June 4, 2018

EPFL, Lausanne, Switzerland

Semi-Blind Inference of Topologies and Dynamical Processes over Graphs

Vassilis N. Ioannidis, Yanning Shen, and Georgios B. Giannakis

Dept. of ECE and DTC, University of Minnesota

UNIVERSITY OF MINNESOTA
Driven to Discover℠

Acknowledgement: NSF 171141, 1500713, 1442686

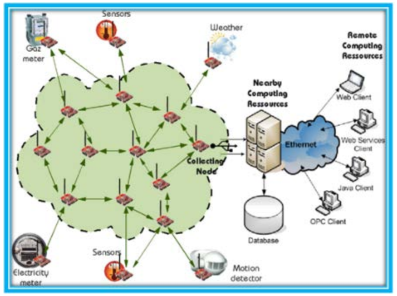
Data science and network science

Q1. What is Data Science (DS)?

A1. Data-driven toolbox discovering patterns/laws *along* and *beyond* physics

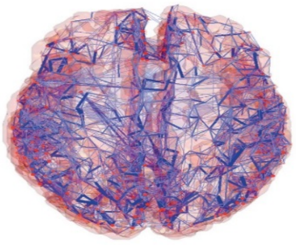
Q2. What about DS ***and*** Network Science?

A2. Networks produce (big) data, and capture data interdependencies

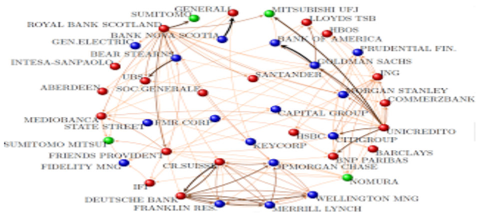


IOT networks

LinkedIn employee network



Brain networks



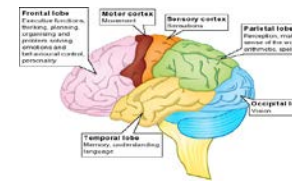
Financial networks

G. B. Giannakis, Y. Shen, and G. V. Karanikolas "Topology Identification and Learning over Graphs: Accounting for Nonlinearities and Dynamics," *Proceedings of the IEEE,* pp. 787-807, May 2018.

Opportunities and challenges

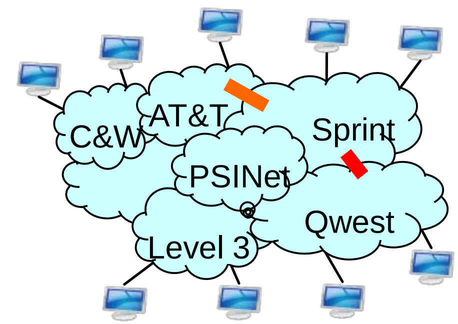
O1. Parsimonious models of network structure

- Identify fake news; Human brain connectome project



O2. Efficient inference algorithms for processes over networks

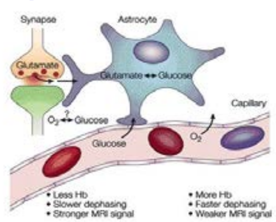
- Predict com. traffic delay; and stock evolution in financial nets



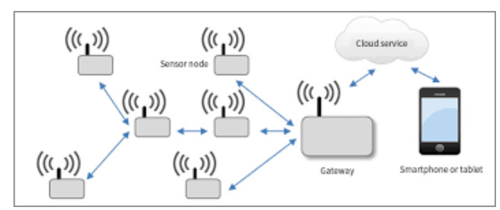
C. Attributes available at subsets of nodes

Privacy

Experimental error



Energy conservation



Desiderata: **J**oint **I**dentification of topologies and **S**ignals over **G**raphs (**JISG**)

V. N. **I**oannidis, Y. **S**hen, and G. B. **G**iannakis, "Semi-Blind Inference of Topologies and Dynamical Processes over Graphs," *IEEE Transactions on Signal Processing*, submitted May 2018.

