BSNMani: Bayesian Scalar-on-Network Regression with Manifold Learning

Jian Kang

University of Michigan, Ann Arbor

IMSI, September 2026

The Annals of Applied Statistics
2026, Vol. 20, No. 2, 1496–1515
https://doi.org/10.1214/26-AOAS2140
© Institute of Mathematical Statistics, 2026

BSNMANI: BAYESIAN SCALAR-ON-NETWORK REGRESSION WITH MANIFOLD LEARNING

BY YIJUN LI[1,a], KI SUENG CHOI[2,c], BOADIE W. DUNLOP[3,e], W. EDWARD CRAIGHEAD[3,f], HELEN S. MAYBERG[2,d], LANA GARMIRE[4,g], YING GUO[5,h] AND JIAN KANG[1,b]

[1]*Department of Biostatistics, University of Michigan,* [a]*liyijun@umich.edu,* [b]*jiankang@umich.edu*

[2]*Center of Advanced Circuit Therapeutics, Icahn School of Medicine at Mount Sinai,* [c]*kisueng.choi@mssm.edu,* [d]*helen.mayberg@mssm.edu*

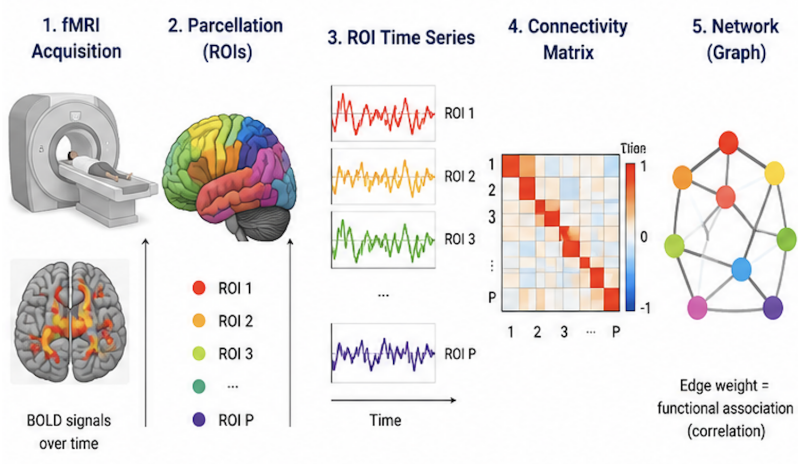
[3]*Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine,* [e]*bdunlop@emory.edu,* [f]*ecraigh@emory.edu*

[4]*Department of Computational Medicine and Bioinformatics, University of Michigan,* [g]*lgarmire@med.umich.edu*

[5]*Department of Biostatistics and Bioinformatics, Emory University,* [h]*yguo2@emory.edu*

- NIH R01DA048993; NIH R01MH105561; NSF IIS2123777

Brain connectivity from neuroimaging



A motivating problem

Dunlop *et al. Trials* 2012, **13**:106
http://www.trialsjournal.com/content/13/1/106

TRIALS

STUDY PROTOCOL Open Access

Predictors of remission in depression to individual and combined treatments (PReDICT): study protocol for a randomized controlled trial

Boadie W Dunlop[1*], Elisabeth B Binder[2,1], Joseph F Cubells[1], Mark M Goodman[6], Mary E Kelley[3], Becky Kinkead[1], Michael Kutner[3], Charles B Nemeroff[4], D Jeffrey Newport[1], Michael J Owens[1], Thaddeus W W Pace[1], James C Ritchie[5], Vivianne Aponte Rivera[1], Drew Westen[1], W Edward Craighead[1,7†] and Helen S Mayberg[1†]

Can baseline brain connectivity help explain treatment response in depression?

- Resting-state fMRI gives a subject-specific connectivity network.
- Clinical response is a scalar outcome, e.g., change in Hamilton Depression Rating Scale (HDRS) after treatment.
- We want both prediction and biological interpretation.

PReDICT study

- Adults with major depressive disorder
- Antidepressant medication or cognitive behavioral therapy (CBT)
- Baseline and week-12 HDRS
- Baseline rs-fMRI connectivity

Goal: learn connectivity patterns that are clinically predictive and neuroscientifically interpretable.

Why is this statistically challenging?

