SUPPLEMENTARY MATERIAL of "The Hierarchical Stochastic Block Model for Multiple Networks"

A Details on the MCMC sampler

The MCMC sampler described in Section 4.1 is a Gibbs Sampler based on the Restaurant Franchise representation of the the HPY, equations (1) and (2) in the main article. In principle, in the Gibbs sampler, we re-sample each classification variable $Z_{s,w}$, $s = 1, \dots, S$, $w = 1, \dots, n_{s\cdot\cdot}$, from their full conditional distribution $\pi(Z_{s,w}|\mathbf{A}, \mathbf{Z}^{-Z_{s,w}})$. However, as in the sampler of Teh (2006) for the HDP, instead of sampling directly $Z_{s,w}|\mathbf{A}, \mathbf{Z}^{-Z_{s,w}}$, we are going to re-sample the within graph vertex to cluster allocations $t_{s,w}$, and the cluster to block allocations $k_{s,t}$. Given $t_{s,w}$ and $k_{s,t}$, and the known labels $(Z_1^{**}, \dots, Z_{K_n}^{**})$, $Z_{s,w}$ can then computed as $Z_{s,w} = Z^{**}_{k_{s,t_{s,w}}}$.

A schematic description of the sampler is provided in Algorithm 1. Initialization of $\{t_{s,w} : s = 1, \dots, S;\ w = 1, \dots, n_{s\cdot\cdot}\}$ and $\{k_{s,t} : s = 1, \dots, S;\ t = 1, \dots, m_{s\cdot}\}$ can be done by drawing values from the prior, i.e. by sampling the variables from a Chinese Restaurant Franchise (1) and (2). Even though the prior clustering allocation from the prior can have much fewer components than the posterior allocation, in our experiments the algorithm seems to converge quickly to its stationary distribution with a much higher number of components.

In the next three subsections, we provide additional details to Section 4.1 on how to update: 1-2) $t_{s,w}$ and $k_{s,t}$ from their full conditional distributions; 3) the hyperparameters of the model.

Algorithm: HSMB - MCMC

Initialise $\sigma_s, \theta_s, \alpha, \gamma, a, b$;

Initialize $t_{s,w}, k_{s,t}$ using the CRF;

for *t in 1:iterations* **do**

  for *s in 1:S* **do**

    Update cluster allocations:

    for *w in* $1 : n_{s\cdot\cdot}$ **do**

      Sample $t_{s,w}$ from $\pi(t_{s,w}|\text{rest})$;

    end

    Update block allocations:

    for *t in 1:*$m_{s\cdot}$ **do**

      Sample $k_{s,t}$ from $\pi(k_{s,t}|\text{rest})$;

    end

    Update Local hyperparameters:

    Sample σ_s, θ_s from $\pi(\sigma_s, \theta_s|\text{rest})$;

  end

  Update Global hyperparameters:

  Sample α, γ from $\pi(\alpha, \gamma|\text{rest})$;

  Sample a, b from $\pi(a, b|\text{rest})$;

  Store all $Z_{s,w} = Z^{**}_{k_{s,t_{s,w}}}$;

end

Update of $t_{s,w}$

The update of $t_{s,w}$ from $\pi(t_{s,w}|\text{rest})$ can be performed by following the four steps described in Section 4.1.

For each $t_{s,w}$, the algorithm first computes $\mathbf{E}_{(s)}^{-t_{s,w}}$ and $\mathbf{N}_{(s)}^{-t_{s,w}}$, i.e. the matrices $\mathbf{E}_{(s)}$ and $\mathbf{N}_{(s)}$ computed without using $t_{s,w}$. Then, for each possible block component $k' \in \{1, \ldots, K_n^{-s,w} + 1\}$, the algorithm needs to compute the likelihood probability of vertex w from graph s being associated with block k'. This is

$$p_{k'} = \frac{\pi(\mathbf{A}|\mathbf{Z}^{-t_{s,w}}, Z_{s,w} = Z_{k'}^{**})}{\pi(\mathbf{A}|\mathbf{Z}^{-t_{s,w}})} \tag{1}$$

where $\pi(\mathbf{A}|\mathbf{Z})$ is the collapsed likelihood (6) in the main article, computed using $\mathbf{E}_{(s)}^{-t_{s,w}}$ and $\mathbf{N}_{(s)}^{-t_{s,w}}$ at the denominator, while using $\mathbf{E}_{(s)}$ and $\mathbf{N}_{(s)}$ recalculated after assigning $Z_{s,w}$ to block k' at the numerator.

For $k' \in \{1, \ldots, K_n^{-t_{s,w}}\}$, (1) can be computed as follows

$$\begin{aligned}
p_{k'} &= \frac{\pi(\mathbf{A}|\mathbf{Z}^{-Z_{s,w}}, Z_{s,w} = k')}{\pi(\mathbf{A}|\mathbf{Z}^{-Z_{s,w}})} \\
&= \frac{(\text{Beta}(a,b))^{-\frac{K_n(K_n+1)}{2}} \prod_{s=1}^{S} \prod_{i=1}^{K_n} \prod_{j=i}^{K_n} \text{Beta}(E_{s,i,j} + a, N_{s,i,j} + b)}{(\text{Beta}(a,b))^{-\frac{K_n^{-t_{s,w}}(K_n^{-t_{s,w}}+1)}{2}} \prod_{s=1}^{S} \prod_{i=1}^{K_n^{-t_{s,w}}} \prod_{j=i}^{K_n^{-t_{s,w}}} \text{Beta}(E_{s,i,j}^{-t_{s,w}} + a, N_{s,i,j}^{-t_{s,w}} + b)} \\
&= \frac{\prod_{i=1}^{K_n} \prod_{j=i}^{K_n} \text{Beta}(E_{s,i,j} + a, N_{s,i,j} + b)}{\prod_{i=1}^{K_n} \prod_{j=i}^{K_n} \text{Beta}(E_{s,i,j}^{-ts,w} + a, N_{s,i,j}^{-ts,w} + b)} \\
&= \frac{\prod_{i=1}^{k'} \text{Beta}(E_{s,i,k'} + a, N_{s,i,k'} + b) \prod_{j=k'+1}^{K_n} \text{Beta}(E_{s,k',j} + a, N_{s,k',j} + b)}{\prod_{i=1}^{k'} \text{Beta}(E_{s,i,k'}^{-ts,w} + a, N_{s,i,k'}^{-ts,w} + b) \prod_{j=k'+1}^{K_n} \text{Beta}(E_{s,k',j}^{-ts,w} + a, N_{s,k',j}^{-ts,w} + b)} \\
&= \frac{\prod_{j=1}^{K_n} \text{Beta}(E_{s,k',j} + a, N_{s,k',j} + b)}{\prod_{j=1}^{K_n} \text{Beta}(E_{s,k',j}^{-ts,w} + a, N_{s,k',j}^{-ts,w} + b)} \\
&= \prod_{j=1}^{K_n} \left(\frac{\text{Beta}(E_{s,k',j} + a, N_{s,k',j} + b)}{\text{Beta}(E_{s,k',j}^{-ts,w} + a, N_{s,k',j}^{-ts,w} + b)} \right)
\end{aligned}$$

where in the third equality, we have used the fact that $K_n^{-t_{s,w}} = K_n$ (since we are considering $k' \in \{1, \ldots, K_n^{-t_{s,w}}\}$), and in the last equality, we have used the fact that $E_{(s)}$ and $N_{(s)}$ are both symmetric.

For $k' = K_n^{-s,w} + 1$, (1) becomes,

$$\begin{aligned}
p_{k'} &= \frac{\pi(\mathbf{A}|\mathbf{Z}^{-t_{s,w}}, Z_{s,w} = Z_{k'}^{**})}{\pi(\mathbf{A}|\mathbf{Z}^{-ts,w})} \\
&= \frac{(\text{Beta}(a,b))^{-\frac{K_n(K_n+1)}{2}} \prod_{s=1}^{S} \prod_{i=1}^{K_n} \prod_{j=i}^{K_n} \text{Beta}(E_{s,i,j} + a, N_{s,i,j} + b)}{(\text{Beta}(a,b))^{-\frac{K_n^{-ts,w}(K_n^{-ts,w}+1)}{2}} \prod_{s=1}^{S} \prod_{i=1}^{K_n^{-ts,w}} \prod_{j=i}^{K_n^{-ts,w}} \text{Beta}(E_{s,i,j}^{-ts,w} + a, N_{s,i,j}^{-t_{s,w}} + b)} \\
&= \frac{(\text{Beta}(a,b))^{-\frac{(K_n^{-t_{s,w}}+1)(K_n^{-t_{s,w}}+2)}{2}} \prod_{j=1}^{K_n^{-t_{s,w}}+1} \text{Beta}(E_{s,K_n^{-t_{s,w}}+1,j} + a, N_{s,K_n^{-t_{s,w}}+1,j} + b)}{(\text{Beta}(a,b))^{-\frac{K_n^{-t_{s,w}}(K_n^{-t_{s,w}}+1)}{2}}} \\
&= \frac{\prod_{j=1}^{K_n^{-t_{s,w}}+1} \text{Beta}(E_{s,K_n^{-t_{s,w}}+1,j} + a, N_{s,K_n^{-t_{s,w}}+1,j} + b)}{\text{Beta}(a,b)^{K_n^{-t_{s,w}}+1}} \\
&= \prod_{j=1}^{K_n^{-t_{s,w}}} \left(\frac{\text{Beta}(E_{s,K_n^{-t_{s,w}}+1,j} + a, N_{s,K_n^{-t_{s,w}}+1,j} + b)}{\text{Beta}(a,b)} \right)
\end{aligned}$$

where the third equality follow from the fact that $K_n = K_n^{t_{s,w}} + 1$, and the last equality follow from the fact that $E_{s,K_n^{-s,w}+1,K_n^{-s,w}+1} = N_{s,K_n^{-s,w}+1,K_n^{-s,w}+1} = 0$, since there is only one vertex assigned to this block and we do not allow self-loops.

By setting, $E^{-t_{s,w}}_{s,K_n^{-t_{s,w}}+1,j} = N^{-t_{s,w}}_{s,K_n^{-t_{s,w}}+1,j} = 0$ (i.e. adding an additional 0 row to $\mathbf{E}_{(s)}^{-t_{s,w}}$ and $\mathbf{N}_{(s)}^{-t_{s,w}}$), formula (1) can be written for all $k' \in \{1, \ldots, K_n^{-t_{s,w}} + 1\}$ as

$$p_{k'} = \prod_{j=1}^{K_n^{-t_{s,w}}} \left(\frac{\text{Beta}(E_{s,k',j} + a, N_{s,k',j} + b)}{\text{Beta}(E_{s,k',j}^{-t_{s,w}} + a, N_{s,k',j}^{-t_{s,w}} + b)} \right).$$

Update of $k_{s,t}$

The update of $k_{s,t}$ from $\pi(k_{s,t}|\text{rest})$ can be performed by following the four steps described in Section 4.1.