Prior art

- Learning processes over graphs
  - Laplacian kernels — [Smola-Kondor'03], [Forero et al' 14]
  - Multi-kernel learning — [Shen etal'17], [Romero et al'17]
  - Graph-bandlimited model — [Narang et al'13], [Shuman et al'13], [Segarra et al' 16], [Di Lorenzo et al' 16]
  - Graph filters — [Isufi et al'17]

  (Laplacian kernels through graph-bandlimited model: Topology known)

  - Kalman filters on graphs — [Rajawat et al'14], [Ioannidis et al'18]

- Topology identification
  - Structural equation models — [Blalock' 61], [Cai et al' 13], [Baingana et al' 14]
  - Modeling nonlinear relations — [Shen et al'17]
  - Capturing diffusions — [Thanou et al'17]

  (Observations at all nodes)

Q: What if topology is unknown/partially known, and just a subset of data is available?

Semi-blind problem formulation

- Network $\mathcal{G} := (\mathcal{V}, \mathbf{A})$ $\quad \mathcal{V} := \{v_1, \ldots, v_N\}$
- Network process $\mathbf{y}_t = [y_{1t}, \ldots, y_{Nt}]$

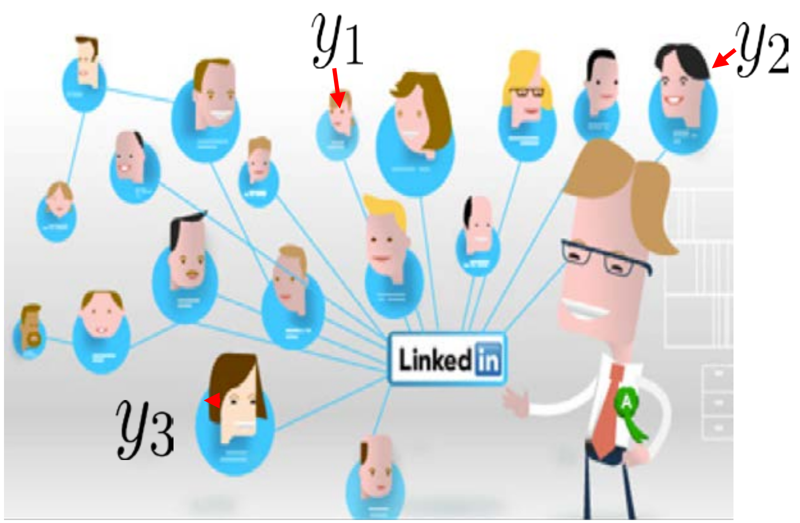
y_{nt} stock price, salary, protein expression, rating, network delay

- Only measure $M_t < N$ nodes (privacy, communication, misses)

$$z_{mt} = y_{n_m t} + \epsilon_{mt}$$

$$\mathbf{z}_t = \mathbf{M}_t \mathbf{y}_t + \boldsymbol{\epsilon}_t$$

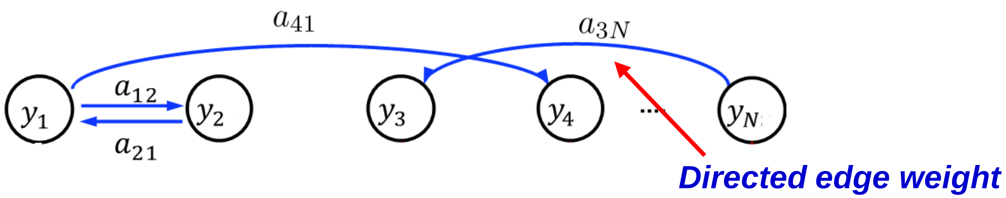
selection matrix $\mathbf{M}_t \in \{0, 1\}^{M_t \times N}$



