Dynamic Bipartite Network: Mathematical Derivations

Ruofan Ma and Qi Liu

September 7, 2023

1 Setup

Let $G_t = (V_{1,t}, V_{2,t}, E_t)$ represent a dynamic bipartite graph observed at time t, where $V_{1\cdot}$ and $V_{2\cdot}$ denote two disjoint families of nodes and $E_\cdot$ represents the set of undirected edges between two nodes of different families. Suppose that (at t) family 1 has $N_{1,t} = |V_{1,t}|$ nodes whereas the number of nodes in family 2 is $N_{2,t} = |V_{2,t}|$. For a pair of nodes p, q, let $z_{pq,t} \in \{1, \ldots, K_1\}$ denote the latent group that node $p \in V_{1,t}$ of family 1 instantiates when interacting with node $q \in V_{2,t}$ whose latent group membership is denoted by $u_{pqt} \in \{1, \ldots, K_2\}$. Further, let $y_{pqt} = 1$ if there exists a directed edge from node p to q for $(p, q) \in E_t$, and $y_{pqt} = 0$ otherwise.

Assume further that the network at time t is in one of M latent states, and that a Markov process governs transitions from one state to the next. We then assume the mixed-membership vectors are generated according to Markov-dependent mixtures with Dirichlet distributions whose concentration parameters are functions of node covariates:

$$\boldsymbol{\pi}_{pt} \mid \boldsymbol{\beta}_1 \sim \sum_{m=1}^{M} \mathbb{P}\left(S_t = m | S_{t-1}\right) \times \text{Dirichlet}\left(\left\{\exp\left(\mathbf{x}_{pt}^\top \boldsymbol{\beta}_{1gm}\right)\right\}_{g=1}^{K_1}\right)$$

$$\boldsymbol{\psi}_{qt} \mid \boldsymbol{\beta}_2 \sim \sum_{m=1}^{M} \mathbb{P}\left(S_t = m | S_{t-1}\right) \times \text{Dirichlet}\left(\left\{\exp\left(\mathbf{x}_{qt}^\top \boldsymbol{\beta}_{2hm}\right)\right\}_{h=1}^{K_2}\right)$$

, where the vector of predictors $\boldsymbol{x}_{it}$ is allowed to vary over time, vectors $\boldsymbol{\beta}_{1gm}$ and $\boldsymbol{\beta}_{2hm}$, indexed by state m in the Markov process, contain regression coefficients associated with the gth and hth groups of vertex families 1 and 2. Further, the random states are generated according to:

$$S_t \mid S_{t-1} \sim \text{Categorical}(\mathrm{A}_n)$$

, where A is the transition matrix. We define a uniform prior over the initial state S_1 and independent symmetric Dirichlet prior distribution for the rows of A. Then, as common in mixed-membership SBMs, we define a categorical sampling model for the dyad-state specific group memberships, $z_{pq,t}$ and $u_{pq,t}$ as:

$$z_{pq,t} \mid \boldsymbol{\pi}_{pt} \sim \text{Categorical}(\boldsymbol{\pi}_{pt}), \quad u_{pq,t} \mid \boldsymbol{\psi}_{qt} \sim \text{Categorical}(\boldsymbol{\psi}_{qt})$$

The model is then completed by defining a $K_1 \times K_2$ blockmodel B, with its B_{gh} element giving the log odds of forming an edge between any two latent group members. Therefore:

$$y_{pqt} \mid z_{pqt}, u_{pqt}, \mathbf{B}, \boldsymbol{\gamma} \overset{\text{indep.}}{\sim} \text{Bernoulli}\left(\text{logit}^{-1}(B_{z_{pqt}, u_{pqt}} + \mathbf{d}_{pqt}^\top \gamma)\right)$$

In sum, the data-generating process can be described as follows:

1. For each time period $t > 1$, draw a historical state $S_t \mid S_{t-1} = n \sim \text{Categorical}(\mathbf{A}_n)$
2. For each node $p \in V_1$ and $q \in V_2$ at time t, draw state-dependent mixed-membership vectors $\boldsymbol{\pi}_{pt} \mid \boldsymbol{\beta}_1, S_t = m \sim \text{Dirichlet}\left(\left\{\exp(\mathbf{x}_{pt}^\top \boldsymbol{\beta}_{1gm})\right\}_{g=1}^{K_1}\right)$ and $\boldsymbol{\psi}_{qt} \mid \boldsymbol{\beta}_2, S_t = m \sim \text{Dirichlet}\left(\left\{\exp(\mathbf{x}_{qt}^\top \boldsymbol{\beta}_{2hm})\right\}_{h=1}^{K_2}\right)$

3. For each pair of nodes $(p, q) \in E_t$ at time t,
   - Sample a group indicator $z_{pq,t} \mid \boldsymbol{\pi}_{pt} \sim \text{Categorical}(\boldsymbol{\pi}_{pt})$
   - Sample a group indicator $u_{pq,t} \mid \boldsymbol{\psi}_{qt} \sim \text{Categorical}(\boldsymbol{\psi}_{qt})$
   - Sample a link between them $y_{pqt} \mid z_{pqt}, u_{pqt}, \mathbf{B}, \boldsymbol{\gamma} \overset{\text{indep.}}{\sim} \text{Bernoulli}\left(\text{logit}^{-1}(B_{z_{pqt},u_{pqt}} + \mathbf{d}_{pqt}^{\top}\gamma)\right)$

Therefore, the DGP gives the full joint distribution of data and latent variables in the model given a set of global hyper-parameters $(\boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{B})$ and covariates $(\mathbf{D}, \mathbf{X})$ as:

$$
\begin{aligned}
f(\mathbf{Y}, \mathbf{Z}, \mathbf{U}, \boldsymbol{\Pi}, \boldsymbol{\Psi}, \mathbf{A} \mid \mathbf{B}, \boldsymbol{\beta}, \boldsymbol{\gamma}) = P(S_1) &\left[\prod_{t=2}^{T} P(S_t \mid S_{t-1}, \mathbf{A})\right] \prod_{m=1}^{M} P(\mathbf{A}_m) \\
&\times \prod_{t=1}^{T} \prod_{p \in V_{1t}} f\left(\boldsymbol{\pi}_{pt} \mid \mathbf{X}_1, \boldsymbol{\beta}_1, S_t\right) \prod_{q \in V_{2t}} f\left(\boldsymbol{\psi}_{qt} \mid \mathbf{X}_2, \boldsymbol{\beta}_2, S_t\right) \\
&\times \prod_{t=1}^{T} \prod_{p,q \in V_{1t} \times V_{2t}} f\left(y_{pqt} \mid z_{pqt}, u_{pqt}, \mathbf{B}, \mathbf{D}, \boldsymbol{\gamma}\right) f\left(z_{pqt} \mid \boldsymbol{\pi}_{pt}\right) f\left(u_{pqt} \mid \boldsymbol{\psi}_{qt}\right)
\end{aligned}
\tag{1}
$$

