Variational Bayesian Weighted Complex Network Reconstruction

Shuang Xu (徐爽)
School of Mathematics and Statistics
Xi’an Jiaotong University

2019.06.08

Content

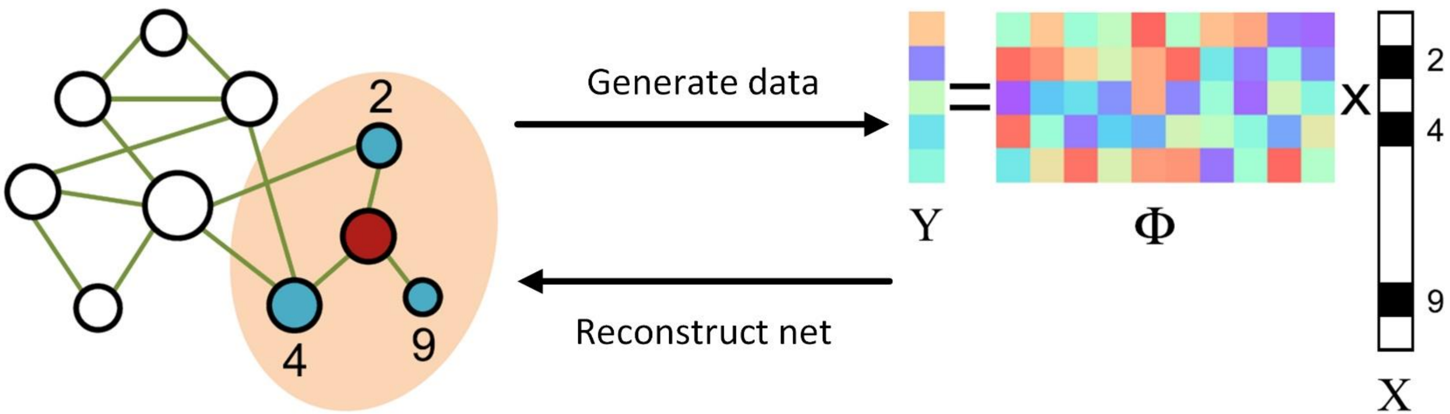
- Introduction
- New framework
- Experiments
- Conclusion and discussion

Content

- Introduction
- New framework
- Experiments
- Conclusion and discussion

Introduction

Network reconstruction



×
Not available

√
Available

Modified from: Han, et al. *Physical Review Letters* 114.2(2015):028701.

Introduction

Example

- Consider electrical current transportation (ECT) in power network
- Based on Kirchhoff's law, there is

$$\sum_{j=1}^{N} \underbrace{\frac{a_{ij}}{r_{ij}}}_{\text{Resistance}} \overbrace{(V_i - V_j)}^{\text{Voltage}} = \underbrace{I_i}_{\text{Electrical current}}, (i = 1,2,\cdots,N)$$

Introduction

Example

- If data is collected at M time points, we have

$$\begin{bmatrix} I_{i,1} \\ \vdots \\ I_{i,M} \end{bmatrix}_{M\times 1} = \begin{bmatrix} V_{i,1} - V_{j,1} & \cdots & V_{i,1} - V_{N,1} \\ \vdots & \ddots & \vdots \\ V_{i,M} - V_{j,M} & \cdots & V_{i,M} - V_{N,M} \end{bmatrix}_{M\times N} \begin{bmatrix} a_{i,1}/r_{i,1} \\ \vdots \\ a_{i,N}/r_{i,N} \end{bmatrix}_{N\times 1}$$

Introduction

Example

- If data is collected at M time points, we have

$$\begin{bmatrix} I_{i,1} \\ \vdots \\ I_{i,M} \end{bmatrix}_{M\times 1} = \begin{bmatrix} V_{i,1} - V_{j,1} & \cdots & V_{i,1} - V_{N,1} \\ \vdots & \ddots & \vdots \\ V_{i,M} - V_{j,M} & \cdots & V_{i,M} - V_{N,M} \end{bmatrix}_{M\times N} \underbrace{\begin{bmatrix} a_{i,1}/r_{i,1} \\ \vdots \\ a_{i,N}/r_{i,N} \end{bmatrix}_{N\times 1}}_{\text{Net Topology}}$$

$$\text{Data} \xrightarrow[\text{Network Reconstruction}]{} \text{Net Topology}$$

Introduction

Related work

- Granger causality analysis (GCA)
  - Partial GCA [Guo, et al., 2008. *J. Neurosci. Meth.* 172, 79–93.]
  - Piecewise GCA [Wu, et al., 2012. *Phys. Rev. E* 86, 046106.]
- Lasso / Compressive sensing
  - Complex network [Han, et al., 2015. *Phys. Rev. Lett.* 114, Art. No. 028701.]
  - Multilayer networks [Mei, et al., 2018. *IEEE T. Cybern.* 48, 754–764.]
- Statistical inference
  - Binary time series [Ma, et al., 2018. *Phys. Rev. E* 97, 022301.]
  - Ising model [Xiang, et al. 2018. *Chaos* 28, 123117.]