Goal: Given $\{\mathbf{z}_t, \mathbf{M}_t\}_{t=1}^T$ learn $\{\mathbf{y}_t\}_{t=1}^T$ and $\mathbf{A}$

Structural models

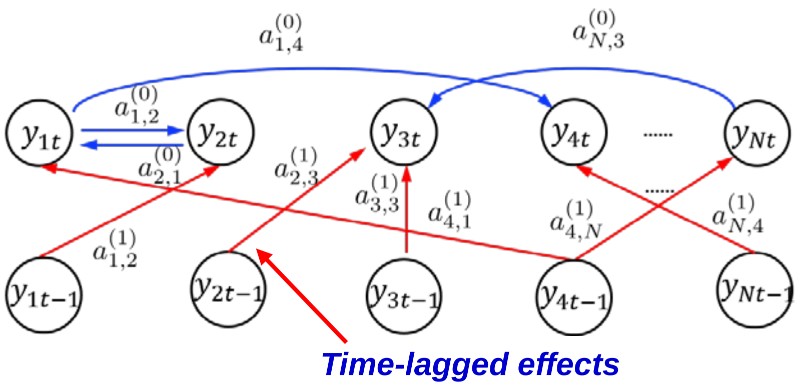
- Structural equation models (SEMs)



$$y_{nt} = \sum_{n' \neq n} a_{n,n'} y_{n't} + e_{nt}$$

$$\mathbf{y}_t = \mathbf{A}\mathbf{y}_t + \mathbf{e}_t$$

- Structural vector autoregressive models (SVARMs)



$$y_{nt} = \sum_{n' \neq n} a^{(0)}_{n,n'} y_{n't} + \sum_{n'=1}^{N} a^{(1)}_{n,n'} y_{n'(t-1)} + e_{nt}$$

$$\mathbf{y}_t = \mathbf{A}^{(0)}\mathbf{y}_t + \mathbf{A}^{(1)}\mathbf{y}_t + \mathbf{e}_t$$

R. Hoyle, *Handbook of Structural Equation Modeling*, New York, NY, Guilford Press, 2012.

JISG over time

$$\min_{\substack{\mathbf{A}^{(0)},\mathbf{A}^{(1)}\in\mathcal{A},\\ \{\mathbf{y}_t\}_{t=1}^T}} \sum_{t=1}^{T} \|\mathbf{y}_t - \mathbf{A}^{(0)}\mathbf{y}_t - \mathbf{A}^{(1)}\mathbf{y}_{t-1}\|_2^2 + \|\mathbf{y}_0 - \boldsymbol{\mu}_0\|_2^2 + \mu \sum_{t=1}^{T} \|\mathbf{z}_t - \mathbf{M}_t\mathbf{y}_t\|_2^2 + \rho(\mathbf{A}^{(0)}) + \rho(\mathbf{A}^{(1)})$$

- Elastic net regularizer $\rho(\mathbf{A}) = \lambda_1\|\mathbf{A}\| + \lambda_2\|\mathbf{A}\|_F^2$; $\boldsymbol{\mu}_0$ initial process value
- Application-specific constraints on ***A*** (positive edges, partially known edge weights)

- BCD solver alternates between topology ID and inferring the process
  - Guaranteed convergence with complexity linear in *T*

- ADMM solver for identifying $\mathbf{A}^{(0)}, \mathbf{A}^{(1)}$

Efficient graph signal estimator

- Given $\hat{\mathbf{A}}^{(0)}$, and $\hat{\mathbf{A}}^{(1)}$ find

$$\{\hat{\mathbf{y}}_{t|T}\}_{t=0}^{T} := \arg\min_{\{\mathbf{y}_t\}_{t=0}^{T}} \sum_{t=1}^{T} \|\mathbf{y}_t - \hat{\mathbf{A}}^{(0)}\mathbf{y}_t - \hat{\mathbf{A}}^{(1)}\mathbf{y}_{t-1}\|_2^2 + \|\mathbf{y}_0 - \boldsymbol{\mu}_0\|_2^2 + \sum_{t=1}^{T} \frac{\mu}{M_t} \|\mathbf{z}_t - \mathbf{M}_t\mathbf{y}_t\|_2^2 \qquad \textbf{(P1)}$$

