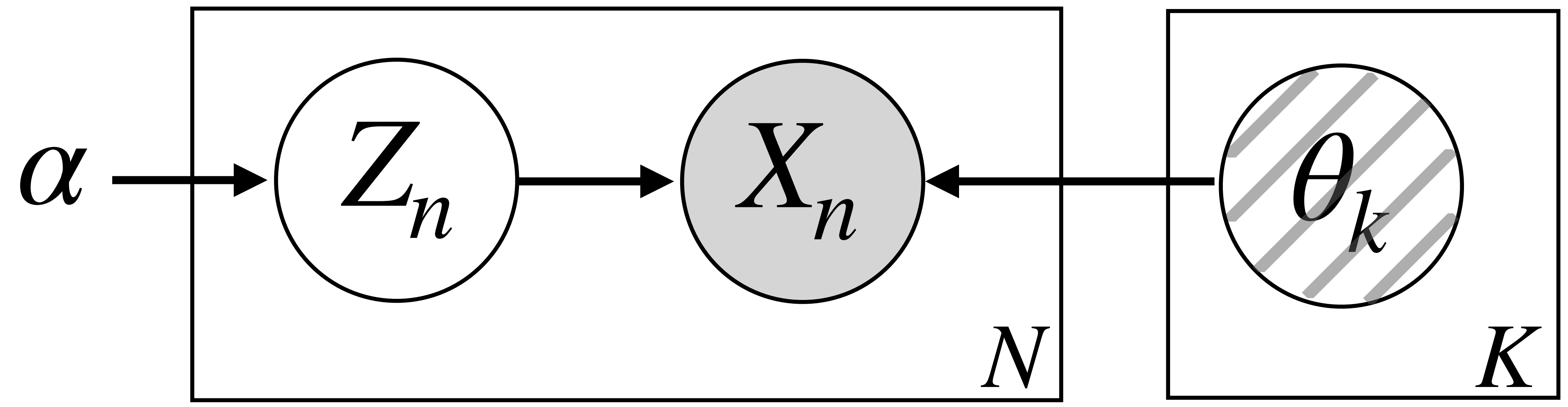
Random Function Priors for Correlation Modeling

Aonan Zhang
John Paisley

Columbia University

Setup

Model exchangeable data $X = [X_1, \ldots, X_N]$



$\theta = (\theta_k)_{k \in K}$ *collection of features*

$Z_n = [Z_{n1}, \ldots, Z_{nk}, \ldots, Z_{nK}] \in \mathbb{R}_+^K$

→ *the extent θ_k is used to express X_n.*

E.g. **Sparse factor models:** $Z_n \in \{0,1\}^K$

Topic models: $Z_n \in \Delta^{K-1}$

Problem: model flexible correlations among $Z_{n1}, \ldots, Z_{nK}$

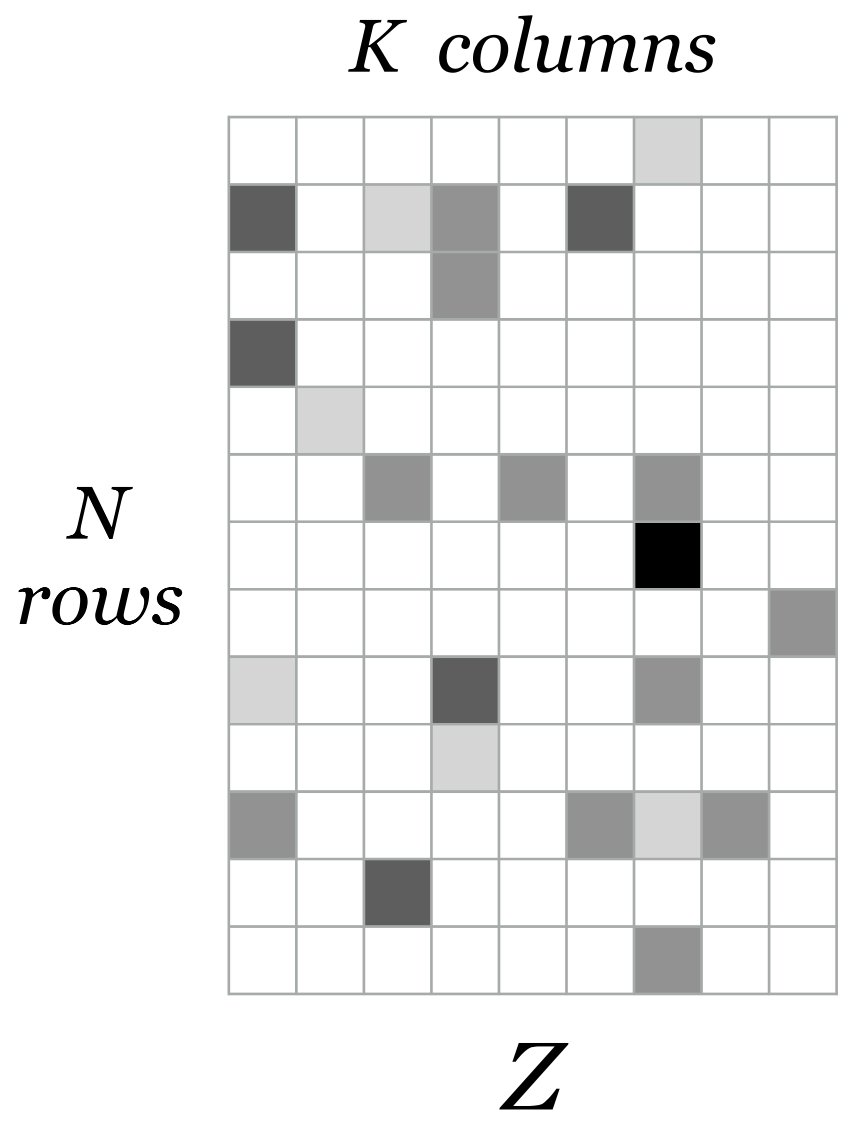
Complexity: $2^{O(K)}$ **Exponential family?** **Solution: random function priors**

