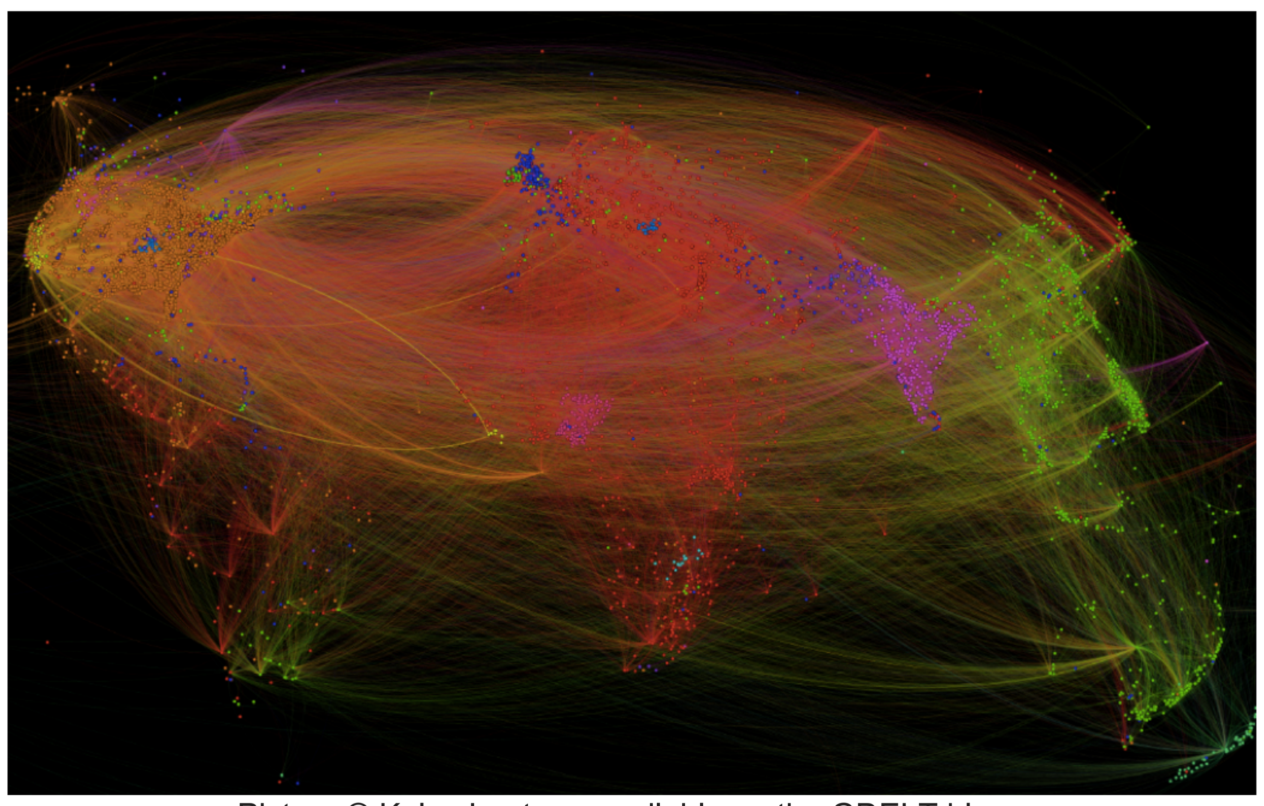
Bayesian Poisson Tucker Decomposition for Learning the Structure of International Relations

Aaron Schein — UMass Amherst
Mingyuan Zhou — Univ. of Texas Austin
David M. Blei — Columbia Univ.
Hanna Wallach — Microsoft Research