Introduction

Related work

- Few researchers pay attention to **large-scale weighted** network reconstruction.

1. Zhang, et al., 2018. **Reconstruction of complex time-varying weighted networks based on lasso**, in: 2018 37th *Chinese Control Conference (CCC)*, pp. 6417–6422.
2. Liu, et al., 2018. **Robust reconstruction of continuously time-varying topologies of weighted networks**. *IEEE T. Circ. & Sys*. I 65, 2970–2982.

Very small-scale!

Introduction

Difficulty

1. Simultaneously estimate linkages and weights.
2. If use lasso, we have to carefully tune hyper-parameter

Goal

To propose a new framework for large-scale weighted networks with accurate reconstruction and fast speed.

Content

- Introduction
- New framework
- Experiments
- Conclusion and discussion

New framework

Problem formulation

$$\begin{bmatrix} I_{i,1} \\ \vdots \\ I_{i,M} \end{bmatrix}_{M\times 1} = \begin{bmatrix} V_{i,1}-V_{1,1} & \cdots & V_{i,1}-V_{N,1} \\ \vdots & \ddots & \vdots \\ V_{i,M}-V_{1,M} & \cdots & V_{i,M}-V_{N,M} \end{bmatrix}_{M\times N} \begin{bmatrix} a_{i,1}/r_{i,1} \\ \vdots \\ a_{i,N}/r_{i,N} \end{bmatrix}_{N\times 1}$$

$$\begin{bmatrix} y_{i,1} \\ \vdots \\ y_{i,M} \end{bmatrix}_{M\times 1} = \begin{bmatrix} x_{1,1}^{(i)} & \cdots & x_{N,1}^{(i)} \\ \vdots & \ddots & \vdots \\ x_{1,M}^{(i)} & \cdots & x_{N,M}^{(i)} \end{bmatrix}_{M\times N} \begin{bmatrix} a_{i,1}w_{i,1} \\ \vdots \\ a_{i,N}w_{i,N} \end{bmatrix}_{N\times 1}$$

$$\boldsymbol{y} = \boldsymbol{X}(\boldsymbol{a} \odot \boldsymbol{w}) = \boldsymbol{X}\mathbb{D}(\boldsymbol{a})\boldsymbol{w}$$

Least Square ?

(Binary)Linkage (Continuous)Weight

New framework

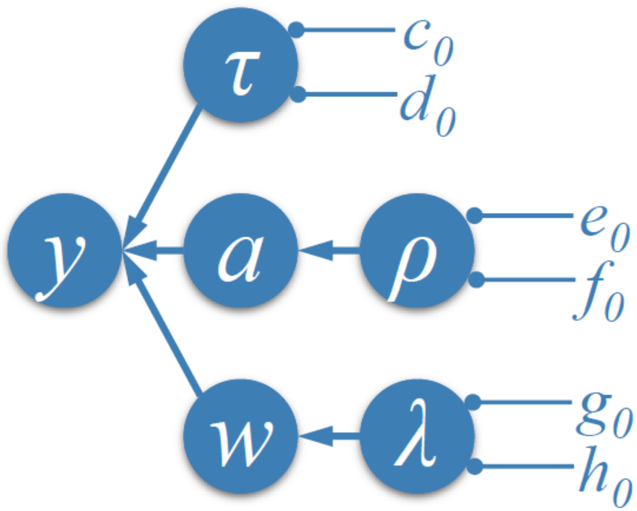
Hierarchical Bayesian Model

$$\boldsymbol{y} = \boldsymbol{X}\mathbb{D}(\boldsymbol{a})\boldsymbol{w} + \boldsymbol{\epsilon}$$

- Noise term: $\epsilon_i \sim \mathcal{N}(0, \tau^{-1})$
- Continuous coefficient: $w_j \sim \mathcal{N}(0, \lambda_j^{-1})$
- Binary coefficient: $a_j \sim \mathrm{Bernoulli}(\rho)$
- Conjugate prior:
  - $\tau \sim \mathrm{Gamma}(c_0, d_0)$
  - $\lambda_j \sim \mathrm{Gamma}(g_0, h_0)$
  - $\rho \sim \mathrm{Beta}(e_0, f_0)$