High-dimensional network predictor

$$454 \text{ ROIs} \quad \Rightarrow \quad \frac{454 \times 453}{2} > 100{,}000 \text{ edges}$$

- Edges are highly correlated.
- Signal is distributed across subnetworks.
- Direct edge-wise regression is unstable and hard to interpret.

Scientific target

Not just "which edges?"

Which latent connectivity patterns matter?

network	$\longrightarrow$	**subnetworks**
Y_i	$\longrightarrow$	$\lambda_{i1}, \ldots, \lambda_{iq}$

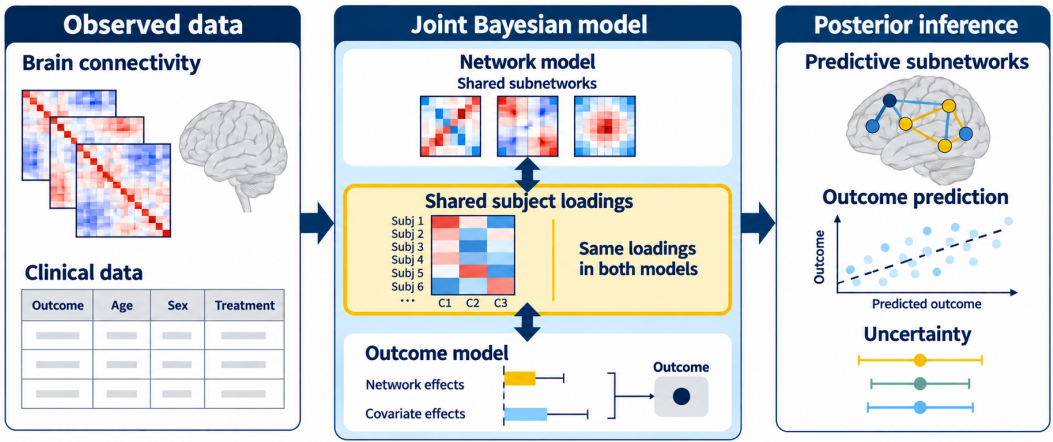
Existing methods: strengths and limitations

Representative methods	Outcome guided	Population subnetworks	Joint learning
Edge-wise scalar-on-network regression			
Arroyo Relión et al. (2019); Jiang et al. (2020); Morris et al. (2022)	✓	✗	✓
Low-rank network decomposition			
Durante et al. (2017); Wang et al. (2019); Wang & Guo (2023)	✗	✓	✗
Two-stage latent feature regression			
Ma, Kundu & Stevens (2022)	✓	✓	✗
Network manifold / tangent-space methods			
Brzyski et al. (2023); Ju, Park & Tarpey (2025)	✓	△	△
BSNMani	✓	✓	✓

△ = partially satisfies the property.

Gap: Existing methods typically achieve either prediction or interpretable network representation, but not both through a unified Bayesian model.

BSNMani: workflow



Shared subject loadings link the network and outcome models.

Models

Network manifold model

$$Y_i = \sum_{\ell=1}^{q} \lambda_{i\ell} u_\ell u_\ell^T + \epsilon_i$$

- $u_\ell u_\ell^T$: population-level subnetwork
- $\lambda_{i\ell}$: subject-specific loading
- $U = (u_1, \ldots, u_q) \in \mathcal{V}_{q,N}$.

$$\boxed{\mathcal{V}_{q,N} = \{U \in \mathbb{R}^{N \times q} : \; U^T U = I_q\}}$$

Stiefel manifold: the space of all orthonormal q-dimensional bases in $\mathbb{R}^N$.

Clinical model

$$C_i = \sum_{\ell=1}^{q} \beta_\ell \lambda_{i\ell} + \alpha^T z_i + \delta_i$$

- C_i: clinical outcome
- z_i: demographic or treatment covariates
- β_ℓ: clinical effect of subnetwork ℓ

Each latent subnetwork is both a network pattern and a clinical predictor.

Why the Stiefel manifold?

