Supplementary Material for "Bayesian Poisson Tucker Decomposition for Learning the Structure of International Relations"

Proceedings of the 33^{rd} International Conference on Machine Learning, New York, NY, USA, 2016.
JMLR: W&CP volume 48.

Aaron Schein Mingyuan Zhou David M. Blei Hanna Wallach

1 Proposition 1

In the limit as $C, K, R \to \infty$, the expected sum of the core tensor elements is finite and equal to

$$\mathbb{E}\left[\sum_{c=1}^{\infty}\sum_{k=1}^{\infty}\sum_{r=1}^{\infty}\left(\lambda^{(r)}_{c\circlearrowleft^{k}} + \sum_{d\neq c}\lambda^{(r)}_{c\xrightarrow{k}d}\right)\right] = \frac{1}{\delta}\left(\frac{\gamma_0^3}{\zeta^3} + \frac{\gamma_0^4}{\zeta^4}\right).$$

The proof is very similar to that of Zhou (2015, Lemma 1). By the law of total expectation,

$$\begin{aligned}
\mathbb{E}\left[\sum_{c=1}^{\infty}\sum_{k=1}^{\infty}\sum_{r=1}^{\infty}\left(\lambda^{(r)}_{c\circlearrowleft^{k}} + \sum_{d\neq c}\lambda^{(r)}_{c\xrightarrow{k}d}\right)\right] &= \sum_{c=1}^{\infty}\sum_{k=1}^{\infty}\sum_{r=1}^{\infty}\left(\mathbb{E}\left[\lambda^{(r)}_{c\circlearrowleft^{k}}\right] + \sum_{d\neq c}\mathbb{E}\left[\lambda^{(r)}_{c\xrightarrow{k}d}\right]\right)\\
&= \sum_{c=1}^{\infty}\sum_{k=1}^{\infty}\sum_{r=1}^{\infty}\left(\mathbb{E}\left[\frac{\eta_c^{\circlearrowleft}\eta_c^{\leftrightarrow}\nu_k\rho_r}{\delta}\right] + \sum_{d\neq c}\mathbb{E}\left[\frac{\eta_c^{\leftrightarrow}\eta_d^{\leftrightarrow}\nu_k\rho_r}{\delta}\right]\right)\\
&= \frac{1}{\delta}\sum_{c=1}^{\infty}\sum_{k=1}^{\infty}\sum_{r=1}^{\infty}\left(\mathbb{E}\left[\eta_c^{\circlearrowleft}\eta_c^{\leftrightarrow}\nu_k\rho_r\right] + \sum_{d\neq c}\mathbb{E}\left[\eta_c^{\leftrightarrow}\eta_d^{\leftrightarrow}\nu_k\rho_r\right]\right)\\
&= \frac{1}{\delta}\,\mathbb{E}\left[\sum_{k=1}^{\infty}\nu_k\right]\mathbb{E}\left[\sum_{r=1}^{\infty}\rho_r\right]\sum_{c=1}^{\infty}\left(\mathbb{E}\left[\eta_c^{\circlearrowleft}\eta_c^{\leftrightarrow}\right] + \sum_{d\neq c}\mathbb{E}\left[\eta_c^{\leftrightarrow}\eta_d^{\leftrightarrow}\right]\right)\\
&= \frac{1}{\delta}\left(\frac{\gamma_0}{\zeta}\right)\left(\frac{\gamma_0}{\zeta}\right)\sum_{c=1}^{\infty}\left(\mathbb{E}\left[\eta_c^{\circlearrowleft}\eta_c^{\leftrightarrow}\right] + \sum_{d\neq c}\mathbb{E}\left[\eta_c^{\leftrightarrow}\eta_d^{\leftrightarrow}\right]\right)\\
&= \frac{1}{\delta}\left(\frac{\gamma_0}{\zeta}\right)^2\left(\sum_{c=1}^{\infty}\mathbb{E}\left[\eta_c^{\circlearrowleft}\right]\mathbb{E}\left[\eta_c^{\leftrightarrow}\right] + \mathbb{E}\left[\sum_{c=1}^{\infty}\sum_{d\neq c}\eta_c^{\leftrightarrow}\eta_d^{\leftrightarrow}\right]\right).
\end{aligned}$$

The marks $\eta_c^{\circlearrowleft}$ are gamma distributed with mean 1, so

$$\begin{aligned}
&= \frac{1}{\delta}\left(\frac{\gamma_0}{\zeta}\right)^2\left(\mathbb{E}\left[\sum_{c=1}^{\infty}\eta_c^{\leftrightarrow}\right] + \mathbb{E}\left[\sum_{c=1}^{\infty}\sum_{d\neq c}\eta_c^{\leftrightarrow}\eta_d^{\leftrightarrow}\right]\right)\\
&= \frac{1}{\delta}\left(\frac{\gamma_0}{\zeta}\right)^2\left(\frac{\gamma_0}{\zeta} + \mathbb{E}\left[\sum_{c=1}^{\infty}\sum_{d\neq c}\eta_c^{\leftrightarrow}\eta_d^{\leftrightarrow}\right]\right)\\
&= \frac{1}{\delta}\left(\frac{\gamma_0}{\zeta}\right)^2\left(\frac{\gamma_0}{\zeta} + \mathbb{E}\left[\sum_{c=1}^{\infty}\sum_{d=1}^{\infty}\eta_c^{\leftrightarrow}\eta_d^{\leftrightarrow}\right] - \mathbb{E}\left[\sum_{c=1}^{\infty}\eta_c^{\leftrightarrow}\eta_c^{\leftrightarrow}\right]\right)\\
&= \frac{1}{\delta}\left(\frac{\gamma_0}{\zeta}\right)^2\left(\frac{\gamma_0}{\zeta} + \mathbb{E}\left[\left(\sum_{c=1}^{\infty}\eta_c^{\leftrightarrow}\right)\left(\sum_{d=1}^{\infty}\eta_d^{\leftrightarrow}\right)\right] - \mathbb{E}\left[\sum_{c=1}^{\infty}\eta_c^{\leftrightarrow}\eta_c^{\leftrightarrow}\right]\right).
\end{aligned}$$

Using $\mathbb{E}\left[\left(\sum_{c=1}^{\infty} \eta_c^{\leftrightarrow}\right)\left(\sum_{d=1}^{\infty} \eta_d^{\leftrightarrow}\right)\right] = \frac{\gamma_0^2}{\zeta^2} + \frac{\gamma_0}{\zeta^2}$, we can write

$$= \frac{1}{\delta}\left(\frac{\gamma_0}{\zeta}\right)^2 \left(\frac{\gamma_0}{\zeta} + \frac{\gamma_0^2}{\zeta^2} + \frac{\gamma_0}{\zeta^2} - \mathbb{E}\left[\sum_{c=1}^{\infty} \eta_c^{\leftrightarrow} \eta_c^{\leftrightarrow}\right]\right).$$