  - $\hat{\mathbf{y}}_{t|T}$ the estimate of $\mathbf{y}_t$ given $\{\mathbf{z}_\tau\}_{\tau=1}^{T}$
  - Directly solving **(P1)** incurs complexity *O(N^3T^3)*

- Rauch-Tung-Striebel (RTS) smoother yields $\{\hat{\mathbf{y}}_{t|T}\}_{t=0}^{T}$ iteratively
  - RTS is a forward-backward (filtering-smoothing) algorithm
  - Solves **(P1)** at complexity *O(TN^3)*
  - Scalability in *N* can be effected by distributed, e.g., consensus-based solvers

H. E. Rauch, C. T. Striebel, and F. Tung, "Maximum likelihood estimates of linear dynamic systems," *American Institute Aeronautics AstronauticsJ.*, vol. 3, no. 8, pp. 1445–1450, 1965

Fixed-lag solver for online JISG

- Smooth estimates up to $t_w := t + w$, for $\{\hat{\mathbf{A}}^{(1)}_{t_w}, \hat{\mathbf{A}}^{(0)}_{t_w}, \{\hat{\mathbf{y}}_{\tau|t_w}\}_{\tau=t}^{t_w}\}$

$$\arg\min_{\substack{\mathbf{A}^{(0)},\mathbf{A}^{(1)}\in\mathcal{A},\\ \{\mathbf{y}_\tau\}_{\tau=t}^{t_w}}} \sum_{\tau=t+1}^{t_w} \|\mathbf{y}_\tau - \mathbf{A}^{(0)}\mathbf{y}_\tau - \mathbf{A}^{(1)}\mathbf{y}_{\tau-1}\|_2^2 + \|\mathbf{y}_t - \hat{\mathbf{y}}_{t|t}\|^2_{\mathbf{\Sigma}_{t|t}} + \sum_{\tau=t+1}^{t_w} \frac{\mu}{M_\tau}\|\mathbf{z}_\tau - \mathbf{M}_\tau \mathbf{y}_\tau\|_2^2$$
$$+ \rho_e(\mathbf{A}^{(0)}) + \rho_e(\mathbf{A}^{(1)}) + \mu_A\|\mathbf{A}^{(0)} - \hat{\mathbf{A}}^{(0)}_{t_w-1}\|_F^2 + \mu_A\|\mathbf{A}^{(1)} - \hat{\mathbf{A}}^{(1)}_{t_w-1}\|_F^2$$

  - Available from RTS at $t_w - 1$: $\hat{\mathbf{A}}^{(0)}_{t_w-1}, \hat{\mathbf{A}}^{(1)}_{t_w-1}$, $\hat{\mathbf{y}}_{t|t}$, $\mathbf{\Sigma}_{t|t}$
  - LS terms, $\{\|\mathbf{A}^{(l)} - \hat{\mathbf{A}}^{(l)}_{t_w-1}\|_F^2\}_{l=0,1}$ promote slow-varying topologies

- BCD solver for the fixed-lag objective with provable convergence
  - Generates estimates within a time-window
  - Tracks time-varying topologies capturing dependencies of non-stationary processes

