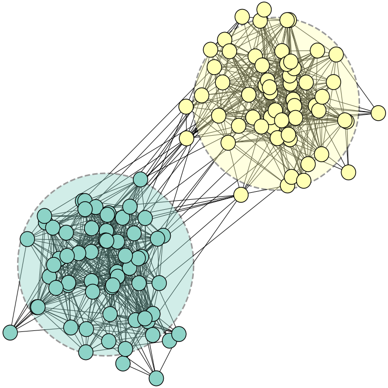
A Spike-and-Slab Prior for Dimension Selection in Generalized Linear Network Eigenmodels

Joshua Daniel Loyal

jloyal@fsu.edu

Department of Statistics
Florida State University



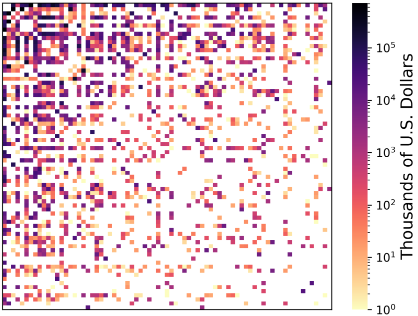
ICSA 2023
June 12th, 2023

Joint work with Yuguo Chen @ UIUC

Network Data

A symmetric $n \times n$ adjacency matrix $\mathbf{Y}$ with entries $\{Y_{ij} : 1 \leq i, j \leq n\}$ that describe the relations between pairs of entities, or nodes. The edge variables Y_{ij} can be binary (0/1) or real-valued, that is, weighted.

International Trade of Bananas in 2018[1]



Y_{ij} = amount of trade in bananas between nation i or nation j in 2018.

[1]Data taken from the BACI database curated by the CEPII.

The Statistical Problem (Edge-Variable Regression)

Goal: Understand the relationship between the edge-variables Y_{ij} and dyadic covariates $\{\mathbf{x}_{ij} \in \mathbb{R}^p : 1 \leq i, j \leq n\}$ by modeling $\mathbb{E}[Y_{ij} \mid \mathbf{x}_{ij}]$, e.g.,

The Gravity Model of Trade (Tinbergen, 1962; Anderson, 1979):

$$\log(\mathbb{E}[Y_{ij} \mid \mathbf{x}_{ij}]) = \beta_1 [\log(\mathtt{GDP}_i) + \log(\mathtt{GDP}_j)] + \beta_2 \log(\mathtt{Dist}_{ij}) + \sum_{k=3}^{p} \beta_k \, x_{ij,k}.$$

The Statistical Problem (Edge-Variable Regression)

Goal: Understand the relationship between the edge-variables Y_{ij} and dyadic covariates $\{\mathbf{x}_{ij} \in \mathbb{R}^p : 1 \leq i, j \leq n\}$ by modeling $\mathbb{E}[Y_{ij} \mid \mathbf{x}_{ij}]$, e.g.,

The Gravity Model of Trade (Tinbergen, 1962; Anderson, 1979):

$$\log(\mathbb{E}[Y_{ij} \mid \mathbf{x}_{ij}]) = \beta_1 [\log(\texttt{GDP}_i) + \log(\texttt{GDP}_j)] + \beta_2 \log(\texttt{Dist}_{ij}) + \sum_{k=3}^{p} \beta_k \, x_{ij,k}.$$

Issues: Traditional conditional independence assumptions always breakdown due to strong network dependencies such as degree, transitivity, and clustering effects.

Solution: Introduce latent variables in the form of latent variable network models that capture residual network structure.

Latent Space Models (LSMs) for Networks (Hoff et al., 2002)

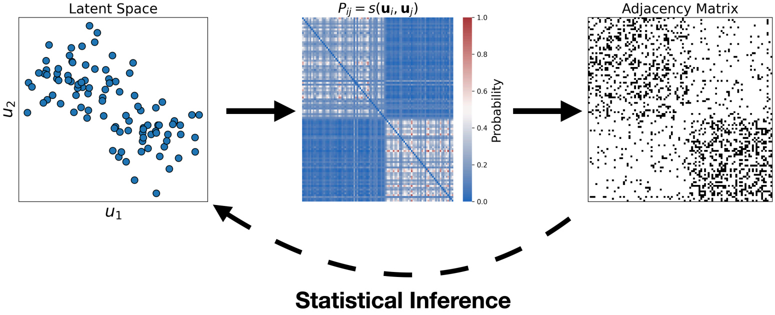
- Nodes are represented with latent positions in $\mathbb{R}^d$

$$\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_n)^\top \in \mathbb{R}^{n \times d}.$$

- Edges are conditionally independent given latent positions

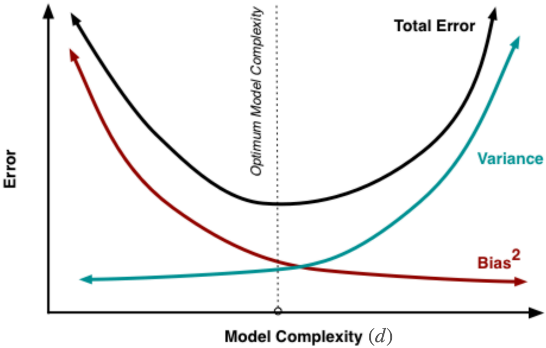
$$Y_{ij} \mid \mathbf{U} \overset{\text{ind.}}{\sim} Q(s(\mathbf{u}_i, \mathbf{u}_j)).$$

- Example: $Y_{ij} \mid \mathbf{U} \overset{\text{ind.}}{\sim} \mathrm{Bernoulli}(s(\mathbf{u}_i, \mathbf{u}_j))$:



How Do We Choose the Dimension of the Latent Space?

