{\boldsymbol{v}}_{\boldsymbol{a}}{}^T \big(\frac{1}{2\lambda_1^2} + \sum_{i=1}^{N} \sum_{i=a}^{M} y_{ia} - \frac{1}{2} \boldsymbol{H_A}(\tilde{\boldsymbol{v}}_{\boldsymbol{a}})\big) \tilde{\boldsymbol{v}}_{\boldsymbol{a}} - 2\tilde{\boldsymbol{v}}_{\boldsymbol{a}} \big(\sum_{i=1}^{N} \sum_{i=a}^{M} y_{ia} \tilde{\boldsymbol{u}}_{\boldsymbol{i}} - \frac{1}{2} \boldsymbol{G_A}(\tilde{\boldsymbol{v}}_{\boldsymbol{a}})\big).$$

With the Taylor-expansions of the log functions, we can obtain the closed form update rule of $\tilde{\boldsymbol{v}}_{\boldsymbol{a}}$ by setting the partial derivative of KL equal to 0. Then, we have

$$\tilde{\boldsymbol{v}}_{\boldsymbol{a}} = \left[\left(\frac{1}{2\lambda_1^2} + \sum_{i=1}^{N} y_{ia}\right) \boldsymbol{I} - \frac{1}{2} \boldsymbol{H_A}(\tilde{\boldsymbol{v}}_{\boldsymbol{a}})\right]^{-1}$$
$$\left[\sum_{i=1}^{N} y_{ia} \tilde{\boldsymbol{u}}_{\boldsymbol{i}} - \frac{1}{2} \boldsymbol{G_A}(\tilde{\boldsymbol{v}}_{\boldsymbol{a}})\right]$$

To find the closed form updates of $\tilde{\Lambda}_0$ and $\tilde{\Lambda}_1$ we used the first-order Taylor-expansions of $\boldsymbol{F_i}$ and $\boldsymbol{F_{ia}}$. The gradients of $\boldsymbol{F_i}$ and $\boldsymbol{F_{ia}}$ with respect to $\tilde{\Lambda}_0$ are:

$$\boldsymbol{G_I}(\tilde{\Lambda}_0) = \sum_{i=1}^{N} \sum_{j=1, j \neq i}^{N} \left[1 + \frac{\det(\mathbf{I} + 4\tilde{\Lambda}_0)^{1/2}}{\exp(\tilde{\alpha}_0)} \exp\Big((\tilde{\boldsymbol{u}}_{\boldsymbol{i}} - \tilde{\boldsymbol{u}}_{\boldsymbol{j}})^T (\mathbf{I} + 4\tilde{\Lambda}_0)^{-1} (\tilde{\boldsymbol{u}}_{\boldsymbol{i}} - \tilde{\boldsymbol{u}}_{\boldsymbol{j}})\Big)\right]^{-1}$$

$$4(\mathbf{I}+4\tilde{\Lambda}_0)^{-1}\Big((\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{u}}_{\boldsymbol{j}})(\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{u}}_{\boldsymbol{j}})^T(\mathbf{I}+4\tilde{\Lambda}_0)^{-1}-\frac{1}{2}\mathbf{I}\Big)$$

$$\boldsymbol{G}_{\boldsymbol{A}}(\tilde{\Lambda}_0)=\sum_{i=1}^{N}\sum_{a=1}^{M}\left[1+\frac{\det(\mathbf{I}+2\tilde{\Lambda}_0+2\tilde{\Lambda}_1)^{1/2}}{\exp(\tilde{\alpha}_1)}\exp\Big((\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{v}}_{\boldsymbol{a}})^T(\mathbf{I}+2\tilde{\Lambda}_0+2\tilde{\Lambda}_1)^{-1}(\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{v}}_{\boldsymbol{a}})\Big)\right]^{-1}$$
$$2(\mathbf{I}+2\tilde{\Lambda}_0+2\tilde{\Lambda}_1)^{-1}\Big((\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{v}}_{\boldsymbol{a}})(\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{v}}_{\boldsymbol{a}})^T(\mathbf{I}+2\tilde{\Lambda}_0+2\tilde{\Lambda}_1)^{-1}-\frac{1}{2}\mathbf{I}\Big)$$

The gradients of $\boldsymbol{F}_{\boldsymbol{a}}$ and $\boldsymbol{F}_{\boldsymbol{ia}}$ with respect to $\tilde{\Lambda}_1$ are:

$$\boldsymbol{G}_{\boldsymbol{A}}(\tilde{\Lambda}_1)=\sum_{i=1}^{N}\sum_{a=1}^{M}\left[1+\frac{\det(\mathbf{I}+2\tilde{\Lambda}_0+2\tilde{\Lambda}_1)^{1/2}}{\exp(\tilde{\alpha}_1)}\exp\Big((\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{v}}_{\boldsymbol{a}})^T(\mathbf{I}+2\tilde{\Lambda}_0+2\tilde{\Lambda}_1)^{-1}(\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{v}}_{\boldsymbol{a}})\Big)\right]^{-1}$$
$$2(\mathbf{I}+2\tilde{\Lambda}_0+2\tilde{\Lambda}_1)^{-1}\Big((\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{v}}_{\boldsymbol{a}})(\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{v}}_{\boldsymbol{a}})^T(\mathbf{I}+2\tilde{\Lambda}_0+2\tilde{\Lambda}_1)^{-1}-\frac{1}{2}\mathbf{I}\Big)$$