Finally, using Campbell's Theorem (Kingman, 1972), we know that $\mathbb{E}\left[\sum_{c=1}^{\infty} \eta_c^{\leftrightarrow} \eta_c^{\leftrightarrow}\right] = \frac{\gamma_0}{\zeta^2}$, so

$$\begin{aligned}
&= \frac{1}{\delta}\left(\frac{\gamma_0}{\zeta}\right)^2 \left(\frac{\gamma_0}{\zeta} + \frac{\gamma_0^2}{\zeta^2} + \frac{\gamma_0}{\zeta^2} - \frac{\gamma_0}{\zeta^2}\right) \\
&= \frac{1}{\delta}\left(\frac{\gamma_0}{\zeta}\right)^2 \left(\frac{\gamma_0}{\zeta} + \frac{\gamma_0^2}{\zeta^2}\right) \\
&= \frac{1}{\delta}\left(\frac{\gamma_0^3}{\zeta^3} + \frac{\gamma_0^4}{\zeta^4}\right).
\end{aligned}$$

2 Proposition 2

For an M-dimensional core tensor with $D_1 \times \ldots \times D_M$ elements, computing the normalizing constant using non-compositional allocation requires $1 \leq \pi < \infty$ times the number of operations required by compositional allocation. When $D_1 = \ldots = D_M = 1$, $\pi = 1$. As $D_m, D_{m'} \to \infty$ for any m and $m' \neq m$, $\pi \to \infty$.

Each event token occurs in an M-dimensional discrete coordinate space—i.e., $e_n = \boldsymbol{p}$, where $\boldsymbol{p} = (p_1, \ldots, p_M)$ is a multi-index. Similarly, each event token's latent class assignment also occurs in an M-dimensional discrete coordinate space—i.e., $z_n = \boldsymbol{q}$, where $\boldsymbol{q} = (q_1, \ldots, q_M)$ is a multi-index.

Assuming M factor matrices $\Theta^{(1)}, \ldots, \Theta^{(M)}$ and an M-dimensional core tensor $\boldsymbol{\Lambda}$,

$$P(z_n = \boldsymbol{q} \mid e_n = \boldsymbol{p}) \propto \lambda_{\boldsymbol{q}} \prod_{m=1}^{M} \theta_{p_m q_m}^{(m)}.$$

The computational bottleneck in MCMC inference is computing the normalizing constant

$$Z_{\boldsymbol{p}} = \sum_{\boldsymbol{q}} \lambda_{\boldsymbol{q}} \prod_{m=1}^{M} \theta_{p_m q_m}^{(m)}.$$

If we use a naïve non-compositional approach, then (assuming each latent dimension m has cardinality D_m) the sum over $\boldsymbol{q}$ involves $\prod_{m=1}^{M} D_m$ terms and each term requires M multiplications. Thus, computing $Z_{\boldsymbol{p}}$ requires a total of $M \prod_{m=1}^{M} D_m$ multiplications and $\prod_{m=1}^{M} D_m$ additions.[1]

However, we can also compute $Z_{\boldsymbol{p}}$ using a compositional approach—i.e.,

$$Z_{\boldsymbol{p}} = \sum_{q_1=1}^{D_1} \theta_{p_1 q_1}^{(1)} \sum_{q_2=1}^{D_2} \theta_{p_2 q_2}^{(2)} \ldots \sum_{q_M=1}^{D_M} \theta_{p_M q_M}^{(M)} \lambda_{\boldsymbol{q}}.$$

[1]Computing a sum of N terms requires either N or $N-1$ additions, depending on whether or not you add the first term to zero. We assume the former definition and say that computing a sum of N terms requires N additions.

This approach requires a total of $\sum_{m=1}^{M} D_m$ multiplications and $1 + \sum_{m=1}^{M}(D_m - 1)$ additions.

The ratio π of the number of operations (i.e., multiplications and additions) required by the non-compositional approach to the number of operations required by the compositional approach is

$$\begin{aligned}\pi &= \frac{\left(M \prod_{m=1}^{M} D_m\right) + \left(\prod_{m=1}^{M} D_m\right)}{\left(\sum_{m=1}^{M} D_m\right) + \left(1 + \sum_{m=1}^{M}(D_m - 1)\right)} \\ &= \frac{(M+1)\prod_{m=1}^{M} D_m}{\left(2\sum_{m=1}^{M} D_m\right) - M + 1}.\end{aligned}$$

As the cardinalities $D_1, \ldots, D_M$ of the latent dimensions grow, the numerator grows at a faster rate than the denominator. Therefore π achieves its lower bound when $D_1 = \ldots = D_M = 1$:

$$\Omega(\pi) = \frac{(M+1)}{(2M) - M + 1}.$$

Because the numerator grows at a faster rate than the denominator, we can find the upper bound by taking the limit as one or more cardinalities tend to infinity. We work with the inverse ratio

$$\begin{aligned}\pi^{-1} &= \frac{\left(2\sum_{m=1}^{M} D_m\right) - M + 1}{(M+1)\prod_{m=1}^{M} D_m} \\ &= \frac{2}{M+1}\left(\sum_{m=1}^{M} \frac{D_m}{\prod_{m=1}^{M} D_m}\right) - \frac{M-1}{M+1}\left(\frac{1}{\prod_{m=1}^{M} D_m}\right) \\ &= \frac{2}{M+1}\left(\sum_{m=1}^{M} \frac{1}{\prod_{m' \neq m} D_{m'}}\right) - \frac{M-1}{M+1}\left(\frac{1}{\prod_{m=1}^{M} D_m}\right).\end{aligned}$$

First, we take the limit of π^{-1} as a single cardinality $D_m \to \infty$:

$$\begin{aligned}\lim_{D_m \to \infty} \pi^{-1} &= \lim_{D_m \to \infty} \frac{2}{M+1}\left(\sum_{m=1}^{M} \frac{1}{\prod_{n \neq m} D_n}\right) - \lim_{D_m \to \infty} \frac{M-1}{M+1}\left(\frac{1}{\prod_{m=1}^{M} D_m}\right) \\ &= \lim_{D_m \to \infty} \frac{2}{M+1}\left(\sum_{m=1}^{M} \frac{1}{\prod_{n \neq m} D_n}\right) \\ &= \frac{2}{M+1}\left(\frac{1}{\prod_{n \neq m} D_n}\right).\end{aligned}$$

However, as any second cardinality $D_{m'} \to \infty$,

$$\lim_{D_m, D_{m'} \to \infty} \pi^{-1} = \lim_{D_{m'} \to \infty} \frac{2}{M+1}\left(\frac{1}{\prod_{n \neq m} D_n}\right) \to 0.$$

Therefore, $\pi \to \infty$ as any two (or more) cardinalities tend to infinity.

3 Inference

Gibbs sampling repeatedly resamples the value of each latent variable from its conditional posterior. In this section, we provide the conditional posterior for each latent variable in BPTD.

We start by defining the Chinese restaurant table (CRT) distribution (Zhou & Carin, 2015): If $l \sim \text{CRT}(m, r)$ is a CRT-distributed random variable, then, we can equivalently say that

$$l \sim \sum_{n=1}^{m} \text{Bern}\left(\frac{r}{r+n-1}\right).$$

We also define $g(x) \equiv \ln(1+x)$.