Model complexity is controlled by $d \Longrightarrow$ a bias-variance trade-off.



As d increases, the model can capture more network structure (lower bias), but results in more estimable parameters (higher variance).

Previous Approaches

- **Information Criterion**: AIC, BIC, DIC, WAIC, etc. Computationally intensive. No theoretical guarantees or post-selection uncertainty quantification.
- **Data Splitting**: K-fold cross-validation (Hoff, 2005; Li et al., 2020). Computationally intensive. Restricted theoretical guarantees (no covariates). No post-selection uncertainty quantification.
- **Bayesian Priors** (Durante and Dunson, 2014; Guhaniyogi and Rodriguez, 2020; Guha and Rodriguez, 2021; Gwee et al., 2022): No theoretical guarantees. Penalization of increasing model complexity only holds in prior expectation. Does this penalization penetrate through to the posterior? How to set hyperparameters?

Outline of Contributions

A Bayesian LSM with theoretical grounded dimension selection for many edge-variable types (binary, ordinal, non-negative continous).

1. Generalized Linear Network Eigenmodels (GLNEMs)
2. The Non-Homogeneous Spike-and-Slab Indian Buffet Process
3. Theoretical Results on Dimension Selection
4. Simulation Study
5. Application to the International Banana Trade Network

Generalized Linear Network Eigenmodels (GLNEMs)

Generalized Linear Network Eigenmodels (Systematic Component)

For $1 \leq i \leq j \leq n$ and some strictly increasing link function g,

$$g(\mathbb{E}[Y_{ij} \mid \mathbf{x}_{ij}]) = \boldsymbol{\beta}^\top \mathbf{x}_{ij} + \left[\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top\right]_{ij} = \boldsymbol{\beta}^\top \mathbf{x}_{ij} + \mathbf{u}_i^\top \boldsymbol{\Lambda} \mathbf{u}_j.$$

- Covariate effects: $\boldsymbol{\beta} \in \mathbb{R}^p$.
- Latent positions: $\mathbf{U} \in \bar{\mathcal{V}}_{d,n} = \{\mathbf{U} \in \mathbb{R}^{n \times d} : \mathbf{U}^\top \mathbf{U} = \mathbf{I}_d, \, \mathbf{U}^\top \mathbf{1}_n = \mathbf{0}_d\}$. $\bar{\mathcal{V}}_{d,n}$ is the set of *centered semi-orthogonal matrices*.
- Assortativity matrix: $\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \ldots, \lambda_d) \in \mathbb{R}^{d \times d}$.
- β_k and λ_h quantify the amount of assortative (or disassortativity) associated with the k-th dyadic covariate and h-th latent feature.

Generalized Linear Network Eigenmodels (GLNEMs)

Generalized Linear Network Eigenmodels (Random Component)

For $1 \leq i \leq j \leq n$,

$$Y_{ij} = Y_{ji} \mid \mathbf{x}_{ij} \overset{\text{ind.}}{\sim} Q\left\{\cdot \mid g^{-1}\left(\boldsymbol{\beta}^\top \mathbf{x}_{ij} + [\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top]_{ij}\right), \phi\right\},$$

where $Q(\cdot \mid \mu, \phi)$ is a member of the exponential dispersion family with mean μ and dispersion factor ϕ. That is, Y_{ij} has a density

$$q(y_{ij}; \theta_{ij}, \phi) = \exp\left\{\frac{y_{ij}\theta_{ij} - b(\theta_{ij})}{\phi} + k(y_{ij}, \phi)\right\},$$

where θ_{ij} is the natural parameter and b and k are known functions, such that, $g(b'(\theta_{ij})) = g(\mathbb{E}[Y_{ij} \mid \mathbf{x}_{ij}]) = \boldsymbol{\beta}^\top \mathbf{x}_{ij} + \mathbf{u}_i^\top \boldsymbol{\Lambda} \mathbf{u}_j$.

Note: GLNEMs allows for non-canonical link functions and dispersion.