$$\mathrm{KL}_{\Lambda_0}=\mathrm{tr}(\tilde{\Lambda}_0)\big(\frac{N}{2\lambda_0^2}+\sum_{i=1}^{N}\sum_{a=1}^{M}y_{ia}+2\sum_{i=1}^{N}\sum_{j=1}^{N}x_{ij}\big)-\frac{N}{2}\log(\det(\tilde{\Lambda}_0))+\boldsymbol{G}_{\boldsymbol{I}}(\tilde{\Lambda}_0)\tilde{\Lambda}_0+\boldsymbol{G}_{\boldsymbol{A}}(\tilde{\Lambda}_0)\tilde{\Lambda}_0$$
$$\mathrm{KL}_{\Lambda_1}=\mathrm{tr}(\tilde{\Lambda}_1)\big(\frac{M}{2\lambda_0^2}+\sum_{i=1}^{N}\sum_{a=1}^{M}y_{ia}\big)-\frac{M}{2}\log(\det(\tilde{\Lambda}_1))+\boldsymbol{G}_{\boldsymbol{A}}(\tilde{\Lambda}_1)\tilde{\Lambda}_1$$

With the Taylor-expansions of the log functions, we can obtain the closed form update rule of $\tilde{\Lambda}_0$ $\tilde{\Lambda}_1$ by setting the partial derivative of KL equal to 0. Then, we have

$$\tilde{\Lambda}_0=\frac{N}{2}\left[\left(\frac{N}{2}\frac{1}{\lambda_0^2}+2\sum_{i=1}^{N}\sum_{j=1}^{N}x_{ij}+\sum_{i=1}^{N}\sum_{a=1}^{M}y_{ia}\right)\boldsymbol{I}+\boldsymbol{G}_{\boldsymbol{I}}(\tilde{\Lambda}_0)+\boldsymbol{G}_{\boldsymbol{A}}(\tilde{\Lambda}_0)\right]^{-1}$$
$$\tilde{\Lambda}_1=\frac{M}{2}\left[\left(\frac{M}{2}\frac{1}{\lambda_1^2}+\sum_{i=1}^{N}\sum_{a=1}^{M}y_{ia}\right)\boldsymbol{I}+\boldsymbol{G}_{\boldsymbol{A}}(\tilde{\Lambda}_1)\right]^{-1}$$

M-step: Estimate $\tilde{\alpha}_0$, $\tilde{\alpha}_1$ and $\tilde{\alpha}_2$ by minimizing the KL divergence. To find the closed form updates of $\tilde{\alpha}_0$, $\tilde{\alpha}_1$ and $\tilde{\alpha}_2$, we used second-order Taylor-expansions of the log functions and set the partial derivatives of KL with respects to $\tilde{\alpha}_0$, $\tilde{\alpha}_1$ and $\tilde{\alpha}_2$ as zeros. Then we have

$$\begin{aligned}\tilde{\alpha}_0 =& \frac{\sum_{i=1}^{N}\sum_{j\neq i,j=1}^{N} x_{ij} - g_I(\tilde{\alpha}_0) + \tilde{\alpha}_0 h_I(\tilde{\alpha}_0)}{h_i(\tilde{\alpha}_0)} \\ \tilde{\alpha}_1 =& \frac{\sum_{i=1}^{N}\sum_{a=1}^{M} y_{ia} - g_A(\tilde{\alpha}_2) + \tilde{\alpha}_2 h_A(\tilde{\alpha}_2)}{h_{ia}(\tilde{\alpha}_2)}\end{aligned}$$