Throughout this section, we use, e.g., $(\theta_{ic} \,|\, -)$ to denote θ_{ic} conditioned on $\boldsymbol{Y}$, ϵ_0, γ_0, and the current values of the other latent variables. We assume that $\boldsymbol{Y}$ is partially observed and include a binary mask $\boldsymbol{B}$, where $b^{(t)}_{i \xrightarrow{a} j} = 0$ means that $y^{(t)}_{i \xrightarrow{a} j} = 0$ is unobserved, not an observed zero.

Action–Topic Factors:

$$\begin{aligned}
y^{(\cdot)}_{\cdot \overset{ak}{\leftrightarrow} \cdot} &\equiv \sum_{i=1}^{V} \sum_{c=1}^{C} \sum_{j \neq i} \sum_{d=1}^{C} \sum_{t=1}^{T} \sum_{r=1}^{R} y^{(tr)}_{ic \xrightarrow{ak} dj} \\
\xi_{ak} &\equiv \sum_{i=1}^{V} \sum_{j \neq i} \sum_{t=1}^{T} b^{(t)}_{i \xrightarrow{a} j} \sum_{c=1}^{C} \theta_{ic} \sum_{d=1}^{C} \theta_{jd} \sum_{r=1}^{R} \psi_{tr} \, \lambda^{(r)}_{c \xrightarrow{k} d} \\
(\phi_{ak} \,|\, -) &\sim \Gamma\left(\epsilon_0 + y^{(\cdot)}_{\cdot \overset{ak}{\leftrightarrow} \cdot}, \epsilon_0 + \xi_{ak}\right)
\end{aligned}$$

Time-Step–Regime Factors:

$$\begin{aligned}
y^{(tr)}_{\cdot \xrightarrow{\cdot} \cdot} &\equiv \sum_{i=1}^{V} \sum_{c=1}^{C} \sum_{j \neq i} \sum_{d=1}^{C} \sum_{a=1}^{A} \sum_{k=1}^{K} y^{(tr)}_{ic \xrightarrow{ak} dj} \\
\xi_{tr} &\equiv \sum_{i=1}^{V} \sum_{j \neq i} \sum_{a=1}^{A} b^{(t)}_{i \xrightarrow{a} j} \sum_{c=1}^{C} \theta_{ic} \sum_{d=1}^{C} \theta_{jd} \sum_{k=1}^{K} \phi_{ak} \, \lambda^{(r)}_{c \xrightarrow{k} d} \\
(\psi_{tr} \,|\, -) &\sim \Gamma\left(\epsilon_0 + y^{(tr)}_{\cdot \xrightarrow{\cdot} \cdot}, \epsilon_0 + \xi_{tr}\right)
\end{aligned}$$

Country–Community Factors:

$$y^{(\cdot)}_{ic\leftrightarrow\cdot} \equiv \sum_{j\neq i}\sum_{d=1}^{C}\sum_{a=1}^{A}\sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{r=1}^{R}\left(y^{(tr)}_{ic\xrightarrow{ak}dj} + y^{(tr)}_{jd\xrightarrow{ak}ci}\right)$$

$$\xi_{ic} \equiv \sum_{j\neq i}\sum_{a=1}^{A}\sum_{t=1}^{T}\left(b^{(t)}_{i\xrightarrow{a}j}\sum_{d=1}^{C}\theta_{jd}\sum_{k=1}^{K}\phi_{ak}\sum_{r=1}^{R}\psi_{tr}\,\lambda^{(r)}_{c\xrightarrow{k}d} + b^{(t)}_{j\xrightarrow{a}i}\sum_{d=1}^{C}\theta_{jd}\sum_{k=1}^{K}\phi_{ak}\sum_{r=1}^{R}\psi_{tr}\,\lambda^{(r)}_{d\xrightarrow{k}c}\right)$$

$$(\theta_{ic}\,|\,-) \sim \Gamma\left(\alpha_i + y^{(\cdot)}_{ic\leftrightarrow\cdot},\, \beta_i + \xi_{ic}\right)$$

Auxiliary Latent Country–Community Counts:

$$(\ell_{ic}\,|\,-) \sim \mathrm{CRT}\left(y^{(\cdot)}_{ic\leftrightarrow\cdot},\, \alpha_i\right)$$

Per-Country Shape Parameters:

$$(\alpha_i\,|\,-) \sim \Gamma\left(\epsilon_0 + \sum_{c=1}^{C}\ell_{ic},\, \epsilon_0 + \sum_{c=1}^{C} g\left(\xi_{ic}\,\beta_i^{-1}\right)\right)$$

Per-Country Rate Parameters:

$$(\beta_i\,|\,-) \sim \Gamma\left(\epsilon_0 + C\alpha_i,\, \epsilon_0 + \sum_{c=1}^{C}\theta_{ic}\right)$$

Diagonal Elements of the Core Tensor:

$$\omega^{(r)}_{c\circlearrowleft k} \equiv \eta^{\circlearrowleft}_{c}\eta^{\leftrightarrow}_{c}\nu_k\rho_r$$

$$y^{(r)}_{c\circlearrowleft k} \equiv \sum_{i=1}^{V}\sum_{j\neq i}\sum_{a=1}^{A}\sum_{t=1}^{T} y^{(tr)}_{ic\xrightarrow{ak}cj}$$

$$\xi^{(r)}_{c\circlearrowleft k} \equiv \sum_{i=1}^{V}\theta_{ic}\sum_{j\neq i}\theta_{jc}\sum_{a=1}^{A}\phi_{ak}\sum_{t=1}^{T}\psi_{tr}\,b^{(t)}_{i\xrightarrow{a}j}$$

$$\left(\lambda^{(r)}_{c\circlearrowleft k}\,|\,-\right) \sim \Gamma\left(\omega^{(r)}_{c\circlearrowleft k} + y^{(r)}_{c\circlearrowleft k},\, \delta + \xi^{(r)}_{c\circlearrowleft k}\right)$$

Off-Diagonal Elements of the Core Tensor:

$$\omega^{(r)}_{c \xrightarrow{k} d} \equiv \eta_c^{\leftrightarrow} \eta_d^{\leftrightarrow} \nu_k \rho_r \qquad c \neq d$$

$$y^{(r)}_{c \xrightarrow{k} d} \equiv \sum_{i=1}^{V} \sum_{j \neq i} \sum_{a=1}^{A} \sum_{t=1}^{T} y^{(tr)}_{ic \xrightarrow{ak} dj} \qquad c \neq d$$

$$\xi^{(r)}_{c \xrightarrow{k} d} \equiv \sum_{i=1}^{V} \theta_{ic} \sum_{j \neq i} \theta_{jd} \sum_{a=1}^{A} \phi_{ak} \sum_{t=1}^{T} \psi_{tr} \, b^{(t)}_{i \xrightarrow{a} j} \qquad c \neq d$$

$$\left(\lambda^{(r)}_{c \xrightarrow{k} d} \,|\, -\right) \sim \Gamma\left(\omega^{(r)}_{c \xrightarrow{k} d} + + y^{(r)}_{c \xrightarrow{k} d}, \delta + \xi^{(r)}_{c \xrightarrow{k} d}\right) \qquad c \neq d$$