Interaction event data

who did what to whom, when

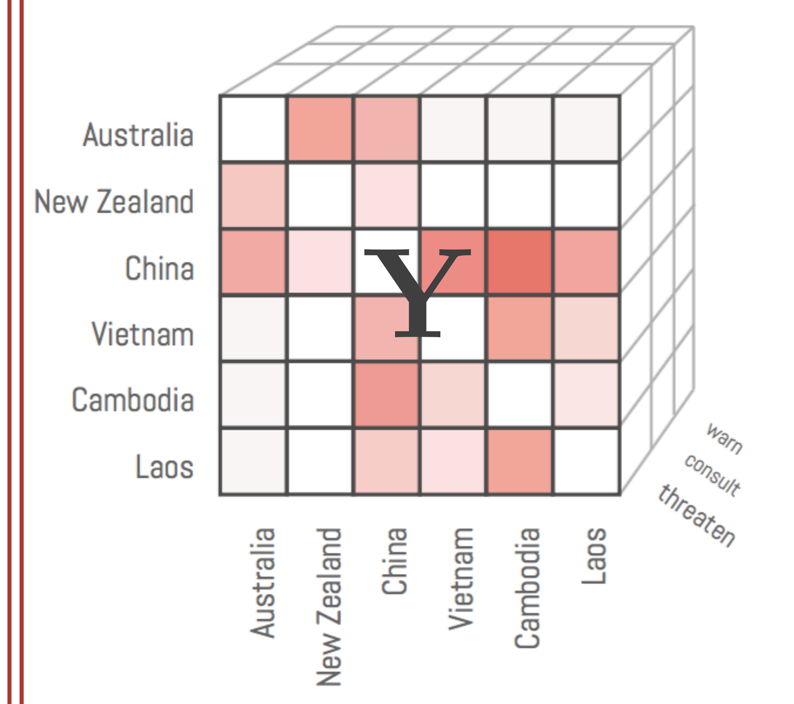


Picture © Kalev Leetaru, available on the GDELT blog

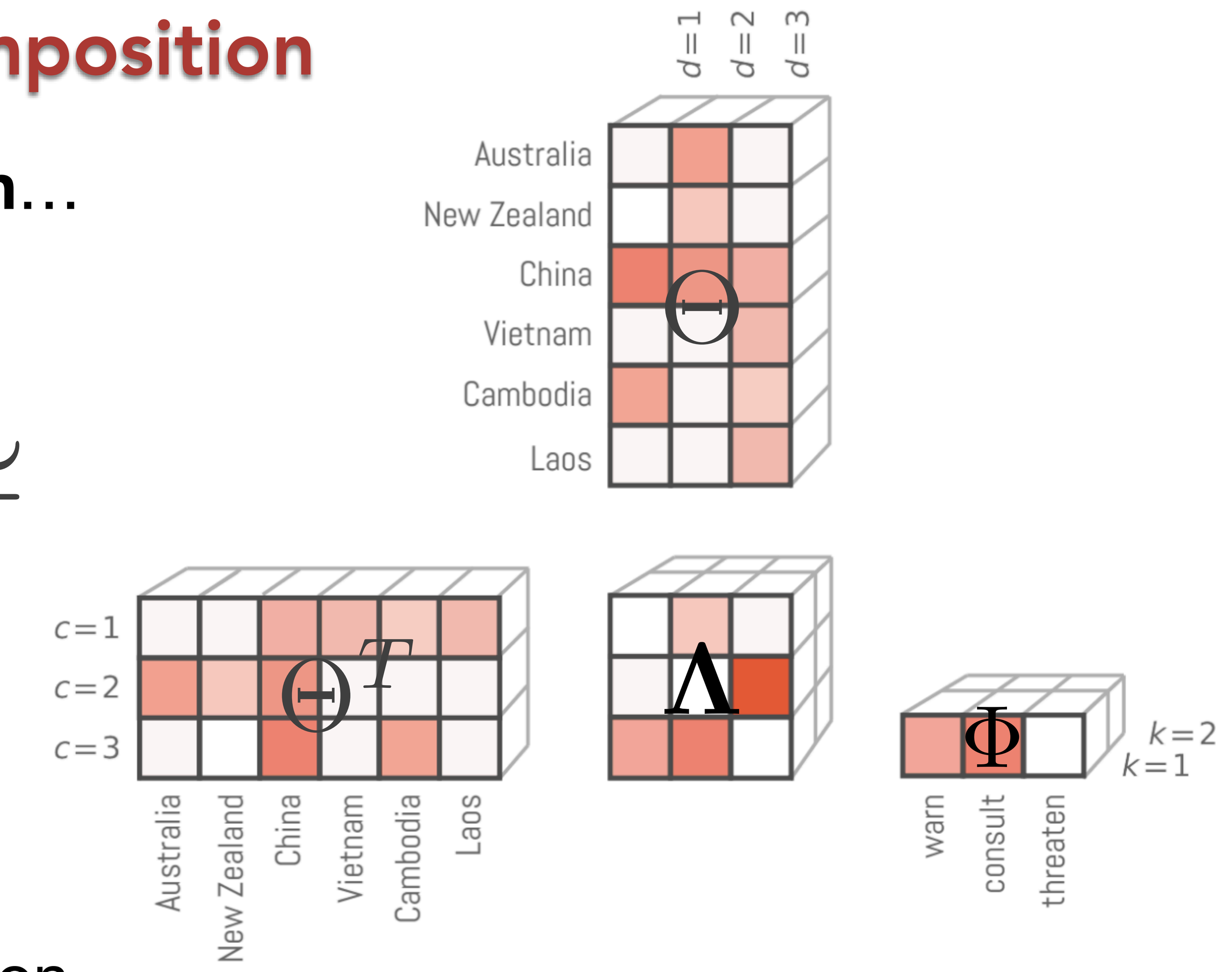
$y^{(t)}_{i \xrightarrow{a} j}$: number of instances country i took action a towards country j during time t

