Supplemental Material

1 Derivation of the conditional distribution in Gibbs sampling of BBTM-II

To simplify the notations, we ignore the predefined parameters $\alpha, \beta, \boldsymbol{\eta}$ in the following equations. Moreover, when $e_i = 0$, we let $z_i = 0$. Thus, we have $P(z_i = 0|e_i = 0) = 1$ and $P(e_i = 0|\mathbf{e}^{\neg i}, \mathbf{z}^{\neg i}, \mathbb{B}) = P(e_i = 0, z_i = 0|\mathbf{e}^{\neg i}, \mathbf{z}^{\neg i}, \mathbb{B})$.

Now we show the derivation of $P(e_i, z_i|\mathbf{e}^{\neg i}, \mathbf{z}^{\neg i}, \mathbb{B})$.Using the chain rule, $P(e_i, z_i|\mathbf{e}^{\neg i}, \mathbf{z}^{\neg i}, \mathbb{B})$ can be rewritten as:

$$P(e_i, z_i|\mathbf{e}^{\neg i}, \mathbf{z}^{\neg i}, \mathbb{B}) = \frac{P(\mathbf{e}, \mathbf{z}, \mathbb{B})}{P(\mathbf{e}^{\neg i}, \mathbf{z}^{\neg i}, \mathbb{B})} \propto \frac{P(\mathbb{B}|\mathbf{z})P(\mathbf{z}|\mathbf{e})P(e_i)}{P(\mathbb{B}^{\neg i}|\mathbf{z}^{\neg i})P(\mathbf{z}^{\neg i}|\mathbf{e}^{\neg i})}. \tag{1}$$

In Eq.(1), $P(\mathbb{B}|\mathbf{z})$ can be obtained by integrating out $\boldsymbol{\Phi} = \{\boldsymbol{\phi}_0, ..., \boldsymbol{\phi}_K\}$:

$$\begin{aligned}
P(\mathbb{B}|\mathbf{z}) &= \int P(\mathbb{B}|\mathbf{z}, \boldsymbol{\Phi})P(\boldsymbol{\Phi})\mathrm{d}\boldsymbol{\Phi} \\
&= \int \left(\prod_{i=1}^{N_B} P(b_i|z_i, \boldsymbol{\phi}_{z_i})\right) P(\boldsymbol{\Phi})\mathrm{d}\boldsymbol{\Phi} \\
&= \int \prod_{k=0}^{K} \left(\frac{\Gamma(W\beta)}{\Gamma(\beta)^W} \prod_{w=1}^{W} \boldsymbol{\phi}_{k,w}^{n_{k,w}+\beta-1} \mathrm{d}\boldsymbol{\phi_k}\right) \\
&= \left(\frac{\Gamma(W\beta)}{\Gamma(\beta)^W}\right)^K \prod_{k=0}^{K} \frac{\prod_{w=1}^{W} \Gamma(n_{k,w} + \beta)}{\Gamma(n_{k,\cdot} + W\beta)},
\end{aligned} \tag{2}$$

where $\Gamma(\cdot)$ is the standard Gamma function, $n_{k,w}$ is the number of times that word w assigned to topic k, and $n_{k,\cdot} = \sum_{w=1}^{W} n_{k,w}$.

$P(\mathbf{z}|\mathbf{e})$ can be obtained by:

$$\begin{aligned}
P(\mathbf{z}|\mathbf{e}) &= \left(\prod_{i:e_i=0} P(z_i = 0|e_i = 0)\right) \left(\int \prod_{j:e_j=1} P(z_j|\boldsymbol{\theta})P(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta}\right) \\
&= \int \frac{\Gamma(K\alpha)}{\Gamma(\alpha)^K} \prod_{k=1}^{K} \theta_k^{n_k+\alpha-1} \mathrm{d}\boldsymbol{\theta} \\
&= \frac{\Gamma(K\alpha)}{\Gamma(\alpha)^K} \frac{\prod_{k=1}^{K} \Gamma(n_k + \alpha)}{\Gamma(n_\cdot + K\alpha)}.
\end{aligned} \tag{3}$$

where n_k $(k > 0)$ is the number of biterms assigned to bursty topic k, and $n_\cdot = \sum_{k=1}^{K} n_k$ is the total number of biterms assigned to bursty topics.

$P(\mathbb{B}^{\neg i}|\mathbf{z}^{\neg i})$ and $P(\mathbf{z}^{\neg i}|\mathbf{e}^{\neg i})$ can be worked out in the same way:

$$P(\mathbb{B}_{\neg i}|\mathbf{z}_{\neg i}) = \left(\frac{\Gamma(W\beta)}{\Gamma(\beta)^W}\right)^K \prod_{k=1}^{K} \frac{\prod_{w=1}^{W} \Gamma(n_{k,w}^{\neg i} + \beta)}{\Gamma(n_{k,\cdot}^{\neg i} + W\beta)}, \tag{4}$$

$$P(\mathbf{z}^{\neg i}|\mathbf{e}^{\neg i}) = \frac{\Gamma(K\alpha)}{\Gamma(\alpha)^K} \frac{\prod_{k=1}^{K} \Gamma(n_k^{\neg i} + \alpha)}{\Gamma(n_\cdot - 1 + K\alpha)}. \tag{5}$$

Since the Gamma function satisfies $\Gamma(x+1) = x\Gamma(x)$, we obtain the final conditional distribution by replacing terms in Eq.(1) with those in Eqs.(2-5):

$$
\begin{aligned}
P(e_i = 0|\mathbf{e}^{\neg i}, \mathbf{z}^{\neg i}, \mathbb{B}) &\propto (1 - \eta_{b_i}) \cdot \frac{(n_{0,w_{i,1}}^{\neg i} + \beta)(n_{0,w_{i,2}}^{\neg i} + \beta)}{(n_{0,\cdot}^{\neg i} + W\beta)(n_{0,\cdot}^{\neg i} + 1 + W\beta)} \\
P(e_i = 1, z_i = k|\mathbf{e}^{\neg i}, \mathbf{z}^{\neg i}) &\propto \eta_{b_i} \cdot \frac{(n_k^{\neg i} + \alpha)}{(n_\cdot^{\neg i} + K\alpha)} \cdot \frac{(n_{k,w_{i,1}}^{\neg i} + \beta)(n_{k,w_{i,2}}^{\neg i} + \beta)}{(n_{k,\cdot}^{\neg i} + W\beta)(n_{k,\cdot}^{\neg i} + 1 + W\beta)}
\end{aligned}
$$