Core Rate Parameter:

$$\omega^{(\cdot)}_{\cdot \overset{\cdot}{\leftrightarrow} \cdot} \equiv \sum_{c=1}^{C} \sum_{k=1}^{K} \sum_{r=1}^{R} \left(\omega^{(r)}_{c \circlearrowleft k} + \sum_{d \neq c} \omega^{(r)}_{c \xrightarrow{k} d}\right)$$

$$\lambda^{(\cdot)}_{\cdot \overset{\cdot}{\leftrightarrow} \cdot} \equiv \sum_{c=1}^{C} \sum_{k=1}^{K} \sum_{r=1}^{R} \left(\lambda^{(r)}_{c \circlearrowleft k} + \sum_{d \neq c} \lambda^{(r)}_{c \xrightarrow{k} d}\right)$$

$$(\delta \,|\, -) \sim \Gamma\left(\epsilon_0 + \omega^{(\cdot)}_{\cdot \overset{\cdot}{\leftrightarrow} \cdot}, \epsilon_0 + \lambda^{(\cdot)}_{\cdot \overset{\cdot}{\leftrightarrow} \cdot}\right)$$

Diagonal Auxiliary Latent Core Counts:

$$\ell^{(r)}_{c \circlearrowleft k} \sim \mathrm{CRT}\left(y^{(r)}_{c \circlearrowleft k}, \omega^{(r)}_{c \circlearrowleft k}\right)$$

Off-Diagonal Auxiliary Latent Core Counts:

$$\ell^{(r)}_{c \xrightarrow{k} d} \sim \mathrm{CRT}\left(y^{(r)}_{c \xrightarrow{k} d}, \omega^{(r)}_{c \xrightarrow{k} d}\right) \qquad c \neq d$$

Within-Community Weights:

$$\ell^{(\cdot)}_{c \circlearrowleft \cdot} \equiv \sum_{k=1}^{K} \sum_{r=1}^{R} \ell^{(r)}_{c \circlearrowleft k}$$

$$\xi_c^{\circlearrowleft} \equiv \sum_{r=1}^{R} \rho_r \sum_{k=1}^{K} \nu_k \sum_{d \neq c} \eta_d^{\leftrightarrow} \left(g\left(\xi^{(r)}_{c \xrightarrow{k} d} \delta^{-1}\right) + g\left(\xi^{(r)}_{d \xrightarrow{k} c} \delta^{-1}\right)\right)$$

$$(\eta_c^{\circlearrowleft} \,|\, -) \sim \Gamma\left(\frac{\gamma_0}{C} + \ell^{(\cdot)}_{c \circlearrowleft \cdot}, \zeta + \xi_c^{\circlearrowleft}\right)$$

Between-Community Weights:

$$
\begin{aligned}
\ell^{(\cdot)}_{c\overset{\cdot}{\leftrightarrow}\cdot} &\equiv \ell^{(\cdot)}_{c\circlearrowleft\cdot} + \sum_{d\neq c}\sum_{k=1}^{K}\sum_{r=1}^{R}\left(\ell^{(r)}_{c\xrightarrow{k}d} + \ell^{(r)}_{d\xrightarrow{k}c}\right) \\
\xi^{\leftrightarrow}_{c} &\equiv \sum_{r=1}^{R}\rho_r\sum_{k=1}^{K}\nu_k\left[\eta^{\circlearrowleft}_{c}\, g\left(\xi^{(r)}_{c\circlearrowleft^{k}}\,\delta^{-1}\right) + \sum_{d\neq c}\eta^{\leftrightarrow}_{d}\left(g\left(\xi^{(r)}_{c\xrightarrow{k}d}\,\delta^{-1}\right) + g\left(\xi^{(r)}_{d\xrightarrow{k}c}\,\delta^{-1}\right)\right)\right] \\
(\eta^{\leftrightarrow}_{c}\,|\,-) &\sim \Gamma\left(\frac{\gamma_0}{C} + \ell^{(\cdot)}_{c\overset{\cdot}{\leftrightarrow}\cdot},\, \zeta + \xi^{\leftrightarrow}_{c}\right)
\end{aligned}
$$

Topic Weights:

$$
\begin{aligned}
\ell^{(\cdot)}_{\cdot\xrightarrow{k}\cdot} &\equiv \sum_{c=1}^{C}\sum_{d=1}^{C}\sum_{r=1}^{R}\ell^{(r)}_{c\xrightarrow{k}d} \\
\xi_k &\equiv \sum_{r=1}^{R}\rho_r\sum_{c=1}^{C}\eta^{\leftrightarrow}_{c}\left[\eta^{\circlearrowleft}_{c}\, g\left(\xi^{(r)}_{c\circlearrowleft^{k}}\,\delta^{-1}\right) + \sum_{d\neq c}\eta^{\leftrightarrow}_{d}\left(g\left(\xi^{(r)}_{c\xrightarrow{k}d}\,\delta^{-1}\right) + g\left(\xi^{(r)}_{d\xrightarrow{k}c}\,\delta^{-1}\right)\right)\right] \\
(\nu_k\,|\,-) &\sim \Gamma\left(\frac{\gamma_0}{K} + \ell^{(\cdot)}_{\cdot\xrightarrow{k}\cdot},\, \zeta + \xi_k\right)
\end{aligned}
$$

Regime Weights:

$$
\begin{aligned}
\ell^{(r)}_{\cdot\overset{\cdot}{\rightarrow}\cdot} &\equiv \sum_{c=1}^{C}\sum_{d=1}^{C}\sum_{k=1}^{K}\ell^{(r)}_{c\xrightarrow{k}d} \\
\xi_r &\equiv \sum_{k=1}^{K}\nu_k\sum_{c=1}^{C}\eta^{\leftrightarrow}_{c}\left[\eta^{\circlearrowleft}_{c}\, g\left(\xi^{(r)}_{c\circlearrowleft^{k}}\,\delta^{-1}\right) + \sum_{d\neq c}\eta^{\leftrightarrow}_{d}\left(g\left(\xi^{(r)}_{c\xrightarrow{k}d}\,\delta^{-1}\right) + g\left(\xi^{(r)}_{d\xrightarrow{k}c}\,\delta^{-1}\right)\right)\right] \\
(\rho_r\,|\,-) &\sim \Gamma\left(\frac{\gamma_0}{R} + \ell^{(r)}_{\cdot\overset{\cdot}{\rightarrow}\cdot},\, \zeta + \xi_r\right)
\end{aligned}
$$

Weights Rate Parameter:

$$
\begin{aligned}
\omega &\equiv \sum_{c=1}^{C}\eta^{\circlearrowleft}_{c} + \sum_{c=1}^{C}\eta^{\leftrightarrow}_{c} + \sum_{k=1}^{K}\nu_k + \sum_{r=1}^{R}\rho_r \\
(\zeta\,|\,-) &\sim \Gamma\left(\epsilon_0 + 4\gamma_0,\, \epsilon_0 + \omega\right)
\end{aligned}
$$