B. D. Anderson and J. B. Moore, *Optimal Filtering*. Prentice Hall, 1979

SEM identifiability in semi-blind setup

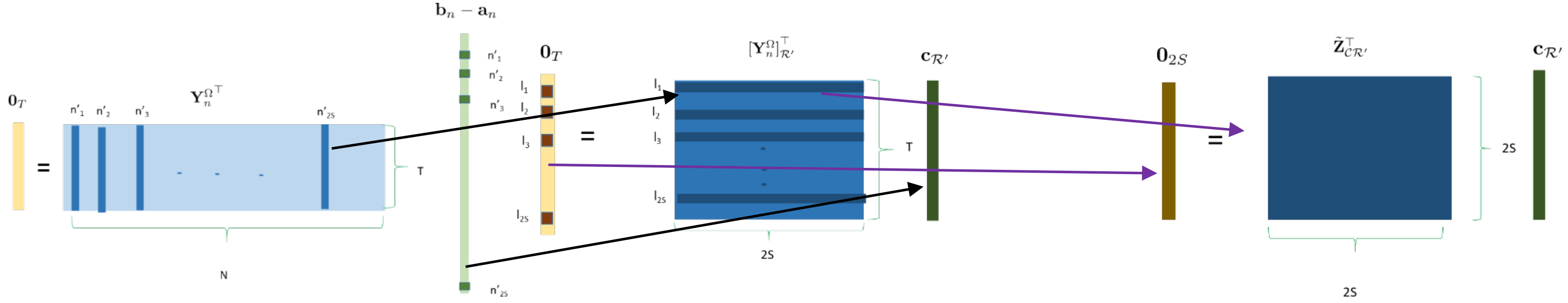
$$\mathbf{y}_t = \mathbf{A}\mathbf{y}_t \qquad (1)$$

$$\mathbf{z}_t = \mathbf{M}_t\mathbf{y}_t \qquad (2)$$

as1. The adjacency matrix **A** has at most *S* non-zero entries per row

as2. For any subset of nodes $\mathcal{R} := \{n_1, n_2, \ldots, n_{2S}\}$ there exists a subset of time-slots $\mathcal{C} := \{t_1, t_2, \ldots, t_{2S}\}$ where the *2Sx2S* observation matrix $\tilde{\mathbf{Z}}_{\mathcal{CR}}$ is fully observable and satisfies $\text{Kruskal}(\tilde{\mathbf{Z}}_{\mathcal{CR}}^\top) = 2S$.

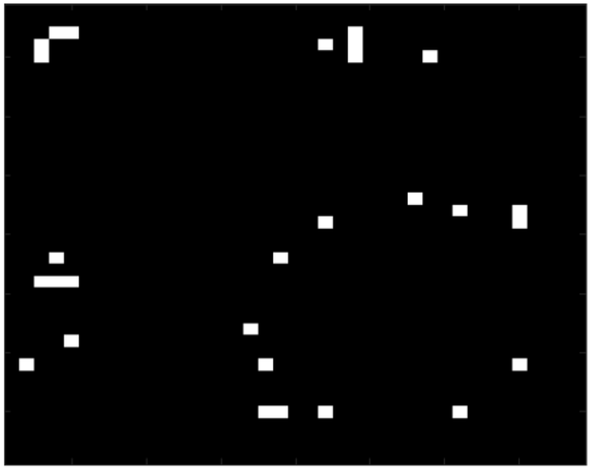
Theorem. Under (as1) and (as2), **A** is uniquely identifiable from (1), (2)

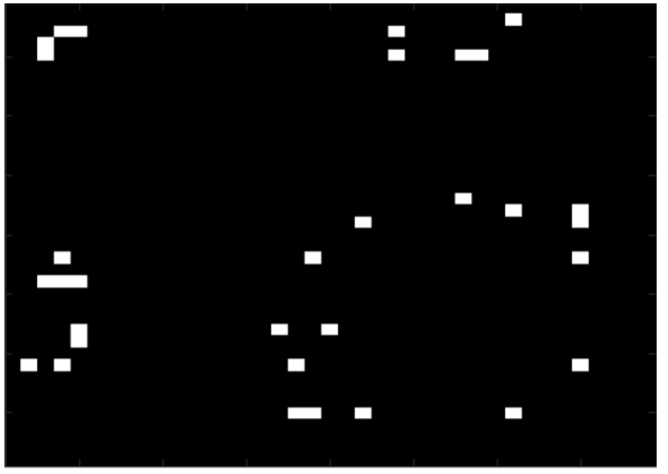


J. A. Bazerque, B. Baingana, and G. B. Giannakis, "Identifiability of Sparse Structural Equation Models for Directed, Cyclic, and Time-varying Networks," *Proc. of Global Conf. on Signal and Info. Processing*, Austin, TX, December 3-5, 2013.