Orthogonal latent basis

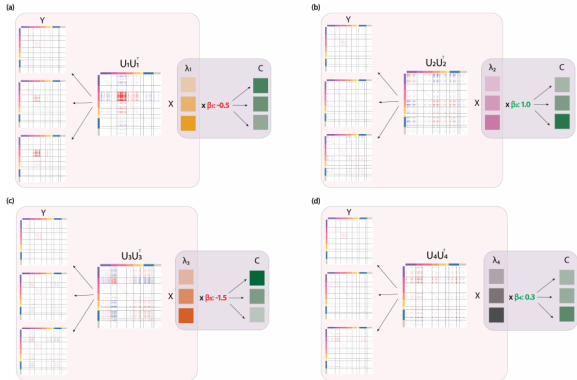
$$U^T U = I_q, \qquad U \in \mathcal{V}_{q,N}$$



- Separates distinct subnetworks.
- Preserves low-rank network structure.
- Reduces rotation ambiguity.
- Gives interpretable subject coordinates λ_i.

Manifold structure turns a huge edge vector into a small set of interpretable network coordinates.

How to interpret a latent subnetwork



- $\lambda_{i\ell}$ **measures the strength of subnetwork** ℓ **in subject** i**;**
- β_ℓ **measures how that strength relates to outcome.**

Bayesian Inference with Prior specification

Network manifold model

$$Y_i = \sum_{\ell=1}^{q} \lambda_{i\ell} u_\ell u_\ell^T + \epsilon_i, \qquad \epsilon_{i,jk} \sim N(0, \sigma^2)$$

$$U \sim \mathsf{U}(\mathcal{V}_{q,N}), \qquad \lambda_i \sim N(0, \tau_\lambda^2 I_q)$$

$$\sigma^{-2} \sim \mathsf{G}(\nu_0/2, \nu_0\sigma_0^2/2), \qquad \tau_\lambda^{-2} \sim \mathsf{G}(\eta_0/2, \eta_0\tau_0^2/2)$$

Clinical model

$$C_i = \beta^T \lambda_i + \alpha^T z_i + \delta_i, \qquad \delta_i \sim N(0, \tau^2)$$

$$\beta \sim N(0, \tau_\beta^2 I_q), \qquad \alpha \sim N(0, \tau_\alpha^2 I_r)$$

$$\tau^{-2} \sim \mathsf{G}(\rho_0/2, \rho_0\psi_0^2/2)$$

$$\tau_\beta^{-2} \sim \mathsf{G}(\gamma_0/2, \gamma_0\kappa_0^2/2), \qquad \tau_\alpha^{-2} \sim \mathsf{G}(\omega_0/2, \omega_0\phi_0^2/2)$$

The prior encourages a low-dimensional orthogonal network representation while allowing shrinkage for subject loadings and clinical effects.

Full Bayesian inferences for the joint model

Sequential pipeline

Network decomposition
$\Downarrow$
Features λ_i
$\Downarrow$
Clinical regression

Risk: learned features may not be clinically relevant.

BSNMani joint posterior

Networks and clinical outcomes
$\Downarrow$
Joint posterior
$\Downarrow$
Predictive subnetworks

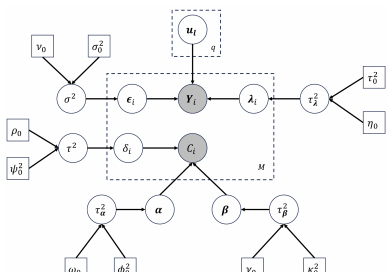
Clinical information guides the learned network representation.

The subnetworks are learned for structure and prediction simultaneously.

Posterior computation

Challenge

- Most parameters have conjugate updates.
- U lives on the Stiefel manifold.
- Standard MCMC does not directly respect $U^T U = I$.



Solution (Jauch, Hoff & Dunson, 2021)

Instead of sampling directly on the constrained Stiefel manifold, introduce an unconstrained matrix

$$\boxed{X \in \mathbb{R}^{N \times q}} \Longrightarrow \boxed{U = X(X^T X)^{-1/2}}$$

$$\log \pi(U) \propto \frac{1}{\sigma^2} \sum_{i=1}^{M} \operatorname{tr}\left(U \Lambda_i U^T Y_i\right),$$

$$\underbrace{\nabla_X \log \pi(X)}_{\text{MALA}} = -X + \underbrace{\frac{\partial \log \pi(U)}{\partial U}}_{\text{Likelihood}} \underbrace{\frac{\partial U}{\partial X}}_{\text{Polar expansion}}$$

Jauch, M., Hoff, P. D., and Dunson, D. B. (2021). *Monte Carlo Simulation on the Stiefel Manifold via Polar Expansion*. Journal of Computational and Graphical Statistics, 30(3):622–631.

