ComE+ - derivations

Based on $q(W, K)$, we can formally define the truncated generative process of ComE+ as:

$$\begin{aligned} q_{\gamma_{k,1},\gamma_{k,2}}(\pi_k): &\quad \pi_k \sim \mathcal{B}(\gamma_{k,1}, \gamma_{k,2}), \text{ for } k = 1, ..., K; && (1) \\ q_{\boldsymbol{\tau}_k}(\boldsymbol{\psi}_k): &\quad \boldsymbol{\psi}_k \sim \mathcal{N}(\boldsymbol{\tau}_k, \mathbf{I}), \text{ for } k = 1, ..., K; && (2) \\ q_{B_k,c_k}(\Sigma_k^{-1}): &\quad \Sigma_k^{-1} \sim \mathcal{W}(B_k^{-1}, c_k), \text{ for } k = 1, ..., K; && (3) \\ q_{\boldsymbol{\xi}_i}(z_i): &\quad z_i \sim Multi(\boldsymbol{\xi}_i), \text{ for } i = 1, ..., |V|. && (4) \\ & && (5) \end{aligned}$$

Then, for inference we define $O_3^{ComE+'}$ as:

$$\begin{aligned} O_3^{ComE+'}(\Phi, q) = &-\tfrac{\beta}{K}\left(\mathbb{E}_q[\log p(W, \Phi|\rho, \nu)] - \mathbb{E}_q[\log q(W, K)]\right) \\ = &-\tfrac{\beta}{K}\Big(\textstyle\sum_{k=1}^{\infty}\big(\mathbb{E}_q[\log p(\pi_k|\rho)] - \mathbb{E}_q[\log q_{\gamma_{k,1},\gamma_{k,2}}(\pi_k)]\big) \\ &+ \textstyle\sum_{k=1}^{\infty}\left(\mathbb{E}_q[\log p(\boldsymbol{\psi}_k)] - \mathbb{E}_q[\log q_{\boldsymbol{\tau}_k}(\boldsymbol{\psi}_k)]\right) \\ &+ \textstyle\sum_{k=1}^{\infty}\big(\mathbb{E}_q[\log p(\Sigma_k^{-1}|\nu)] - \mathbb{E}_q[\log q_{B_k,c_k}(\Sigma_k^{-1})]\big) \\ &+ \textstyle\sum_{i=1}^{|V|}\big(\mathbb{E}_q[\log p(z_i|\boldsymbol{\pi})] - \mathbb{E}_q[\log q_{\boldsymbol{\xi}_i}(z_i)]\big) \\ &+ \textstyle\sum_{i=1}^{|V|}\mathbb{E}_q[\log p(\boldsymbol{\phi}_i|\boldsymbol{\psi}_{z_i}, \Sigma_{z_i})]\Big). \end{aligned} \quad (6)$$

Finally, we derive the updates for each variational parameter in $q(W, K)$ as:

$$\gamma_{k,1} = 1 + \textstyle\sum_{i=1}^{|V|}\xi_{i,k}, \quad (7)$$

$$\gamma_{k,2} = \rho + \textstyle\sum_{i=1}^{|V|}\sum_{j=k+1}^{K}\xi_{i,j}, \quad (8)$$

$$\boldsymbol{\tau}_k = \Big(\mathbf{I} + B_k^{-1}c_k\textstyle\sum_{i=1}^{|V|}\xi_{i,k}\Big)^{-1}\Big(B_k^{-1}c_k\textstyle\sum_{i=1}^{|V|}\xi_{i,k}\boldsymbol{\phi}_i\Big), \quad (9)$$

$$B_k = \Big(\textstyle\sum_{i=1}^{|V|}\xi_{i,k} + 1\Big)\mathbf{I} + \textstyle\sum_{i=1}^{|V|}\xi_{i,k}(\boldsymbol{\phi}_i - \boldsymbol{\tau}_k)(\boldsymbol{\phi}_i - \boldsymbol{\tau}_k)^T, \quad (10)$$

$$c_k = 2 + \nu + \textstyle\sum_{i=1}^{|V|}\xi_{i,k} \quad (11)$$

$$\xi_{i,k} \propto e^{\Lambda(\gamma_{k,1},\gamma_{k,2}) + \sum_{j=1}^{k}\Lambda(\gamma_{j,1},\gamma_{j,2}) + \mathbb{E}_q[\log p(\boldsymbol{\phi}_i|z_i=k)]}, \quad (12)$$

where we define $\Lambda(\gamma_{l,1}, \gamma_{l,2}) = \Psi(\gamma_{l,1}) - \Psi(\gamma_{l,1} + \gamma_{l,2})$, given $\Psi(\cdot)$ as a digamma function.