Under the truncated model, η is a $F \times \bar{K}$ matrix of regression coefficients, and u is a $\bar{K} \times D$ matrix satisfying $u_{:d} \sim N(\eta^T \phi_d, I_{\bar{K}})$. Similarly, A is a $\bar{K} \times \bar{K}$ lower triangular matrix, and $v_{:d} \sim N(Au_{:d}, \lambda_v^{-1} I_{\bar{K}})$. The probabilities π_{kd} for the first $\bar{K}$ topics are set as in eq. (2), with the final topic set so that a valid distribution is ensured: $\pi_{Kd} = 1 - \sum_{k=1}^{K-1} \pi_{kd} = \prod_{k=1}^{K-1} \psi(-v_{kd})$.

3.1 Gibbs Updates for Topic Assignments, Correlation Parameters, and Hyperparameters

The precision parameter λ_f controls the variability of the feature weights associated with each topic. As in many regression models, the gamma prior is conjugate so that

$$
\begin{aligned}
p(\lambda_f \mid \eta, a_f, b_f) &\propto \mathrm{Gam}(\lambda_f \mid a_f, b_f) \prod_{k=1}^{\bar{K}} N(\eta_{fk} \mid \mu_f, \lambda_f^{-1}) \\
&\propto \mathrm{Gam}\Big(\lambda_f \mid \frac{1}{2}\bar{K} + a_f,\ \frac{1}{2}\sum_{k=1}^{\bar{K}} (\eta_{fk} - \mu_f)^2 + b_f\Big).
\end{aligned} \tag{3}
$$

Similarly, the precision parameter λ_v has a gamma prior and posterior:

$$
\begin{aligned}
p(\lambda_v \mid v, a_v, b_v) &\propto \mathrm{Gam}(\lambda_v \mid a_v, b_v) \prod_{d=1}^{D} N(v_{:d} \mid Au_{:d}, L^{-1}) \\
&\propto \mathrm{Gam}\Big(\lambda_v \mid \frac{1}{2}\bar{K}D + a_v,\ \frac{1}{2}\sum_{d=1}^{D} (v_{:d} - Au_{:d})^T (v_{:d} - Au_{:d}) + b_v\Big).
\end{aligned} \tag{4}
$$

Entries of the regression matrix A have a rescaled Gaussian prior $A_{k\ell} \sim N(0, (k\lambda_A)^{-1})$. With a gamma prior, the precision parameter λ_A nevertheless has the following gamma posterior:

$$
\begin{aligned}
p(\lambda_A \mid A, a_A, b_A) &\propto \mathrm{Gam}(\lambda_A \mid a_A, b_A) \prod_{k=1}^{\bar{K}} \prod_{\ell=1}^{k} N(A_{k\ell} \mid 0, (k\lambda_A)^{-1}) \\
&\propto \mathrm{Gam}\Big(\lambda_A \mid \frac{1}{2}\bar{K}(\bar{K}-1) + a_A,\ \frac{1}{2}\sum_{k=1}^{\bar{K}} \sum_{\ell=1}^{k} kA_{k\ell}^2 + b_A\Big).
\end{aligned} \tag{5}
$$

Conditioning on the feature regression weights η, the mean weight μ_f in our hierarchical prior for each feature f has a Gaussian posterior:

$$
\begin{aligned}
p(\mu_f \mid \eta) &\propto N(\mu_f \mid 0, \gamma_\mu) \prod_{k=1}^{\bar{K}} N(\eta_{fk} \mid \mu_f, \lambda_f^{-1}) \\
&\propto N\Big(\mu_f \mid \frac{\gamma_\mu}{\bar{K}\gamma_\mu + \lambda_f^{-1}} \sum_{k=1}^{\bar{K}} \eta_{fk},\ (\gamma_\mu^{-1} + \bar{K}\lambda_f)^{-1}\Big)
\end{aligned} \tag{6}
$$

To sample $\eta_{:k}$, the linear function relating metadata to topic k, we condition on all documents $u_{k:}$ as well as ϕ, μ, and Λ. Columns of η are conditionally independent, with Gaussian posteriors:

$$
\begin{aligned}
p(\eta_{:k} \mid u, \phi, \mu, \Lambda) &\propto N(\eta_{:k} \mid \mu, \Lambda^{-1}) N(u_{k:}^T \mid \phi^T \eta_{:k}, I_D) \\
&\propto N(\eta_{:k} \mid (\Lambda + \phi\phi^T)^{-1}(\phi u_{k:}^T + \Lambda\mu),\ (\Lambda + \phi\phi^T)^{-1}).
\end{aligned} \tag{7}
$$

Similarly, the scores $u_{:d}$ for each document are conditionally independent with Gaussian posteriors:

$$
\begin{aligned}
p(u_{:d} \mid v_{:d}, \eta, \phi_d, L) &\propto N(u_{:d} \mid \eta^T \phi_d, I_{\bar{K}}) N(v_{:d} \mid Au_{:d}, L^{-1}) \\
&\propto N(u_{:d} \mid (I_{\bar{K}} + A^T L A)^{-1}(A^T L v_{:d} + \eta^T \phi_d),\ (I_{\bar{K}} + A^T L A)^{-1}).
\end{aligned} \tag{8}
$$

To resample A, we note that its rows are conditionally independent. The posterior of the k entries $A_{k:}$ in row k depends on $v_{k:}$ and $\hat{U}_k \triangleq u_{1:k,:}$, the first k entries of $u_{:d}$ for each document d:

$$
\begin{aligned}
p(A_{k:}^T \mid v_{k:}, \hat{U}_k, \lambda_A, \lambda_v) &\propto \prod_{j=1}^{k} N(A_{kj} \mid 0, (k\lambda_A)^{-1}) N(v_{k:}^T \mid \hat{U}_k^T A_{k:}^T, \lambda_v^{-1} I_D) \\
&\propto N(A_{k:}^T \mid (k\lambda_A\lambda_v^{-1} I_k + \hat{U}_k \hat{U}_k^T)^{-1} \hat{U}_k v_{k:}^T, (k\lambda_A I_k + \lambda_v \hat{U}_k \hat{U}_k^T)^{-1}).
\end{aligned} \tag{9}
$$