In order to compute

$$p_{k'} = \frac{\pi(\mathbf{A}|\mathbf{Z}^{-k_{s,t}}, Z_{s,w} = Z_{k'}^{**})}{\pi(\mathbf{A}|\mathbf{Z}^{-k_{s,t}})} \tag{2}$$

we can follow similar calculations as for $t_{s,w}$. Specifically, for $k' \in \{1, \ldots, K_n^{-k_{s,t}}\}$, (2) becomes

$$p_{k'} = \prod_{j=1}^{K_n} \left(\frac{\text{Beta}(E_{s,k',j} + a, N_{s,k',j} + b)}{\text{Beta}(E^{-k_{s,t}}_{s,k',j} + a, N^{-k_{s,t}}_{s,k',j} + b)} \right)$$

while, for $k' = K_n^{-k_{s,t}} + 1$, it is

$$p_{k'} = \prod_{j=1}^{K_n^{-k_{s,t}}+1} \left(\frac{\text{Beta}(E_{s,K_n^{-k_{s,t}}+1,j} + a, N_{s,K_n^{-k_{s,t}}+1,j} + b)}{\text{Beta}(a, b)} \right),$$

By setting $E^{-k_{s,t}}_{s,K_n^{-k_{s,t}}+1,j} = N^{-k_{s,t}}_{s,K_n^{-k_{s,t}}+1,j} = 0$ (i.e. adding a row of 0s to $\mathbf{E}^{-k_{s,t}}_{(s)}$ and $\mathbf{N}^{-k_{s,t}}_{(s)}$), these two formulas can be rewritten for all $k' \in \{1, \ldots, K_n^{-k_{s,t}} + 1\}$, as

$$p_{k'} = \prod_{j=1}^{K_n^{-k_{s,t}}+1} \left(\frac{\text{Beta}(E_{s,k',j} + a, N_{s,k',j} + b)}{\text{Beta}(E^{-k_{s,t}}_{s,k',j} + a, N^{-k_{s,t}}_{s,k',j} + b)} \right).$$

Update of the HPY hyperparameters

The hyperparameters of the HPY can also been randomized and included in MCMC sampler. Assuming independent priors for hyperparameters of different Pitman-Yor processes, the full conditional distributions can be derived from

$$\pi(\alpha, \gamma | (m_{sk} : j \in \{1, \ldots, S\}, k \in \{1, \ldots, K_n\}), (\sigma_s, \theta_s : s \in \{1, \ldots, S\}), \mathbf{A}) =$$

$$= \pi(\alpha, \gamma | m_{\cdot\cdot}, K_n) \propto \frac{\Gamma(\frac{\gamma}{\alpha} + K_n)\Gamma(\gamma)\mathcal{C}(m_{\cdot\cdot}, K_n, \alpha)}{\Gamma(\frac{\gamma}{\alpha})\Gamma(\gamma + m_{\cdot\cdot})} \pi^{prior}(\alpha, \gamma)$$

and, for each couple $((\sigma_s, \theta_s) : s \in \{1, \ldots S\})$, from

$$\pi(\sigma_s, \theta_s | (m_{st} : s \in \{1, \ldots, S\}, t \in \{1, \ldots, m_{s\cdot}\}), \sigma_{-s}, \theta_{-s}, \alpha, \gamma, \mathbf{A}_{(s)}) =$$

$$= \pi(\sigma_s, \theta_s | n_{s\cdot\cdot}, m_{s\cdot}) \propto \frac{\Gamma(\frac{\theta_s}{\sigma_s} + m_{s\cdot})\Gamma(\theta_s)\mathcal{C}(n_{s\cdot\cdot}, m_{s\cdot}, \sigma_s)}{\Gamma(\frac{\theta_s}{\sigma_s})\Gamma(\theta_s + n_{s\cdot\cdot})} \pi^{prior}(\sigma_s, \theta_s)$$

where $\mathcal{C}$ is the generalized factorial coefficient, defined for all $n \in \mathbb{N}, k \leq n, 0 \leq \sigma \leq 1$ as $\mathcal{C}(n, k; \sigma) = (1/k!) \cdot \sum_{0 \leq j \leq k} (-1)^j \binom{k}{j} (-j\sigma)_n$, with the proviso $\mathcal{C}(0, 0; \sigma) = 1$ and $\mathcal{C}(n, 0; \sigma) = 0 \; \forall n$ and where $(\theta + 1)_{n-1} = (\theta + 1)(\theta + 2) \cdots (\theta + n - 1)$ is the rising factorial coefficient.

The hyperparameters a and b can be updated using Metropolis-Hastings and the collapsed likelihood (6) in the main article. In our real data examples, we have used Random Walk Metropolis-Hastings for a and b with proposal variance 0.5. If needed, this proposal variance can be easily tuned to a specific dataset using the posterior variance estimate of a and b obtained from a short initial MCMC run.

B Derivation of the Variatonal Inference updates

In this section, we derive the updates for VI. In VI, we need to derive the values of the free parameters that maximise the ELBO function, which is defined as

$$\mathrm{ELBO}(q(\boldsymbol{\beta}', \boldsymbol{\pi}', \boldsymbol{k}, \boldsymbol{t}, \mathbf{C})) := \mathbb{E}(\log \pi(A, \boldsymbol{\beta}', \boldsymbol{\pi}', \boldsymbol{k}, \boldsymbol{t}, \mathbf{C})) - \mathbb{E}(\log q(\boldsymbol{\beta}', \boldsymbol{\pi}', \boldsymbol{k}, \boldsymbol{t}, \mathbf{C})) \tag{3}$$

Given our choice of the class $\mathcal{Q}$, described in formula (8) in the main article, we first derive a formula for the ELBO function, hence computing the expectations in (3), and then maximize it with respect to the free variational parameters using Coordinate Ascent and Stochastic Gradient Descent. Throughout the derivations, we consider the hyperparameters of the HPY, $\sigma_{1:S}, \theta_{1:S}, \alpha, \gamma$, as fixed. These hyperparameters could potentially be included in the maximization using Variational EM.

The ELBO function can be written as

$$\begin{aligned}
\mathrm{ELBO}(q) &= \mathbb{E}_q[\log p(\mathbf{A}, \boldsymbol{\beta}', \boldsymbol{\pi}', \boldsymbol{t}, \boldsymbol{k}, \mathbf{C})] - \mathbb{E}_q[\log q(\boldsymbol{\beta}', \boldsymbol{\pi}', \boldsymbol{t}, \boldsymbol{k}, \mathbf{C})] \\
&= \sum_{s=1}^{S} \left\{ \mathbb{E}_q[\log p(A_{(s)} | \mathbf{Z}_{(s)}, \mathbf{C}) p(\boldsymbol{k}_s | \boldsymbol{\beta}') p(\boldsymbol{t}_s | \boldsymbol{\pi}'_s) p(\boldsymbol{\pi}'_s)] - \mathbb{E}_q[\log q(\boldsymbol{k}_s) q(\boldsymbol{t}_s) q(\boldsymbol{\pi}'_s)] \right\} \\
&\quad + \mathbb{E}_q[\log p(\boldsymbol{\beta}') p(\mathbf{C})] - \mathbb{E}_q[\log q(\boldsymbol{\beta}') q(\mathbf{C})] \\
&= \sum_{s=1}^{S} \left\{ \mathbb{E}_q[\log p(A_{(s)} | \mathbf{Z}_{(s)}, \mathbf{C})] + \mathbb{E}_q[\log p(\boldsymbol{k}_s | \boldsymbol{\beta}')] + \mathbb{E}_q[\log p(\boldsymbol{t}_s | \boldsymbol{\pi}'_s)] + \mathbb{E}_q[\log p(\boldsymbol{\pi}'_s)] \right. \\
&\qquad \left. - \mathbb{E}_q[\log q(\boldsymbol{k}_s)] - \mathbb{E}_q[\log q(\boldsymbol{t}_s)] - \mathbb{E}_q[\log q(\boldsymbol{\pi}'_s)] \right\} \\
&\quad + \mathbb{E}_q[\log p(\boldsymbol{\beta}')] + \mathbb{E}_q[\log p(\mathbf{C})] - \mathbb{E}_q[\log q(\boldsymbol{\beta}')] - \mathbb{E}_q[\log q(\mathbf{C})] \\
&= \sum_{s=1}^{S} \left\{ \sum_{p=1}^{n_{s\cdot\cdot}} \sum_{q=p+1}^{n_{s\cdot\cdot}} \mathbb{E}_q[\log p(A_{s,p,q} | Z_{s,p}, Z_{s,q}, \mathbf{C})] \right. \\
&\qquad + \sum_{t=1}^{\infty} \mathbb{E}_q[\log p(k_{s,t} | \boldsymbol{\beta}')] + \sum_{p=1}^{n_{s\cdot\cdot}} \mathbb{E}_q[\log p(t_{s,p} | \boldsymbol{\pi}'_s)] + \sum_{t=1}^{\infty} \mathbb{E}_q[\log p(\pi'_{s,t})] \\
&\qquad \left. - \sum_{t=1}^{\infty} \mathbb{E}_q[\log q(k_{s,t})] - \sum_{p=1}^{n_{s\cdot\cdot}} \mathbb{E}_q[\log q(t_{s,p})] - \sum_{t=1}^{\infty} \mathbb{E}_q[\log q(\pi'_{s,t})] \right\} \\
&\quad + \sum_{k=1}^{\infty} \mathbb{E}_q[\log p(\beta'_k)] + \sum_{k=1}^{\infty} \sum_{k'=1}^{\infty} \mathbb{E}_q[\log p(C_{k,k'})]
\end{aligned}$$