2 Marginalization

2.1 Marginalizing $\boldsymbol{\Pi}$

Collect and integrate all terms that contain π:

$$
\int \cdots \int \prod_{t=1}^{T} \prod_{p \in V_{1t}} \left[P\left(\boldsymbol{\pi}_{pt} \mid \mathbf{X}, \boldsymbol{\beta}_1, S_t\right)\right] \prod_{qt \in V_{2t}} P\left(\mathbf{z}_{pqt} \mid \boldsymbol{\pi}_{pt}\right) \mathrm{d}\boldsymbol{\pi}_{1t} \ldots \mathrm{d}\boldsymbol{\pi}_{N_{1t}}
$$

Denote $\alpha_{ptgm} = \exp\left(\mathbf{x}_{pt}^{\top}\boldsymbol{\beta}_{1gm}\right)$, and $\xi_{ptm} = \sum_{g=1}^{K_1} \alpha_{pgm}$, then $\boldsymbol{\pi}_{pt} \mid \xi_{ptm} \sim \text{Dir}(\xi_{ptm})$. Therefore, plugging in the PDF for Dirichlet yields:

$$
\prod_{t=1}^{T} \prod_{p \in V_{1t}} \int \prod_{m=1}^{M} \left[\frac{\Gamma\left(\xi_{ptm}\right)}{\prod_{g=1}^{K_1} \Gamma\left(\alpha_{ptgm}\right)} \prod_{g=1}^{K_1} \pi_{ptg}^{\alpha_{ptgm}-1}\right]^{s_{tm}} \prod_{q \in V_{2t}} \prod_{g=1}^{K_1} \pi_{ptg}^{z_{pqt,g}} \, \mathrm{d}\boldsymbol{\pi}_{pt}
$$

, where $z_{pqt,g} = \mathbb{I}(z_{pqt} = g)$. Define $C_{ptg} = \sum_{q \in V_{2t}} z_{pqt,g}$, applying the trick that for indicator function $s_{tm} = \mathbb{I}(S_t = m)$, $\sum_m s_{tm} x = \prod_m x^{s_{tm}}$ and taking the constant terms out of the integral:

$$
\prod_{t=1}^{T} \prod_{p \in V_{1t}} \prod_{m=1}^{M} \left[\frac{\Gamma\left(\xi_{ptm}\right)}{\prod_{g=1}^{K_1} \Gamma\left(\alpha_{ptgm}\right)}\right]^{s_{tm}} \int \prod_{g=1}^{K_1} \pi_{ptg}^{\sum_{m=1}^{M} s_{tm}\alpha_{ptgm} + C_{ptg} - 1} \, \mathrm{d}\boldsymbol{\pi}_{pt}
$$

The integrand can be recognized as the kernel of a Dirichlet distribution. As the integral is over the entire support of this Dirichlet and must integrate to one, we can compute it as the inverse of the corresponding normalizing constant:

$$
\prod_{t=1}^{T} \prod_{p \in V_{1t}} \prod_{m=1}^{M} \left[\frac{\Gamma\left(\xi_{ptm}\right)}{\prod_{g=1}^{K_1} \Gamma\left(\alpha_{ptgm}\right)}\right]^{s_{tm}} \frac{\prod_{g=1}^{K_1} \Gamma\left(\sum_{m=1}^{M} s_{tm}\alpha_{ptgm} + C_{ptg}\right)}{\Gamma\left(\sum_{m=1}^{M} s_{tm}\xi_{ptm} + N_{2t}\right)}
$$

, where N_{2t} is the number of nodes in family 2 at time t. Apply the indicator trick again and rearranging the factorals:

$$
\prod_{t=1}^{T} \prod_{p \in V_{1t}} \prod_{m=1}^{M} \left[\frac{\Gamma\left(\xi_{ptm}\right)}{\Gamma\left(\xi_{ptm} + N_{2t}\right)} \prod_{g=1}^{K_1} \frac{\Gamma\left(\alpha_{ptgm} + C_{ptg}\right)}{\Gamma\left(\alpha_{ptgm}\right)}\right]^{s_{tm}}
\tag{2}
$$

2.2 Marginalizing $\mathbf{\Psi}$

Collect and integrate all terms that contain ψ:

$$\int \cdots \int \prod_{t=1}^{T} \prod_{q \in V_{2t}} \left[P\left(\boldsymbol{\psi}_{qt} \mid \mathbf{X}, \boldsymbol{\beta}_2, S_t\right) \right] \prod_{pt \in V_{1t}} P\left(\mathbf{u}_{pqt} \mid \boldsymbol{\psi}_{qt}\right) \mathrm{d}\boldsymbol{\psi}_{1t} \ldots \mathrm{d}\boldsymbol{\psi}_{N_{2t}}$$

Following a similar strategy as 2.1 yields:

$$\prod_{t=1}^{T} \prod_{q \in V_{2t}} \prod_{m=1}^{M} \left[\frac{\Gamma\left(\xi_{qtm}\right)}{\Gamma\left(\xi_{qtm} + N_{1t}\right)} \prod_{h=1}^{K_2} \frac{\Gamma\left(\alpha_{qthm} + C_{qth}\right)}{\Gamma\left(\alpha_{qthm}\right)} \right]^{s_{tm}} \tag{3}$$

, where $C_{qth} = \sum_{p \in V_{1t}} \mathbb{I}(u_{qp,t} = h)$, $\alpha_{qthm} = \exp\left(\mathbf{x}_{qt}^{\top} \boldsymbol{\beta}_{2hm}\right)$, and $\xi_{qtm} = \sum_{h=1}^{K_2} \alpha_{qhm}$.