where

$$\begin{aligned}g_I(\tilde{\alpha}_0) =& \sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\left[1 + \frac{\det(\mathbf{I}+4\tilde{\Lambda}_0)^{1/2}}{\exp(\tilde{\alpha}_0)}\exp\Big((\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{u}}_{\boldsymbol{j}})^T(\mathbf{I}+4\tilde{\Lambda}_0)^{-1}(\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{u}}_{\boldsymbol{j}})\Big)\right]^{-1} \\ h_I(\tilde{\alpha}_0) =& \sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N}\left[1 + \frac{\det(\mathbf{I}+4\tilde{\Lambda}_0)^{1/2}}{\exp(\tilde{\alpha}_0)}\exp\Big((\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{u}}_{\boldsymbol{j}})^T(\mathbf{I}+4\tilde{\Lambda}_0)^{-1}(\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{u}}_{\boldsymbol{j}})\Big)\right]^{-1} \\ & \left[1 + \frac{\exp(\tilde{\alpha}_0)}{\det(\mathbf{I}+4\tilde{\Lambda}_0)^{1/2}}\exp\Big(-(\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{u}}_{\boldsymbol{j}})^T(\mathbf{I}+4\tilde{\Lambda}_0)^{-1}(\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{u}}_{\boldsymbol{j}})\Big)\right]^{-1} \\ g_A(\tilde{\alpha}_1) =& \sum_{i=1}^{N}\sum_{a=1}^{M}\left[1 + \frac{\det(\mathbf{I}+2\tilde{\Lambda}_0+2\tilde{\Lambda}_1)^{1/2}}{\exp(\tilde{\alpha}_1)}\exp\Big((\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{v}}_{\boldsymbol{a}})^T(\mathbf{I}+2\tilde{\Lambda}_0+2\tilde{\Lambda}_1)^{-1}(\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{v}}_{\boldsymbol{a}})\Big)\right]^{-1} \\ h_A(\tilde{\alpha}_1) =& \sum_{i=1}^{N}\sum_{a=1}^{M}\left[1 + \frac{\det(\mathbf{I}+2\tilde{\Lambda}_0+2\tilde{\Lambda}_1)^{1/2}}{\exp(\tilde{\alpha}_1)}\exp\Big((\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{v}}_{\boldsymbol{a}})^T(\mathbf{I}+2\tilde{\Lambda}_0+2\tilde{\Lambda}_1)^{-1}(\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{v}}_{\boldsymbol{a}})\Big)\right]^{-1} \\ & \left[1 + \frac{\exp(\tilde{\alpha}_1)}{\det(\mathbf{I}+2\tilde{\Lambda}_0+2\tilde{\Lambda}_1)^{1/2}}\exp\Big(-(\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{v}}_{\boldsymbol{a}})^T(\mathbf{I}+2\tilde{\Lambda}_0+2\tilde{\Lambda}_1)^{-1}(\tilde{\boldsymbol{u}}_{\boldsymbol{i}}-\tilde{\boldsymbol{v}}_{\boldsymbol{a}})\Big)\right]^{-1}\end{aligned}$$

References

Airoldi, E. M., Blei, D. M., Fienberg, S. E., and Xing, E. P. (2008). Mixed membership stochastic blockmodels. *Journal of Machine Learning Research*, 9(Sep):1981–2014.

Attias, H. (1999). Inferring parameters and structure of latent variable models by variational bayes. In *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*, pages 21–30. Morgan Kaufmann Publishers Inc.

Baum, L. E., Petrie, T., Soules, G., and Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. *The annals of mathematical statistics*, 41(1):164–171.

Beal, M. J. et al. (2003). *Variational algorithms for approximate Bayesian inference*. university of London London.

Beal, M. J., Ghahramani, Z., et al. (2006). Variational bayesian learning of directed graphical models with hidden variables. *Bayesian Analysis*, 1(4):793–831.

Blei, D. M., Kucukelbir, A., and McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112(518):859–877.

Celisse, A., Daudin, J. J., and Pierre, L. (2012). Consistency of maximum-likelihood and variational estimators in the stochastic block model. *Electronic Journal of Statistics*, 6:1847–1899.

Daudin, J. J., Picard, F., and Robin, S. (2008). A mixture model for random graphs. *Stat Comput*, 18:173–183.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–38.

Ferrara, E., Interdonato, R., and Tagarelli, A. (2014). Online popularity and topical interests through the lens of instagram. In *Proceedings of the 25th ACM conference on Hypertext and social media*, pages 24–34. ACM.

Friel, N., Rastelli, R., Wyse, J., and Raftery, A. E. (2016). Interlocking directorates in irish companies using a latent space model for bipartite networks. *Proceedings of the National Academy of Sciences*, 113(24):6629–6634.

Gollini, I. and Murphy, T. B. (2016). Joint modeling of multiple network views. *Journal of Computational and Graphical Statistics*, 25(1):246–265.

Henderson, H. V. and Searle, S. R. (1981). On deriving the inverse of a sum of matrices. *Siam Review*, 23(1):53–60.

Jensen, J. L. W. V. (1906). Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta mathematica*, 30(1):175–193.

Jordan, M. I., Ghahramani, Z., Jaakkola, T. S., and Saul, L. K. (1999). An introduction to variational methods for graphical models. *Machine learning*, 37(2):183–233.

Kadushin, C. (1995). Friendship among the french financial elite. *American Sociological Review*, pages 202–221.

Krivitsky, P. N. and Handcock, M. S. (2008). Fitting position latent cluster models for social networks with latentnet. *Journal of statistical software*, 24.

Matias, C. and Miele, V. (2016). Statistical clustering of temporal networks through a dynamic stochastic block model. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*.

Paul, S. and Chen, Y. (2016). Consistent community detection in multi-relational data through restricted multi-layer stochastic blockmodel. *Electronic Journal of Statistics*, 10(2):3807–3870.

Salter-Townshend, M. and Murphy, T. B. (2013). Variational bayesian inference for the latent position cluster model for network data. *Computational Statistics & Data Analysis*, 57(1):661–671.

Xu, K. S., Kliger, M., and Hero Iii, A. O. (2014). Adaptive evolutionary clustering. *Data Mining and Knowledge Discovery*, 28(2):304–336.