$$
\begin{aligned}
&-\sum_{k=1}^{\infty}\mathbb{E}_q[\log q(\beta'_k)]-\sum_{k=1}^{\infty}\sum_{k'=1}^{\infty}\mathbb{E}_q[\log q(C_{k,k'})] \\
&=\sum_{s=1}^{S}\Bigg\{\sum_{p=1}^{n_{s\cdot\cdot}}\sum_{q=p+1}^{n_{s\cdot\cdot}}\sum_{k=1}^{\infty}\sum_{k'=k}^{\infty}\mathbb{E}_q\Big[\log C_{k,k'}^{A_{s,p,q}\mathbb{I}(Z_{s,p}=Z^{**}_k)\mathbb{I}(Z_{s,q}=Z^{**}_{k'})} \\
&\qquad\qquad\qquad\qquad\qquad\times(1-C_{k,k'})^{(1-A_{s,p,q})\mathbb{I}(Z_{s,p}=Z^{**}_k)\mathbb{I}(Z_{s,q}=Z^{**}_{k'})}\Big] \\
&\quad+\sum_{t=1}^{\infty}\mathbb{E}_q\left[\log\prod_{k=1}^{\infty}(1-\beta'_k)^{\mathbb{I}(k_{s,t}>k)}\beta_k'^{\mathbb{I}(k_{s,t}=k)}\right] \\
&\quad+\sum_{p=1}^{n_{s\cdot\cdot}}\mathbb{E}_q\left[\log\prod_{t=1}^{\infty}(1-\pi'_{s,t})^{\mathbb{I}(t_{s,p}>t)}\pi_{s,t}'^{\mathbb{I}(t_{s,p}=t)}\right] \\
&\quad+\sum_{t=1}^{\infty}\mathbb{E}_q\left[\log\frac{\Gamma(1-\sigma_s+\theta_s+\sigma_s t)}{\Gamma(1-\sigma_s)\Gamma(\theta_s+\sigma_s t)}{\pi'_{s,t}}^{1-\sigma_s-1}(1-\pi'_{s,t})^{\theta_s+\sigma_s t-1}\right] \\
&\quad-\sum_{t=1}^{\infty}\mathbb{E}_q\left[\log\prod_{k=1}^{\infty}\hat{\phi}_{s,t,k}^{\mathbb{I}(k_{s,t}=k)}\right]-\sum_{p=1}^{n_{s\cdot\cdot}}\mathbb{E}_q\left[\log\prod_{t=1}^{\infty}\hat{\chi}_{s,p,t}^{\mathbb{I}(t_{s,p}=t)}\right] \\
&\quad-\sum_{t=1}^{\infty}\mathbb{E}_q\left[\log\frac{\Gamma(\hat{\lambda}_{s,t}+\hat{\delta}_{s,t})}{\Gamma(\hat{\lambda}_{s,t})\Gamma(\hat{\delta}_{s,t})}{\pi'_{s,t}}^{\hat{\lambda}_{s,t}-1}(1-\pi'_{s,t})^{\hat{\delta}_{s,t}-1}\right]\Bigg\} \\
&+\sum_{k=1}^{\infty}\mathbb{E}_q\left[\log\frac{\Gamma(1-\alpha+\gamma+\alpha k)}{\Gamma(1-\alpha)\Gamma(\gamma+\alpha k)}{\beta'_k}^{1-\alpha-1}(1-\beta'_k)^{\gamma+\alpha k-1}\right] \\
&+\sum_{k=1}^{\infty}\sum_{k'=k}^{\infty}\mathbb{E}_q\left[\log\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}C_{k,k'}^{a-1}(1-C_{k,k'})^{b-1}\right] \\
&-\sum_{k=1}^{\infty}\mathbb{E}_q\left[\log\frac{\Gamma(\hat{u}_k+\hat{v}_k)}{\Gamma(\hat{u}_k)\Gamma(\hat{v}_k)}{\beta'_k}^{\hat{u}_k-1}(1-\beta'_k)^{\hat{v}_k-1}\right] \\
&-\sum_{k=1}^{\infty}\sum_{k'=k}^{\infty}\mathbb{E}_q\left[\log\frac{\Gamma(\hat{a}_{k,k'}+\hat{b}_{k,k'})}{\Gamma(\hat{a}_{k,k'})\Gamma(\hat{b}_{k,k'})}C_{k,k'}^{\hat{a}_{k,k'}-1}(1-C_{k,k'})^{\hat{b}_{k,k'}-1}\right]
\end{aligned}
$$

Before turning the expectations into analytical expressions, we consider one that appears multiple times, with the random variable r following $\text{Beta}(c,d)$ under q:

$$
\begin{aligned}
g(\alpha,\beta,c,d):&=\mathbb{E}_q\left[\log\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}r^{\alpha-1}(1-r)^{\beta-1}\right] \\
&=\mathbb{E}_q[\log\Gamma(\alpha+\beta)]-\mathbb{E}_q[\log\Gamma(\alpha)]-\mathbb{E}_q[\log\Gamma(\beta)] \\
&\quad+(\alpha-1)\mathbb{E}_q[\log r]+(\beta-1)\mathbb{E}_q[\log(1-r)] \\
&=\log\Gamma(\alpha+\beta)-\log\Gamma(\alpha)-\log\Gamma(\beta) \\
&\quad+(\alpha-1)\left[\Psi(c)-\Psi(c+d)\right]+(\beta-1)\left[\Psi(d)-\Psi(c+d)\right]
\end{aligned}
$$

Now, the expression of ELBO(q) can be simplified:

$$
\begin{aligned}
\mathrm{ELBO}(q) = \sum_{s=1}^{S} \Bigg\{ & \sum_{p=1}^{n_{s\cdot\cdot}} \sum_{q=p+1}^{n_{s\cdot\cdot}} \sum_{k=1}^{\infty} \sum_{k'=k}^{\infty} \mathbb{E}_q \Big[\log C_{k,k'}^{A_{s,p,q}\mathbb{I}(Z_{s,p}=Z_k^{**})\mathbb{I}(Z_{s,q}=Z_{k'}^{**})} \\
& \qquad\qquad \times (1-C_{k,k'})^{(1-A_{s,p,q})\mathbb{I}(Z_{s,p}=Z_k^{**})\mathbb{I}(Z_{s,q}=Z_{k'}^{**})} \Big] \\
& + \sum_{t=1}^{\infty} \mathbb{E}_q \left[\log \prod_{k=1}^{\infty} (1-\beta'_k)^{\mathbb{I}(k_{s,t}>k)} \beta_k'^{\mathbb{I}(k_{s,t}=k)} \right] \\
& + \sum_{p=1}^{n_{s\cdot\cdot}} \mathbb{E}_q \left[\log \prod_{t=1}^{\infty} (1-\pi'_{s,t})^{\mathbb{I}(t_{s,p}>t)} \pi_{s,t}'^{\mathbb{I}(t_{s,p}=t)} \right] \\
& - \sum_{t=1}^{\infty} \mathbb{E}_q \left[\log \prod_{k=1}^{\infty} \hat{\phi}_{s,t,k}^{\mathbb{I}(k_{s,t}=k)} \right] - \sum_{p=1}^{n_{s\cdot\cdot}} \mathbb{E}_q \left[\log \prod_{t=1}^{\infty} \hat{\chi}_{s,p,t}^{\mathbb{I}(t_{s,p}=t)} \right] \\
& + \sum_{t=1}^{\infty} \left[g\left(1-\sigma_s, \theta_s+\sigma_s t, \hat{\lambda}_{s,t}, \hat{\delta}_{s,t}\right) - g\left(\hat{\lambda}_{s,t}, \hat{\delta}_{s,t}, \hat{\lambda}_{s,t}, \hat{\delta}_{s,t}\right) \right] \Bigg\} \\
& + \sum_{k=1}^{\infty} \left[g\left(1-\alpha, \gamma+\alpha k, \hat{u}_k, \hat{v}_k\right) - g\left(\hat{u}_k, \hat{v}_k, \hat{u}_k, \hat{v}_k\right) \right] \\
& + \sum_{k=1}^{\infty} \sum_{k'=k}^{\infty} \left[g\left(a, b, \hat{a}_{k,k'}, \hat{b}_{k,k'}\right) - g\left(\hat{a}_{k,k'}, \hat{b}_{k,k'}, \hat{a}_{k,k'}, \hat{b}_{k,k'}\right) \right]
\end{aligned}
$$

The last three lines are due to the posterior of $\pi'_{s,t}$, β'_k and $C_{k,k'}$ being approximated by Beta$(\hat{\lambda}_{s,t}, \hat{\delta}_{s,t})$, Beta$(\hat{u}_k, \hat{v}_k)$ and Beta$(\hat{a}_{k,k'}, \hat{b}_{k,k'})$, respectively.

Thanks to the mean field assumption and definition of $\mathbf{Z}$, the first line can be written as

$$
\begin{aligned}
&\mathbb{E}_q \left[\log C_{k,k'}^{A_{s,p,q}\mathbb{I}(Z_{s,p}=Z_k^{**})\mathbb{I}(Z_{s,q}=Z_{k'}^{**})} \right] \\
&\quad = A_{s,p,q} \mathbb{E}_q[\mathbb{I}(k_{s,t_{s,p}}=k)] \mathbb{E}_q[\mathbb{I}(k_{s,t_{s,q}}=k')] \mathbb{E}_q[\log C_{k,k'}] \\
&\quad = \sum_{t=1}^{\infty} \sum_{t'=1}^{\infty} A_{s,p,q} \mathbb{E}_q[\mathbb{I}(k_{s,t}=k)] \mathbb{E}_q[\mathbb{I}(t_{s,p}=t)] \mathbb{E}_q[\mathbb{I}(k_{s,t'}=k')] \mathbb{E}_q[\mathbb{I}(t_{s,q}=t')] \mathbb{E}_q[\log C_{k,k'}]
\end{aligned}
$$

and similarly for the part depending on $(1-C_{k,k'})$.

Also, under the variational approximation q, the distribution are truncated at K and T, hence the elements greater than these values in the infinite sums are all zeros, since the excpected values of the indicators under q are zero, $\mathbb{E}_q[\mathbb{I}(k_{s,t}=k)] = 0$ for $k > K$, $\mathbb{E}_q[\mathbb{I}(t_{s,p}=t)] = 0$ for $t > T$.

Now we continue the calculations of ELBO:

$$
\begin{aligned}
\mathrm{ELBO}(q) = \sum_{s=1}^{S} \Bigg\{ & \sum_{p=1}^{n_{s\cdot\cdot}} \sum_{q=p+1}^{n_{s\cdot\cdot}} \sum_{k=1}^{K} \sum_{k'=k}^{K} \sum_{t=1}^{T} \sum_{t'=1}^{T} \left[\mathbb{E}_q[\mathbb{I}(k_{s,t}=k)\mathbb{I}(t_{s,p}=t)\mathbb{I}(k_{s,t'}=k')\mathbb{I}(t_{s,q}=t')] \right. \\
& \qquad \left. \times \left(A_{s,p,q}\mathbb{E}_q[\log C_{k,k'}] + (1-A_{s,p,q})\mathbb{E}_q[\log(1-C_{k,k'})] \right) \right] \\
& + \sum_{t=1}^{T} \sum_{k=1}^{K} \left[\mathbb{E}_q[\mathbb{I}(k_{s,t}>k)]\mathbb{E}_q[\log(1-\beta'_k)] + \mathbb{E}_q[\mathbb{I}(k_{s,t}=k)]\mathbb{E}_q[\log \beta'_k] \right] \\
& + \sum_{p=1}^{n_{s\cdot\cdot}} \sum_{t=1}^{T} \left[\mathbb{E}_p[\mathbb{I}(t_{s,p}>t)]\mathbb{E}_q[\log(1-\pi'_{s,t})] + \mathbb{E}_p[\mathbb{I}(t_{s,p}=t)]\mathbb{E}_p[\log \pi'_{s,t}] \right] \\
& - \sum_{t=1}^{T} \sum_{k=1}^{K} \mathbb{E}_q[\mathbb{I}(k_{s,t}=k)] \log \hat{\phi}_{s,t,k} - \sum_{p=1}^{n_{s\cdot\cdot}} \sum_{t=1}^{T} \mathbb{E}_q[\mathbb{I}(t_{s,p}=t)] \log \hat{\chi}_{s,p,t} \\
& + \sum_{t=1}^{T} \left[g\left(1-\sigma_s, \theta_s+\sigma_s t, \hat{\lambda}_{s,t}, \hat{\delta}_{s,t}\right) - g\left(\hat{\lambda}_{s,t}, \hat{\delta}_{s,t}, \hat{\lambda}_{s,t}, \hat{\delta}_{s,t}\right) \right] \Bigg\} \\
+ & \sum_{k=1}^{T} \left[g\left(1-\alpha, \gamma+\alpha k, \hat{u}_k, \hat{v}_k\right) - g\left(\hat{u}_k, \hat{v}_k, \hat{u}_k, \hat{v}_k\right) \right] \\
+ & \sum_{k=1}^{K} \sum_{k'=k}^{K} \left[g\left(a, b, \hat{a}_{k,k'}, \hat{b}_{k,k'}\right) - g\left(\hat{a}_{k,k'}, \hat{b}_{k,k'}, \hat{a}_{k,k'}, \hat{b}_{k,k'}\right) \right]
\end{aligned}
$$