Dimension Selection Through Sparsity

Assume under the true model:

$$g(\mathbb{E}[Y_{ij} \mid \mathbf{x}_{ij}]) = \boldsymbol{\beta}_0^\top \mathbf{x}_{ij} + [\mathbf{U}_0 \boldsymbol{\Lambda}_0 \mathbf{U}_0^\top]_{ij}, \quad \text{with}$$

$$\boldsymbol{\beta}_0 \in \mathbb{R}^p, \quad \mathbf{U}_0 \in \bar{\mathcal{V}}_{d_0,n}, \quad \boldsymbol{\Lambda}_0 = \mathrm{diag}(\boldsymbol{\lambda}_0) \in \mathbb{R}^{d_0 \times d_0} \text{ with } \|\boldsymbol{\Lambda}_0\|_0 = d_0.$$

We can embed this model in a higher dimensional model with $d \geq d_0$.

Dimension Selection Through Sparsity

Assume under the true model:

$$g(\mathbb{E}[Y_{ij} \mid \mathbf{x}_{ij}]) = \boldsymbol{\beta}_0^\top \mathbf{x}_{ij} + [\mathbf{U}_0 \boldsymbol{\Lambda}_0 \mathbf{U}_0^\top]_{ij}, \quad \text{with}$$

$$\boldsymbol{\beta}_0 \in \mathbb{R}^p, \quad \mathbf{U}_0 \in \bar{\mathcal{V}}_{d_0,n}, \quad \boldsymbol{\Lambda}_0 = \mathrm{diag}(\boldsymbol{\lambda}_0) \in \mathbb{R}^{d_0 \times d_0} \text{ with } \|\boldsymbol{\Lambda}_0\|_0 = d_0.$$

We can embed this model in a higher dimensional model with $d \geq d_0$.

Let $\mathbf{U} = \begin{bmatrix} \mathbf{U}_0 & \mathbf{U}_1 \end{bmatrix} \in \bar{\mathcal{V}}_{d,n}$ and $\boldsymbol{\Lambda} = \mathrm{diag}(\boldsymbol{\lambda}_0, \mathbf{0}_{(d-d_0)})$, then

$$\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top = \begin{bmatrix} \mathbf{U}_0 & \mathbf{U}_1 \end{bmatrix} \begin{bmatrix} \boldsymbol{\Lambda}_0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{U}_0^\top \\ \mathbf{U}_1^\top \end{bmatrix} = \mathbf{U}_0 \boldsymbol{\Lambda}_0 \mathbf{U}_0^\top.$$

Note: Likelihood is invariant to column permutations.

Prior Structure for Bayesian Inference

Idea: Take a Bayesian approach and construct a prior that induces posterior zeros in $\mathbf{\Lambda}$. Two remaining challenges:

1. $\mathbf{\Lambda} \sim \Pi(\mathbf{\Lambda})$:
   What prior induces an ordering constraint and posterior zeros?
2. $\mathbf{U} \sim \Pi(\mathbf{U})$:
   What is an appropriate prior on $\bar{\mathcal{V}}_{d,n}$ that allows for computationally efficient inference for a variety of GLNEMs?

A Spike-and-Slab Indian Buffet Process (SS-IBP) Prior

Propose the following prior for a collection of random variables $\{\eta_h\}_{h=1}^{d}$.
Similar priors in Ročková and George (2016) and Ohn and Kim (2022).

The Non-Homogeneous SS-IBP Truncated at d:
$\text{SS-IBP}_d(\alpha, \kappa, \mathbb{P}_{spike}, \mathbb{P}_{slab})$

$$\eta_h \mid \theta_h \overset{\text{ind.}}{\sim} \theta_h \, \mathbb{P}_{slab} + (1 - \theta_h) \, \mathbb{P}_{spike}, \quad \theta_h = \prod_{\ell=1}^{h} \nu_\ell, \quad h = 1, \ldots, d,$$

$$\nu_1 \overset{\text{ind.}}{\sim} \text{Beta}(\alpha, \kappa + 1), \quad \nu_\ell \overset{\text{iid}}{\sim} \text{Beta}(\alpha, 1), \quad \ell = 2, \ldots, d,$$

where $\alpha > 0$ and $\kappa \geq 0$. Forces $\theta_1 > \theta_2 > \cdots > \theta_d$.

A Spike-and-Slab Indian Buffet Process (SS-IBP) Prior

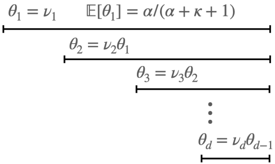
Propose the following prior for a collection of random variables $\{\eta_h\}_{h=1}^{d}$. Similar priors in Ročková and George (2016) and Ohn and Kim (2022).

The Non-Homogeneous SS-IBP Truncated at d: $\text{SS-IBP}_d(\alpha, \kappa, \mathbb{P}_{spike}, \mathbb{P}_{slab})$

$$\eta_h \mid \theta_h \overset{\text{ind.}}{\sim} \theta_h \, \mathbb{P}_{slab} + (1 - \theta_h) \, \mathbb{P}_{spike}, \quad \theta_h = \prod_{\ell=1}^{h} \nu_\ell, \quad h = 1, \ldots, d,$$

$$\nu_1 \overset{\text{ind.}}{\sim} \text{Beta}(\alpha, \kappa + 1), \quad \nu_\ell \overset{\text{iid}}{\sim} \text{Beta}(\alpha, 1), \quad \ell = 2, \ldots, d,$$

where $\alpha > 0$ and $\kappa \geq 0$. Forces $\theta_1 > \theta_2 > \cdots > \theta_d$.