4 Baseline Models

BPTF (Schein et al., 2015):

$$
\begin{aligned}
y^{(t)}_{i\xrightarrow{a}j} &\sim \mathrm{Po}\Bigg(\sum_{q=1}^{Q} \theta^{\rightarrow}_{iq}\,\theta^{\leftarrow}_{jq}\,\phi_{aq}\,\psi_{tq}\,\lambda_q\Bigg)\\
\theta^{\rightarrow}_{iq} &\sim \Gamma\left(\epsilon_0, \beta_1\right)\\
\theta^{\leftarrow}_{jq} &\sim \Gamma\left(\epsilon_0, \beta_2\right)\\
\phi_{aq} &\sim \Gamma\left(\epsilon_0, \beta_3\right)\\
\psi_{tq} &\sim \Gamma\left(\epsilon_0, \beta_4\right)\\
\lambda_q &\sim \Gamma\left(\frac{\gamma_0}{Q}, \delta\right)\\
\beta_1, \cdots, \beta_4, \delta &\sim \Gamma\left(\epsilon_0, \epsilon_0\right)
\end{aligned}
$$

GPIRM (Schmidt & Mørup, 2013):

$$
\begin{aligned}
y^{(t)}_{i\xrightarrow{a}j} &\sim \mathrm{Po}\Big(\lambda^{(z_t)}_{z_i\xrightarrow{z_a}z_j}\Big)\\
z_i &\sim \mathrm{Cat}\left(\frac{\eta_1}{\sum_c \eta_c}, \ldots, \frac{\eta_C}{\sum_c \eta_c}\right)\\
z_a &\sim \mathrm{Cat}\left(\frac{\nu_1}{\sum_k \nu_k}, \ldots, \frac{\nu_K}{\sum_k \nu_k}\right)\\
z_t &\sim \mathrm{Cat}\left(\frac{\rho_1}{\sum_r \rho_r}, \ldots, \frac{\rho_R}{\sum_r \rho_r}\right)\\
\eta_c &\sim \Gamma\left(\frac{\gamma_0}{C}, \zeta\right)\\
\nu_k &\sim \Gamma\left(\frac{\gamma_0}{K}, \zeta\right)\\
\rho_r &\sim \Gamma\left(\frac{\gamma_0}{R}, \zeta\right)\\
\lambda^{(r)}_{c\xrightarrow{k}d}, \zeta &\sim \Gamma\left(\epsilon_0, \epsilon_0\right)
\end{aligned}
$$

DCGPIRM:

$$
\begin{aligned}
y^{(t)}_{i\xrightarrow{a}j} &\sim \mathrm{Po}\Big(\theta_i\,\theta_j\,\phi_a\,\psi_t\,\lambda^{(z_t)}_{z_i\xrightarrow{z_a}z_j}\Big)\\
\theta_i, \phi_a, \psi_t &\sim \Gamma(\epsilon_0, \epsilon_0)
\end{aligned}
$$

The rest of the generative process is the same as that of the GPIRM.

5 Supplementary Plots

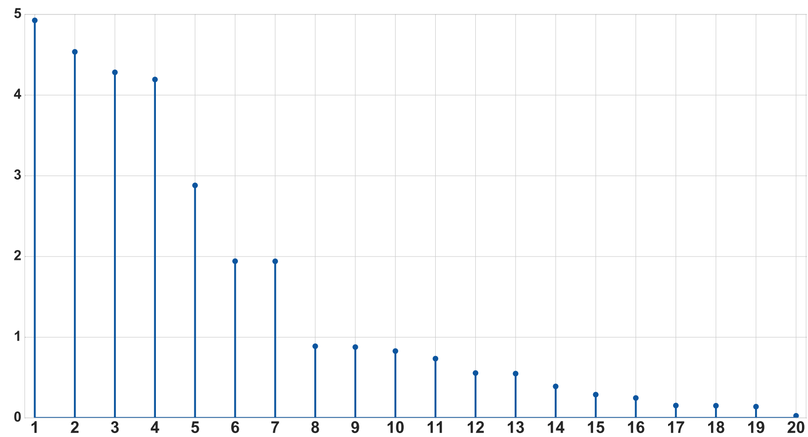

Figure 1: Inferred community weights $\eta_1^{\leftrightarrow}, \ldots, \eta_C^{\leftrightarrow}$. We use the between-community weights to interpret shrinkage because they are used for the on- and off-diagonal elements of the core tensor.

References

Kingman, J. F. C. *Poisson Processes*. Oxford University Press, 1972.

Schein, A., Paisley, J., Blei, D. M., and Wallach, H. Bayesian Poisson tensor factorization for inferrring multilateral relations from sparse dyadic event counts. In *Proceedings of the Twenty-First ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1045–1054, 2015.

Schmidt, M. N. and Mørup, M. Nonparametric Bayesian modeling of complex networks: An introduction. *IEEE Signal Processing Magazine*, 30(3):110–128, 2013.

Zhou, M. Infinite edge partition models for overlapping community detection and link prediction. In *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics*, pp. 1135–1143, 2015.

Zhou, M. and Carin, L. Negative binomial process count and mixture modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(2):307–320, 2015.