In the first line, when $t = t'$, $k = k'$, there is one fewer indicator and no squared terms, as

$$
\begin{aligned}
& \mathbb{E}_q[\mathbb{I}(k_{s,t}=k)\mathbb{I}(t_{s,p}=t)\mathbb{I}(k_{s,t'}=k')\mathbb{I}(t_{s,q}=t')] \\
= \; & \mathbb{E}_q[\mathbb{I}(k_{s,t}=k)\mathbb{I}(t_{s,p}=t)\mathbb{I}(k_{s,t}=k)\mathbb{I}(t_{s,q}=t)] \\
= \; & \mathbb{E}_q[\mathbb{I}(k_{s,t}=k)\mathbb{I}(t_{s,p}=t)\mathbb{I}(t_{s,q}=t)] = \hat{\phi}_{s,t,k}\hat{\chi}_{s,p,t}\hat{\chi}_{s,q,t}
\end{aligned}
$$

Also, when $t = t'$ but $k \neq k'$, either $\mathbb{I}(k_{s,t} = k)$ or $\mathbb{I}(k_{s,t'} = k')$ has to be zero. This prompts us to split the sum in the first two lines into three sums:

$$\begin{aligned}
&\sum_{k=1}^{K}\sum_{k'=k}^{K}\sum_{t=1}^{T}\sum_{t'=1}^{T} f_{s,p,q}(k,k',t,t',\cdots) \\
&= \sum_{k=1}^{K}\sum_{k'=k}^{K}\sum_{t=1}^{T}\sum_{t'\neq t}^{T} f_{s,p,q}(k,k',t,t',\cdots) + \sum_{k=1}^{K}\sum_{k'=k}^{K}\sum_{t=1}^{T} f_{s,p,q}(k,k',t,t'=t,\cdots) \\
&= \sum_{k=1}^{K}\sum_{k'=k}^{K}\sum_{t=1}^{T}\sum_{t'\neq t}^{T} f_{s,p,q}(k,k',t,t',\cdots) + \sum_{k=1}^{K}\sum_{t=1}^{T} f_{s,p,q}(k,k'=k,t,t'=t,\cdots)
\end{aligned}$$

where $f_{s,p,q}(k,k',t,t',\cdots)$ is the term within the (outer) pair of square brackets in the first two lines. From there we can continue our calculations:

$$\begin{aligned}
&\mathrm{ELBO}(q) \\
&= \sum_{s=1}^{S}\sum_{p=1}^{n_{s\cdot\cdot}}\sum_{q=p+1}^{n_{s\cdot\cdot}}\sum_{k=1}^{K}\sum_{k'=k}^{K}\sum_{t=1}^{T}\sum_{t'\neq t}^{T} \hat{\phi}_{s,t,k}\hat{\chi}_{s,p,t}\hat{\phi}_{s,t',k'}\hat{\chi}_{s,q,t'}h(s,p,q,k,k') \\
&+ \sum_{s=1}^{S}\sum_{p=1}^{n_{s\cdot\cdot}}\sum_{q=p+1}^{n_{s\cdot\cdot}}\sum_{k=1}^{K}\sum_{t=1}^{T} \hat{\phi}_{s,t,k}\hat{\chi}_{s,p,t}\hat{\chi}_{s,q,t}h(s,p,q,k,k) \\
&+ \sum_{s=1}^{S}\sum_{t=1}^{T}\sum_{k=1}^{K}\left[\sum_{l=k+1}^{K}\hat{\phi}_{s,t,l}(\Psi(\hat{v}_k) - \Psi(\hat{u}_k + \hat{v}_k)) + \hat{\phi}_{s,t,k}(\Psi(\hat{u}_k) - \Psi(\hat{v}_k + \hat{u}_k))\right] \\
&+ \sum_{s=1}^{S}\sum_{p=1}^{n_{s\cdot\cdot}}\sum_{t=1}^{T}\left[\sum_{l=t+1}^{T}\hat{\chi}_{s,p,l}(\Psi(\hat{\delta}_{s,t}) - \Psi(\hat{\lambda}_{s,t} + \hat{\delta}_{s,t})) + \hat{\chi}_{s,p,t}(\Psi(\hat{\lambda}_{s,t}) - \Psi(\hat{\lambda}_{s,t} + \hat{\delta}_{s,t}))\right] \\
&- \sum_{s=1}^{S}\sum_{t=1}^{T}\sum_{k=1}^{K}\hat{\phi}_{s,t,k}\log\hat{\phi}_{s,t,k} - \sum_{s=1}^{S}\sum_{p=1}^{n_{s\cdot\cdot}}\sum_{t=1}^{T}\hat{\chi}_{s,p,t}\log\hat{\chi}_{s,p,t} \\
&+ \sum_{s=1}^{S}\sum_{t=1}^{T}\left[g\left(1-\sigma_s, \theta_s + \sigma_s t, \hat{\lambda}_{s,t}, \hat{\delta}_{s,t}\right) - g\left(\hat{\lambda}_{s,t}, \hat{\delta}_{s,t}, \hat{\lambda}_{s,t}, \hat{\delta}_{s,t}\right)\right] \\
&+ \sum_{k=1}^{K}\left[g\left(1-\alpha, \gamma + \alpha k, \hat{u}_k, \hat{v}_k\right) - g\left(\hat{u}_k, \hat{v}_k, \hat{u}_k, \hat{v}_k\right)\right] \\
&+ \sum_{k=1}^{K}\sum_{k'=k}^{K}\left[g\left(a, b, \hat{a}_{k,k'}, \hat{b}_{k,k'}\right) - g\left(\hat{a}_{k,k'}, \hat{b}_{k,k'}, \hat{a}_{k,k'}, \hat{b}_{k,k'}\right)\right]
\end{aligned}$$

where

$$h(s,p,q,k,k') = A_{s,p,q}(\Psi(\hat{a}_{k,k'}) - \Psi(\hat{a}_{k,k'} + \hat{b}_{k,k'})) + (1 - A_{s,p,q})(\Psi(\hat{b}_{k,k'}) - \Psi(\hat{a}_{k,k'} + \hat{b}_{k,k'})).$$

We need to optimize $\mathrm{ELBO}(q)$ with respect to all variational parameters:

- $\phi_{s,t,k}$: $s = 1, \ldots, S$, $t = 1, \ldots, T$, $k = 1, \ldots, K$ s.t. $\sum_{k=1}^{K} \hat{\phi}_{s,t,k} = 1$;

- $\hat{\chi}_{s,p,t}$: $s = 1, \ldots, S$, $p = 1, \ldots, N$, $t = 1, \ldots, T$ s.t. $\sum_{t=1}^{T} \hat{\chi}_{s,p,t} = 1$;

- $(\hat{u}_k, \hat{v}_k)$: $k = 1, \ldots, K$;

- $(\hat{\lambda}_{s,t}, \hat{\delta}_{s,t})$: $s = 1, \ldots, S$, $t = 1, \ldots, T$;

- $(\hat{a}_{k,k'}, \hat{b}_{k,k'})$: $k = 1, \ldots, K$, $k' = k, \ldots, K$.

Before doing so, we first calculate the following derivatives:

$$\frac{\partial}{\partial c} g(\alpha, \beta, c, d) = (\alpha - 1)\left[\Psi'(c) - \Psi'(c+d)\right] - (\beta - 1)\Psi'(c+d) \tag{4}$$

$$\frac{\partial}{\partial d} g(\alpha, \beta, c, d) = (\beta - 1)\left[\Psi'(d) - \Psi'(c+d)\right] - (\alpha - 1)\Psi'(c+d) \tag{5}$$

The derivatives in (4) and (5) still hold if $\alpha = c$ and $\beta = d$ as the extra terms will be cancelled out.

We first work with optimizing $(\hat{u}_k, \hat{v}_k)$ given all other free parameters:

$$\begin{aligned}
\text{ELBO}\left(\hat{u}_k, \hat{v}_k\right) &= \sum_{s=1}^{S}\sum_{t=1}^{T}\sum_{l=k+1}^{K} \hat{\phi}_{s,t,l}\left(\Psi(\hat{v}_k) - \Psi(\hat{u}_k + \hat{v}_k)\right) \\
&\quad + \sum_{s=1}^{S}\sum_{t=1}^{T} \hat{\phi}_{s,t,k}\left(\Psi(\hat{u}_k) - \Psi(\hat{u}_k + \hat{v}_k)\right) \\
&\quad + g\left(1 - \alpha, \gamma + \alpha k, \hat{u}_k, \hat{v}_k\right) - g\left(\hat{u}_k, \hat{v}_k, \hat{u}_k, \hat{v}_k\right) + \text{const} \\
\frac{\partial \text{ELBO}\left(\hat{u}_k, \hat{v}_k\right)}{\partial \hat{u}_k} &= -\sum_{s=1}^{S}\sum_{t=1}^{T}\sum_{l=k+1}^{K} \hat{\phi}_{s,t,l}\Psi'(\hat{u}_k + \hat{v}_k) \\
&\quad + \sum_{s=1}^{S}\sum_{t=1}^{T} \hat{\phi}_{s,t,k}\left(\Psi'(\hat{u}_k) - \Psi'(\hat{u}_k + \hat{v}_k)\right) \\
&\quad + (-\alpha)\left(\Psi'(\hat{u}_k) - \Psi'(\hat{u}_k + \hat{v}_k)\right) - (\gamma + \alpha k - 1)\Psi'(\hat{u}_k + \hat{v}_k) \\
&\quad - (\hat{u}_k - 1)\left(\Psi'(\hat{u}_k) - \Psi'(\hat{u}_k + \hat{v}_k)\right) + (\hat{v}_k - 1)\Psi'(\hat{u}_k + \hat{v}_k) \\
&= -\Psi'(\hat{u}_k + \hat{v}_k)\left[\left(\sum_{s=1}^{S}\sum_{t=1}^{T}\sum_{l=k+1}^{K} \hat{\phi}_{s,t,l}\right) + \gamma + \alpha k - \hat{v}_k\right] \\
&\quad + \left(\Psi'(\hat{u}_k) - \Psi'(\hat{u}_k + \hat{v}_k)\right)\left[\left(\sum_{s=1}^{S}\sum_{t=1}^{T} \hat{\phi}_{s,t,k}\right) - \alpha - \hat{u}_k + 1\right]
\end{aligned}$$