A Spike-and-Slab Indian Buffet Process (SS-IBP) Prior

Propose the following prior for a collection of random variables $\{\eta_h\}_{h=1}^{d}$.
Similar priors in Ročková and George (2016) and Ohn and Kim (2022).

The Non-Homogeneous SS-IBP Truncated at d:
$\text{SS-IBP}_d(\alpha, \kappa, \mathbb{P}_{spike}, \mathbb{P}_{slab})$

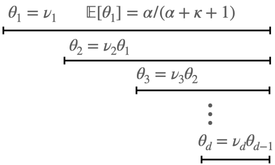
$$\eta_h \mid \theta_h \overset{\text{ind.}}{\sim} \theta_h \, \mathbb{P}_{slab} + (1 - \theta_h) \, \mathbb{P}_{spike}, \quad \theta_h = \prod_{\ell=1}^{h} \nu_\ell, \quad h = 1, \ldots, d,$$

$$\nu_1 \overset{\text{ind.}}{\sim} \text{Beta}(\alpha, \kappa + 1), \quad \nu_\ell \overset{\text{iid}}{\sim} \text{Beta}(\alpha, 1), \quad \ell = 2, \ldots, d,$$

where $\alpha > 0$ and $\kappa \geq 0$. Forces $\theta_1 > \theta_2 > \cdots > \theta_d$.

α, κ controls size of $\mathbb{E}[\theta_1] = \alpha/(\alpha + \kappa + 1)$.
α controls rate of shrinkage:

$$\mathbb{E}[\theta_h] = \mathbb{E}[\theta_1] \times [\alpha/(\alpha + 1)]^{h-1}.$$



Stochastic Ordering Under the SS-IBP$_d(\alpha, \kappa, \mathbb{P}_{spike}, \mathbb{P}_{slab})$

Ordering of $\{\theta_h\}_{h=1}^d$ induces a stochastic ordering of $\{\eta_h\}_{h=1}^d$.

Proposition 1

For $\epsilon > 0$ and fixed $\eta_0 \in \mathbb{R}$, let $\mathbb{B}_\epsilon(\eta_0) = \{\eta : |\eta - \eta_0| \leq \epsilon\}$ denote the ϵ-ball centered at η_0. Under the SS-IBP$_d(\alpha, \kappa, \mathbb{P}_{spike}, \mathbb{P}_{slab})$, if

$$\mathbb{P}_{slab}(\mathbb{B}_\epsilon(\eta_0)) < \mathbb{P}_{spike}(\mathbb{B}_\epsilon(\eta_0)),$$

then

$$\mathbb{P}(|\eta_h - \eta_0| \leq \epsilon) < \mathbb{P}(|\eta_{h+1} - \eta_0| \leq \epsilon).$$

Stochastic Ordering Under the SS-IBP$_d(\alpha, \kappa, \mathbb{P}_{spike}, \mathbb{P}_{slab})$

Ordering of $\{\theta_h\}_{h=1}^d$ induces a stochastic ordering of $\{\eta_h\}_{h=1}^d$.

Proposition 1

For $\epsilon > 0$ and fixed $\eta_0 \in \mathbb{R}$, let $\mathbb{B}_\epsilon(\eta_0) = \{\eta : |\eta - \eta_0| \leq \epsilon\}$ denote the ϵ-ball centered at η_0. Under the SS-IBP$_d(\alpha, \kappa, \mathbb{P}_{spike}, \mathbb{P}_{slab})$, if

$$\mathbb{P}_{slab}(\mathbb{B}_\epsilon(\eta_0)) < \mathbb{P}_{spike}(\mathbb{B}_\epsilon(\eta_0)),$$

then

$$\mathbb{P}(|\eta_h - \eta_0| \leq \epsilon) < \mathbb{P}(|\eta_{h+1} - \eta_0| \leq \epsilon).$$

Remark: Set $\eta_0 = 0$, then $|\eta_{h+1}|$ is stochastically less than $|\eta_h|$:

$$\mathbb{P}(|\eta_1| \leq \epsilon) < \mathbb{P}(|\eta_2| \leq \epsilon) < \cdots < \mathbb{P}(|\eta_d| \leq \epsilon).$$

An SS-IBP Prior for Λ in GLNEMs

We place a non-homogeneous spike-and-slab IBP prior on $\mathbf{\Lambda}$:

$$(\lambda_1, \ldots, \lambda_d) \sim \text{SS-IBP}_d(\alpha, \kappa, \mathbb{P}_{spike}, \mathbb{P}_{slab}) \qquad \text{with}$$
$$\mathbb{P}_{spike} = \delta_0, \qquad \mathbb{P}_{slab} = \text{Laplace}(b).$$

Corollary 1. For any $\epsilon > 0$, $\mathbb{P}(|\lambda_h| \leq \epsilon) < \mathbb{P}(|\lambda_{h+1}| \leq \epsilon)$.