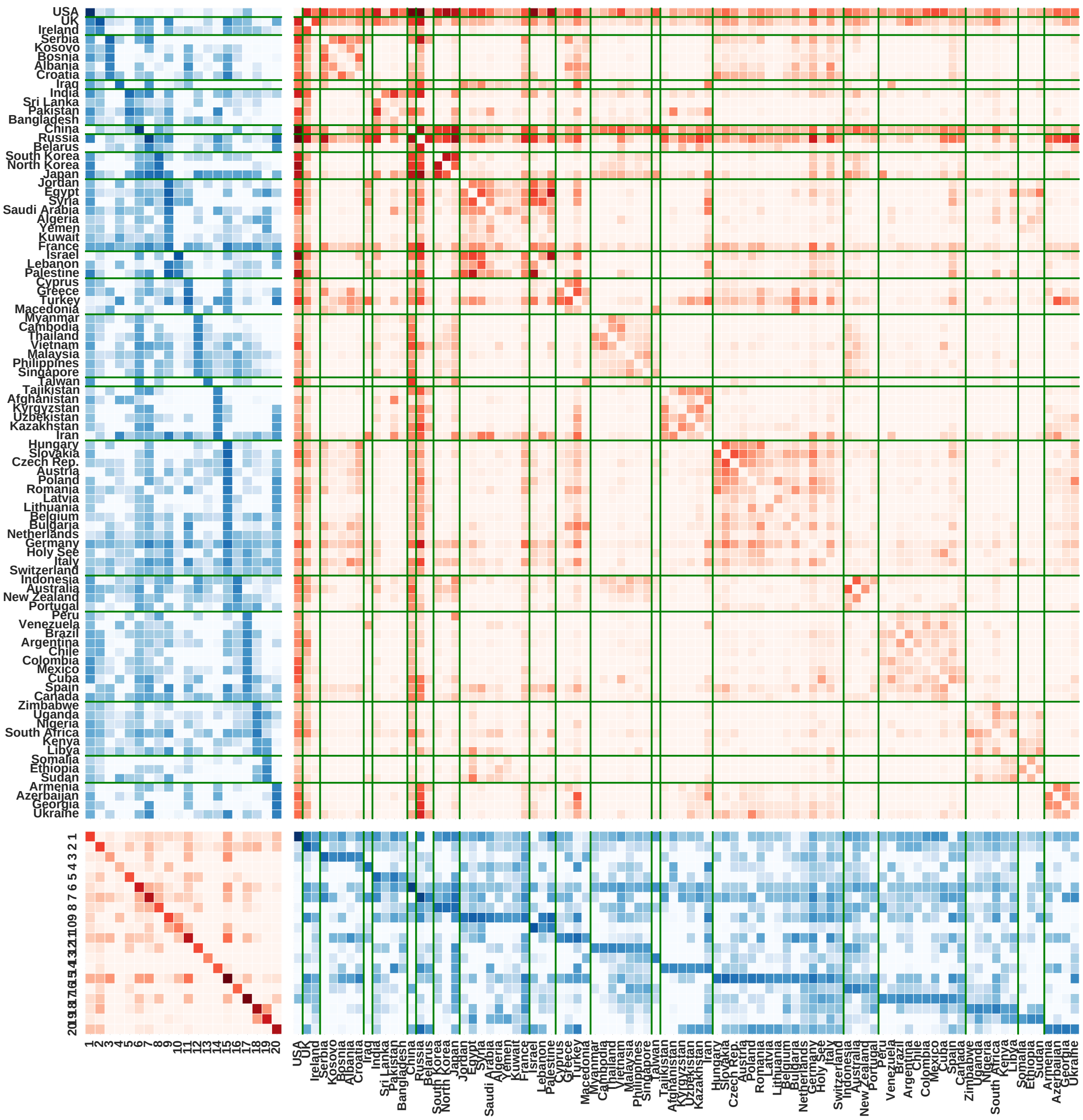

Figure 2: Latent structure discovered by BPTD for topic $k = 1$ (mostly Verbal Cooperation action types) and the most active regime, including the community–community interaction network (bottom left), the rate at which each country acts as a sender (top left) and a receiver (bottom right) in each community, and the number of times each country i took an action associated with topic k toward each country j during regime r (top right). We show only the most active 100 countries.

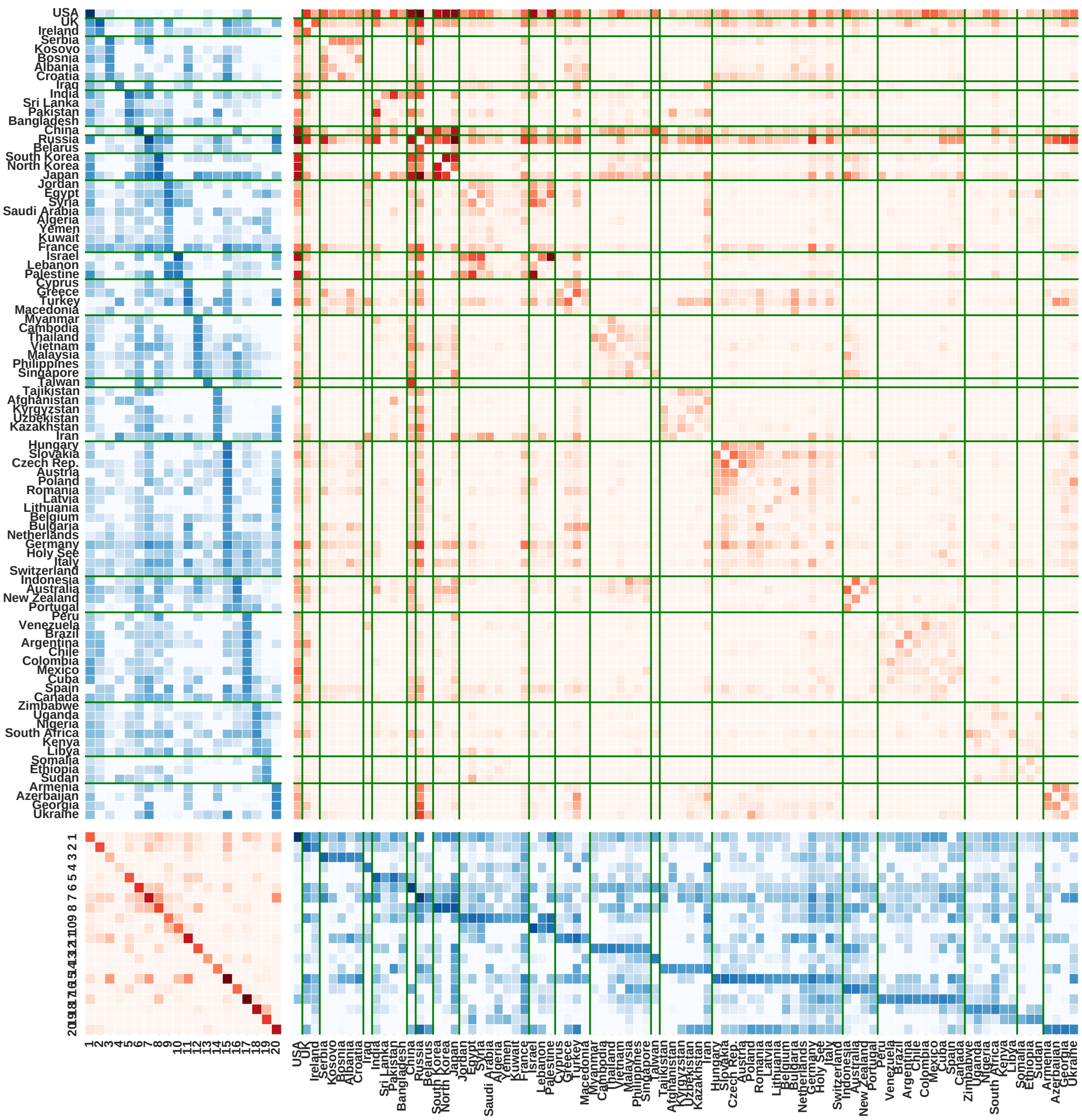

Figure 3: Latent structure discovered by BPTD for topic $k=2$ (Verbal Cooperation).