This is equal to zero if

$$\hat{u}_k = \left(\sum_{s=1}^{S}\sum_{t=1}^{T}\hat{\phi}_{s,t,k}\right) - \alpha + 1, \qquad \hat{v}_k = \left(\sum_{s=1}^{S}\sum_{t=1}^{T}\sum_{l=k+1}^{K}\hat{\phi}_{s,t,l}\right) + \gamma + \alpha k \tag{6}$$

We double check that at these two values the derivative with respect to $\hat{v}_k$ is also 0:

$$\begin{aligned}
\frac{\partial \mathrm{ELBO}\left(\hat{u}_k, \hat{v}_k\right)}{\partial \hat{v}_k} = &\sum_{s=1}^{S}\sum_{t=1}^{T}\sum_{l=k+1}^{K}\hat{\phi}_{s,t,l}\left(\Psi'(\hat{v}_k) - \Psi'(\hat{u}_k + \hat{v}_k)\right) \\
&- \sum_{s=1}^{S}\sum_{t=1}^{T}\hat{\phi}_{s,t,k}\Psi'(\hat{u}_k + \hat{v}_k) \\
&+ (\gamma + \alpha k - 1)\left(\Psi'(\hat{v}_k) - \Psi'(\hat{u}_k + \hat{v}_k)\right) - (-\alpha)\Psi'(\hat{u}_k + \hat{v}_k) \\
&- (\hat{v}_k - 1)\left(\Psi'(\hat{v}_k) - \Psi'(\hat{u}_k + \hat{v}_k)\right) + (\hat{u}_k - 1)\Psi'(\hat{u}_k + \hat{v}_k) \\
= &-\Psi'(\hat{u}_k + \hat{v}_k)\left[\left(\sum_{s=1}^{S}\sum_{t=1}^{T}\hat{\phi}_{s,t,k}\right) - \alpha - \hat{u}_k + 1\right] \\
&+ \left(\Psi'(\hat{v}_k) - \Psi'(\hat{u}_k + \hat{v}_k)\right)\left[\left(\sum_{s=1}^{S}\sum_{t=1}^{T}\sum_{l=k+1}^{K}\hat{\phi}_{s,t,l}\right) + \gamma + \alpha k - \hat{v}_k\right],
\end{aligned}$$

which is 0 if (6) holds.

Next, we work with $(\hat{a}_{k,k'}, \hat{b}_{k,k'})$, first considering $k' > k$:

$$\text{ELBO}\left(\hat{a}_{k,k'}, \hat{b}_{k,k'}\right) = \sum_{s=1}^{S}\sum_{p=1}^{n_{s\cdot\cdot}}\sum_{q=p+1}^{n_{s\cdot\cdot}}\sum_{t=1}^{T}\sum_{t'\neq t}^{T} \hat{\phi}_{s,t,k}\hat{\chi}_{s,p,t}\hat{\phi}_{s,t',k'}\hat{\chi}_{s,q,t'} h(s,p,q,k,k')$$

$$+ g(a, b, \hat{a}_{k,k'}, \hat{b}_{k,k'}) - g(\hat{a}_{k,k'}, \hat{b}_{k,k'}, \hat{a}_{k,k'}, \hat{b}_{k,k'}) + \text{const}$$

$$\frac{\partial \text{ELBO}\left(\hat{a}_{k,k'}, \hat{b}_{k,k'}\right)}{\partial \hat{a}_{k,k'}}$$

$$= \sum_{s=1}^{S}\sum_{p=1}^{n_{s\cdot\cdot}}\sum_{q=p+1}^{n_{s\cdot\cdot}}\sum_{t=1}^{T}\sum_{t'\neq t}^{T} \hat{\phi}_{s,t,k}\hat{\chi}_{s,p,t}\hat{\phi}_{s,t',k'}\hat{\chi}_{s,q,t'} A_{s,p,q}\left(\Psi'(\hat{a}_{k,k'}) - \Psi'(\hat{a}_{k,k'} + \hat{b}_{k,k'})\right)$$

$$- \sum_{s=1}^{S}\sum_{p=1}^{n_{s\cdot\cdot}}\sum_{q=p+1}^{n_{s\cdot\cdot}}\sum_{t=1}^{T}\sum_{t'\neq t}^{T} \hat{\phi}_{s,t,k}\hat{\chi}_{s,p,t}\hat{\phi}_{s,t',k'}\hat{\chi}_{s,q,t'}(1 - A_{s,p,q})\ \Psi'(\hat{a}_{k,k'} + \hat{b}_{k,k'})$$

$$+ (a-1)\left(\Psi'(\hat{a}_{k,k'}) - \Psi'(\hat{a}_{k,k'} + \hat{b}_{k,k'})\right) - (b-1)\Psi'(\hat{a}_{k,k'} + \hat{b}_{k,k'})$$

$$- (\hat{a}_{k,k'} - 1)\left(\Psi'(\hat{a}_{k,k'}) - \Psi'(\hat{a}_{k,k'} + \hat{b}_{k,k'})\right) + (\hat{b}_{k,k'} - 1)\Psi'(\hat{a}_{k,k'} + \hat{b}_{k,k'})$$

$$= \left(\sum_{s=1}^{S} A_{k,k'}(s) + a - \hat{a}_{k,k'}\right)\left(\Psi'(\hat{a}_{k,k'}) - \Psi'(\hat{a}_{k,k'} + \hat{b}_{k,k'})\right)$$

$$- \left(\sum_{s=1}^{S} B_{k,k'}(s) + b - \hat{b}_{k,k'}\right)\Psi'(\hat{a}_{k,k'} + \hat{b}_{k,k'})$$

where

$$A_{k,k'}(s) = \sum_{p=1}^{n_{s\cdot\cdot}}\sum_{q=p+1}^{n_{s\cdot\cdot}}\sum_{t=1}^{T}\sum_{t'\neq t}^{T} \hat{\phi}_{s,t,k}\hat{\chi}_{s,p,t}\hat{\phi}_{s,t',k'}\hat{\chi}_{s,q,t'} A_{s,p,q}, \qquad \text{and}$$

$$B_{k,k'}(s) = \sum_{p=1}^{n_{s\cdot\cdot}}\sum_{q=p+1}^{n_{s\cdot\cdot}}\sum_{t=1}^{T}\sum_{t'\neq t}^{T} \hat{\phi}_{s,t,k}\hat{\chi}_{s,p,t}\hat{\phi}_{s,t',k'}\hat{\chi}_{s,q,t'}(1 - A_{s,p,q})$$

The derivative is equal to zero if

$$\hat{a}_{k,k'} = \sum_{s=1}^{S} A_{k,k'}(s) + a, \qquad \hat{b}_{k,k'} = \sum_{s=1}^{S} B_{k,k'}(s) + b \tag{7}$$

We double check that at these two values the derivative with respect to $\hat{b}_{k,k'}$ is also 0:

$$\frac{\partial \text{ELBO}\left(\hat{a}_{k,k'}, \hat{b}_{k,k'}\right)}{\partial \hat{b}_{k,k'}}$$

$$= -\sum_{s=1}^{S}\sum_{p=1}^{n_{s\cdot\cdot}}\sum_{q=p+1}^{n_{s\cdot\cdot}}\sum_{t=1}^{T}\sum_{t'\neq t}^{T} \hat{\phi}_{s,t,k}\hat{\chi}_{s,p,t}\hat{\phi}_{s,t',k'}\hat{\chi}_{s,q,t'} A_{s,p,q}\ \Psi'(\hat{a}_{k,k'} + \hat{b}_{k,k'})$$

$$+ \sum_{s=1}^{S}\sum_{p=1}^{n_{s\cdot\cdot}}\sum_{q=p+1}^{n_{s\cdot\cdot}}\sum_{t=1}^{T}\sum_{t'\neq t}^{T} \hat{\phi}_{s,t,k}\hat{\chi}_{s,p,t}\hat{\phi}_{s,t',k'}\hat{\chi}_{s,q,t'}(1 - A_{s,p,q})\left(\Psi'(\hat{b}_{k,k'} - \Psi'(\hat{a}_{k,k'} + \hat{b}_{k,k'})\right)$$

$$
\begin{aligned}
&+ (b-1)\left(\Psi'(\hat{b}_{k,k'}) - \Psi'(\hat{a}_{k,k'} + \hat{b}_{k,k'})\right) - (a-1)\Psi'(\hat{a}_{k,k'} + \hat{b}_{k,k'}) \\
&- (\hat{b}_{k,k'} - 1)\left(\Psi'(\hat{b}_{k,k'}) - \Psi'(\hat{a}_{k,k'} + \hat{b}_{k,k'})\right) + (\hat{a}_{k,k'} - 1)\Psi'(\hat{a}_{k,k'} + \hat{b}_{k,k'}) \\
= &-(\sum_{s=1}^{S} A_{k,k'}(s) + a - \hat{a}_{k,k'})\Psi'(\hat{a}_{k,k'} + \hat{b}_{k,k'}) + \\
&(\sum_{s=1}^{S} B_{k,k'}(s) + b - \hat{b}_{k,k'})\left(\Psi'(\hat{b}_{k,k'}) - \Psi'(\hat{a}_{k,k'} + \hat{b}_{k,k'})\right),
\end{aligned}
$$

which is equal to 0 if (7) holds.