Note: In practice, we represent this process as an exponential scale mixture (Park and Casella, 2008) with binary indicator variables $Z_1, \ldots, Z_d$.

Theoretical Results on Dimension Selection

Assume $\mathbf{Y}$ is drawn from a GLNEM with true latent space dimension d_0 and true parameters $\{\boldsymbol{\beta}_0, \mathbf{U}_0, \boldsymbol{\Lambda}_0\}$, i.e.,

$$Y_{ij} = Y_{ji} \overset{\text{ind.}}{\sim} Q\left\{\cdot \mid g^{-1}\left(\boldsymbol{\beta}_0^\top \mathbf{x}_{ij} + \left[\mathbf{U}_0 \boldsymbol{\Lambda}_0 \mathbf{U}_0^\top\right]_{ij},\right), \phi\right\},\ 1 \leq i \leq j \leq n$$

$$\boldsymbol{\beta}_0 \in \mathbb{R}^p, \quad \mathbf{U}_0 \in \bar{\mathcal{V}}_{d_0,n}, \quad \boldsymbol{\Lambda}_0 \in \mathbb{R}^{d_0 \times d_0} \text{ with } \|\boldsymbol{\Lambda}_0\|_0 = d_0.$$

Let $\mathbb{E}_0^{(n)}$ denote the expectation under this model.

Setup: Since d_0 often grows with n, we allow $d \to \infty$ in the SS-IBP$_d$ prior with the hope that the posterior $\mathbb{P}(\|\boldsymbol{\Lambda}\|_0 \mid \mathbf{Y})$ concentrates near d_0.

Question: Any theoretical guarantee that the posterior will not overfit?

Some Assumptions

A1. (Growth of d with n) $d = \lceil n^\gamma \rceil$ for some $\gamma \in (0, 1]$,

A2. (Growth of d_0) $d_0 = o(\log d)$,

A3. (Bounded scale parameter) $b = O(d)$,

A4. (Bounded $\mathbf{\Lambda}_0$) $\|\mathbf{\Lambda}_0\|_\infty \leq K_\lambda$ for some $K_\lambda > 0$,

A5. (Bounded latent space) $\max_{1 \leq i \leq n} \|\mathbf{u}_{0,i}\|_2 \leq K_u$ for some $K_u > 0$,

A6. (Bounded covariate effects) $\|\boldsymbol{\beta}\|_2 \leq K_\beta$ for some $K_\beta > 0$,

A7. (Bounded covariates) $\max_{1 \leq i \leq j \leq n} \|\mathbf{x}_{ij}\|_2 \leq K_x$ for some $K_x > 0$,

A8. (Bounded variance) For any compact subset $\mathcal{K} \subset \Theta$, there exists positive constants $K_{b,1}, K_{b,2}$ such that $K_{b,1} \leq \inf_{\theta \in \mathcal{K}} b''(\theta) \leq \sup_{\theta \in \mathcal{K}} b''(\theta) \leq K_{b,2}$,

A9. (Inverse link has a bounded derivative) $\sup_{\{\eta : |\eta| \leq M\}} (g^{-1})'(\eta) \leq K_g$ for some $K_g > 0$.

Note: The proof of the following theorem is based on machinery developed in Goshal and van der Vaart (2007) and Jeong and Ghoshal (2021) for posterior concentration in sparse generalized linear models.

The Posterior Concentrates on Low Dimensions on Average

Theorem 1

Assume $\mathbf{Y}$ comes from a GLNEM with non-zero latent space dimension d_0 and true parameters $\{\boldsymbol{\beta}_0, \boldsymbol{\Lambda}_0, \mathbf{U}_0\}$ such that $\|\boldsymbol{\Lambda}_0\|_0 = d_0$. Assume the following prior: $\boldsymbol{\lambda} \sim \text{SS-IBP}_d(1/d, d^{1+\delta}, \mathbb{P}_{spike}, \mathbb{P}_{slab})$ for $\delta > 0$, $\mathbb{P}_{spike} = \delta_0$, $\mathbb{P}_{slab} = \text{Laplace}(b)$ for $b \geq 1$, $\boldsymbol{\beta} \sim N(\mathbf{0}_p, \sigma_\beta^2 \mathbf{I}_p)$, and $\mathbf{U} \in \bar{\mathcal{V}}_{d,n}$ with prior probability one.

If (**A1**) - (**A9**) hold, then

$$\lim_{n\to\infty} \mathbb{E}_0^{(n)} \, \Pi\Big(\|\boldsymbol{\Lambda}\|_0 > C d_0 \mid \mathbf{Y}\Big) = 0,$$

for some $C > 0$ that only depends on δ and K_λ.

Estimation

Approximate the posterior with samples obtained using Markov Chain Monte Carlo (MCMC).