Gradient computation for MALA

Step 1. Polar expansion: $U = X(X^TX)^{-1/2}$
Define

$$R(X) = (X^TX)^{-1/2},$$

so that

$$U = XR.$$

Step 2. Product rule

$$dU = d(XR)$$

$$\boxed{dU = dX\,R + X\,dR.}$$

Step 3. Matrix chain rule
Substitute the differential of R:

$$\boxed{\frac{\partial U}{\partial X} = (I_N \otimes R) + (X \otimes I_q)\frac{\partial R}{\partial X}.}$$

The derivative of the inverse square root is

$$\begin{aligned}\frac{\partial R}{\partial X} = &-\left[I_q \otimes R + R \otimes I_q\right]^{-1} \\ &\times \left[(X^TX)^{-1} \otimes (X^TX)^{-1}\right]\left[(I_q \otimes X^T)(I_q \boxtimes I_N) + (X^T \otimes I_q)I_{Nq}\right],\end{aligned}$$

Therefore,

$$\boxed{\nabla_X \log \pi(X) = -X + \frac{\partial \log \pi(U)}{\partial U}\frac{\partial U}{\partial X}.}$$

The polar expansion converts constrained optimization on the Stiefel manifold into unconstrained MALA with an analytic gradient.

Hybrid MALA-Gibbs sampler

Gibbs Sampler	$\Rightarrow$	update $\lambda_i, \beta, \alpha, \sigma^2, \tau^2$ and variance components
MALA on X	$\Rightarrow$	update unconstrained representation of the orthogonal basis
Polar map	$\Rightarrow$	obtain $U = X(X^T X)^{-1/2}$
Repeat	$\Rightarrow$	posterior samples of subnetworks, loadings, and clinical effects

The sampler avoids Hessian calculations and handles the orthogonality constraint through a Euclidean expansion.

Two-stage sampling

Why a two-stage proposal?

- Networks can contain hundreds of ROIs.
- Loadings couple the network and clinical models.
- A factorized proposal can simplify exploration.

Proposal and correction

1. Propose network parameters and subject loadings.
2. Propose clinical parameters conditional on those loadings.
3. Apply an independence Metropolis–Hastings correction.

$$a(\theta, \theta') = \min\left\{1, \frac{\pi(\theta')g(\theta)}{\pi(\theta)g(\theta')}\right\}, \qquad \theta' \sim g.$$

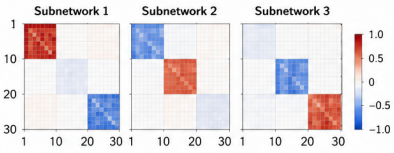
The proposal has two stages. The corrected target remains the joint posterior.

Simulation results I: synthetic networks

Data generation (low-rank network model)

- $N = 30$ regions (ROIs), true $q = 3$ subnetworks
  Block-diagonal ground-truth subnetworks

True subnetworks (U)



Subject-specific loadings (Λ_i)

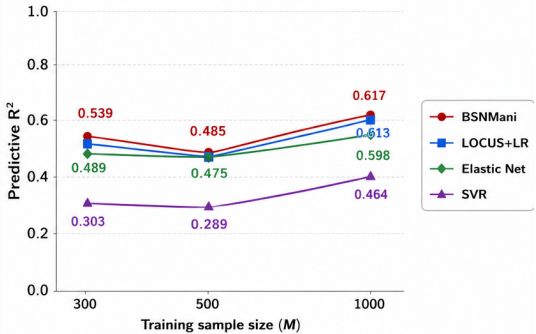


Simulation settings

- Sample size: $M = 390, 650, 1300$
- Network SNR (SNR_Y) $\in \{0.5, 1\}$ Clinical SNR (SNR_C) $\in \{3, 6\}$
- Covariates: treatment indicator + age + gender + site

Predictive performance (test-set R^2)

Effect of increasing sample size ($\mathrm{SNR}_Y = 0.5$, $\mathrm{SNR}_C = 3$)



Takeaway: Across different sample sizes and SNR settings, BSNMani achieves the highest predictive accuracy.