Model the matrix Z

Joint distribution

$$p(X, Z, \theta) = \color{red}{p(Z)} \cdot \prod_{k=1}^{K} p(\theta_k) \cdot \prod_{n=1}^{N} p(X_n \,|\, Z_n, \theta)$$

???   *i.i.d.*   *application dependent*



Workflow to derive *p(Z)*

Exchangeability assumptions on p(Z) ——**representation theorems**——> *p(Z) is a random function model*

Representation theorem

Trick: Transform Z (random matrix) to ξ (random measure) on S.

$$\xi = \sum_{n,k} Z_{nk} \delta_{\tau_n, \sigma_k}$$

Assumption: ξ is separately exchangeable.

Proposition. *A discrete random measure ξ on S is separately exchangeable, if and only if almost surely,*

$$\xi = \sum_{n,k} f_n(\vartheta_k) \delta_{\tau_n, \sigma_k} + \textit{ trivial terms}$$

Poisson process on $\mathbf{R}_+^2$

random functions on $\mathbf{R}_+$

$$Z_{nk} = f_n(\vartheta_k)$$

The power of random function priors

Prototype to applicable models

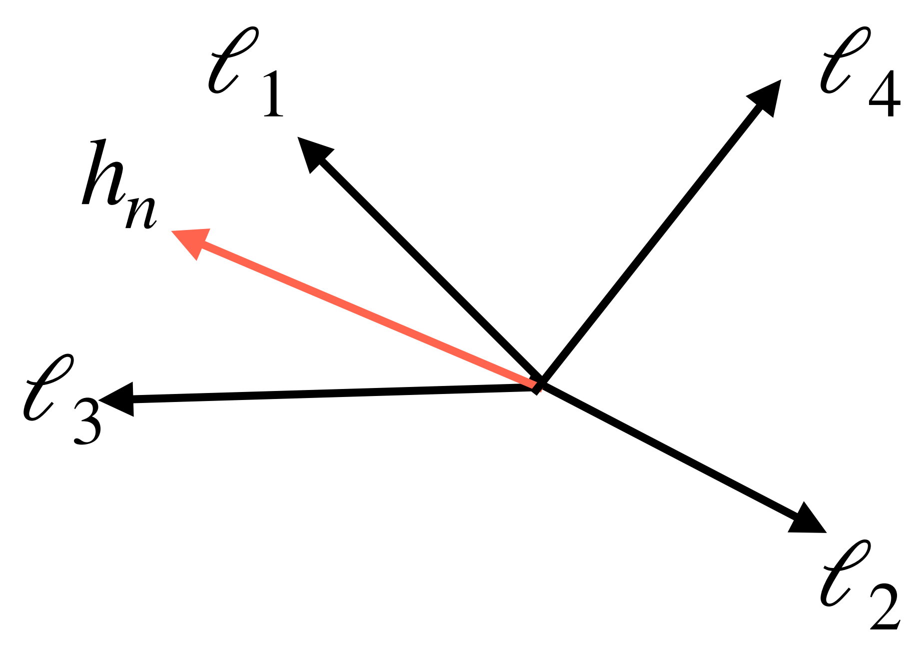
1. $f_n(\vartheta_k) \to f(h_n, \vartheta_k)$
   - h_n → *learned via inference networks* $h_n = g(X_n)$
   - f → *decoder network*

2. $f(h_n, \vartheta_k) \to f(h_n, \vartheta_k, \ell_k)$
   - ℓ_k → *augment the 2d Poisson process* (ϑ_k, σ_k) *to higher dimension* $(\vartheta_k, \sigma_k, \ell_k)$