Next, we consider $(\hat{a}_{k,k}, \hat{b}_{k,k})$:

$$\mathrm{ELBO}\left(\hat{a}_{k,k}, \hat{b}_{k,k}\right) = \sum_{s=1}^{S} \sum_{p=1}^{n_{s\cdot\cdot}} \sum_{q=p+1}^{n_{s\cdot\cdot}} \sum_{t=1}^{T} \hat{\phi}_{s,t,k} \hat{\chi}_{s,p,t} \hat{\chi}_{s,q,t} h(s,p,q,k,k)$$
$$+ g(a, b, \hat{a}_{k,k}, \hat{b}_{k,k}) - g(\hat{a}_{k,k}, \hat{b}_{k,k}, \hat{a}_{k,k}, \hat{b}_{k,k}) + \text{const}$$

This leads to a derivative with respect to $\hat{a}_{k,k}$ similar to when $k > k'$:

$$\frac{\partial \mathrm{ELBO}\left(\hat{a}_{k,k}, \hat{b}_{k,k}\right)}{\partial \hat{a}_{k,k}} = \left(\sum_{s=1}^{S} A_{k,k}(s) + a - \hat{a}_{k,k}\right) \left(\Psi'(\hat{a}_{k,k}) - \Psi'(\hat{a}_{k,k} + \hat{b}_{k,k})\right) +$$
$$- \left(\sum_{s=1}^{S} B_{k,k}(s) + b - \hat{b}_{k,k}\right) \Psi'(\hat{a}_{k,k} + \hat{b}_{k,k}),$$

where

$$A_{k,k}(s) = \sum_{s=1}^{S} \sum_{p=1}^{n_{s\cdot\cdot}} \sum_{q=p+1}^{n_{s\cdot\cdot}} \sum_{t=1}^{T} \hat{\phi}_{s,t,k} \hat{\chi}_{s,p,t} \hat{\chi}_{s,q,t} A_{s,p,q}, \qquad \text{and}$$
$$B_{k,k}(s) = \sum_{s=1}^{S} \sum_{p=1}^{n_{s\cdot\cdot}} \sum_{q=p+1}^{n_{s\cdot\cdot}} \sum_{t=1}^{T} \hat{\phi}_{s,t,k} \hat{\chi}_{s,p,t} \hat{\chi}_{s,q,t} \left(1 - A_{s,p,q}\right)$$

The derivative is equal to zero if

$$\hat{a}_{k,k} = \sum_{s=1}^{S} A_{k,k}(s) + a, \qquad \hat{b}_{k,k} = \sum_{s=1}^{S} B_{k,k}(s) + b \tag{8}$$

We can verify that at these two values the derivative with respect to $\hat{b}_{k,k}$ is also 0, as

$$\frac{\partial \mathrm{ELBO}\left(\hat{a}_{k,k}, \hat{b}_{k,k}\right)}{\partial \hat{b}_{k,k}}$$
$$= - \left(\sum_{s=1}^{S} A_{k,k}(s) + a - \hat{a}_{k,k}\right) \Psi'(\hat{a}_{k,k} + \hat{b}_{k,k})$$
$$+ \left(\sum_{s=1}^{S} B_{k,k}(s) + b - \hat{b}_{k,k}\right) \left(\Psi'(\hat{a}_{k,k}) - \Psi'(\hat{a}_{k,k} + \hat{b}_{k,k})\right)$$

Next, we consider $\hat{\phi}_{s,t,k}$. As $\hat{\phi}_{s,t,k}$ has to sum over k to 1, we incorporate a term $\mu_{s,t}$ for such constrained optimization:

$$
\begin{aligned}
&\mathrm{ELBO}(\hat{\phi}_{s,t,k}) \\
&= \sum_{p=1}^{n_{s\cdot\cdot}} \sum_{q=p+1}^{n_{s\cdot\cdot}} \sum_{k=1}^{K} \sum_{k'=k}^{K} \sum_{t=1}^{T} \sum_{t'\neq t}^{T} \hat{\phi}_{s,t,k} \hat{\chi}_{s,p,t} \hat{\phi}_{s,t',k'} \hat{\chi}_{s,q,t'} h(s,p,q,k,k') \\
&\quad + \sum_{p=1}^{n_{s\cdot\cdot}} \sum_{q=p+1}^{n_{s\cdot\cdot}} \sum_{k=1}^{K} \sum_{t=1}^{T} \hat{\phi}_{s,t,k} \hat{\chi}_{s,p,t} \hat{\chi}_{s,q,t} h(s,p,q,k,k) \\
&\quad + \sum_{t=1}^{T} \sum_{k=1}^{K} \left[\sum_{l=k+1}^{K} \hat{\phi}_{s,t,l} (\Psi(\hat{v}_k) - \Psi(\hat{u}_k + \hat{v}_k)) + \hat{\phi}_{s,t,k} (\Psi(\hat{u}_k) - \Psi(\hat{v}_k + \hat{u}_k)) \right] \\
&\quad - \sum_{t=1}^{T} \sum_{k=1}^{K} \hat{\phi}_{s,t,k} \log \hat{\phi}_{s,t,k} + \mu_{s,t} \left(\sum_{k=1}^{K} \hat{\phi}_{s,t,k} - 1 \right) + \mathrm{const}_1
\end{aligned}
$$

The first line of the previous formula can be rewritten up to an additive constant (not depending on $\hat{\phi}_{s,t,k}$) as

$$
\begin{aligned}
&= \sum_{p=1}^{n_{s\cdot\cdot}} \sum_{q=p+1}^{n_{s\cdot\cdot}} \sum_{k'=k}^{K} \sum_{t'\neq t}^{T} \hat{\phi}_{s,t,k} \hat{\chi}_{s,p,t} \hat{\phi}_{s,t',k'} \hat{\chi}_{s,q,t'} h(s,p,q,k,k') \\
&\quad + \sum_{p=1}^{n_{s\cdot\cdot}} \sum_{q=p+1}^{n_{s\cdot\cdot}} \sum_{k'=k}^{K} \sum_{t'\neq t}^{T} \hat{\phi}_{s,t',k} \hat{\chi}_{s,p,t'} \hat{\phi}_{s,t,k'} \hat{\chi}_{s,q,t} h(s,p,q,k,k') \\
&\quad + \sum_{p=1}^{n_{s\cdot\cdot}} \sum_{q=p+1}^{n_{s\cdot\cdot}} \sum_{k'=1}^{k-1} \sum_{t'\neq t}^{T} \hat{\phi}_{s,t,k'} \hat{\chi}_{s,p,t} \hat{\phi}_{s,t',k} \hat{\chi}_{s,q,t'} h(s,p,q,k',k) \\
&\quad + \sum_{p=1}^{n_{s\cdot\cdot}} \sum_{q=p+1}^{n_{s\cdot\cdot}} \sum_{k'=1}^{k-1} \sum_{t'\neq t}^{T} \hat{\phi}_{s,t',k'} \hat{\chi}_{s,p,t'} \hat{\phi}_{s,t,k} \hat{\chi}_{s,q,t} h(s,p,q,k',k)
\end{aligned}
$$

The third line does not depend on $\hat{\phi}_{s,t,k}$. Also, the second line depends on $\hat{\phi}_{s,t,k}$ only for $k' = k$. Therefore,

$$
\begin{aligned}
&= \sum_{p=1}^{n_{s\cdot\cdot}} \sum_{q=p+1}^{n_{s\cdot\cdot}} \sum_{k'=k}^{K} \sum_{t'\neq t}^{T} \hat{\phi}_{s,t,k} \hat{\chi}_{s,p,t} \hat{\phi}_{s,t',k'} \hat{\chi}_{s,q,t'} h(s,p,q,k,k') \\
&\quad + \sum_{p=1}^{n_{s\cdot\cdot}} \sum_{q=p+1}^{n_{s\cdot\cdot}} \sum_{k'=1}^{k-1} \sum_{t'\neq t}^{T} \hat{\phi}_{s,t',k'} \hat{\chi}_{s,p,t'} \hat{\phi}_{s,t,k} \hat{\chi}_{s,q,t} h(s,p,q,k',k) \\
&\quad + \sum_{p=1}^{n_{s\cdot\cdot}} \sum_{q=p+1}^{n_{s\cdot\cdot}} \sum_{t'\neq t}^{T} \hat{\phi}_{s,t',k} \hat{\chi}_{s,p,t'} \hat{\phi}_{s,t,k} \hat{\chi}_{s,q,t} h(s,p,q,k,k) \\
&= \hat{\phi}_{s,t,k} \sum_{p=1}^{n_{s\cdot\cdot}} \sum_{q=p+1}^{n_{s\cdot\cdot}} (\sum_{k'=k}^{K} \sum_{t'\neq t}^{T} \hat{\chi}_{s,p,t} \hat{\phi}_{s,t',k'} \hat{\chi}_{s,q,t'} h(s,p,q,k,k')
\end{aligned}
$$

$$
+ \sum_{k'=1}^{k-1} \sum_{t' \neq t}^{T} \hat{\phi}_{s,t',k'} \hat{\chi}_{s,p,t'} \hat{\chi}_{s,q,t} h(s,p,q,k',k)
$$

$$
+ \sum_{t' \neq t}^{T} \hat{\phi}_{s,t',k} \hat{\chi}_{s,p,t'} \hat{\chi}_{s,q,t} h(s,p,q,k,k))
$$

So, the ELBO is

$$
\begin{aligned}
&= \hat{\phi}_{s,t,k} \sum_{p=1}^{n_{s\cdot\cdot}} \sum_{q=p+1}^{n_{s\cdot\cdot}} \Big(\sum_{k'=k}^{K} \sum_{t' \neq t}^{T} \hat{\chi}_{s,p,t} \hat{\phi}_{s,t',k'} \hat{\chi}_{s,q,t'} h(s,p,q,k,k') \\
&\quad + \sum_{k'=1}^{k-1} \sum_{t' \neq t}^{T} \hat{\phi}_{s,t',k'} \hat{\chi}_{s,p,t'} \hat{\chi}_{s,q,t} h(s,p,q,k',k) \\
&\quad + \sum_{t' \neq t}^{T} \hat{\phi}_{s,t',k} \hat{\chi}_{s,p,t'} \hat{\chi}_{s,q,t} h(s,p,q,k,k) \\
&\quad + \hat{\chi}_{s,p,t} \hat{\chi}_{s,q,t} h(s,p,q,k,k)\Big) \\
&\quad + \sum_{l=1}^{k-1} \hat{\phi}_{s,t,k} (\Psi(\hat{v}_l) - \Psi(\hat{u}_l + \hat{v}_l)) + \hat{\phi}_{s,t,k} (\Psi(\hat{u}_k) - \Psi(\hat{v}_k + \hat{u}_k)) \\
&\quad - \hat{\phi}_{s,t,k} \log \hat{\phi}_{s,t,k} + \mu_{s,t} \hat{\phi}_{s,t,k} + \mathrm{const}_2
\end{aligned}
$$

By setting $\hat{a}_{k,k'} := \hat{a}_{k',k}$ and $\hat{b}_{k,k'} := \hat{b}_{k',k}$ for $k > k'$, the function $h(s,p,q,k,k')$ can be defined for all $k, k' = 1, \ldots, K$, and is symmetric in the last two terms. Therefore, we can rewrite the term multiplying $\hat{\phi}_{s,t,k}$ in the first four lines of the last formula as

$$
\begin{aligned}
\Phi_{s,t,k} := &\sum_{p=1}^{n_{s\cdot\cdot}} \sum_{q=p+1}^{n_{s\cdot\cdot}} \Big(\sum_{k'=k}^{K} \sum_{t' \neq t}^{T} \hat{\chi}_{s,p,t} \hat{\phi}_{s,t',k'} \hat{\chi}_{s,q,t'} h(s,p,q,k,k') \\
&\quad + \sum_{k'=1}^{k-1} \sum_{t' \neq t}^{T} \hat{\chi}_{s,p,t'} \hat{\phi}_{s,t',k'} \hat{\chi}_{s,q,t} h(s,p,q,k,k') \\
&\quad + \sum_{t' \neq t}^{T} \hat{\phi}_{s,t',k} \hat{\chi}_{s,p,t'} \hat{\chi}_{s,q,t} h(s,p,q,k,k) \\
&\quad + \hat{\chi}_{s,p,t} \hat{\chi}_{s,q,t} h(s,p,q,k,k)\Big)
\end{aligned}
$$

Now, differentiating the ELBO with respect to $\phi_{s,t,k}$, we obtain

$$
\frac{\partial \mathrm{ELBO}(\hat{\phi}_{s,t,k})}{\partial (\hat{\phi}_{s,t,k})} = \Phi_{s,t,k} + \sum_{l=1}^{k-1} (\Psi(\hat{v}_l) - \Psi(\hat{u}_l + \hat{v}_l)) + (\Psi(\hat{u}_k) - \Psi(\hat{v}_k + \hat{u}_k)) - 1 - \log \hat{\phi}_{s,t,k} + \mu_{s,t}
$$

Setting this formula equal to zero, moving the $\log \hat{\phi}_{s,t,k}$ term on the other side and exponentiating both sides, we can see that the update for $\hat{\phi}_{s,t,k}$ is

$$\hat{\phi}_{s,t,k} \propto \exp \left\{ \Phi_{s,t,k} + \sum_{l=1}^{k-1} (\Psi(\hat{v}_l) - \Psi(\hat{u}_l + \hat{v}_l)) + (\Psi(\hat{u}_k) - \Psi(\hat{v}_k + \hat{u}_k)) \right\}$$

This proportionality is over k, meaning that the right-hand side has to sum over k to 1.