New framework

Hierarchical Bayesian Model



$$p(\boldsymbol{y} \mid \boldsymbol{a}, \boldsymbol{w}, \tau) = \left(\frac{\tau}{2\pi}\right)^{M/2} \exp\left(-\frac{\tau}{2}\|\boldsymbol{y} - \boldsymbol{X}\mathrm{D}(\boldsymbol{a})\boldsymbol{w}\|^2\right), p(\tau) = \frac{d_0^{c_0}}{\Gamma(c_0)}\tau^{c_0-1}\exp(-d_0\tau)$$

$$p(\boldsymbol{a} \mid \rho) = \prod_{j=1}^{N}\rho^{a_j}(1-\rho)^{1-a_j}, p(\rho) = \frac{1}{\mathrm{B}(e_0, f_0)}\rho^{e_0-1}(1-\rho)^{f_0-1}$$

$$p(\boldsymbol{w} \mid \boldsymbol{\lambda}) = \prod_{j=1}^{N}\left(\frac{\lambda_j}{2\pi}\right)^{1/2}\exp\left(-\frac{\lambda_j}{2}w_j^2\right), p(\boldsymbol{\lambda}) = \prod_{j=1}^{N}\frac{d_0^{c_0}}{\Gamma(c_0)}\lambda_j^{c_0-1}\exp(-d_0\lambda_j)$$

New framework

Hierarchical Bayesian Model

- Joint distribution

$$p(\boldsymbol{y},\boldsymbol{a},\boldsymbol{w},\lambda,\tau,\rho)=p(\boldsymbol{y}|\boldsymbol{a},\boldsymbol{w},\tau)p(\tau)p(\rho)\prod_j p(a_j|\rho)p(w_j|\lambda_j)p(\lambda_j)$$

computationally infeasible

- Posterior distribution (Goal)

$$p(\boldsymbol{a},\boldsymbol{w},\lambda,\tau,\rho|\boldsymbol{y})=p(\boldsymbol{y},\boldsymbol{a},\boldsymbol{w},\lambda,\tau,\rho)/p(\boldsymbol{y})$$

$$p(\boldsymbol{y})=\int p(\boldsymbol{y}|\boldsymbol{a},\boldsymbol{w},\lambda,\tau,\rho)p(\boldsymbol{a},\boldsymbol{w},\lambda,\tau,\rho)\,\mathrm{d}\boldsymbol{a}\mathrm{d}\boldsymbol{w}\mathrm{d}\lambda\mathrm{d}\tau\mathrm{d}\rho$$

New framework

Variational Inference

- Use variational distribution

$$q(\boldsymbol{a}, \boldsymbol{w}, \lambda, \tau, \rho) = q(\boldsymbol{a})q(\boldsymbol{w})q(\lambda)q(\tau)q(\rho)$$

to approximate posterior $p(\boldsymbol{a}, \boldsymbol{w}, \lambda, \tau, \rho|\boldsymbol{y})$

- In formula, we employ KL divergence measure the difference,

$$q^* = \min_{q(\boldsymbol{a}, \boldsymbol{w}, \lambda, \tau, \rho)} KL(q(\boldsymbol{a}, \boldsymbol{w}, \lambda, \tau, \rho)||p(\boldsymbol{a}, \boldsymbol{w}, \lambda, \tau, \rho|\boldsymbol{y}))$$

New framework

Variational Inference

- ***Theorem 1***. The optimal variational posterior distributions of our model are

$$
\begin{aligned}
q(\boldsymbol{w}) &= \mathcal{N}(\boldsymbol{w} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}), \\
q(\tau) &= \text{Gamma}(\tau \mid c, d), \\
q(\rho) &= \text{Beta}(\rho \mid e, f), \\
q(a_j) &= \text{Bernoulli}(a_j \mid \theta_j), j = 1, \cdots, N, \\
q(\lambda_j) &= \text{Gamma}(\lambda_j \mid g_j, h_j), j = 1, \cdots, N,
\end{aligned}
$$

(See proof at https://arxiv.org/abs/1812.04369)

New framework

Variational Inference

Algorithm 1 $(\boldsymbol{\mu}, \boldsymbol{\theta}) = \mathsf{vbr}(\boldsymbol{X}, \boldsymbol{y})$

1: Initialize $g_0 = c_0 = 10^{-2}, d_0 = h_0 = 10^{-4}$, $e_0 = 1, f_0 = N$, $\boldsymbol{\theta}^{(0)} = \mathbf{1}, c = c_0 + M/2, g_j = g_0 + 1/2, (j = 1, \cdots, p)$.
2: **while** the convergence criterion does not satisfy **do**
3: Update parameters according to theorem 1;
4: **end while**

$$\boldsymbol{\Omega} = \boldsymbol{\theta}\boldsymbol{\theta}^{\mathrm{T}} + \mathbb{D}(\boldsymbol{\theta}) \odot (\boldsymbol{I}_N - \mathbb{D}(\boldsymbol{\theta})),$$