Poisson Tucker decomposition

A **Tucker decomposition**…



$\simeq$



…with a **Poisson** assumption.

$$y^{(t)}_{i \xrightarrow{a} j} \sim \text{Pois}\left(\sum_{c=1}^{C} \theta_{ic} \sum_{d=1}^{C} \theta_{jd} \sum_{k=1}^{K} \phi_{ak} \sum_{r=1}^{R} \psi_{rt}\, \lambda^{(r)}_{c \xrightarrow{k} d}\right)$$

Compositional allocation

$$P\Big(z_n = (c \xrightarrow{k} d, r) \,|\, e_n = (i \xrightarrow{a} j, t), \cdots\Big) \propto \theta_{ic}\,\theta_{jd}\,\phi_{ak}\,\psi_{tr}\,\lambda^{(r)}_{c \xrightarrow{k} d}$$

Event type:

$$(i \xrightarrow{a} j, t)$$

Event token:

$$e_n = (i \xrightarrow{a} j, t)$$

For N tokens:

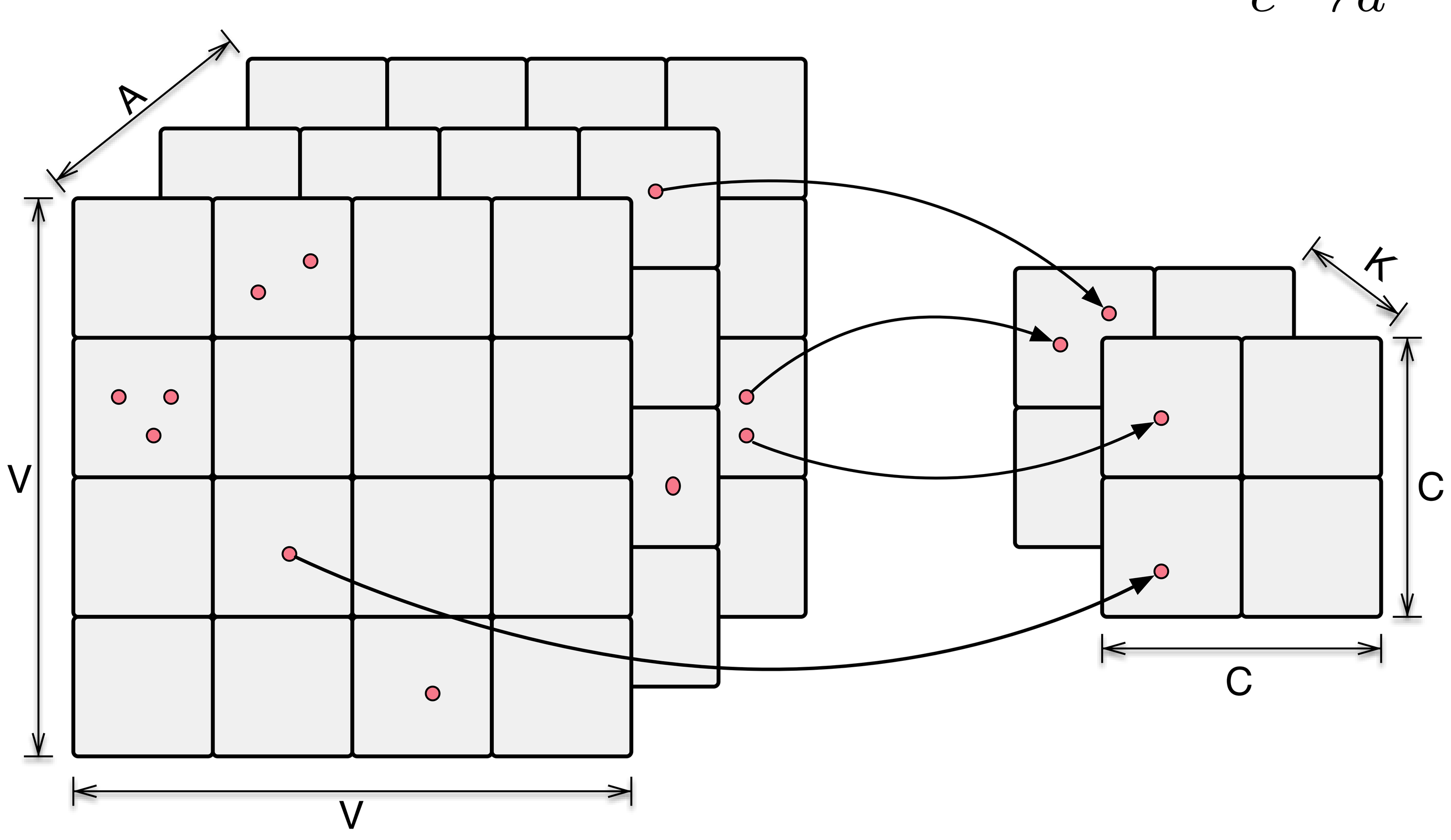
$$y^{(t)}_{i \xrightarrow{a} j} = \sum_{n=1}^{N} \delta[e_n = (i \xrightarrow{a} j, t)]$$

Latent event class:

$$(c \xrightarrow{k} d, r)$$

Token assignment:

$$z_n = (c \xrightarrow{k} d, r)$$



Multirelational Gamma process

$$\lambda^{(r)}_{c \circlearrowleft k} \sim \Gamma(\eta^{\circlearrowleft}_c \eta^{\leftrightarrow}_c \nu_k \rho_r, \delta)$$

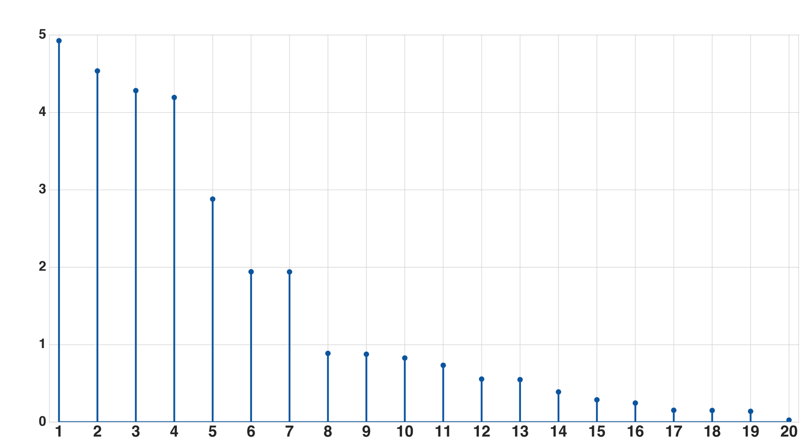
$$\lambda^{(r)}_{c \xrightarrow{k} d} \sim \Gamma(\eta^{\leftrightarrow}_c \eta^{\leftrightarrow}_d \nu_k \rho_r, \delta) \quad c \neq d$$