2.3 Marginalizing A

Since the transition probabilities have independent Dirichlet priors, and they are conjugate to the multinomial distribution over states at any given time, we can follow a similar strategy when collapsing the rows of $\mathbf{A}$. More specifically, and focusing on the portion of the joint distribution that involves $\mathbf{A}$, we have

$$\begin{aligned} &\int \cdots \int \prod_{t=2}^{T} P\left(s_t \mid s_{t-1}, \mathbf{A}\right) \prod_{m} P\left(\mathbf{A}_m\right) \mathrm{d}\mathbf{A}_1 \cdots \mathrm{d}\mathbf{A}_M = \\ &\qquad \int \cdots \int \prod_{t=2}^{T} \prod_{m} \prod_{n} A_{m,n}^{s_{t,n} \times s_{t-1,m}} \prod_{m} \frac{\Gamma(M\eta)}{\prod_n \Gamma(\eta)} \prod_{n} A_{m,n}^{\eta-1} \, \mathrm{d}\mathbf{A}_1 \cdots \mathrm{d}\mathbf{A}_M \\ &\qquad = \prod_{m} \frac{\Gamma(M\eta)}{\Gamma\left(M\eta + U_{m\cdot}\right)} \prod_{n} \frac{\Gamma\left(\eta + U_{m,n}\right)}{\Gamma(\eta)} \end{aligned} \tag{4}$$

, where $U_{m,n} = \sum_{t=2}^{T} s_{t,n} s_{t-1,m}$ is the number of times the Markov chain transitions from state m to state n, and $U_{m\cdot} = \sum_{t=2}^{T} \sum_{n} s_{t,n} s_{t-1,m}$ is the total number of times the Markov chain transitions from m (potentially to stay at m). η is the hyperprior concentration parameter of a symmetric Dirichlet distribution.

2.4 Marginalized Joint Distribution

Plugging Equations (2) to (4) back into Equation (1), we can get the joint distribution collapsed over the mixed-membership vectors and the transition matrix.

$$\begin{aligned} f(\mathbf{Y}, \mathbf{Z}, \mathbf{U} \mid \mathbf{B}, \boldsymbol{\beta}, \gamma) &= \iiint f(\mathbf{Y}, \mathbf{Z}, \mathbf{U}, \mathbf{\Pi}, \mathbf{\Psi}, \mid \mathbf{B}, \boldsymbol{\beta}, \gamma) d\mathbf{\Pi} d\mathbf{\Psi} d\mathbf{A} \\ &= P(s_1) \left[\prod_{m=1}^{M} \frac{\Gamma(M\eta)}{\Gamma\left(M\eta + U_{m\cdot}\right)} \prod_{n=1}^{M} \frac{\Gamma\left(\eta + U_{m,n}\right)}{\Gamma(\eta)} \right] \\ &\times \prod_{t=2}^{T} \prod_{m=1}^{M} \prod_{p \in V_{1t}} \left[\frac{\Gamma\left(\xi_{ptm}\right)}{\Gamma\left(\xi_{ptm} + N_{2t}\right)} \prod_{g=1}^{K_1} \frac{\Gamma\left(\alpha_{ptgm} + C_{ptg}\right)}{\Gamma\left(\alpha_{ptgm}\right)} \right]^{s_{tm}} \\ &\times \prod_{q \in V_{2t}} \left[\frac{\Gamma\left(\xi_{qtm}\right)}{\Gamma\left(\xi_{qtm} + N_{1t}\right)} \prod_{h=1}^{K_2} \frac{\Gamma\left(\alpha_{qthm} + C_{qth}\right)}{\Gamma\left(\alpha_{qthm}\right)} \right]^{s_{tm}} \\ &\times \prod_{p,q \in V_{1t} \times V_{2t}} \left[\prod_{g=1}^{K_1} \prod_{h=1}^{K_2} \left(\theta_{pqt, z_{pqt}, u_{pqt}}^{y_{pqt}} \left(1 - \theta_{pqt, z_{pqt}, u_{pqt}}\right)^{1 - y_{pqt}} \right)^{z_{pqt,g} \times u_{qpt,h}} \right. \end{aligned} \tag{5}$$

, where $\theta_{pqt,z_{pqt},u_{qpt}} = \text{logit}^{-1}(B_{z_{pqt},u_{pqt}} + \mathbf{d}_{pq}^{\top}\boldsymbol{\gamma})$ is the probability of a tie formation between p in family 1 and q in family 2 at time t

3 Estimation via Variational EM

Define a factorized distribution over the latent variables $\mathbf{L} := \{\mathbf{Z}, \mathbf{U}, \mathbf{S}\}$:

$$\tilde{Q}(\mathbf{L} \mid \boldsymbol{\Phi}, \boldsymbol{\Lambda}, \boldsymbol{\Delta}) = \prod_{t=1}^{T} Q_1(\mathbf{s}_t \mid \boldsymbol{\phi}_t) \prod_{p,q \in V_{1t} \times V_{2t}} Q_2(\mathbf{z}_{pq,t} \mid \boldsymbol{\lambda}_{pq,t}) Q_2(\mathbf{u}_{qp,t} \mid \boldsymbol{\delta}_{qp,t}) \tag{6}$$

, where $\boldsymbol{\phi}_t$, $\boldsymbol{\lambda}_{pq,t}$, and $\boldsymbol{\delta}_{qp.t}$ are variational parameters. We can then find the lower bound for the log marginal probability of the network data $\mathbf{Y}$ by applying Jensen's inequality:

$$P(\mathbf{Y} \mid \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{B}) \geq \mathcal{L} := \mathbb{E}_{\widetilde{Q}}\left[\log P\left(\mathbf{Y}, \mathbf{L} \mid \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{B}\right)\right] - \mathbb{E}_{\widetilde{Q}}\left[\log \widetilde{Q}(\mathbf{L} \mid \boldsymbol{\Phi}, \boldsymbol{\Lambda}, \boldsymbol{\Delta})\right] \tag{7}$$

To approximate the true posterior over the latent variables, we optimize this lower bound by iterating between finding an optimal $\widetilde{Q}$ (the E-step) and optimizing the corresponding lower bound with respect to the hyper-parameters $\mathbf{B}, \boldsymbol{\beta}, \boldsymbol{\gamma}$ (the M-step).

3.1 The E-Steps

3.1.1 E step 1: Z and U

Variational parameters λ_{pqt} and δ_{pqt} are updated by restricting eq. (5) to the terms that only contain $\mathbf{z}_{pqt}$ and $\mathbf{u}_{pqt}$ and taking the logarithm of the resulting expression. First, consider $\mathbf{z}_{pqt}$:

$$\begin{aligned}
&\log P\left(\mathbf{Y}, \mathbf{L} \mid \mathbf{B}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\gamma}, \mathbf{X}_1, \mathbf{X}_2, \mathbf{D}\right) \\
&\quad = z_{pqgt} \sum_{h=1}^{K_2} u_{pqht} \left\{Y_{pqt} \log\left(\theta_{pqght}\right) + \left(1 - Y_{pqt}\right) \log\left(1 - \theta_{pqght}\right)\right\} \\
&\quad + \sum_{m=1}^{M} s_{tm} \log \Gamma\left(\alpha_{pgtm} + C_{pgt}\right) + \text{ const.}
\end{aligned}$$