$$\boldsymbol{\Sigma} = \left[\frac{c}{d}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}) \odot \boldsymbol{\Omega} + \mathbb{D}\left(\frac{\boldsymbol{g}}{\boldsymbol{h}}\right)\right]^{-1}, \boxed{\boldsymbol{\mu} = \frac{c}{d}\boldsymbol{\Sigma}\mathbb{D}(\boldsymbol{\theta})\boldsymbol{X}^{\mathrm{T}}\boldsymbol{y},}$$

$$g_j = g_0 + \frac{1}{2},$$

$$h_j = h_0 + \frac{1}{2}(\Sigma_{jj} + \mu_j^2),$$

$$c = c_0 + \frac{M}{2},$$

$$d = d_0 + \frac{1}{2}\left\{\|\boldsymbol{y}\|^2 - 2\boldsymbol{y}^{\mathrm{T}}\boldsymbol{X}\mathbb{D}(\boldsymbol{\theta})\boldsymbol{\mu} + \mathrm{tr}\left[\left((\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}) \odot \boldsymbol{\Omega}\right)(\boldsymbol{\Sigma} + \boldsymbol{\mu}^{\mathrm{T}}\boldsymbol{\mu})\right]\right\},$$

$$\boxed{\begin{aligned} \theta_j &= \frac{1}{\exp(-u_j) + 1}, \\ u_j &= \psi(e) - \psi(f) + \frac{c}{2d}\Big\{X_j^{\mathrm{T}}X_j[\mu_j^2\mathbb{D}(\theta) - 0.5(\Sigma_{jj} + \mu_j^2)] \\ &\quad + \mu_j X_j^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{X}\mathbb{D}(\theta)\boldsymbol{\mu})\Big\}, \end{aligned}}$$

$$e = e_0 + \sum_{j=1}^{N}\theta_j, f = f_0 + \sum_{j=1}^{N}(1 - \theta_j).$$

New framework

Variational Inference

Algorithm 2 Bayesian complex network reconstruction: $\hat{\boldsymbol{W}} = \textsf{BayesRecon}(V, I)$

1: **for** $i = 1, 2, \cdots, N$ **do**

2: Compute the response vector $\boldsymbol{y}^{(i)} = (y_{t_1}^{(i)}, \cdots, y_{t_M}^{(i)})^{\mathrm{T}}$ and the design matrix $\boldsymbol{X}^{(i)} = (x_{j,t_m}^{(i)})_{M\times(N-1)}$ with $x_{j,t_m}^{(i)} = V_i(t_m) - V_j(t_m)$, where $m = 1, 2, \cdots, M$ and $j = 1, 2, \cdots, i-1, i+1, \cdots, N$.

3: Apply Algorithm 1 to $(\boldsymbol{X}^{(i)}, \boldsymbol{y}^{(i)})$ and let $(\boldsymbol{\mu}_i, \boldsymbol{\theta}_i) = \textsf{vbr}(\boldsymbol{X}^{(i)}, \boldsymbol{y}^{(i)})$.

4: **end for**

5: Let $\hat{w}_{ij} = \mu_{ij}\hat{a}_{ij}$, where $\hat{a}_{ij} = 1$ if $\theta_{ij} > 0.5$ and 0 otherwise.

Content

- Introduction
- New framework
- Experiments
- Conclusion and discussion

Experiments

Environment

- Matlab 2017b
- Intel Core CPU 3.6GHz, 8GB RAM
- Source code (`VBR`): https://github.com/xsxjtu/VBR
- `Lasso`: Statistics and Machine Learning Toolbox of Matlab

Experiments

Metrics

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{P}} = \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij} \hat{a}_{ij}}{\sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij}},$$