Next, we work with $(\hat{\lambda}_{s,t}, \hat{\delta}_{s,t})$:

$$
\begin{aligned}
\text{ELBO}\left(\hat{\lambda}_{s,t}, \hat{\delta}_{s,t}\right) &= \sum_{p=1}^{n_{s\cdot\cdot}} \sum_{l=t+1}^{T} \hat{\chi}_{s,p,l} \left(\Psi(\hat{\delta}_{s,t}) - \Psi(\hat{\lambda}_{s,t} + \hat{\delta}_{s,t})\right) \\
&\quad + \sum_{p=1}^{n_{s\cdot\cdot}} \hat{\chi}_{s,p,t} \left(\Psi(\hat{\lambda}_{s,t}) - \Psi(\hat{\lambda}_{s,t} + \hat{\delta}_{s,t})\right) \\
&\quad + g\left(1 - \sigma_s, \theta_s + \sigma_s t, \hat{\lambda}_{s,t}, \hat{\delta}_{s,t}\right) - g\left(\hat{\lambda}_{s,t}, \hat{\delta}_{s,t}, \hat{\lambda}_{s,t}, \hat{\delta}_{s,t}\right) + \text{const} \\
\frac{\partial \text{ELBO}\left(\hat{\lambda}_{s,t}, \hat{\delta}_{s,t}\right)}{\partial \hat{\lambda}_{s,t}} &= -\sum_{p=1}^{n_{s\cdot\cdot}} \sum_{l=t+1}^{T} \hat{\chi}_{s,p,l} \Psi'(\hat{\lambda}_{s,t} + \hat{\delta}_{s,t}) \\
&\quad + \sum_{p=1}^{n_{s\cdot\cdot}} \hat{\chi}_{s,p,t} \left(\Psi'(\hat{\lambda}_{s,t}) - \Psi'(\hat{\lambda}_{s,t} + \hat{\delta}_{s,t})\right) \\
&\quad - \sigma_s \left(\Psi'(\hat{\lambda}_{s,t}) - \Psi'(\hat{\lambda}_{s,t} + \hat{\delta}_{s,t})\right) - (\theta_s + \sigma_s t - 1)\Psi'(\hat{\lambda}_{s,t} + \hat{\delta}_{s,t}) \\
&\quad - (\hat{\lambda}_{s,t} - 1)\left(\Psi'(\hat{\lambda}_{s,t}) - \Psi'(\hat{\lambda}_{s,t} + \hat{\delta}_{s,t})\right) + (\hat{\delta}_{s,t} - 1)\Psi'(\hat{\lambda}_{s,t} + \hat{\delta}_{s,t}) \\
&= -\Psi'(\hat{\lambda}_{s,t} + \hat{\delta}_{s,t}) \left[\left(\sum_{p=1}^{n_{s\cdot\cdot}} \sum_{l=t+1}^{T} \hat{\chi}_{s,p,l}\right) + \theta_s + \sigma_s t - \hat{\delta}_{s,t}\right] \\
&\quad + \left(\Psi'(\hat{\lambda}_{s,t}) - \Psi'(\hat{\lambda}_{s,t} + \hat{\delta}_{s,t})\right) \left[\left(\sum_{p=1}^{n_{s\cdot\cdot}} \hat{\chi}_{s,p,t}\right) - \sigma_s - \hat{\lambda}_{s,t} + 1\right]
\end{aligned}
$$

This is equal to zero if

$$
\hat{\lambda}_{s,t} = \left(\sum_{p=1}^{n_{s\cdot\cdot}} \hat{\chi}_{s,p,t}\right) - \sigma_s + 1, \qquad \hat{\delta}_{s,t} = \left(\sum_{p=1}^{n_{s\cdot\cdot}} \sum_{l=t+1}^{T} \hat{\chi}_{s,p,l}\right) + \theta_s + \sigma_s t \tag{9}
$$

We can check that at these two values the derivative with respect to $\hat{v}_k$ is also 0 as

$$
\begin{aligned}
\frac{\partial \text{ELBO}\left(\hat{\lambda}_{s,t}, \hat{\delta}_{s,t}\right)}{\partial \hat{\lambda}_{s,t}} &= \sum_{p=1}^{n_{s\cdot\cdot}} \sum_{l=t+1}^{T} \hat{\chi}_{s,p,l} \left(\Psi'(\hat{\delta}_{s,t}) - \Psi'(\hat{\lambda}_{s,t} + \hat{\delta}_{s,t})\right) - \sum_{p=1}^{n_{s\cdot\cdot}} \hat{\chi}_{s,p,t} \Psi'(\hat{\lambda}_{s,t} + \hat{\delta}_{s,t}) \\
&\quad + (\theta_s + \sigma_s t - 1)\left(\Psi'(\hat{\lambda}_{s,t}) - \Psi'(\hat{\delta}_{s,t} + \hat{\delta}_{s,t})\right) + \sigma_s \Psi'(\hat{\delta}_{s,t} + \hat{\delta}_{s,t}) \\
&\quad - (\hat{\delta}_{s,t} - 1)\left(\Psi'(\hat{\delta}_{s,t}) - \Psi'(\hat{\lambda}_{s,t} + \hat{\delta}_{s,t})\right) + (\hat{\lambda}_{s,t} - 1)\Psi'(\hat{\lambda}_{s,t} + \hat{\delta}_{s,t}) \\
&= -\Psi'(\hat{\lambda}_{s,t} + \hat{\delta}_{s,t}) \left[\left(\sum_{p=1}^{n_{s\cdot\cdot}} \hat{\chi}_{s,p,t}\right) - \sigma_s - \hat{\lambda}_{s,t} + 1\right] \\
&\quad + \left(\Psi'(\hat{\lambda}_{s,t}) - \Psi'(\hat{\lambda}_{s,t} + \hat{\delta}_{s,t})\right) \left[\left(\sum_{p=1}^{n_{s\cdot\cdot}} \sum_{l=t+1}^{T} \hat{\chi}_{s,p,l}\right) + \theta_s + \sigma_s t - \hat{\delta}_{s,t}\right]
\end{aligned}
$$

Finally, we consider $\hat{\chi}_{s,p,t}$, which has to sum over t to 1, thus requiring a similar constrained optimization to $\hat{\phi}_{s,t,k}$:

$$
\begin{aligned}
&\mathrm{ELBO}(\hat{\chi}_{s,p,t}) \\
&= \sum_{p=1}^{n_{s\cdot\cdot}} \sum_{q=p+1}^{n_{s\cdot\cdot}} \sum_{k=1}^{K} \sum_{k'=k}^{K} \sum_{t=1}^{T} \sum_{t'\neq t}^{T} \hat{\phi}_{s,t,k}\hat{\chi}_{s,p,t}\hat{\phi}_{s,t',k'}\hat{\chi}_{s,q,t'} h(s,p,q,k,k') \\
&\quad + \sum_{p=1}^{n_{s\cdot\cdot}} \sum_{q=p+1}^{n_{s\cdot\cdot}} \sum_{k=1}^{K} \sum_{t=1}^{T} \hat{\phi}_{s,t,k}\hat{\chi}_{s,p,t}\hat{\chi}_{s,q,t} h(s,p,q,k,k) \\
&\quad + \sum_{p=1}^{n_{s\cdot\cdot}} \sum_{t=1}^{T} \left[\sum_{l=t+1}^{T} \hat{\chi}_{s,p,l}(\Psi(\hat{\delta}_{s,t}) - \Psi(\hat{\lambda}_{s,t}+\hat{\delta}_{s,t})) + \hat{\chi}_{s,p,t}(\Psi(\hat{\lambda}_{s,t}) - \Psi(\hat{\lambda}_{s,t}+\hat{\delta}_{s,t})) \right] \\
&\quad - \sum_{p=1}^{n_{s\cdot\cdot}} \sum_{t=1}^{T} \hat{\chi}_{s,p,t}\log\hat{\chi}_{s,p,t} + \mu_{s,p}\left(\sum_{t=1}^{T} \hat{\chi}_{s,p,t} - 1 \right) + \mathrm{const}_1
\end{aligned}
$$

The first line can be decomposed into,

$$
\begin{aligned}
&\sum_{p=1}^{n_{s\cdot\cdot}} \sum_{q=p+1}^{n_{s\cdot\cdot}} \sum_{k=1}^{K} \sum_{k'=k}^{K} \sum_{t=1}^{T} \sum_{t'\neq t}^{T} \hat{\phi}_{s,t,k}\hat{\chi}_{s,p,t}\hat{\phi}_{s,t',k'}\hat{\chi}_{s,q,t'} h(s,p,q,k,k') \\
&= \sum_{k=1}^{K} \sum_{k'=k}^{K} (\sum_{q=1}^{p-1} \sum_{t'\neq t}^{T} \hat{\phi}_{s,t',k}\hat{\chi}_{s,q,t'}\hat{\phi}_{s,t,k'}\hat{\chi}_{s,p,t} h(s,q,p,k,k') \\
&\quad + \sum_{q=p+1}^{n_{s\cdot\cdot}} \sum_{t'\neq t}^{T} \hat{\phi}_{s,t',k}\hat{\chi}_{s,p,t'}\hat{\phi}_{s,t,k'}\hat{\chi}_{s,q,t} h(s,p,q,k,k') \\
&\quad + \sum_{q=1}^{p-1} \sum_{t'\neq t}^{T} \hat{\phi}_{s,t,k}\hat{\chi}_{s,q,t}\hat{\phi}_{s,t',k'}\hat{\chi}_{s,p,t'} h(s,q,p,k,k') \\
&\quad + \sum_{q=p+1}^{n_{s\cdot\cdot}} \sum_{t'\neq t}^{T} \hat{\phi}_{s,t,k}\hat{\chi}_{s,p,t}\hat{\phi}_{s,t',k'}\hat{\chi}_{s,q,t'} h(s,p,q,k,k')) + \mathrm{const}
\end{aligned}
$$

and the second and third line of the previous formula do not depend on $\hat{\chi}_{s,p,t}$.