Gamma process **shrinkage priors:**

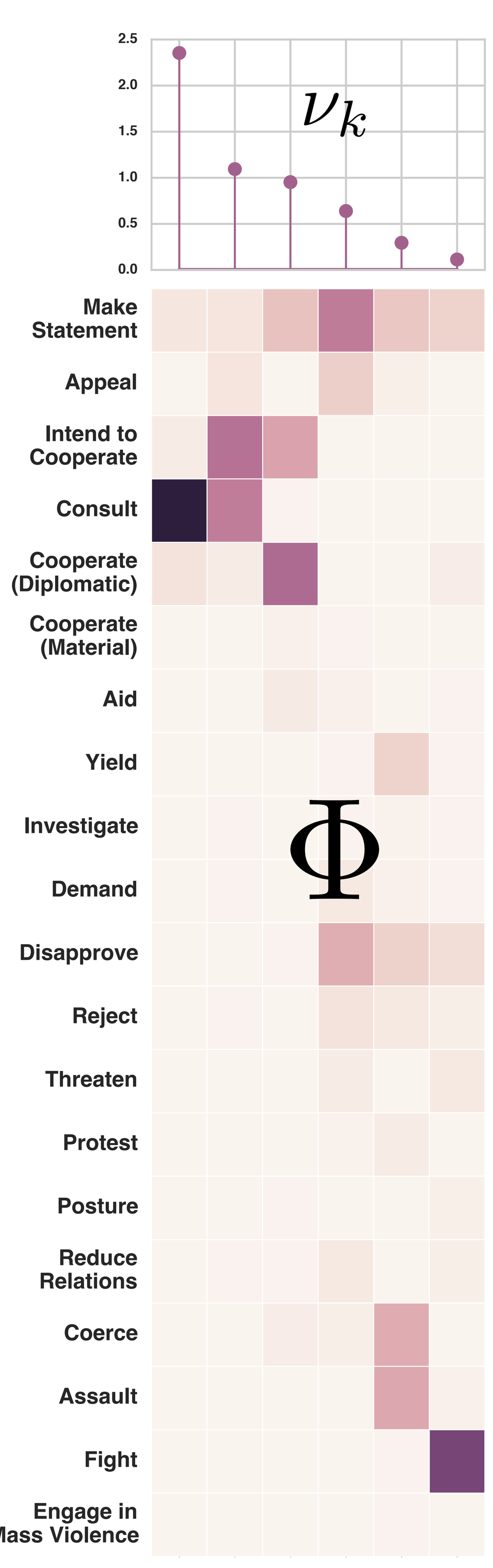
$$\eta^{(\leftrightarrow)}_c \sim \text{Gam}\left(\frac{\gamma_0}{C}, \delta\right)$$