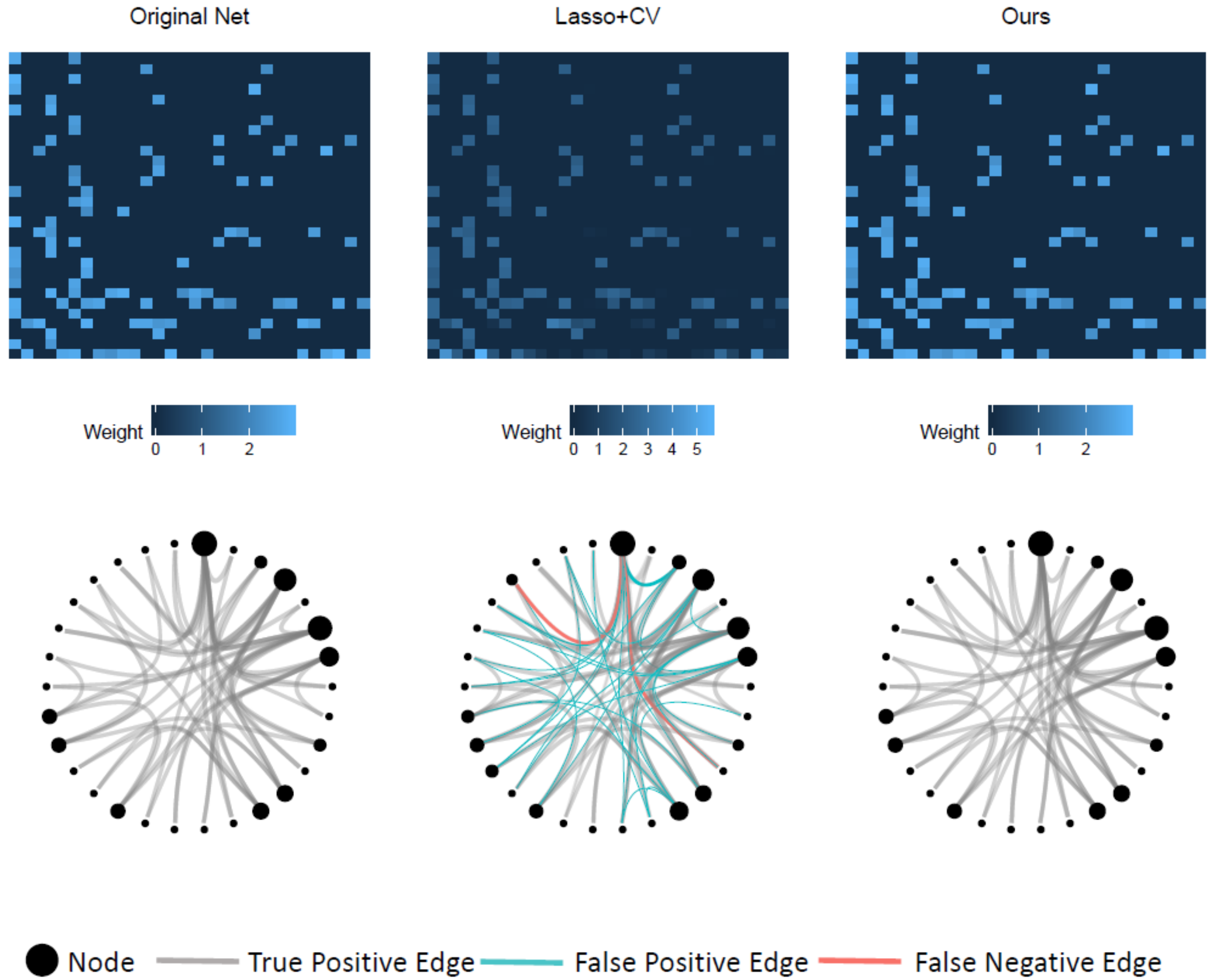
$$\mathrm{TNR} = \frac{\mathrm{TN}}{\mathrm{N}} = \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} (1 - a_{ij})(1 - \hat{a}_{ij})}{\sum_{i=1}^{N} \sum_{j=1}^{N} (1 - a_{ij})},$$

$$\mathrm{Error} = \frac{\sqrt{\sum_{i=1}^{N} \sum_{j=1}^{N} (\hat{w}_{ij} - w_{ij})^2}}{\sqrt{\sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}^2}}.$$