Therefore, the ELBO can be written as

$$
\begin{aligned}
\mathrm{ELBO}(\hat{\chi}_{s,p,t}) = \hat{\chi}_{s,p,t} &\sum_{k=1}^{K} \sum_{k'=k}^{K} (\sum_{q=1}^{p-1} \sum_{t'\neq t}^{T} \hat{\phi}_{s,t',k}\hat{\chi}_{s,q,t'}\hat{\phi}_{s,t,k'} h(s,q,p,k,k') \\
&+ \sum_{q=p+1}^{n_{s\cdot\cdot}} \sum_{t'\neq t}^{T} \hat{\phi}_{s,t,k}\hat{\phi}_{s,t',k'}\hat{\chi}_{s,q,t'} h(s,p,q,k,k'))+ \\
&+ \hat{\chi}_{s,p,t}(\sum_{q=1}^{p-1} \sum_{k=1}^{K} \hat{\phi}_{s,t,k}\hat{\chi}_{s,q,t} h(s,q,p,k,k) + \sum_{q=p+1}^{n_{s\cdot\cdot}} \sum_{k=1}^{K} \hat{\phi}_{s,t,k}\hat{\chi}_{s,q,t} h(s,p,q,k,k))
\end{aligned}
$$

$$+\sum_{l=1}^{t-1}\hat{\chi}_{s,p,t}(\Psi(\hat{\delta}_{s,l})-\Psi(\hat{\lambda}_{s,l}+\hat{\delta}_{s,l}))+\hat{\chi}_{s,p,t}(\Psi(\hat{\lambda}_{s,t})-\Psi(\hat{\lambda}_{s,t}+\hat{\delta}_{s,t}))$$

$$-\hat{\chi}_{s,p,t}\log\hat{\chi}_{s,p,t}+\mu_{s,p}\hat{\chi}_{s,p,t}+\text{const}_2$$

Denoting,

$$X_{s,p,t}=\sum_{k=1}^{K}\sum_{k'=k}^{K}(\sum_{q=1}^{p-1}\sum_{t'\neq t}^{T}\hat{\phi}_{s,t',k}\hat{\chi}_{s,q,t'}\hat{\phi}_{s,t,k'}h(s,q,p,k,k')+\sum_{q=p+1}^{n_{s\cdot\cdot}}\sum_{t'\neq t}^{T}\hat{\phi}_{s,t,k}\hat{\phi}_{s,t',k'}\hat{\chi}_{s,q,t'}h(s,p,q,k,k'))+$$
$$+\sum_{k=1}^{K}(\sum_{q=1}^{p-1}\hat{\phi}_{s,t,k}\hat{\chi}_{s,q,t}h(s,q,p,k,k)+\sum_{q=p+1}^{n_{s\cdot\cdot}}\hat{\phi}_{s,t,k}\hat{\chi}_{s,q,t}h(s,p,q,k,k))$$

$$\frac{\partial\text{ELBO}(\hat{\chi}_{s,p,t})}{\partial\hat{\chi}_{s,p,t}}=X_{s,p,t}+\sum_{l=1}^{t-1}(\Psi(\hat{\delta}_{s,l})-\Psi(\hat{\lambda}_{s,l}+\hat{\delta}_{s,l}))+(\Psi(\hat{\lambda}_{s,t})-\Psi(\hat{\lambda}_{s,t}+\hat{\delta}_{s,t}))$$
$$-1-\log\hat{\chi}_{s,p,t}+\mu_{s,p}$$

Setting this formula equal to zero, moving the $\log\hat{\chi}_{s,p,t}$ on one side and exponentiating both sides, we can see that the update for $\hat{\chi}_{s,p,t}$ is

$$\hat{\chi}_{s,p,t}\propto\exp\left\{X_{s,p,t}+\sum_{l=1}^{t-1}(\Psi(\hat{\delta}_{s,l})-\Psi(\hat{\lambda}_{s,l}+\hat{\delta}_{s,l}))+(\Psi(\hat{\lambda}_{s,t})-\Psi(\hat{\lambda}_{s,t}+\hat{\delta}_{s,t}))\right\}$$

This proportionality is over t, meaning that the right-hand side has to sum over t to 1.

Finally, the Coordinate Ascent Variational Inference (CAVI) algorithm for the HSBM can be summarized as in Algorithm 2. Both in Algorithm 2 and Algorithm 1 (in the main document), the quantities $A_{k,k'}(s)$, $B_{k,k'}(s)$, $\Phi_{s,t,k}$ and $X_{s,p,t}$ are defined as follows,

$$A_{k,k'}(s) = \sum_{p=1}^{n_{s\cdot\cdot}} \sum_{q=p+1}^{n_{s\cdot\cdot}} \sum_{t=1}^{T} \sum_{t'\neq t}^{T} \hat{\phi}_{s,t,k} \hat{\chi}_{s,p,t} \hat{\phi}_{s,t',k'} \hat{\chi}_{s,q,t'} A_{s,p,q}$$

$$B_{k,k'}(s) = \sum_{p=1}^{n_{s\cdot\cdot}} \sum_{q=p+1}^{n_{s\cdot\cdot}} \sum_{t=1}^{T} \sum_{t'\neq t}^{T} \hat{\phi}_{s,t,k} \hat{\chi}_{s,p,t} \hat{\phi}_{s,t',k'} \hat{\chi}_{s,q,t'} (1 - A_{s,p,q})$$

$$\begin{aligned}\Phi_{s,t,k} := \sum_{p=1}^{n_{s\cdot\cdot}} \sum_{q=p+1}^{n_{s\cdot\cdot}} \Big(& \sum_{k'=k}^{K} \sum_{t'\neq t}^{T} \hat{\chi}_{s,p,t} \hat{\phi}_{s,t',k'} \hat{\chi}_{s,q,t'} h(s,p,q,k,k') \\ & + \sum_{k'=1}^{k-1} \sum_{t'\neq t}^{T} \hat{\chi}_{s,p,t'} \hat{\phi}_{s,t',k'} \hat{\chi}_{s,q,t} h(s,p,q,k,k') \\ & + \sum_{t'\neq t}^{T} \hat{\phi}_{s,t',k}^{\mathbb{I}(t'\neq t)} \hat{\chi}_{s,p,t'} \hat{\chi}_{s,q,t} h(s,p,q,k,k) \Big)\end{aligned}$$

$$\begin{aligned}X_{s,p,t} = & \sum_{k=1}^{K} \sum_{k'=k}^{K} \Big(\sum_{q=1}^{p-1} \sum_{t'\neq t}^{T} \hat{\phi}_{s,t',k} \hat{\chi}_{s,q,t'} \hat{\phi}_{s,t,k'} h(s,q,p,k,k') + \sum_{q=p+1}^{n_{s\cdot\cdot}} \sum_{t'\neq t}^{T} \hat{\phi}_{s,t,k} \hat{\phi}_{s,t',k'} \hat{\chi}_{s,q,t'} h(s,p,q,k,k') \Big) + \\ & + \sum_{k=1}^{K} \Big(\sum_{q=1}^{p-1} \hat{\phi}_{s,t,k} \hat{\chi}_{s,q,t} h(s,q,p,k,k) + \sum_{q=p+1}^{n_{s\cdot\cdot}} \hat{\phi}_{s,t,k} \hat{\chi}_{s,q,t} h(s,p,q,k,k) \Big)\end{aligned}$$

$$h(s,p,q,k,k') = A_{s,p,q}(\Psi(\hat{a}_{k,k'}) - \Psi(\hat{a}_{k,k'} + \hat{b}_{k,k'})) + (1 - A_{s,p,q})(\Psi(\hat{b}_{k,k'}) - \Psi(\hat{a}_{k,k'} + \hat{b}_{k,k'})).$$

where $\Psi(s) = \Gamma'(s)/\Gamma(s)$ denotes the Digamma function.

Algorithm 2: HSMB - CAVI

Initialize global parameters $(\hat{u}_k, \hat{v}_k)_{1\le k\le K}$, $(\hat{a}_{k,k'}, \hat{b}_{k,k'})_{1\le k,k'\le K}$

for *t in 1:convergence* **do**

  for *s in 1:S* **do**

    Update local parameters:

    for *t in 1:T* **do**

      $\hat{\phi}_{s,t,k} \propto$

      $\exp\left\{\Phi_{s,t,k} + \sum_{l=1}^{k-1}(\Psi(\hat{v}_l) - \Psi(\hat{u}_l + \hat{v}_l)) + (\Psi(\hat{u}_k) - \Psi(\hat{v}_k + \hat{u}_k))\right\}$

      $\hat{\lambda}_{s,t} = \sum_{p=1}^{n_{s\cdot\cdot}} \hat{\chi}_{s,p,t} - \sigma_s + 1$

      $\hat{\delta}_{s,t} = \sum_{p=1}^{n_{s\cdot\cdot}} \sum_{l=t+1}^{T} \hat{\chi}_{s,p,l} + \theta_s + \sigma_s t$

    end

    for *p in 1:*$n_{s\cdot\cdot}$ **do**

      $\hat{\chi}_{s,p,t} \propto$

      $\exp\left\{X_{s,p,t} + \sum_{l=1}^{t-1}(\Psi(\hat{\delta}_{s,l}) - \Psi(\hat{\lambda}_{s,l} + \hat{\delta}_{s,l})) + (\Psi(\hat{\lambda}_{s,t}) - \Psi(\hat{\lambda}_{s,t} + \hat{\delta}_{s,t}))\right\}$

    end

    Update Global parameters:

    for *k in 1:K* **do**

      $\hat{u}_k = \sum_{s=1}^{S} \sum_{t=1}^{T} \hat{\phi}_{s,t,k} - \alpha + 1$

      $\hat{v}_k = \sum_{s=1}^{S} \sum_{t=1}^{T} \sum_{l=k+1}^{K} \hat{\phi}_{s,t,l} + \gamma + \alpha k$

      for *k' in k:K* **do**

        $\hat{a}_{k,k'} = \sum_{s=1}^{S} A_{k,k'}(s) + a$

        $\hat{b}_{k,k'} = \sum_{s=1}^{S} B_{k,k'}(s) + b$

      end

    end

  end

  Evaluate ELBO$(q(\boldsymbol{\beta'}, \boldsymbol{\pi'}, \boldsymbol{k}, \boldsymbol{t}, \mathbf{C}))$

end

References

Teh, Y. W. (2006), A hierarchical Bayesian language model based on Pitman–Yor processes, *in* ‘Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, Morristown.’, pp. 985–992.