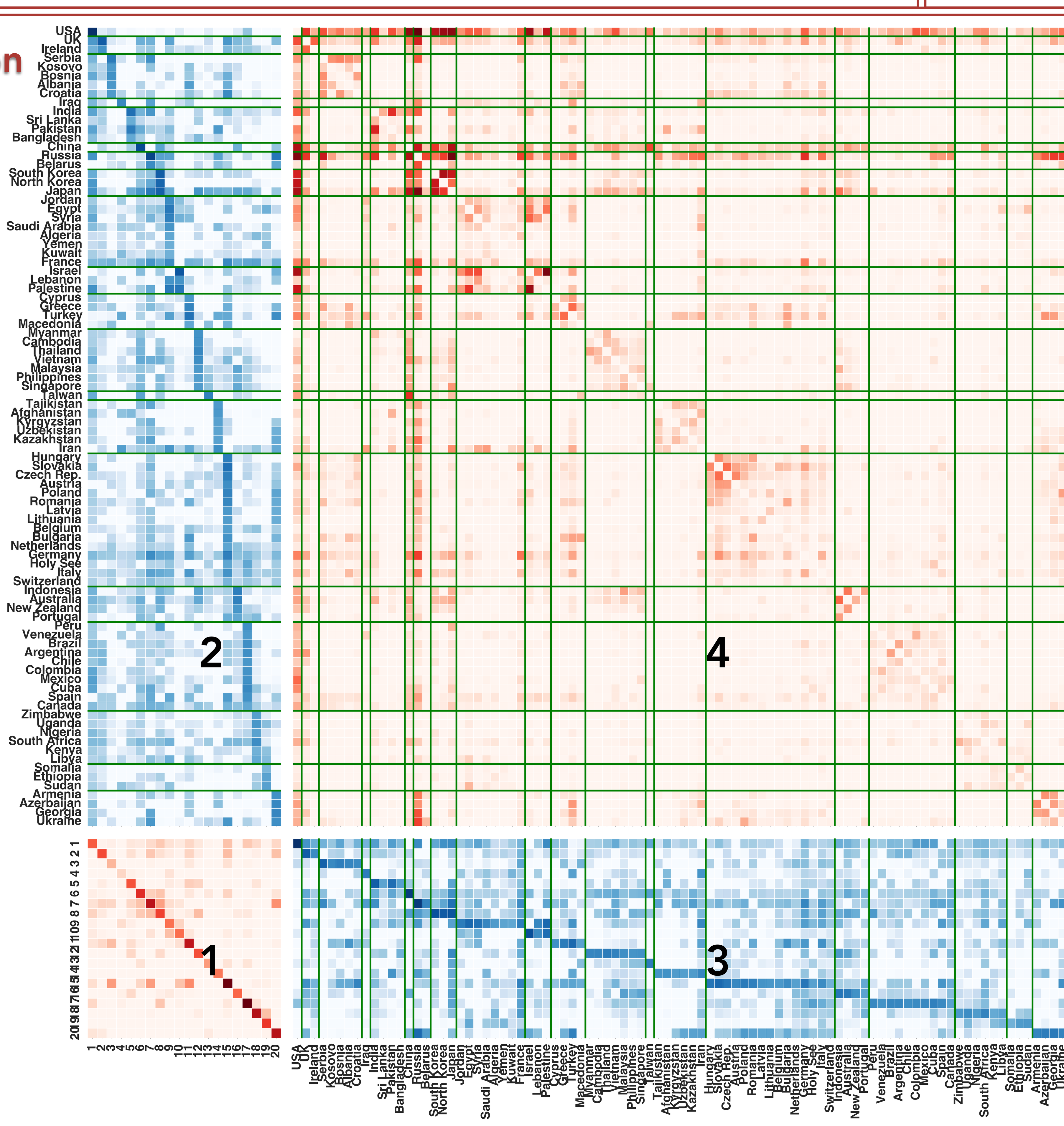
$$\nu_k \sim \text{Gam}\left(\frac{\gamma_0}{K}, \delta\right)$$