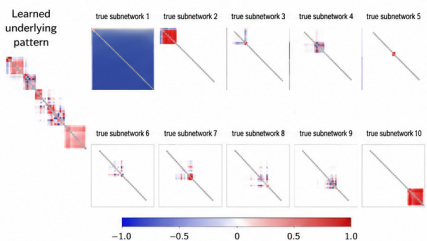
- ✓ Accurately recovers latent subnetworks and subject loadings
- ✓ Produces the best clinical outcome prediction

Simulation results II: data-driven networks

Data-driven simulation design

- Subnetworks U and loadings $\mathbf{\Lambda}_i$ learned from PReDICT data.
- Clinical SNR calibrated to PReDICT: $\mathrm{SNR}_C = 0.38$.
- **Region–pair–specific noise** variances used (heteroscedastic), mis-specified in all methods (including **BSNMani**).
- Mean network SNR varied: $\mathrm{SNR}_Y \in \{3, 1, 0.5, 0.0416\}$.

Learned subnetworks from PReDICT data ($q = 10$)



Evaluation metric

Predictive performance measured by test-set R^2.

Competitors

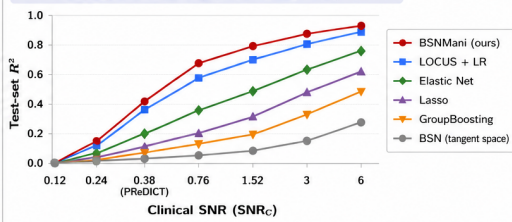
- LOCUS + linear regression
- Lasso, Elastic Net, GroupBoosting (full-edge)
- BSN: scalar-on-network regression in tangent space

Predictive performance (test-set R^2)

Varying network SNR ($\mathrm{SNR}_C = 0.38$)

Network SNR (SNR_Y)	BSNMani (ours)	LOCUS + LR	Elastic Net	Lasso	Group Boosting	BSN (tangent space)
3	0.702	0.700	0.623	0.614	0.638	0.481
1	0.711	0.703	0.598	0.583	0.638	0.205
0.5	0.712	0.696	0.564	0.540	0.618	0.123
0.0416	0.703	0.686	0.596	0.617	0.544	0.055

Varying clinical SNR ($M = 500$, $\mathrm{SNR}_Y = 0.5$)



Application: PReDICT depression study

Data

- $M = 130$ patients with MDD
- Baseline rs-fMRI connectivity
- 454 ROIs: 400 cortical + 54 subcortical
- Outcome: baseline-to-week-12 change in HDRS

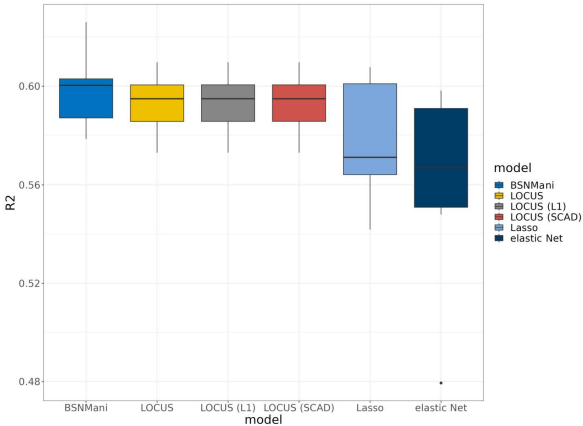
Adjusted covariates

- Age
- Sex
- Treatment type: medication vs CBT
- Response group

Model choice

$q = 4$ selected by predictive performance.

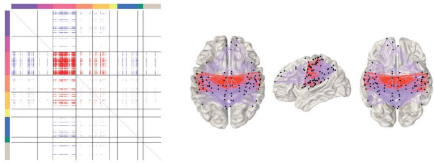
Prediction in PReDICT



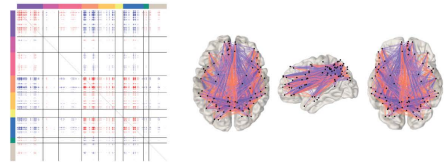
Repeated 5-fold CV

- BSNMani has the highest median predictive R^2.
- Subnetwork methods have less variable performance in this comparison.
- Interpret predictions conditional on the stated covariates.