Experiments

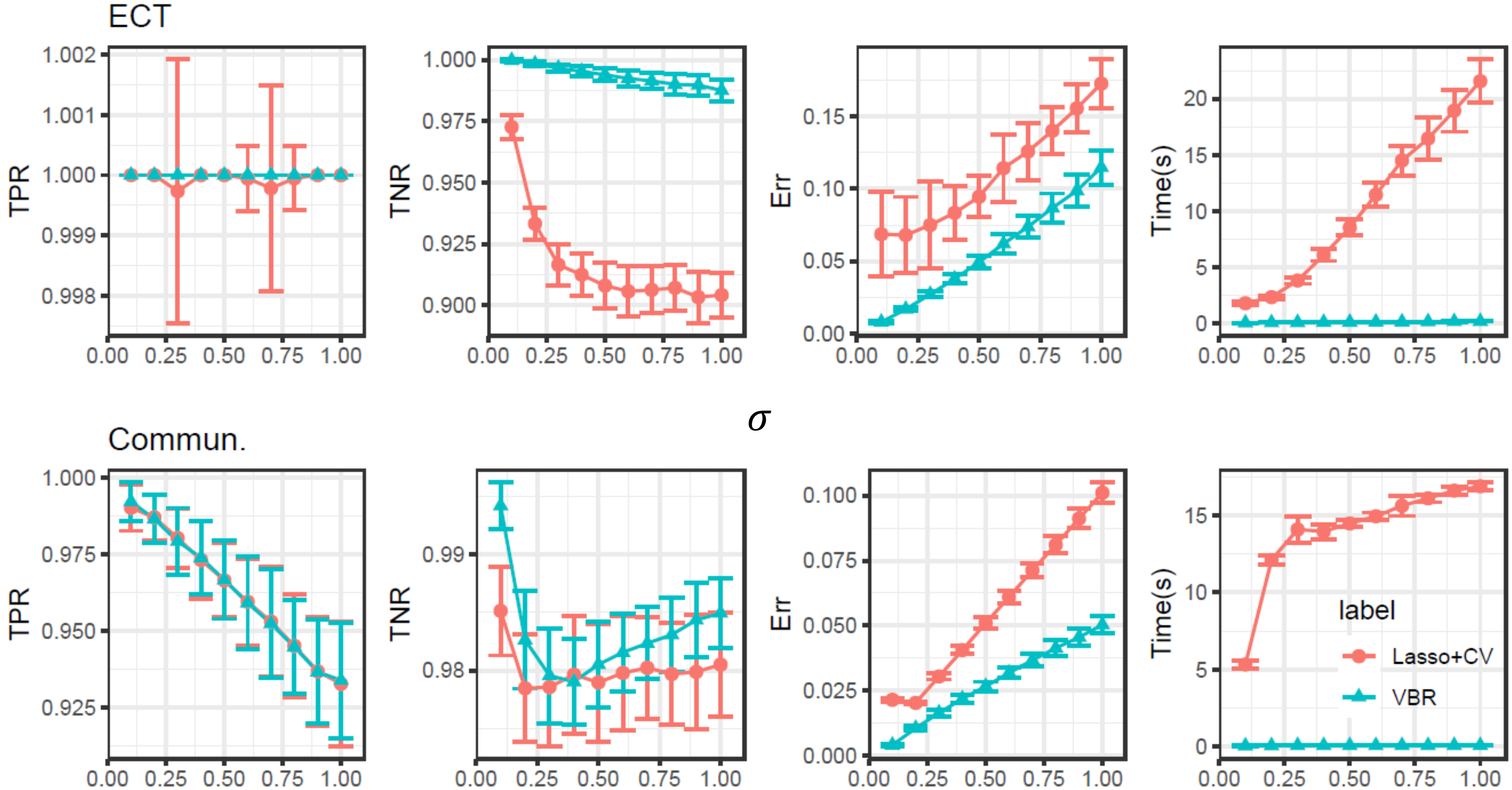
Toy Example

- BA net
- 30 nodes
- 30 time points



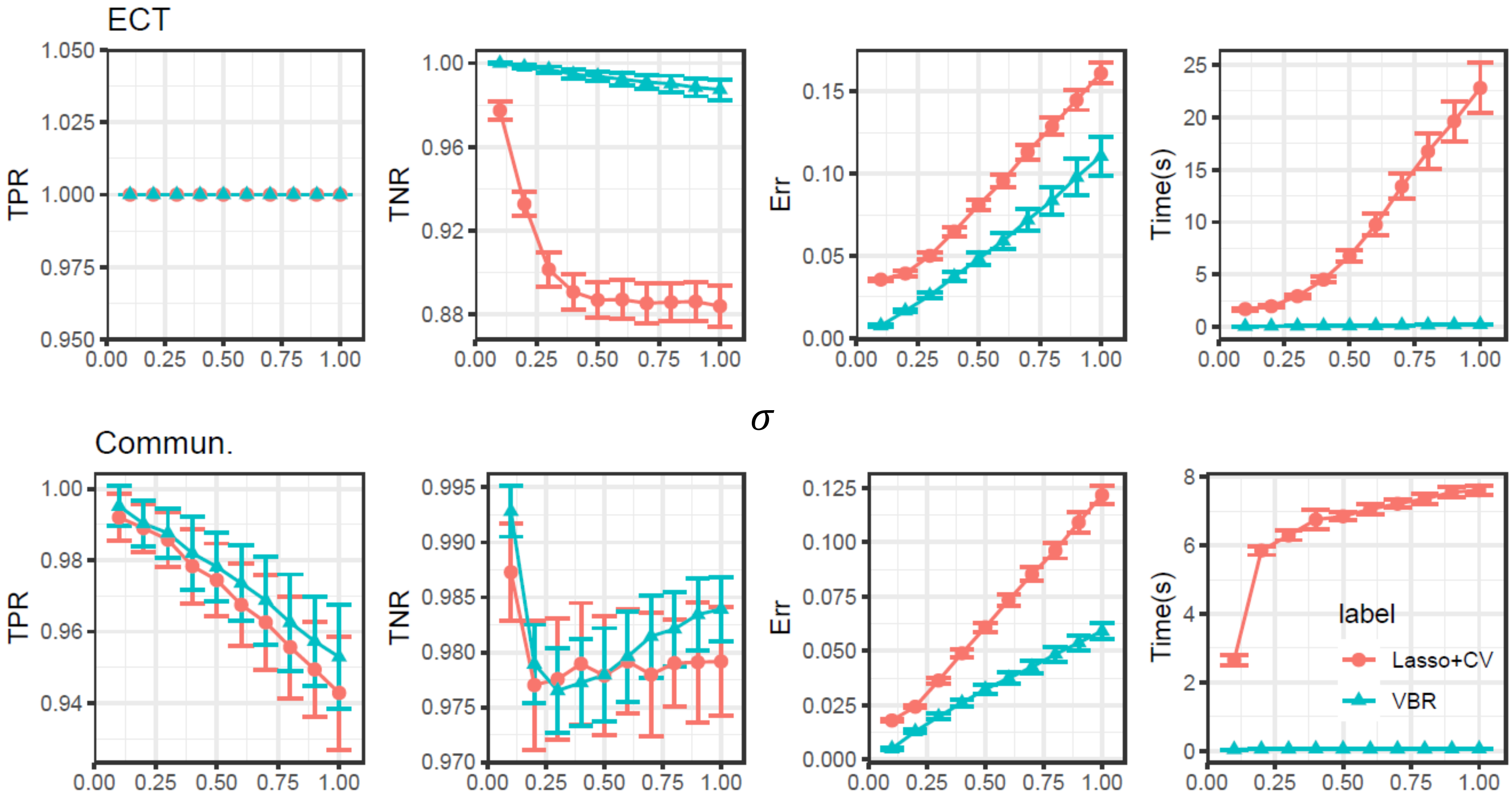
Experiments