Propose a Metropolis-within-Gibbs sampler that alternates between sampling

1. $\boldsymbol{\psi} = \{\boldsymbol{\beta}, \boldsymbol{\Lambda}, \mathbf{U}, \phi, \theta_{1:d}\}$,
2. The dimension indicators $Z_{1:d} = (Z_1, \ldots, Z_d) \in \{0, 1\}^d$.

$\boldsymbol{\psi}$'s high-dimensionalty motivates using a gradient-based sampler.

Challenge: The requirement that $\mathbf{U}$ lie in $\bar{\mathcal{V}}_{d,n}$ poses a challenge for naive gradient updates.

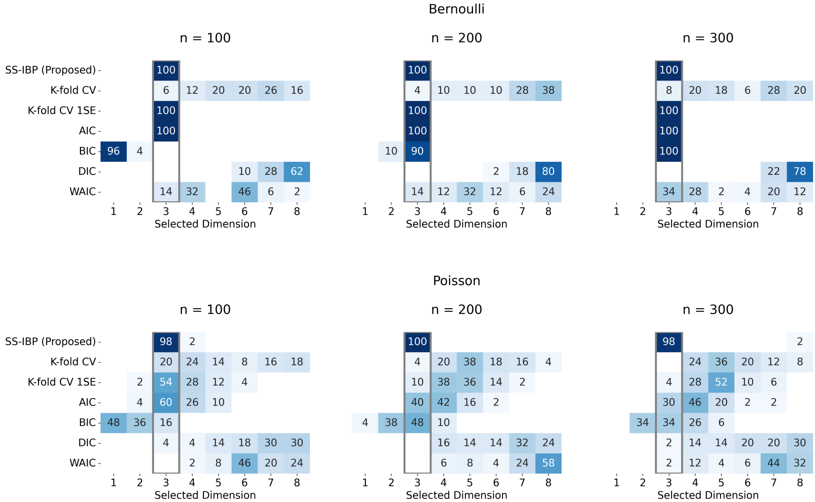
Solution: We introduced a new differentiable parameter expansion strategy based on the QR decomposition that has full support on $\bar{\mathcal{V}}_{d,n}$.

Simulation Study: Dimension Selection

- Compared the SS-IBP to traditional methods for dimension selection.
- Competitors estimated a sequence of GLNEMs with a non-shrinkage prior for $\boldsymbol{\Lambda}$ and selected the dimension according to
  - **Information Criterion:** AIC, BIC, DIC, and WAIC.
  - **Data-Splitting**: K-fold cross-validation (K = 5).
- All models estimated using Metropolis-within-Gibbs or Hamiltonian Monte Carlo with 5,000 samples after 5,000 iterations of burn-in.

Simulation Study: Dimension Selection (Canonical Link)

True dimension $d_0 = 3$. Cells display percentages out of 50 repetitions.



Simulation Study: Dimension Selection (Non-canonical Link)

True dimension $d_0 = 3$. Cells display percentages out of 50 repetitions.

Negative Binomial

n = 100 (Selected Dimension)

	1	2	3	4	5	6	7	8
SS-IBP (Proposed)		2	96	2				
K-fold CV		2	76	14	6			2
K-fold CV 1SE	14	30	52	4				
AIC		4	96					
BIC	90	8	2					
DIC				8	14	16	28	34
WAIC					10	32	8	50

n = 150 (Selected Dimension)

	1	2	3	4	5	6	7	8
SS-IBP (Proposed)			100					
K-fold CV			72	26	2			
K-fold CV 1SE		20	72	8				
AIC			92	8				
BIC	70	20	10					
DIC				10	16	16	28	30
WAIC				2	8	2	26	62

n = 200 (Selected Dimension)

	1	2	3	4	5	6	7	8
SS-IBP (Proposed)			98	2				
K-fold CV								
K-fold CV 1SE								
AIC			78	22				
BIC	52	34	14					
DIC				10	22	28	12	28
WAIC						16	46	38

Tweedie

n = 50 (Selected Dimension)

	1	2	3	4	5	6	7	8
SS-IBP (Proposed)			100					
K-fold CV	4	34	16	20	8	2	8	8
K-fold CV 1SE	38	48	12					2
AIC		6	92	2				
BIC	60	30	10					
DIC			2	2	6	8	52	30
WAIC					2	8	72	18

n = 100 (Selected Dimension)

	1	2	3	4	5	6	7	8
SS-IBP (Proposed)			100					
K-fold CV			10	38	34	12	2	4
K-fold CV 1SE	2		28	56	14			
AIC			46	40	14			
BIC		26	66	8				
DIC				8	2	14	6	70
WAIC			6	30	8	14	10	32

n = 150 (Selected Dimension)

	1	2	3	4	5	6	7	8
SS-IBP (Proposed)			100					
K-fold CV								
K-fold CV 1SE								
AIC			46	32	22			
BIC		8	52	30	10			
DIC				6	4	14	12	64
WAIC			6	20	4	18	22	30

International Trade of Bananas

A network of the international trade of bananas in 2018.[1]

- Y_{ij}: the amount of trade of bananas in thousands of U.S. Dollars between nation i and nation j in 2018.
- Five dyadic covariates:
  - $\log(\texttt{GDP}_i) + \log(\texttt{GDP}_j)$
  - $\log(\texttt{Distance}_{ij})$
  - $\texttt{CommLang}_{ij}$
  - $\texttt{Border}_{ij}$
  - $\texttt{TradeAgreement}_{ij}$
- $n = 75$ countries, $p = 5$ covariates.