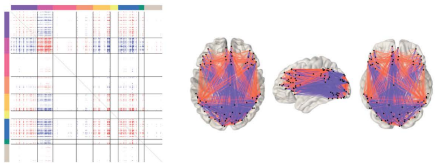
PReDICT study: Recovered predictive subnetworks



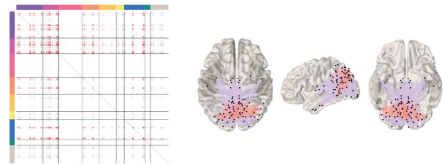
Subnetwork 1
somatomotor and attention networks



Subnetwork 2
default mode with attention/control networks



Subnetwork 3
visual hub with salience/control/temporoparietal networks



Subnetwork 4
default mode and visual networks

Clinical interpretation of subnetworks

Estimated clinical effects

The effect size, 95% credible interval, and interval width of β and α for the PReDICT study

	Effect size	95% credible interval	Credible interval width
β_1: subnetwork 1	−0.053	[−0.723, 0.6091]	1.332
β_2: subnetwork 2	0.684	[0.026, 1.351]	1.325
β_3: subnetwork 3	0.094	[−0.604, 0.789]	1.393
β_4: subnetwork 4	0.520	[−0.193, 1.227]	1.420
α_1: age	−0.101	[−0.783, 0.571]	1.354
α_2: sex (female)	1.076	[−0.240, 2.393]	2.633
α_3: treatment (medication)	−0.886	[−2.325, 0.566]	2.891
α_4: response group (responder)	−10.001	[−11.448, −8.552]	2.896

Clinical covariates

Only the responder indicator shows a strong effect.
Age, sex, and treatment type have credible intervals containing zero.

Subnetwork 1

Somatomotor + Attention

$$\beta_1 = -0.053$$

Higher expression $\Longrightarrow$ a more negative HDRS change, conditional on covariates.

Subnetworks 2–4

Default Mode, Control, Salience, Visual

$$\beta_2 = 0.684,\ \beta_3 = 0.094,\ \beta_4 = 0.520$$

Higher expression $\Longrightarrow$ a more positive HDRS change, conditional on covariates.

Clinical prediction is tied to interpretable connectivity patterns, not anonymous features.

What does BSNMani add?

Statistical

Joint Bayesian scalar-on-network regression with low-rank network decomposition.

Computational

- Hybrid MALA-Gibbs sampler with polar expansion for the Stiefel manifold.
- Two-stage proposals with a joint-posterior correction

Scientific

Predictive subnetworks linked directly to treatment-related symptom change in MDD.

Prediction, interpretation, and uncertainty quantification are learned in a single coherent framework.

A new motivating problem: cognition in ABCD

Scientific question

Which shared brain connectivity patterns are associated with differences in children's general cognitive ability, after adjusting for demographic and family characteristics?

ABCD analysis

- 4,549 children from 16 study sites.
- Resting-state connectivity across 352 brain regions.
- General cognitive ability score and demographic/family covariates.

Why extend BSNMani?

- Retain interpretable subnetworks and subject loadings shared by both models.
- Learn jointly across thousands of children, with uncertainty quantification.
- BSNMani requires repeated MALA gradients through the polar map.

Can we retain joint network–outcome learning while simplifying posterior computation?

BSNMani+: model and prior specification

Network model

$$Y_i = U\Lambda_i V^T + E_i, \qquad \Lambda_i = \mathrm{diag}(\lambda_i)$$

$U, V \in \mathbb{R}^{N\times q}$.
Relax exact orthogonality and $U = V$.

Clinical model

$$C_i = \beta^T \lambda_i + \alpha^T z_i + \delta_i$$

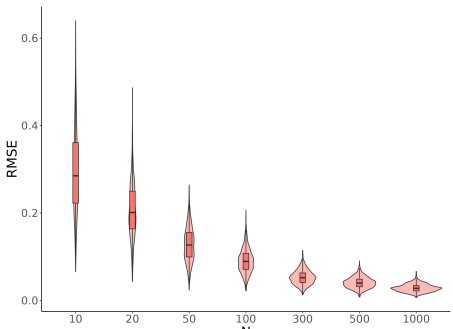
- Same clinical regression structure.