Identifying gene-regulatory network topologies

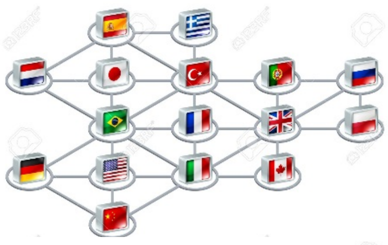
- *N*=39 immune-related genes; *T*=69 unrelated individuals; *y:* gene expression level
  - SEM oracle observes all genes *M*=39 (left); **JISG** with *M*=31 (right); *20%* misses





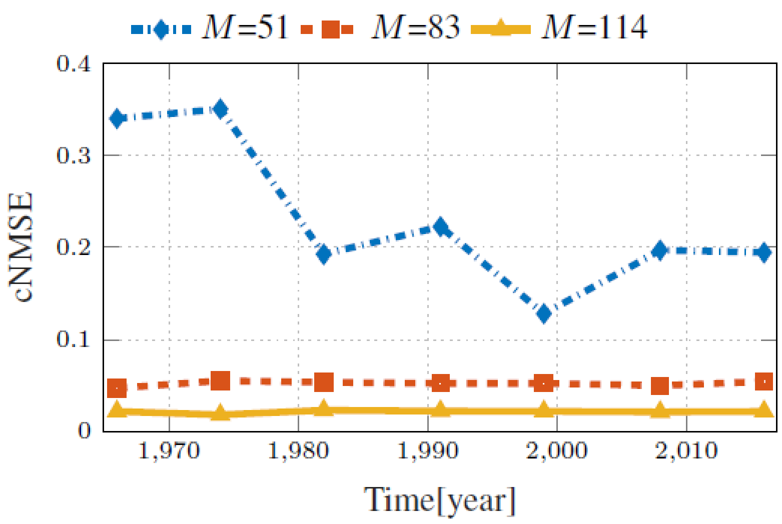
  - NMSE for $\{\hat{\mathbf{y}}_t\}_{t=1}^{T}$ 0.017
  - JISG recovers similar topology as the oracle

GDP prediction

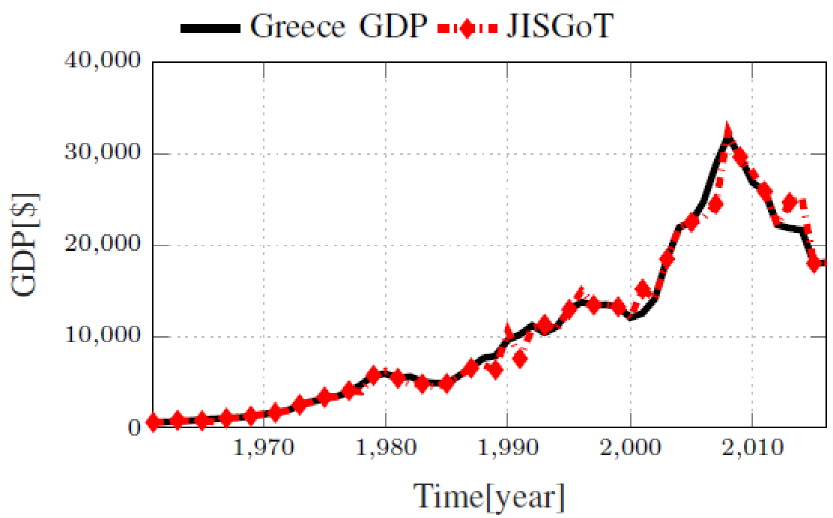


- Gross domestic product (GDP) for *N=127* countries for 1960-2016
- Cumulative normalized mean-square error $$\mathrm{cNMSE}(T) := \frac{\sum_{t=1}^{T} \|\mathbf{y}_t - \hat{\mathbf{y}}_{t|t}\|_2^2}{\sum_{t=1}^{T} \|\mathbf{y}_t\|_2^2}$$