Note that $C_{pgt} = C'_{pgt} + z_{pqtg}$ and that, for $x \in \{0,1\}, \Gamma(y+x) = y^x \Gamma(y)$. Since the $z_{pqtg} \in \{0,1\}$, we can re-express $\log \Gamma\left(\alpha_{ptmg} + C_{ptg}\right) = z_{pqtk} \log\left(\alpha_{ptmg} + C'_{ptg}\right) + \log \Gamma\left(\alpha_{ptmg} + C'_{ptg}\right)$ and thus simplify the expression to

$$\begin{aligned}
&\log P\left(\mathbf{Y}, \mathbf{L} \mid \mathbf{B}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\gamma}, \mathbf{X}_1, \mathbf{X}_2, \mathbf{D}\right) \\
&\quad = z_{pqgt} \sum_{h=1}^{K_2} u_{pqht} \left\{Y_{pqt} \log\left(\theta_{pqght}\right) + \left(1 - Y_{pqt}\right) \log\left(1 - \theta_{pqght}\right)\right\} \\
&\quad + z_{pqgt} \sum_{m=1}^{M} s_{tm} \log \Gamma\left(\alpha_{pgtm} + C'_{pgt}\right) + \text{ const.}
\end{aligned}$$

We proceed by taking the expectation of $\tilde{Q}(-z)$ under the variational distribution $\tilde{Q}$:

$$\begin{aligned}
&\mathbb{E}_{\tilde{Q}}\left[\log P\left(\mathbf{Y}, \mathbf{L} \mid \mathbf{B}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\gamma}, \mathbf{X}_1, \mathbf{X}_2, \mathbf{D}\right)\right. \\
&\quad = z_{pqgt} \sum_{h=1}^{K_2} \mathbb{E}_{\tilde{Q}_2}\left(u_{pqht}\right) \left\{Y_{pqt} \log\left(\theta_{pqght}\right) + \left(1 - Y_{pqt}\right) \log\left(1 - \theta_{pqght}\right)\right\} \\
&\quad + z_{pqgt} \sum_{m=1}^{M} \mathbb{E}_{\tilde{Q}_1}\left(s_{tm}\right) \log \Gamma\left(\alpha_{pgtm} + C'_{pgt}\right) + \text{ const.}
\end{aligned}$$

The exponential of this expression corresponds to the (unnormalized) parameter vector of a multinomial distribution $\tilde{Q}_2$:

$$\hat{\lambda}_{pqtg} \propto \exp\left[z_{pqgt}\sum_{h=1}^{K_2}\mathbb{E}_{\tilde{Q}_2}(u_{pqht})\left\{Y_{pqt}\log\left(\theta_{pqght}\right)+\left(1-Y_{pqt}\right)\log\left(1-\theta_{pqght}\right)\right\}\right]$$
$$\times\exp\left[z_{pqgt}\sum_{m=1}^{M}\mathbb{E}_{\tilde{Q}_1}(s_{tm})\log\Gamma\left(\alpha_{pgtm}+C'_{pgt}\right)\right]$$

Analogously, the update for $\mathbf{u}_{qp}$ is similarly derived:

$$\begin{aligned}
&\mathbb{E}_{\tilde{Q}}\log P\left(\mathbf{Y},\mathbf{L}\mid\mathbf{B},\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\gamma},\mathbf{X}_1,\mathbf{X}_2,\mathbf{D}\right)\\
&\qquad=u_{pqht}\sum_{g=1}^{K_1}\mathbb{E}_{\tilde{Q}_2}(z_{pqgt})\left\{Y_{pqt}\log\left(\theta_{pqght}\right)+\left(1-Y_{pqt}\right)\log\left(1-\theta_{pqght}\right)\right\}\\
&\qquad+u_{pqht}\sum_{m=1}^{M}\mathbb{E}_{\tilde{Q}_1}(s_{tm})\log\Gamma\left(\alpha_{qhtm}+C'_{qht}\right)+\text{ const.}
\end{aligned}$$

The exponential of this expression corresponds to the (unnormalized) parameter vector of a multinomial distribution $\tilde{Q}_2$:

$$\hat{\delta}_{pqth} \propto \exp\left[u_{pqht}\sum_{g=1}^{K_1}\mathbb{E}_{\tilde{Q}_2}(z_{pqgt})\left\{Y_{pqt}\log\left(\theta_{pqght}\right)+\left(1-Y_{pqt}\right)\log\left(1-\theta_{pqght}\right)\right\}\right]$$
$$\times\exp\left[u_{pqht}\sum_{m=1}^{M}\mathbb{E}_{\tilde{Q}_1}(s_{tm})\log\Gamma\left(\alpha_{qhtm}+C'_{qht}\right)\right]$$

3.1.2 E step 2: S

Similar to the last section, collection all terms in eq. (5) that contain s_{tm} for a specific $t>1$ and m:

$$\begin{aligned}
P(\mathbf{Y},\mathbf{L}\mid\mathbf{B},\boldsymbol{\beta},\boldsymbol{\gamma})&=\Gamma\left(M\eta+U_m\right)^{-1}\prod_{m=1}^{M}\prod_{n=1}^{M}\Gamma\left(\eta+U_{mn}\right)\\
&\times\prod_{p\in V_{1t}}\left[\frac{\Gamma\left(\xi_{ptm}\right)}{\Gamma\left(\xi_{ptm}+N_{2t}\right)}\prod_{g=1}^{K_1}\frac{\Gamma\left(\alpha_{ptgm}+C_{ptg}\right)}{\Gamma\left(\alpha_{ptgm}\right)}\right]^{s_{tm}}\\
&\times\prod_{q\in V_{2t}}\left[\frac{\Gamma\left(\xi_{qtm}\right)}{\Gamma\left(\xi_{qtm}+N_{1t}\right)}\prod_{h=1}^{K_2}\frac{\Gamma\left(\alpha_{qthm}+C_{qth}\right)}{\Gamma\left(\alpha_{qthm}\right)}\right]^{s_{tm}}+\text{ const.}
\end{aligned}$$

Isolating terms that depend on $s_{tm}(n\neq m)$, define

$$\begin{aligned}
U'_m&=U_m-s_{tm}\\
U'_{mm}&=U_{mm}-s_{t-1,m}s_{tm}-s_{tm}s_{t+1,m}\\
U'_{nm}&=U_{nm}-s_{t-1,n}s_{tm}\\
U'_{mn}&=U_{mn}-s_{tm}s_{t+1,n}
\end{aligned}$$