Gaussian factor priors

$$U_{jk} \sim N(0, 1/N), \qquad V_{jk} \mid U_{jk} \sim N(U_{jk}, \tau_v^2).$$

Small τ_v^2 encourages $V \approx U$, hence approximate symmetry.

Why does the Gaussian prior approximate orthogonality?



For independent $U_{jk} \sim N(0, 1/N)$,

$$E\|U_{\cdot k}\|^2 = 1,$$
$$\mathrm{Var}(\|U_{\cdot k}\|^2) = 2/N,$$
$$\mathrm{Var}(U_{\cdot k}^T U_{\cdot l}) = 1/N, \quad k \neq l.$$

For fixed q, $U^T U$ approaches I_q under the prior as N grows.

Prior concentration motivates the relaxation. Posterior geometry still needs checking.

Gaussian relaxation: a quantitative guarantee

Let $U \in \mathbb{R}^{N\times q}$, $N \geq q$, with $U_{jk} \overset{\text{iid}}{\sim} N(0, 1/N)$. Define the polar factor $Q = U(U^TU)^{-1/2}$.

Proposition: distance to the Stiefel manifold

The polar factor is exactly uniform: $Q \sim \mathrm{Unif}(\mathcal{V}_{q,N})$.

$$\boxed{E\|U - Q\|_F^2 \;\leq\; E\|U^TU - I_q\|_F^2 \;=\; \frac{q(q+1)}{N}.}$$

Proof idea. Sum q diagonal variances $2/N$ and $q(q-1)$ off-diagonal second moments $1/N$. For each singular value $s \geq 0$, $(s-1)^2 \leq (s^2-1)^2$.

Growing rank. Gaussian concentration also gives

$$\|U^TU - I_q\|_{\text{op}} = O_P\left(\sqrt{q/N}\right).$$

If $q/N \to 0$, the prior becomes orthogonal in operator norm.

Total mean squared Frobenius error vanishes when $q^2/N \to 0$. These are prior guarantees.

Polar factor: Jauch, Hoff & Dunson (2021), Sec. 3. Gaussian concentration: Vershynin (2012), Cor. 5.35.

BSNMani+: Gibbs sampling for posterior computation

Closed-form conditional update

$$U \mid \text{rest} \sim \mathcal{MN}_{N,q}(M_U, I_N, S_U)$$

$$S_U^{-1} = (\tau_v^{-2} + N) I_q + \sigma^{-2} \sum_i \Lambda_i V^T V \Lambda_i.$$

A corresponding matrix-normal update applies to V.

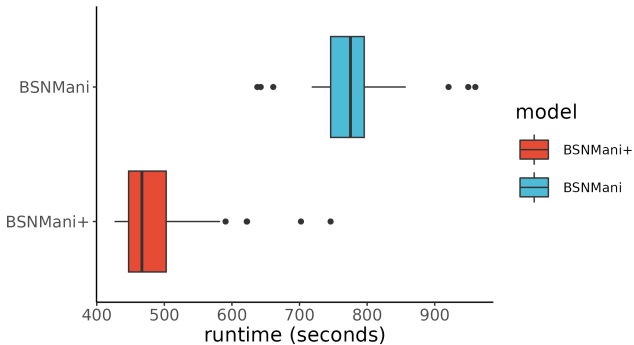
- Factor covariance matrices are only $q \times q$.
- Products such as $Y_i V \Lambda_i$ still access the network data.

Computation

- Update factor matrices, loadings, and clinical parameters by Gibbs sampling.
- Decrease τ_v^2 during annealing.
- Normalize factor columns and adjust loading/coefficient scales afterward.

Conjugate factor updates remove the polar-gradient calculation.

Runtime comparison: faster iterations



Small-scale simulation

- 100 training subjects
- 454 ROIs, $q = 3$
- Time per 1,000 iterations

Displayed medians suggest approximately 1.6–$1.7\times$ faster iterations.

Iteration time measures computational work. Convergence requires separate diagnostics.

BSNMani+: predictive performance

100 training subjects, 30 test subjects, $N = 454$, true $q = 3$.