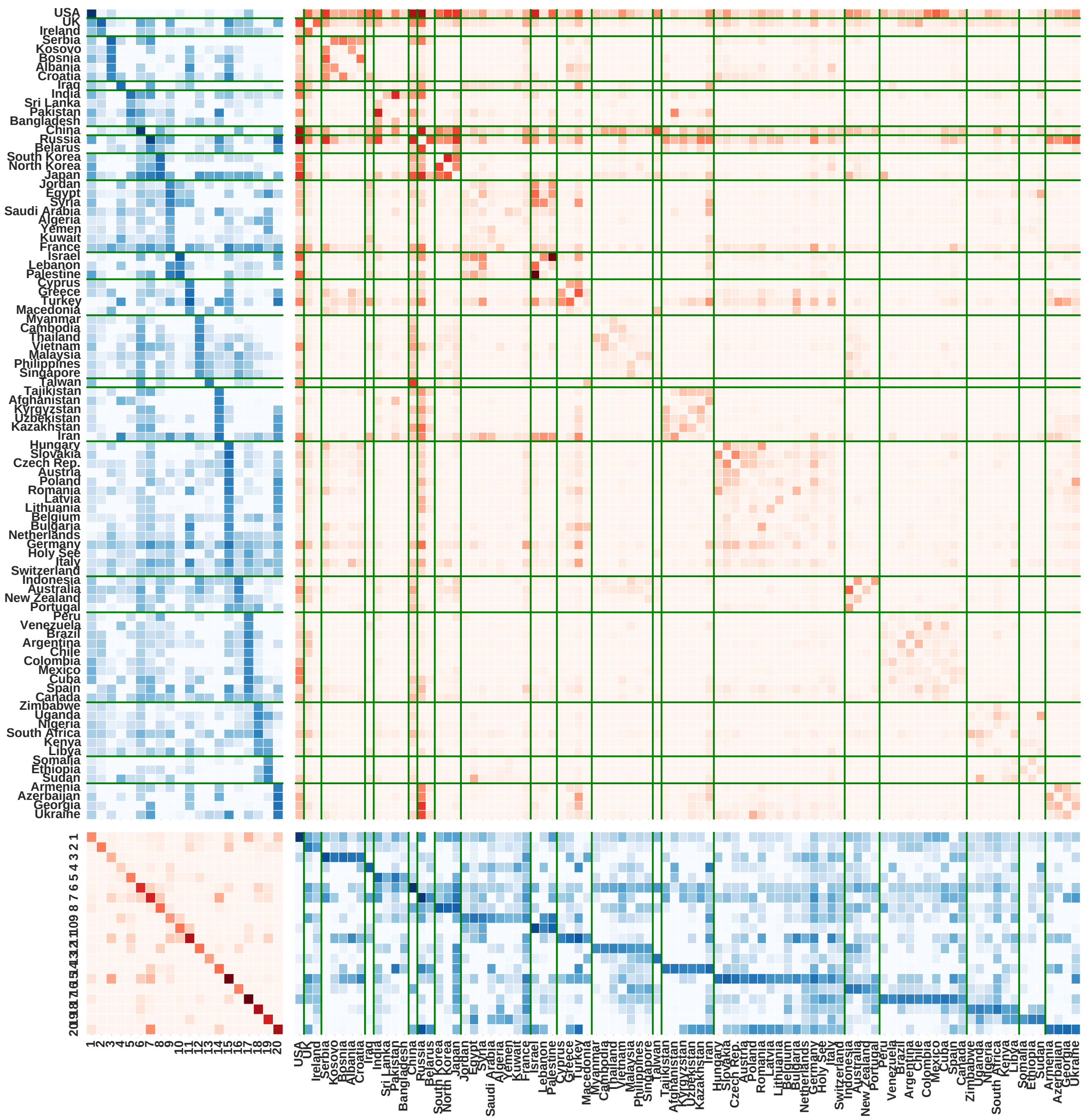

Figure 4: Latent structure discovered by BPTD for topic $k = 3$ (Material Cooperation).

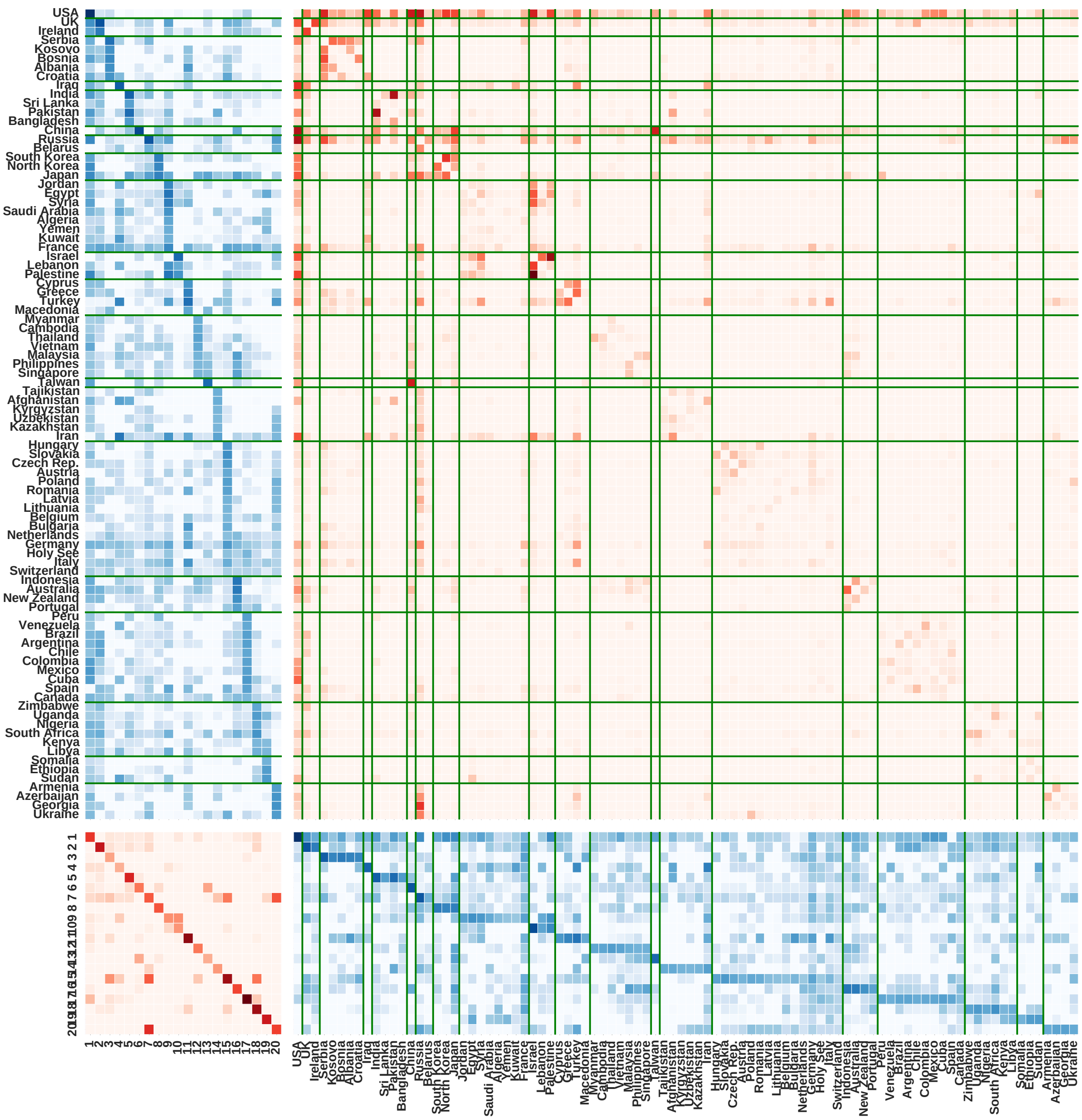

Figure 5: Latent structure discovered by BPTD for topic $k = 4$ (Verbal Conflict).