So that $U'_{ab},a,b\in\{m,n\}$ counts the number of times the hidden Markov process transitions from a to b except for when the transition happens into or out of time t. Therefore, separating the case where $m=n$ and $m\neq n$, the first two terms on the right hand side can be written as:

$$\begin{aligned}
&\Gamma\left(M\eta+s_{tm}+U'_m\right)^{-1}\Gamma\left(\eta+s_{t+1,m}s_{tm}+s_{t-1,m}s_{tm}+U'_{mm}\right)\\
&\quad\times\textstyle\prod_{n\neq m}^{M}\Gamma\left(\eta+s_{t+1,n}s_{tm}+U'_{mn}\right)\Gamma\left(\eta+s_{tm}s_{t-1,n}+U'_{nm}\right)
\end{aligned}$$

Recall that for Gamma function, $\Gamma(y+x) = y^x\Gamma(y)$, for $x \in \{0,1\}$, therefore this expression becomes

$$(M\eta + U'_m)^{-s_{tm}}\,\Gamma\left(M\eta + U'_m\right)^{-1}\left\{(\eta + U'_{mm}+1)^{s_{t+1,m}s_{t-1,m}}\,(\eta + U'_{mm})^{s_{t-1,m}-s_{t-1,m}s_{t+1,m}+s_{t+1,m}}\right\}^{s_{tm}}$$
$$\times\Gamma\left(\eta + U'_{mm}\right)\prod_{n\neq m}^{M}\left(\eta + U'_{mn}\right)^{s_{t+1,n}s_{tm}}\Gamma\left(\eta + U'_{mn}\right)\prod_{n\neq m}^{M}\left(\eta + U'_{nm}\right)^{s_{tm}s_{t-1,n}}\Gamma\left(\eta + U'_{nm}\right)$$

To see why

$$\Gamma\left(\eta + s_{t+1,m}s_{tm} + s_{t-1,m}s_{tm} + U'_{mm}\right) = \left\{(\eta + U'_{mm}+1)^{s_{t+1,m}s_{t-1,m}}\,(\eta + U'_{mm})^{s_{t-1,m}-s_{t-1,m}s_{t+1,m}+s_{t+1,m}}\right\}^{s_{tm}}\Gamma(\eta + U'_{mm}) \quad (*)$$

Recall that $s_{tm} = \mathbb{I}(s_t = m) \in \{0,1\}$, and consider in turn the following cases:

1) When $s_{tm} = 0$, equation (*) simplifies to $\Gamma(\eta + U'_{mm}) = 1 \times \Gamma(\eta + U'_{mm})$, for any values of $s_{t-1,m}$and $s_{t+1,m}$
2) When $s_{tm} = 1$,
   i) $s_{t-1,m} = 0$ and $s_{t+1,m} = 0$, equation (*) simplifies to: $\Gamma(\eta + U'_{mm}) = (1\times 1)^1\Gamma(\eta + U'_{mm})$
   ii) $s_{t-1,m} = 1$ and $s_{t+1,m} = 0$, equation (*) becomes (recall the recursive property of the Gamma function):
   $$\Gamma(\eta + U'_{mm}+1) = \left(1\times(\eta + U'_{mm}+1)\right)^1\Gamma(\eta + U'_{mm})$$
   iii) $s_{t-1,m} = 1$ and $s_{t+1,m} = 0$. This is analgous to ii)
   iv) $s_{t-1,m} = 1$ and $s_{t+1,m} = 1$, equation (*) becomes (applying the recursive property twice):

$$\begin{aligned}\Gamma(\eta + U'_{mm}+1+1) &= (\eta + U'_{mm}+1)\Gamma(\eta + U'_{mm}+1)\\ &= (\eta + U'_{mm}+1)(\eta + U'_{mm})\Gamma(\eta + U'_{mm})\\ &= \left[(\eta + U'_{mm}+1)^1(\eta + U'_{mm})^1\right]^1\Gamma(\eta + U'_{mm})\end{aligned}$$

Again focus on the terms that are specific to a t and m,

$$\begin{aligned}&P(\mathbf{Y},\mathbf{L}\mid\mathbf{B},\boldsymbol{\beta},\boldsymbol{\gamma})\\ &= (M\eta + U'_m)^{-s_{tm}}\left\{(\eta + U'_{mm}+1)^{s_{t+1,m}s_{t-1,m}}\,(\eta + U'_{mm})^{s_{t-1,m}-s_{t-1,m}s_{t+1,m}+s_{t+1,m}}\right\}^{s_{tm}}\\ &\times\prod_{n\neq m}^{M}\left(\eta + U'_{mn}\right)^{s_{t+1,n}s_{tm}}\left(\eta + U'_{nm}\right)^{s_{tm}s_{t-1,n}}\\ &\times\prod_{p\in V_{1t}}\left[\frac{\Gamma\left(\xi_{ptm}\right)}{\Gamma\left(\xi_{ptm}+N_{2t}\right)}\prod_{g=1}^{K_1}\frac{\Gamma\left(\alpha_{ptgm}+C_{ptg}\right)}{\Gamma\left(\alpha_{ptgm}\right)}\right]^{s_{tm}}\\ &\times\prod_{q\in V_{2t}}\left[\frac{\Gamma\left(\xi_{qtm}\right)}{\Gamma\left(\xi_{qtm}+N_{1t}\right)}\prod_{h=1}^{K_2}\frac{\Gamma\left(\alpha_{qthm}+C_{qth}\right)}{\Gamma\left(\alpha_{qthm}\right)}\right]^{s_{tm}} + \text{ const.}\end{aligned}$$

Taking log and expectation under $\tilde{Q}$ w.r.t. variables do not contain s_{tm}:

$$
\begin{aligned}
\log \hat{\phi}_{tm} &= -s_{tm}\mathbb{E}_{\tilde{Q}_1}\left[\log\left(M\eta + U'_m\right)\right] + s_{tm}\phi_{t+1,m}\phi_{t-1,m}\mathbb{E}_{\tilde{Q}_1}\left[\log\left(\eta + U'_{mm} + 1\right)\right] \\
&\quad + s_{tm}\left(\phi_{t-1,m} - \phi_{t-1,m}\phi_{t+1,m} + \phi_{t+1,m}\right)\mathbb{E}_{\tilde{Q}_1}\left[\log\left(\eta + U'_{mm}\right)\right] \\
&\quad + s_{tm}\sum_{n\neq m}^{M}\phi_{t+1,n}\mathbb{E}_{\tilde{Q}_1}\left[\log\left(\eta + U'_{mn}\right)\right] + s_{tm}\sum_{n\neq m}^{M}\phi_{t-1,n}\mathbb{E}_{\tilde{Q}_1}\left[\log\left(\eta + U'_{nm}\right)\right] \\
&\quad + s_{tm}\sum_{p\in V_{1t}}\left[\frac{\Gamma\left(\xi_{ptm}\right)}{\Gamma\left(\xi_{ptm} + N_{2t}\right)}\right] + s_{tm}\sum_{p\in V_{t1}}\sum_{g=1}^{K_1}\mathbb{E}_{\tilde{Q}_2}\left[\log\left[\frac{\Gamma\left(\alpha_{ptmg} + C_{ptg}\right)}{\Gamma\left(\alpha_{ptmg}\right)}\right]\right] \\
&\quad + s_{tm}\sum_{q\in V_{2t}}\left[\frac{\Gamma\left(\xi_{qtm}\right)}{\Gamma\left(\xi_{qtm} + N_{1t}\right)}\right] + s_{tm}\sum_{q\in V_{t2}}\sum_{h=1}^{K_2}\mathbb{E}_{\tilde{Q}_2}\left[\log\left[\frac{\Gamma\left(\alpha_{qtmh} + C_{qth}\right)}{\Gamma\left(\alpha_{qtmh}\right)}\right]\right] + \text{ const}
\end{aligned}
$$