cNMSE evolution

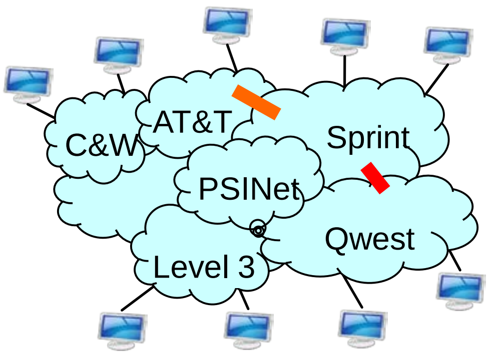


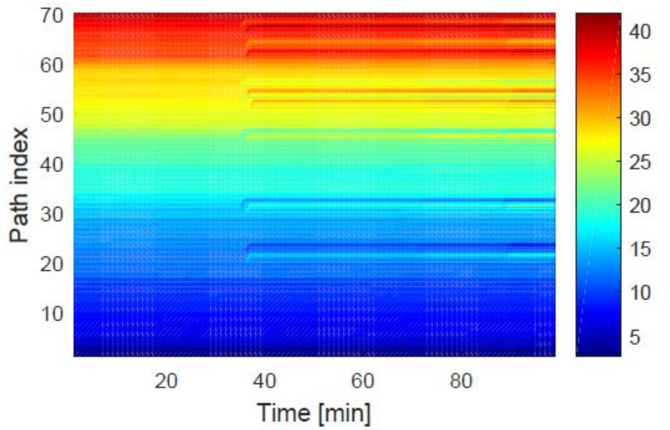
GDP tracking



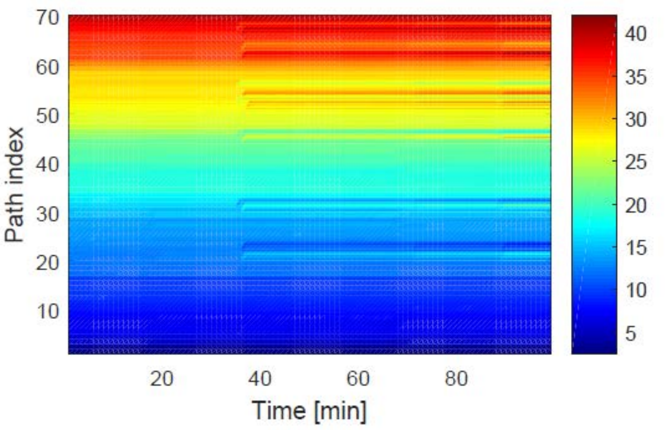
Internet2 backbone network: delay prediction

- Path delay per minute in Internet2 for *N=70* paths
- Delay maps show network delay per path over time





True



JSIG

Conclusions and the road ahead

- Contributions
  - SEMs and SVARMs for JISG
  - Efficient minimization approaches with provable convergence
  - Identifiability conditions in semi-blind setup

- Research outlook
  - Distributed version of the JISG filtering algorithms [Schizas et al'08]
  - Account for nonlinear graph processes [Shen et al'17]

Thank you

V. N. **I**oannidis, Y. **S**hen, and G. B. **G**iannakis, "Semi-Blind Inference of Topologies and Dynamical Processes over Graphs," *IEEE Transactions on Signal Processing*, submitted May 2018, available in arXiv.