Selected settings with network $\mathrm{SNR}_Y = 3$:

Clinical SNR	**BSNMani+**	BSNMani	Lasso	Elastic net	LOCUS+LR
0.38	**0.124**	0.111	0.040	0.049	0.009
0.50	**0.172**	0.159	0.077	0.079	0.145
1.00	**0.358**	0.348	0.246	0.249	0.341
3.00	**0.706**	0.702	0.614	0.623	0.700

Test-set R^2. LOCUS variants have identical reported values in these settings.

BSNMani+ retains strong prediction, with modest gains over BSNMani.

BSNMani+: prediction in larger simulations

$N = 350$ brain regions, true $q = 3$ subnetworks.

Simulation setting			Predictive R^2					
M	SNR_Y	SNR_C	**BSNMani+**	Lasso	Elastic net	LOCUS (SCAD) + LR	LOCUS (L1) + LR	LOCUS + LR
1,300	3	3	**0.759**	0.748	0.754	0.572	0.571	0.572
1,300	6	3	**0.760**	0.749	0.754	0.571	0.573	0.595
1,300	3	0.38	**0.336**	0.286	0.319	0.238	0.238	0.238
1,300	3	6	**0.857**	0.852	0.855	0.651	0.650	0.651
3,900	3	3	**0.772**	0.768	0.770	0.605	0.601	0.605
3,900	6	3	**0.772**	0.769	0.769	0.611	0.611	0.603
3,900	3	0.38	**0.321**	0.314	0.314	0.250	0.248	0.250
3,900	3	6	**0.869**	0.867	0.868	0.681	0.677	0.682

M: total subjects. Training/test splits: 1,000/300 and 3,000/900. Bold: highest R^2 in each row.
Selected settings: $\mathrm{SNR}_Y \in \{3, 6\}$. BSNMani was omitted because of runtime restrictions.

BSNMani+ has the highest predictive R^2 in all eight displayed settings.

Application: ABCD study

Data

- $M = 4{,}549$ children from 16 sites
- Resting-state functional connectivity
- $N = 352$ ROIs, Gordon parcellation
- General cognitive ability score

Adjusted covariates

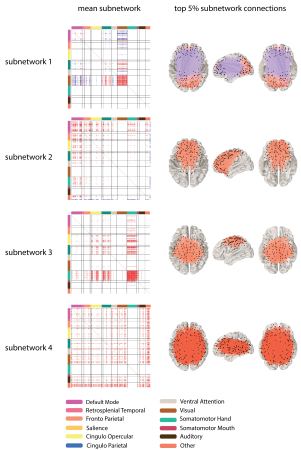
- Age, gender, and ethnicity
- Household marital status and income
- Highest parental education level

Model choice

$q = 4$ selected by deviance information criterion (DIC).

The extension supports a cohort of thousands with shared interpretable subnetworks.

ABCD: recovered connectivity patterns



- **Subnetwork 1:** visual with attention, frontoparietal, and default-mode connections.
- **Subnetwork 2:** default-mode connections across modules.
- **Subnetwork 3:** somatomotor-hand with cognitive-control and attention modules.
- **Subnetwork 4:** distributed cognitive and sensory connections.

Strongest 5% of connections. Blue edges are negative and red edges are positive.

Limitations and future directions

Current limitations

- Large M and N still require repeated network operations.
- BSNMani+ uses approximate orthogonality and symmetry.
- Linear clinical models may be restrictive.

Future directions

- Runtime, memory, and effective sample size as M, N, q grow
- Geometric error and sensitivity to the coupling schedule
- Variational or amortized inference
- Multimodal and longitudinal outcomes

Scalability combines computational efficiency with reliable posterior inference.

BSNMani and its scalable extension

BSNMani: the main framework

Joint Bayesian learning of population subnetworks, subject loadings, and clinical effects, with an orthogonal basis on the Stiefel manifold.

Li et al. (2026). *BSNMani: Bayesian scalar-on-network regression with manifold learning. The Annals of Applied Statistics.*

BSNMani+: the extension

Gaussian factor priors relax the exact constraints and allow conjugate Gibbs updates, with an application to 4,549 participants.

Yijun Li and Jian Kang. *BSNMani+*. Working paper.

Thank you!