So that the mth element of the parameter vector for $\widetilde{Q}_1(s_t \mid \phi_{tm})$ is (so we could treat the expectations w.r.t. to $\tilde{Q}_2$ as constant):

$$
\begin{aligned}
\hat{\phi}_{tm} &\propto \exp\left[-\mathbb{E}_{\tilde{Q}_1}\left[\log\left(M\eta + U'_m\right)\right]\right]\exp\left[\phi_{t+1,m}\phi_{t-1,m}\mathbb{E}_{\tilde{Q}_1}\left[\log\left(\eta + U'_{mm} + 1\right)\right]\right] \\
&\quad \times \exp\left[\left(\phi_{t-1,m} - \phi_{t-1,m}\phi_{t+1,m} + \phi_{t+1,m}\right)\mathbb{E}_{\tilde{Q}_1}\left[\log\left(\eta + U'_{mm}\right)\right]\right] \\
&\quad \times \prod_{n\neq m}\exp\left[\phi_{t+1,n}\mathbb{E}_{\tilde{Q}_1}\left[\log\left(\eta + U'_{mn}\right)\right]\right]\exp\left[\phi_{t-1,n}\mathbb{E}_{\tilde{Q}_1}\left[\log\left(\eta + U'_{nm}\right)\right]\right] \\
&\quad \times \prod_{p\in V_{1t}}\left[\frac{\Gamma\left(\xi_{ptm}\right)}{\Gamma\left(\xi_{ptm} + N_{2t}\right)}\prod_{g=1}^{K_1}\frac{\Gamma\left(\alpha_{ptgm} + C_{ptg}\right)}{\Gamma\left(\alpha_{ptgm}\right)}\right] \\
&\quad \times \prod_{q\in V_{2t}}\left[\frac{\Gamma\left(\xi_{qtm}\right)}{\Gamma\left(\xi_{qtm} + N_{1t}\right)}\prod_{h=1}^{K_2}\frac{\Gamma\left(\alpha_{qthm} + C_{qth}\right)}{\Gamma\left(\alpha_{qthm}\right)}\right]
\end{aligned}
$$

3.2 The M-Steps

3.2.1 The Lower Bound

The full expression of the lower bound can be written as:

$$
\begin{aligned}
\mathcal{L}(\widetilde{Q}) &= \mathbb{E}_{\widetilde{Q}}\left[\log P\left(\mathbf{Y}, \mathbf{L} \mid \boldsymbol{\beta_1}, \boldsymbol{\beta}_2, \boldsymbol{\gamma}, \mathbf{B}, \mathbf{X}_1, \mathbf{X}_2, \mathbf{D}\right)\right] - \mathbb{E}_{\widetilde{Q}}\left[\log \widetilde{Q}(\mathbf{L} \mid \boldsymbol{\Phi}, \boldsymbol{\Lambda}, \boldsymbol{\Delta})\right] \\
&= \log\left(P\left(s_1\right)\right) + \log\Gamma(M\eta) - \sum_m \mathbb{E}_{\tilde{Q}}\left[\log\Gamma\left(M\eta + U_{m\cdot}\right)\right] + \sum_{m,n} \mathbb{E}_{\tilde{Q}}\left[\log\Gamma\left(\eta + U_{m,n}\right)\right] - \log\Gamma(\eta) \\
&+ \sum_{t,m} \phi_{tm} \sum_{p \in V_{1,t}} \left[\Gamma(\xi_{ptm}) - \Gamma(\xi_{ptm} + N_{2t})\right] + \sum_{t,m} \phi_{tm} \sum_{p \in V_{1,t}} \sum_{g=1}^{K_1} \left[\mathbb{E}_{\tilde{Q}}\left[\log\Gamma(\alpha_{ptgm} + C_{ptg})\right] - \log\Gamma(\alpha_{ptgm})\right] \\
&+ \sum_{t,m} \phi_{tm} \sum_{q \in V_{2,t}} \left[\Gamma(\xi_{qtm}) - \Gamma(\xi_{qtm} + N_{1t})\right] + \sum_{t,m} \phi_{tm} \sum_{q \in V_{2,t}} \sum_{h=1}^{K_2} \left[\mathbb{E}_{\tilde{Q}}\left[\log\Gamma(\alpha_{qtgm} + C_{qth})\right] - \log\Gamma(\alpha_{qthm})\right] \\
&+ \sum_t \sum_{(p,q) \in V_{1t} \times V_{2t}} \sum_{g=1}^{K_1} \sum_{h=1}^{K_2} \lambda_{pq,g} \delta_{pq,h} \left\{y_{pqt} \log\theta_{pqght} + (1 - y_{pqt}) \log\left(1 - \theta_{pqght}\right)\right\} \\
&- \sum_{g=1}^{K_1} \sum_{h=1}^{K_2} \frac{\left(B_{gh} - \mu_{gh}\right)^2}{2\sigma_{gh}^2} - \sum_{j=1}^{J_d} \frac{\left(\gamma_j - \mu_\gamma\right)^2}{2\sigma_\gamma^2} - \sum_{g=1}^{K_1} \sum_{j=1}^{J_{1x}} \sum_{m=1}^{M} \frac{\left(\beta_{1gjm} - \mu_{\beta_1}\right)^2}{2\sigma_{\beta_1}^2} - \sum_{h=1}^{K_2} \sum_{j=1}^{J_{2x}} \sum_{m=1}^{M} \frac{\left(\beta_{2hjm} - \mu_{\beta_2}\right)^2}{2\sigma_{\beta_2}^2} \\
&- \sum_{t,m} \phi_{tm} \log(\phi_{t,m}) - \sum_{(p,q) \in V_{1t} \times V_{2t}} \sum_{g=1}^{K_1} \sum_{h=1}^{K_2} \left\{\lambda_{pqgt} \log(\lambda_{pqgt}) - \delta_{qpht} \log\left(\delta_{qpht}\right)\right\}
\end{aligned}
\tag{8}
$$