$$\rho_r \sim \text{Gam}\left(\frac{\gamma_0}{R}, \delta\right)$$



Example results on 1995–2000 data





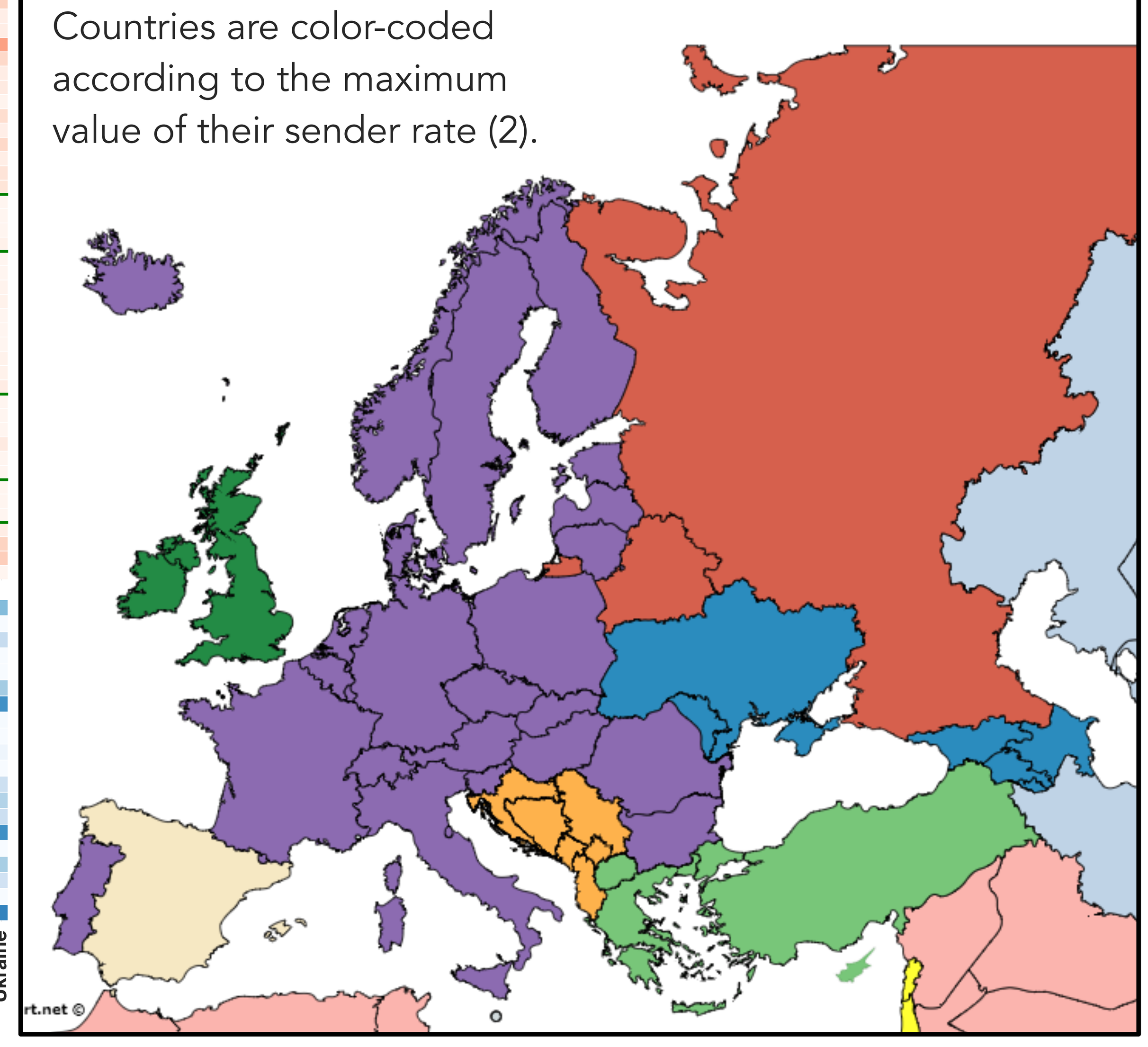
1: $\Lambda^{(r)}_k$ slice of **core** tensor for k=2 and top r

2: $\theta_{ic} \sum_{j=1}^{V} \sum_{d=1}^{C} \theta_{jd}\, \lambda^{(r)}_{c \xrightarrow{k} d}$ **sender** rates for k=2, top r

3: $\theta_{jd} \sum_{i=1}^{V} \sum_{c=1}^{C} \theta_{ic}\, \lambda^{(r)}_{c \xrightarrow{k} d}$ **receiver** rates for k=2, top r

4: $\sum_{n=1}^{N} \sum_{c=1}^{C} \sum_{d=1}^{C} \delta\big[e_n = (i \xrightarrow{a} j, t), z_n = (c \xrightarrow{k} d, r)\big]$ **thinned** counts for k=2, top r

Countries are color-coded according to the maximum value of their sender rate (2).



Comparison to other models

BPTD **generalizes** many previous models for interaction events including:

- Bayesian Poisson CP-decomposition (BPTF)
- Stochastic block model (SBM)
- Infinite relational model (IRM)

BPTD **outperforms** these models when predicting out-of-sample events.

Inverse perplexity on heldout data. Higher is better. *Top:* Events involving most active countries are heldout. *Bottom:* Events for least active are heldout.