[1] Data taken from the BACI database maintained by the CEPII: http://www.cepii.fr/CEPII/en/bdd_modele/bdd_modele_item.asp?id=37

A Tweedie GLNEM for Non-Negative Continous Networks

Systematic Component: A Gravity Model with Latent Network Effects

$$\log(\mathbb{E}[Y_{ij} \mid \mathbf{x}_{ij}]) = \beta_1[\log(\texttt{GDP}_i)+\log(\texttt{GDP}_j)]+\beta_2 \log(\texttt{Dist}_{ij})+\sum_{k=3}^{5} \beta_k x_{ij,k}+\mathbf{u}_i^\top \mathbf{\Lambda} \mathbf{u}_j.$$

Random Component: The Tweedie Distribution (Jørgensen, 1987)

Compound Poisson-gamma: "Total trade is the sum of individual trades."

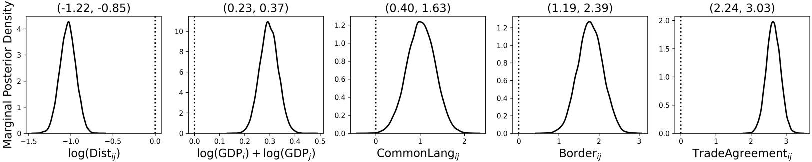
$$Y_{ij} = \begin{cases} \sum_{t=1}^{N_{ij}} Z_{ij,t} & N_{ij} > 0 \\ 0 & N_{ij} = 0. \end{cases}$$

$$N_{ij} \sim \text{Poisson}\left(\frac{\mu_{ij}^{2-\xi}}{\phi(2-\xi)}\right), \quad Z_{ij,t} \overset{\text{ind.}}{\sim} \text{Gamma}\left(\frac{2-\xi}{\xi-1}, \frac{\mu_{ij}^{\xi-1}}{\phi(\xi-1)}\right).$$

Proposed as a distribution for trade by Barabesi et al. (2016).

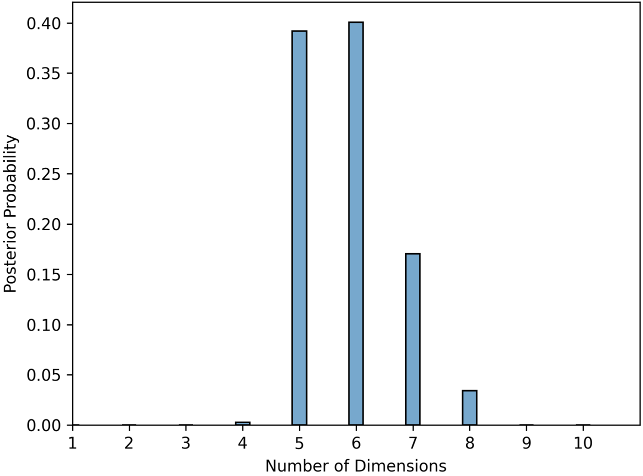
Covariate Effects

All covariates are significant and the sign of the coefficients for $\log(\texttt{GDP}_i) + \log(\texttt{GDP}_j)$ and $\log(\texttt{Dist}_{ij})$ agree with economic theory.

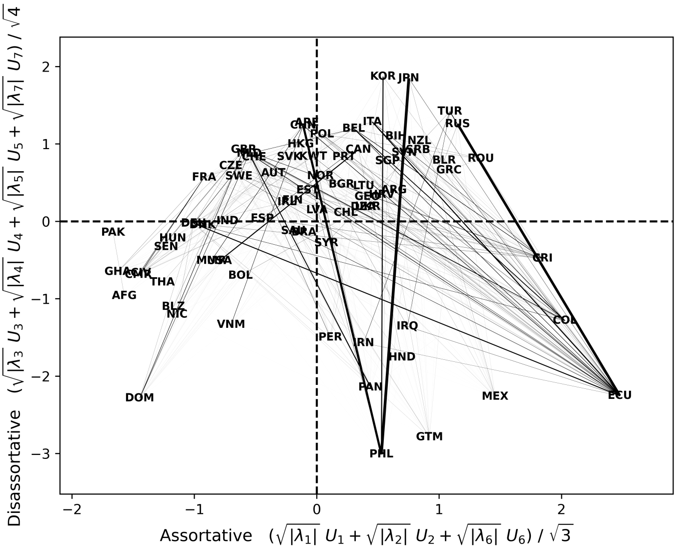


Residual Network Structure

The dimension of the latent space is uncertain:
$\mathbb{P}(d_0 = 5 \mid \mathbf{Y}) = 0.39$, $\mathbb{P}(d_0 = 6 \mid \mathbf{Y}) = 0.40$, $\mathbb{P}(d_0 = 7 \mid \mathbf{Y}) = 0.17$.