3.2.2 M step 1: B

Collect the terms that contain B_{gh} in the lower bound:

$$
\begin{aligned}
\mathcal{L}(\widetilde{Q}) = &\sum_{t=1}^{T} \sum_{p,q \in V_{1t} \times V_{2t}} \sum_{g,h=1}^{K} \lambda_{pqtg} \delta_{qpth} \left\{y_{pqt} \log\theta_{pqtgh} + (1 - y_{pqt}) \log\left(1 - \theta_{pqtgh}\right)\right\} \\
&- \sum_{g=1}^{K_1} \sum_{h=1}^{K_2} \frac{\left(B_{gh} - \mu_{gh}\right)^2}{2\sigma_{gh}^2} + \text{ const.}
\end{aligned}
$$

We optimize this lower bound with respect to $\mathbf{B}_{gh}$ using a gradient-based numerical optimization method. The corresponding gradient is given by,

$$
\frac{\partial \mathcal{L}_{B_{gh}}}{\partial B_{gh}} = \sum_{t=1}^{T} \sum_{p,q \in V_{1t} \times V_{2t}} \lambda_{pqtg} \delta_{qpth} \left(y_{pqt} - \theta_{pqtgh}\right) - \frac{B_{gh} - \mu_{B_{gh}}}{\sigma_{B_{gh}}^2}
$$

3.2.3 M step 2: γ

Collect the terms that contain $\boldsymbol{\gamma}$ in the lower bound (note that $\theta_{pqtgh} = \text{logit}^{-1}(B_{z_{pqt}, u_{pqt}} + \mathbf{d}_{pq}^\top \boldsymbol{\gamma})$ is also a function of $\boldsymbol{\gamma}$), J_d is the number of dyadic covariates:

$$
\begin{aligned}
\mathcal{L}(\tilde{Q}) = &\sum_{t=1}^{T} \sum_{p,q \in V_{1t} \times V_{2t}} \sum_{g=1}^{K_1} \sum_{h=1}^{K_2} \lambda_{pqtg} \delta_{qpth} \left\{y_{pqt} \log\theta_{pqtgh} + (1 - y_{pqt}) \log\left(1 - \theta_{pqtgh}\right)\right\} \\
&- \sum_{j}^{J_d} \frac{\left(\gamma_j - \mu_\gamma\right)^2}{2\sigma_\gamma^2} + \text{ const.}
\end{aligned}
$$

Similarly, we use a numerical optimization algorithm based on the following gradient to optimize this expression with respect to γ_j (the jth element of the $\boldsymbol{\gamma}$ vector). The corresponding gradient is given by,

$$\frac{\partial \mathcal{L}_{\gamma_j}}{\partial \gamma_j} = \sum_{t=1}^{T} \sum_{p,q \in V_{1t} \times V_{2t}} \sum_{g=1}^{K_1} \sum_{h=1}^{K_2} \lambda_{pqtg} \delta_{qpth} \mathbf{d}_{pqtj}^{\top} (y_{pqt} - \theta_{pqtgh}) - \frac{\gamma_j - \mu_\gamma}{\sigma_\gamma^2}$$

3.2.4 M step 3: $\boldsymbol{\beta_{1m}}$ and $\boldsymbol{\beta_{2m}}$

First, collect all terms that contain $\boldsymbol{\beta}_{1gm}$ and roll the rest of the terms into a constant. J_{1x} is the number of monadic covariates for family 1, and J_{2x} is the number of monadic covariates for family 2:

$$\begin{aligned}
\mathcal{L}(\widetilde{Q}) = & \sum_{t=1}^{T} \sum_{m=1}^{M} \phi_{tm} \sum_{p \in V_{1t}} \left[\log \Gamma\left(\xi_{ptm}\right) - \log \Gamma\left(\xi_{ptm} + N_{2t}\right)\right] \\
& + \sum_{t,m} \phi_{tm} \sum_{p \in V_{1t}} \sum_{g=1}^{K_1} \left[\mathbb{E}_{\widetilde{Q}_2}\left[\log \Gamma\left(\alpha_{pgtm} + C_{ptg}\right)\right] - \log \Gamma\left(\alpha_{pgtm}\right)\right] \\
& - \sum_{g=1}^{K_1} \sum_{j=1}^{J_{1x}} \sum_{m=1}^{M} \frac{\left(\beta_{1gjm} - \mu_{\beta_1}\right)^2}{2\sigma_{\beta_1}^2} + \text{ const.}
\end{aligned}$$

No closed-form solution exists for an optimum with respect to β_{1mgj}, but a gradient-based algorithm can be implemented to maximize the above expression. The corresponding gradient with respect to each element in vector $\boldsymbol{\beta}_{1gm}$ is given by:

$$\begin{aligned}
\frac{\partial \mathcal{L}(\tilde{Q})}{\partial \beta_{1mgj}} = & \sum_{t=1}^{T} \phi_{tm} \sum_{p \in V_{1t}} \alpha_{ptmg} x_{1ptj} \left(\mathbb{E}_{\tilde{Q}_2}\left[\breve{\psi}\left(\alpha_{ptmg} + C_{ptg}\right) - \breve{\psi}\left(\alpha_{ptgm}\right)\right]\right. \\
& \left. + \left[\breve{\psi}\left(\xi_{ptm}\right) - \breve{\psi}\left(\xi_{ptm} + N_{2t}\right)\right]\right) - \frac{\beta_{1mgj} - \mu_{\beta_1}}{\sigma_{\beta_1}^2}
\end{aligned}$$

Here, $\breve{\psi}$ is the digamma function. Similarly for $\boldsymbol{\beta_{2hm}}$, collect all the relevant terms yield:

$$\begin{aligned}
\mathcal{L}(\widetilde{Q}) = & \sum_{t=1}^{T} \sum_{m=1}^{M} \phi_{tm} \sum_{q \in V_{2t}} \left[\log \Gamma\left(\xi_{qtm}\right) - \log \Gamma\left(\xi_{qtm} + N_{1t}\right)\right] \\
& + \sum_{q \in V_{2t}} \sum_{h=1}^{K_2} \left[\mathbb{E}_{\widetilde{Q}_2}\left[\log \Gamma\left(\alpha_{qhtm} + C_{qth}\right)\right] - \log \Gamma\left(\alpha_{qhtm}\right)\right] \\
& - \sum_{h=1}^{K_2} \sum_{j=1}^{J_{2x}} \sum_{m=1}^{M} \frac{\left(\beta_{2hjm} - \mu_{\beta_2}\right)^2}{2\sigma_{\beta_2}^2} + \text{ const.}
\end{aligned}$$