Experiment 1: Noise sensitivity (BA-50 nodes)



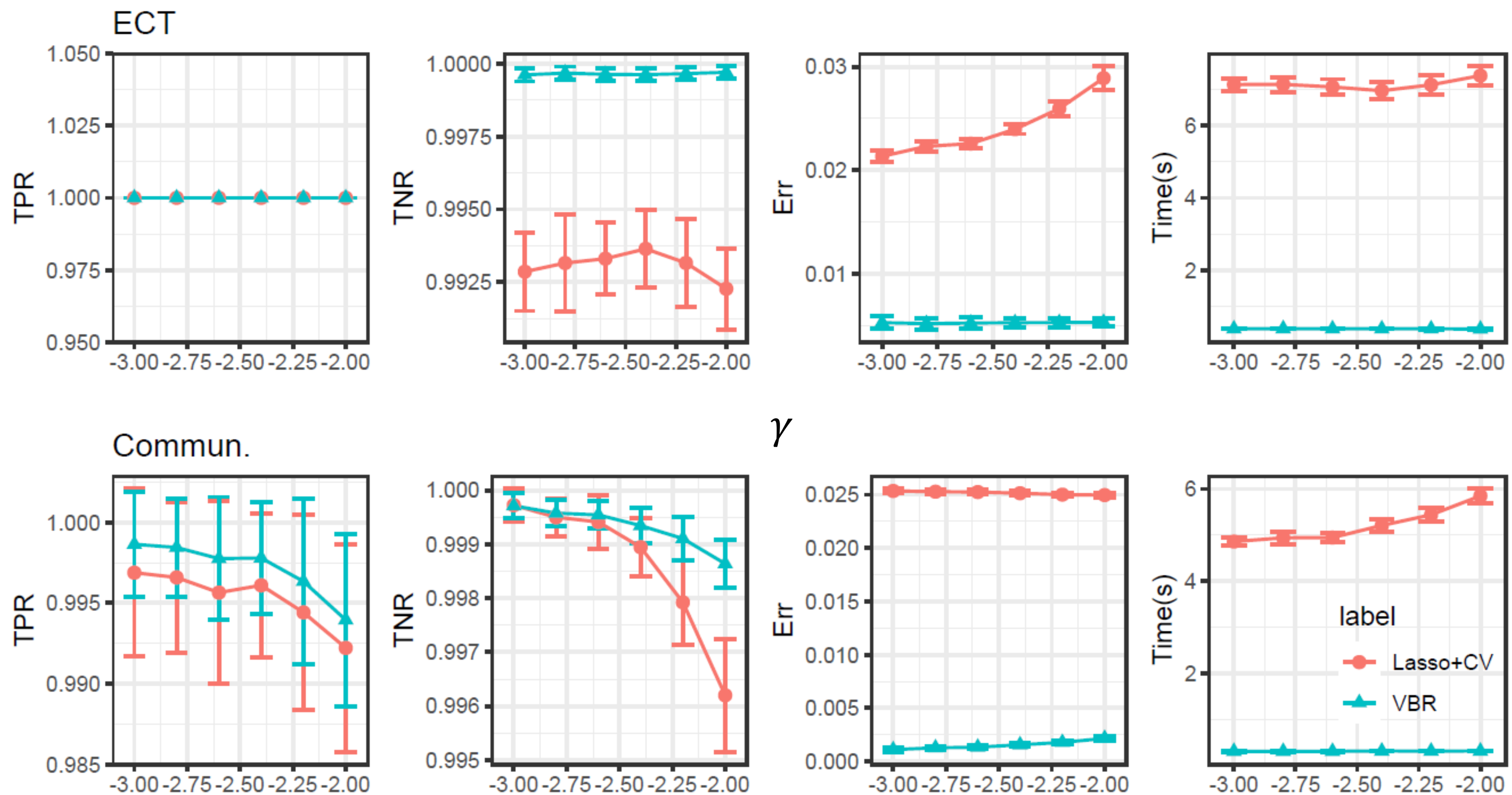
Experiments

Experiment 1: Noise sensitivity (WS-50 nodes)