Model correlations through arbitrary moments

Assume $Z_{nk} = f(h_n^\top \ell_k)$

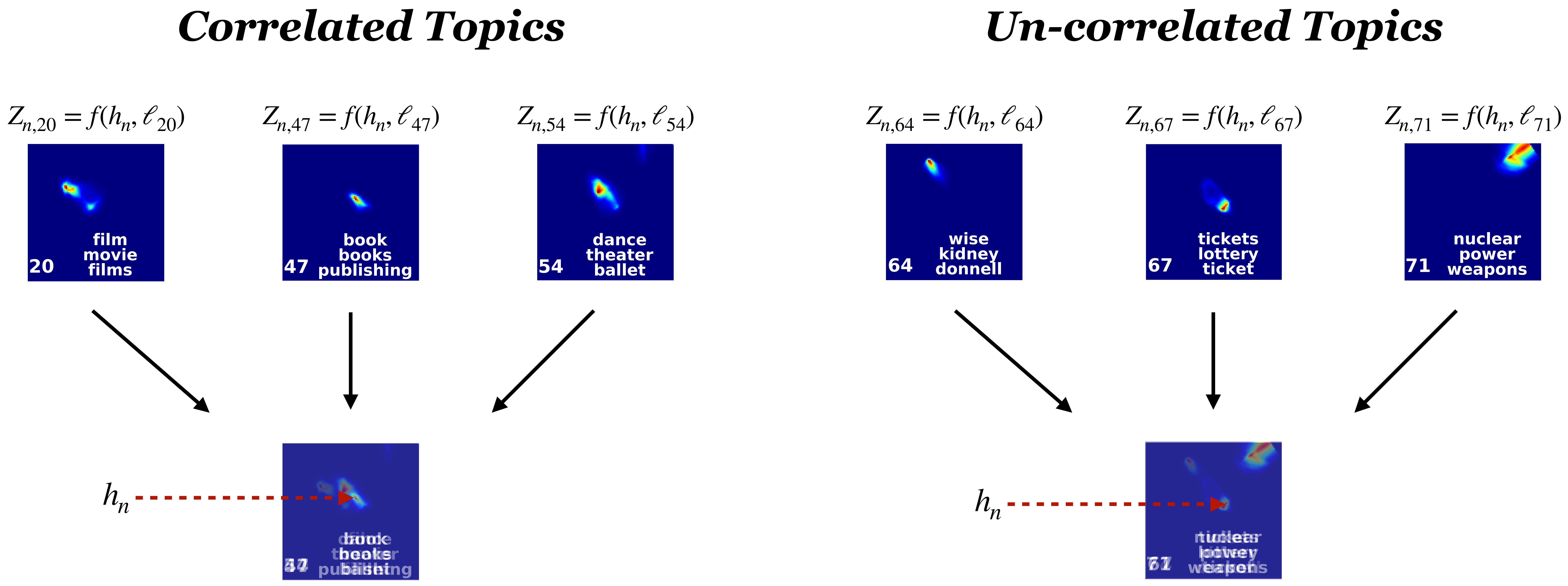
Then $\mathbb{E}[Z_{nk_1} Z_{nk_2} \ldots Z_{nk_j}] = \mathbb{E}[f(h_n^\top \ell_{k_1}) \cdots f(h_n^\top \ell_{k_j})]$



Visualize correlations via paintboxes

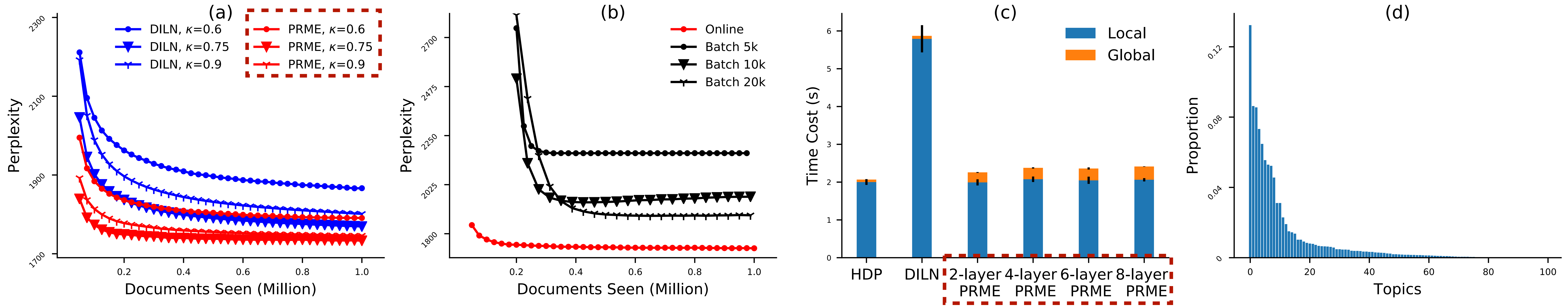
Each paintbox is a heatmap.

Visualize correlations via paintboxes



Model performance

Our model: PRME



Summarize

More details

A representation theorem for correlation modeling.

A deeper understanding of IBP beyond the Beta-Bernoulli process.

A generalization of Kingman's and Broderick's paintbox models.

Connections to random graphs.

...

Poster: #222

Code: https://github.com/zan12/prme