With the corresponding gradient:

$$\begin{aligned}
\frac{\partial \mathcal{L}(\tilde{Q})}{\partial \beta_{2mhj}} = & \sum_{t=1}^{T} \phi_{tm} \sum_{q \in V_{2t}} \alpha_{qtmh} x_{2ptj} \left(\mathbb{E}_{\tilde{Q}_2}\left[\breve{\psi}\left(\alpha_{qtmh} + C_{qth}\right) - \breve{\psi}\left(\alpha_{qthm}\right)\right]\right. \\
& \left. + \left[\breve{\psi}\left(\xi_{qtm}\right) - \breve{\psi}\left(\xi_{qtm} + N_{1t}\right)\right]\right) - \frac{\beta_{2mhj} - \mu_{\beta_2}}{\sigma_{\beta_2}^2}
\end{aligned}$$

3.3 Standard Error Computation

3.3.1 Hessian for $\boldsymbol{\gamma}$

Restricted to terms that involve γ, we have shown that

$$\frac{\partial \mathcal{L}(\tilde{Q})}{\partial \gamma_j} = \sum_{t=1}^{T} \sum_{p,q \in V_{1t} \times V_{2t}} \sum_{g=1}^{K_1} \sum_{h=1}^{K_2} \lambda_{pqtg} \delta_{qpth} \mathbf{d}_{pqtj}^{\top} \left(y_{pqt} - \theta_{pqtgh} \right) - \frac{\gamma_j - \mu_\gamma}{\sigma_\gamma^2}$$

Then,

$$\frac{\partial^2 \mathcal{L}(\tilde{Q})}{\partial \gamma_j \partial \gamma_{j'}} = \sum_{t=1}^{T} \sum_{p,q \in V_{1t} \times V_{2t}} \sum_{g=1}^{K_1} \sum_{h=1}^{K_2} -\mathbf{d}_{pqtj}^{\top} \mathbf{d}_{pqtj\prime} \left[\bar{\theta}_{pqtgh} (1 - \bar{\theta}_{pqtgh}) \right] - \sigma_\gamma^{-2} \delta_{jj\prime}$$

Here, $\delta_{jj\prime}$ is the Kronecker delta function, and

$$\bar{\theta}_{pqtgh} = \mathbb{E}_{\widetilde{Q}} \left[\theta_{pqtgh} \right] = \hat{\lambda}_{pqtg}^{\top} \hat{\mathbf{B}} \hat{\delta}_{qpth} + \mathbf{d}_{pqt}^{\top} \boldsymbol{\gamma}$$

is a closed-form solution to the expectation over $\widetilde{Q}$.

3.3.2 Hessian for $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$

First, we focus on family 1 coefficients, which is $\boldsymbol{\beta_1}$. For coefficients in the same group g:

$$\begin{aligned}\frac{\partial \mathcal{L}(\tilde{Q})}{\partial \beta_{1mgj}} = \sum_{t=1}^{T} \phi_{tm} \sum_{p \in V_{1t}} \alpha_{ptmg} x_{1ptj} \Big(& \mathbb{E}_{\tilde{Q}_2} \left[\breve{\psi}\left(\alpha_{ptmg} + C_{ptg} \right) - \breve{\psi}\left(\alpha_{ptgm} \right) \right] \\ & + \left[\breve{\psi}\left(\xi_{ptm} \right) - \breve{\psi}\left(\xi_{ptm} + N_{2t} \right) \right] \Big) - \frac{\beta_{1mgj} - \mu_{\beta_1}}{\sigma_{\beta_1}^2}\end{aligned}$$

So,

$$\begin{aligned}\frac{\partial^2 \mathcal{L}(\tilde{Q})}{\partial \beta_{1mgj} \partial \beta_{1mgj\prime}} = \sum_{t=1}^{T} \phi_{tm} \sum_{p \in V_{1t}} & x_{1ptj} x_{1ptj\prime} \alpha_{ptmg} \Big\{ \mathbb{E}_{\tilde{Q}_2} \left[\breve{\psi}\left(\alpha_{ptmg} + C_{ptg} \right) - \breve{\psi}\left(\alpha_{ptgm} \right) \right] \\ & + \breve{\psi}\left(\xi_{ptm} \right) - \breve{\psi}\left(\xi_{ptm} + N_{2t} \right) \\ & + \alpha_{ptmg} \left[\mathbb{E}_{\tilde{Q}_2} \left[\breve{\psi}_1\left(\alpha_{ptmg} + C_{ptg} \right) - \breve{\psi}_1\left(\alpha_{ptgm} \right) \right] + \breve{\psi}_1\left(\xi_{ptm} \right) - \breve{\psi}_1\left(\xi_{ptm} + N_{2t} \right) \right] \Big\}\end{aligned}$$

Here, $\breve{\psi}_1$ is the trigamma function. For coefficients in different latent groups g and $g\prime$,

$$\frac{\partial^2 \mathcal{L}(\tilde{Q})}{\partial \beta_{1mgj} \partial \beta_{1mg\prime j\prime}} = \sum_{t=1}^{T} \phi_{tm} \sum_{p \in V_{1t}} x_{1ptj} x_{1ptj\prime} \alpha_{ptmg} \alpha_{ptmg\prime} \left(\breve{\psi}_1\left(\xi_{ptm} \right) - \breve{\psi}_1\left(\xi_{ptm} + N_{2t} \right) \right)$$

The Hessian for $\boldsymbol{\beta_2}$ can be derived similarly.

Unlike $\boldsymbol{\gamma}$, there are no closed-form solutions for the expectations involved in the Hessian for $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$. To approximate them, we take S samples from the Poisson-Binomial distribution of C_{ptg}, and we get $C_{ptg}^{(s)}$ ($s \in 1...S$), and let

$$\mathbb{E}_{\tilde{Q}_2} \left[\breve{\psi}\left(\alpha_{ptmg} + C_{ptg} \right) \right] \approx \frac{1}{S} \sum_{S} \left(\breve{\psi}\left(\alpha_{ptmg} + C_{ptg}^{(s)} \right) \right)$$

$$\mathbb{E}_{\tilde{Q}_2} \left[\breve{\psi}_1\left(\alpha_{ptmg} + C_{ptg} \right) \right] \approx \frac{1}{S} \sum_{S} \left(\breve{\psi}_1\left(\alpha_{ptmg} + C_{ptg}^{(s)} \right) \right)$$