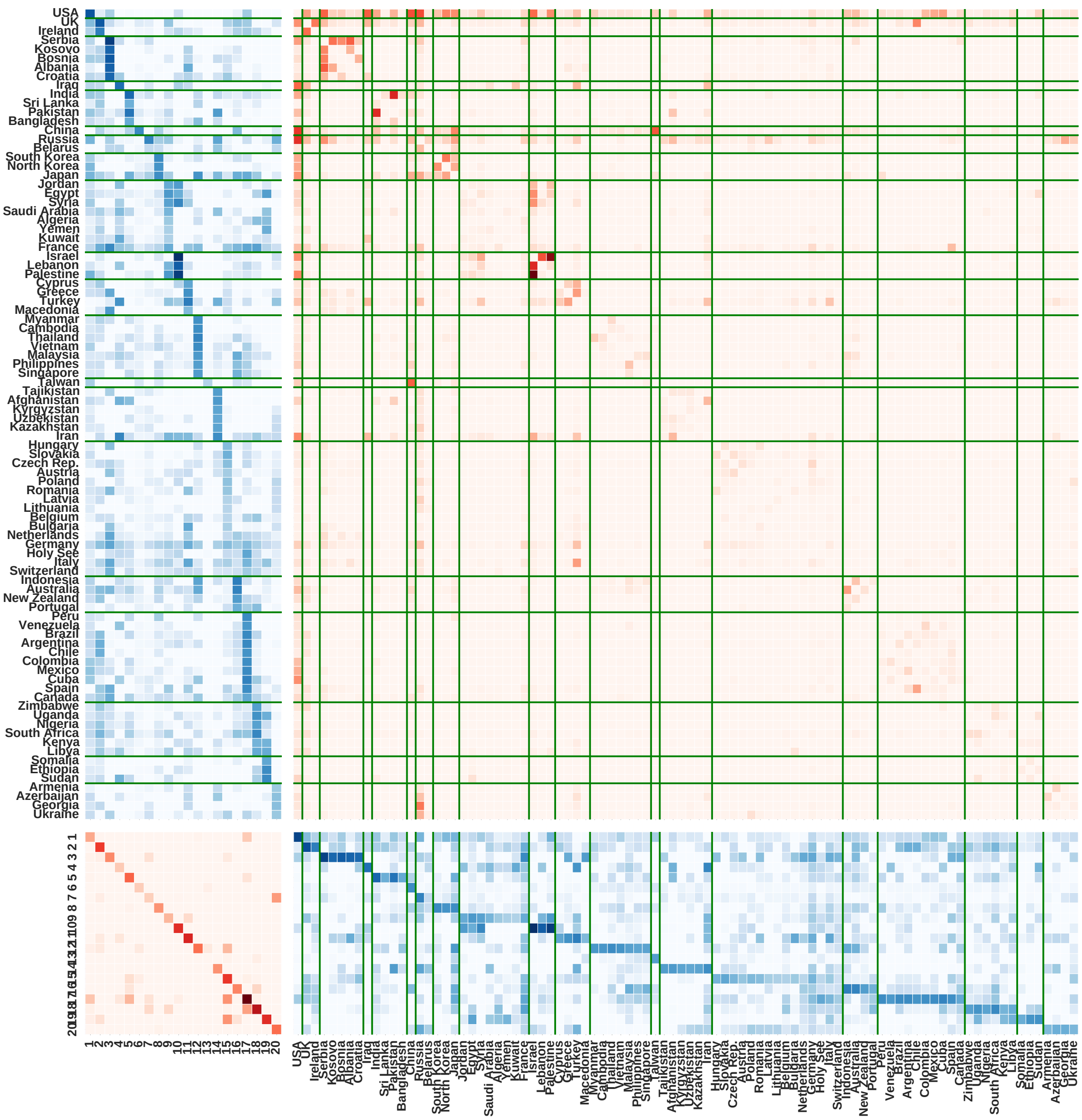

Figure 6: Latent structure discovered by BPTD for topic $k = 5$ (Material Conflict).

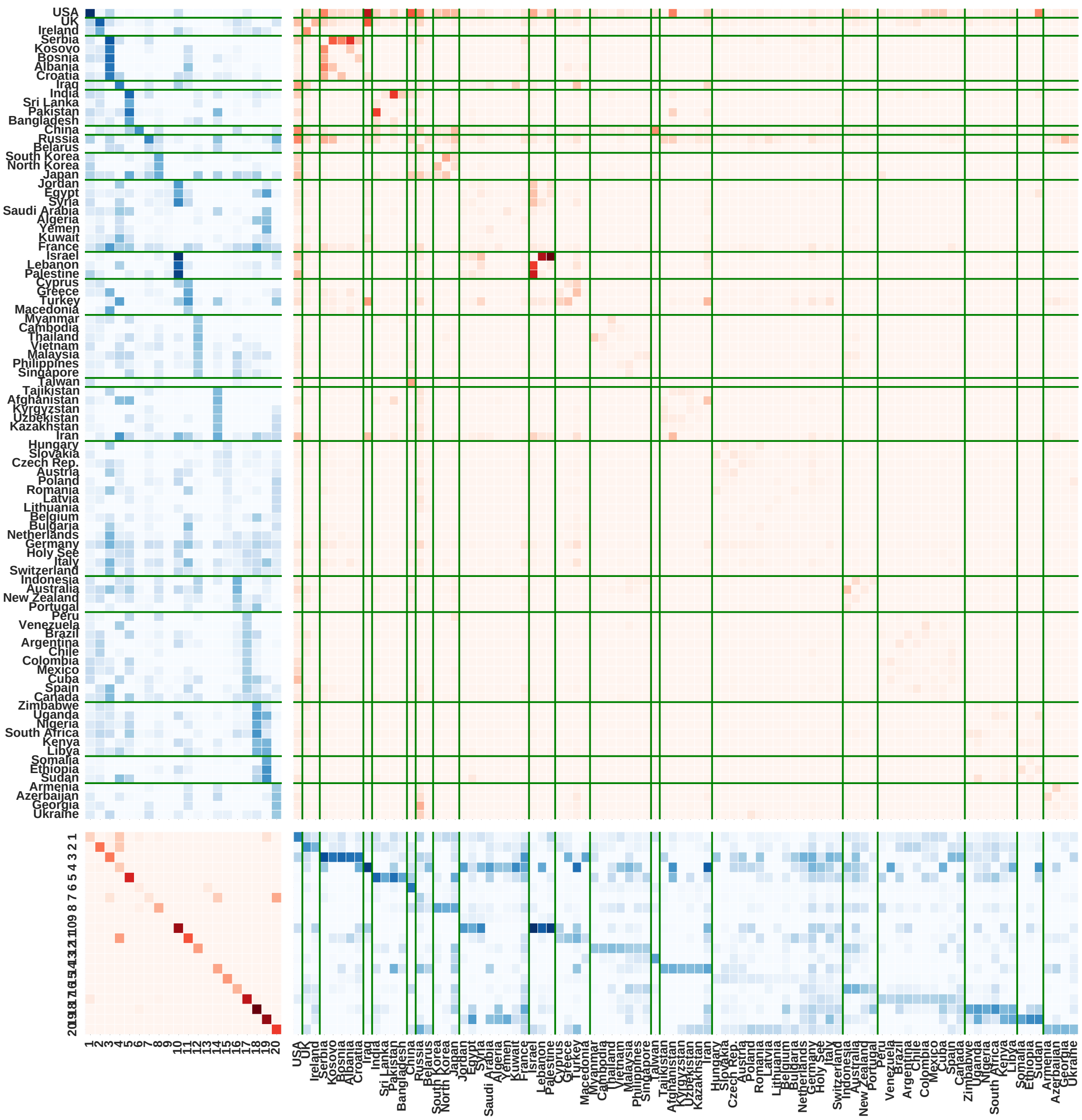

Figure 7: Latent structure discovered by BPTD for topic $k=6$ (Material Conflict).