The Latent Space Reveals Bipartite Structure



Conclusion

- Developed a theoretically supported Bayesian approach to dimension selection for a general class of network models we called GLNEMs.
- Demonstrated that the $\text{SS-IBP}_d(\alpha, \kappa, \mathbb{P}_{spike}, \mathbb{P}_{slab})$ prior adapts to d_0 when used as a prior in a GLNEM.
- Scalable inference for large networks is a challange (currently only tractable for a few hundred nodes).
- Applications to directed networks is an open problem:

$$\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top \longrightarrow \mathbf{U}\mathbf{S}\mathbf{V}^\top,$$
$$\mathbf{U} \in \bar{\mathcal{V}}_{d,n}, \quad \mathbf{V} \in \bar{\mathcal{V}}_{d,n}, \quad \mathbf{S} = \operatorname{diag}(s_1, \ldots, s_d) \succeq 0.$$

- Preprint on website: https://joshloyal.github.io/publications

Anderson, J. E. (1979), "A Theoretical Foundation for the Gravity Equation," *The American Economic Review*, 69, 106–116.

Barabesi, L., Cerasa, A., Perrotta, D., and Cerioli, A. (2016), "Modeling international trade data with the Tweedie distribution for anti-fraud and policy support," *European Journal of Operational Research*, 248.

Durante, D. and Dunson, D. B. (2014), "Nonparametric Bayes Dynamic Modelling of Relational Data," *Biometrika*, 101, 883–898.

Goshal, S. and van der Vaart, A. (2007), "Convergence rates of posterior distributions for noniid observations," *The Annals of Statistics*, 35, 192–223.

Guha, S. and Rodriguez, A. (2021), "Bayesian regression with undirected network predictors with an application to brain connectome data," *Journal of the American Statistical Association*, 116, 581–593.

Guhaniyogi, R. and Rodriguez, A. (2020), "Joint modeling of longitudinal relational data and exogenous variables," *Bayesian Analysis*, 15, 477–503.

Gwee, X. Y., Gormley, I. C., and Flop, M. (2022), "A Latent Shrinkage Position Model for Binary and Count Data," *arXiv preprint arXiv:2211.13034*.

Hoff, P. D. (2005), "Bilinear Mixed-Effects Models for Dyadic Data," *Journal of the American Statistical Association*, 100, 286–295.

Hoff, P. D., Raftery, A. E., and Handcock, M. S. (2002), "Latent Space Approaches to Social Network Analysis," *Journal of the American Statistical Association*, 97, 1090–1098.

Jeong, S. and Ghoshal, S. (2021), "Posterior Contraction in Sparse Generalized Linear Models," *Biometrika*, 108, 367–379.

Jørgensen, B. (1987), "Exponential Dispersion Models," *Journal of the Royal Statistical Society, Series B*, 49, 127–162.

Li, L., Levina, E., and Zhu, J. (2020), "Network cross-validation by edge sampling," *Biometrika*, 107, 257–276.

Ohn, I. and Kim, Y. (2022), "Posterior Consistency for Factor Dimensionality in High-Dimensional Sparse Factor Models," *Bayesian Analysis*, 17, 491–514.

Park, T. and Casella, G. (2008), "The Bayesian Lasso," *Journal of the American Statistical Association*, 103, 681–686.

Ročková, V. and George, E. I. (2016), "Fast Bayesian Factor Analysis via Automatic Rotations to Sparsity," *Journal of the American Statistical Association*, 111, 1608–1622.

Tinbergen, J. (1962), *Shaping the World Economy - Suggestions for an International Economic Policy*, New York: Twentieth Century Fund.

Parameter Identifiability and Marginal Effect Interpretation

Define the node-averaged covariate matrix
$\bar{\mathbf{X}} = (1/n) \sum_{j=1}^{n} (\mathbf{x}_{ij}, \ldots, \mathbf{x}_{nj})^\top = (\bar{\mathbf{x}}_1, \ldots, \bar{\mathbf{x}}_n)^\top$.

Proposition 2

Assume $\mathbf{Y}$ is drawn from a GLNEM with parameters $\{\boldsymbol{\beta}, \mathbf{U}, \boldsymbol{\Lambda}, \phi\}$ such that $\mathbf{U} \in \bar{\mathcal{V}}_{d,n}$ and $\text{rank}(\bar{\mathbf{X}}) = p$, then $\boldsymbol{\beta}$ is identifiable.

Remark: The sum-to-zero constraint on $\mathbf{U}$'s columns allows us to interpret the β_k's as marginal effects since

$$n^{-1} \sum_{j=1}^{n} g(\mathbb{E}[Y_{ij} \mid \mathbf{x}_{ij}]) = \boldsymbol{\beta}^\top \bar{\mathbf{x}}_i,$$

which are not conditioned on keeping the latent positions $\mathbf{u}_i$ fixed.