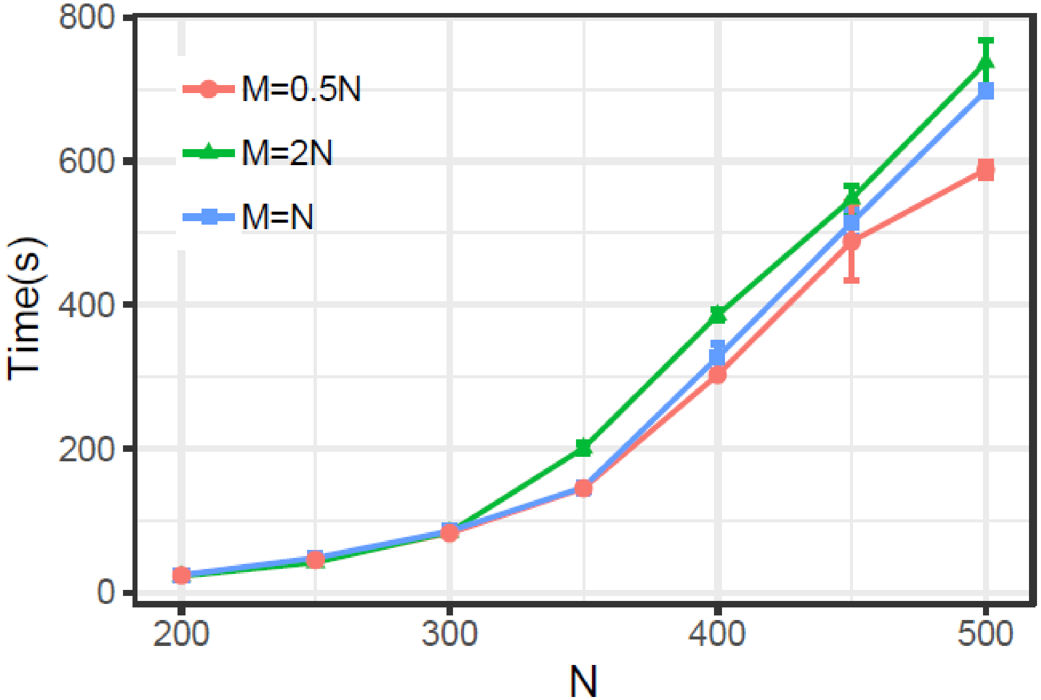
Experiments

Experiment 2: Exponential-law coefficient sensitivity (SF-100 nodes)



Experiments

Experiment 3: Running time $O(N^3)$



Experiments

Experiment 4: Real world nets

Results on empirical networks. ECT is simulated on first two networks and communication is simulated on rest ones.

Networks	N	Method	TPR	TNR	Error	Time
HB494BUS	494	Lasso	0.972	1.000	0.028	238.549
		VBR	0.987	1.000	0.002	73.710
HB1138BUS	1138	Lasso	0.948	1.000	0.027	3381.924
		VBR	0.977	1.000	0.001	1553.495
Jazz	198	Lasso	0.981	0.998	0.031	49.361
		VBR	0.969	1.000	0.003	0.913
Karate	34	Lasso	0.983	0.999	0.031	0.665
		VBR	0.997	1.000	0.000	0.015

Content

- Introduction
- New framework
- Experiments
- Conclusion and discussion

Conclusion and discussion

Lasso

$$\min|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{w}|_2^2 + \lambda_{lasso}|\boldsymbol{w}|_1$$

- Sparsity is derived from L_1 norm penalty!
- We have to fine-tune λ_{lasso} for weighted network!
- Cross validation is time-consuming!
- Bad statistical property: Biased estimation!

VBR

$$\boldsymbol{y} = \boldsymbol{X}\mathbb{D}(\boldsymbol{a})\boldsymbol{w}$$

- Respectively estimate linkages and weights.
- Full Bayesian inference.
- No parameters need fine-tune.
- An interpretable framework.

Conclusion and discussion

Future work

- Reduce computational complexity
- Design models for other dynamics

西安交通大學
XI'AN JIAOTONG UNIVERSITY

Thanks!

See details at https://arxiv.org/abs/1812.04369

Email: shuangxu@stu.xjtu.edu.cn

Homepage: https://xsxjtu.github.io/