Bellairs Research Institute
Barbados, February 10, 2019

Topology ID and Learning over Graphs:
Accounting for Nonlinearities and Dynamics

Georgios B. Giannakis

Acknowledgments: *Drs. B. Baingana, J.-A. Bazerque, P. Forero, G. Mateos, K. Rajawat, D. Romero; and V. Ioannidis, G.-V. Karanikolas, M. Ma, Y. Shen, and P. Traganitis*

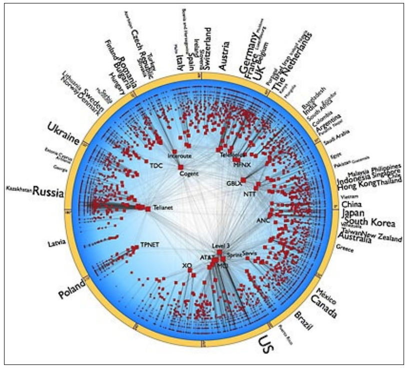
NSF 1514056, 1500713, 1711471

UNIVERSITY OF MINNESOTA
Driven to Discover℠

Networks as graphs … everywhere

Social networks

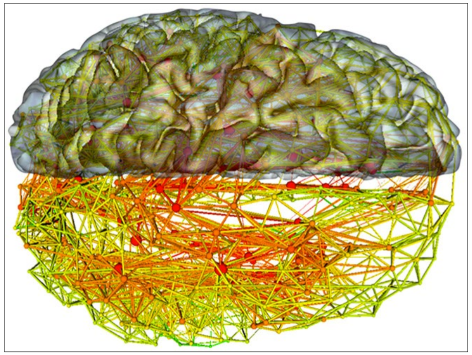
Internet



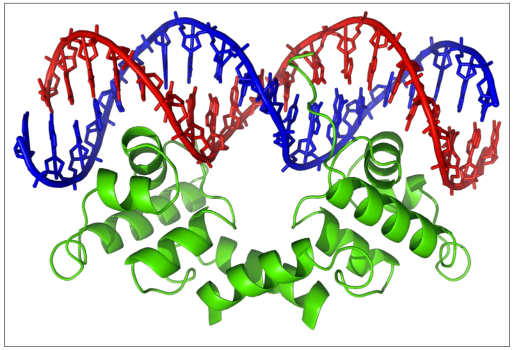
Energy grids

Network Science

Financial markets



Brain networks

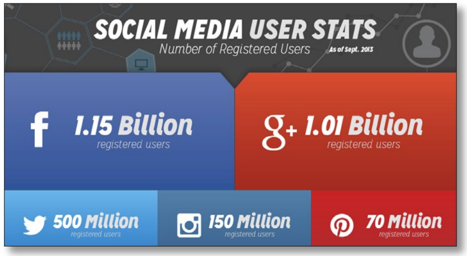


Gene/protein-regulatory nets

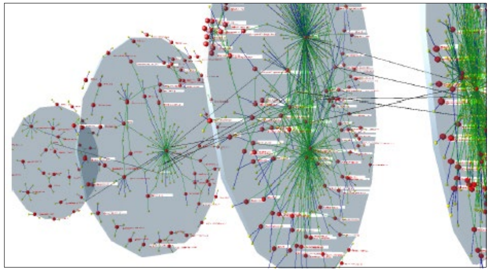
E. D. Kolaczyk, *Statistical Analysis of Network Data*, Springer, 2009.
D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Proc. Mag.*, May 2013.

Challenges and opportunities

- Network-level challenges
  - Massive scale



  - Dynamic topologies



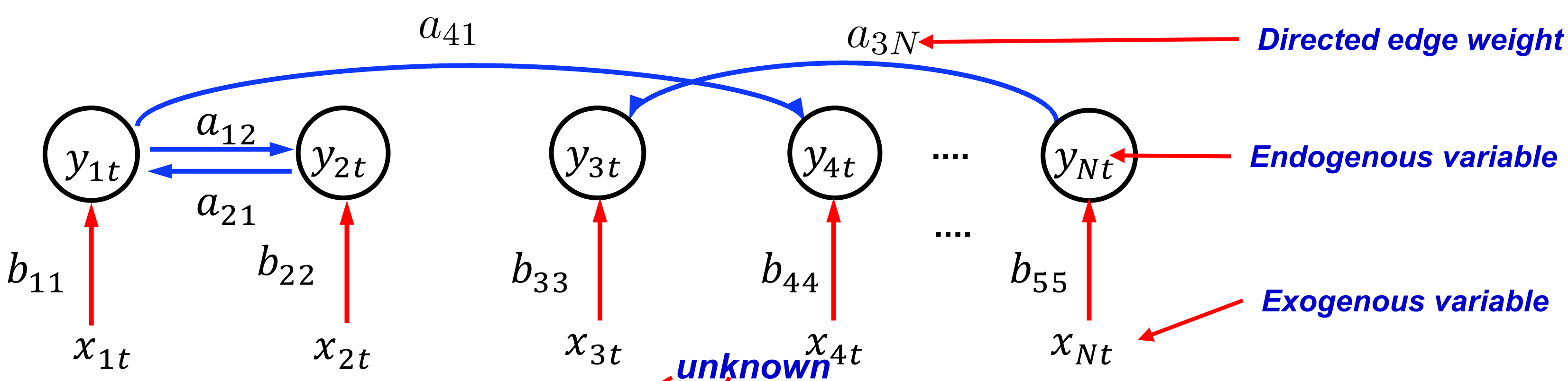
  - Streaming data (~6,000 tweets/sec)

  - Unobservable links

- **Desiderata**
  - Parsimonious models of network structure
  - Efficient inference algorithms over networks
- **Framework:** Cross-roads of machine learning, statistical SP, optimization, networking

Linear structural equation models

- Setting: *N* nodes over which "observable" processes propagate [Goldberger'72,Kaplan'09]



$$y_{it} = \underbrace{\sum_{j \neq i} a_{ij} y_{jt}}_{\text{endogenous}} + \underbrace{b_{ii} x_{it}}_{\text{exogenous}} + e_{it}$$

$$i = 1, \ldots, N$$
$$t = 1, \ldots, T$$

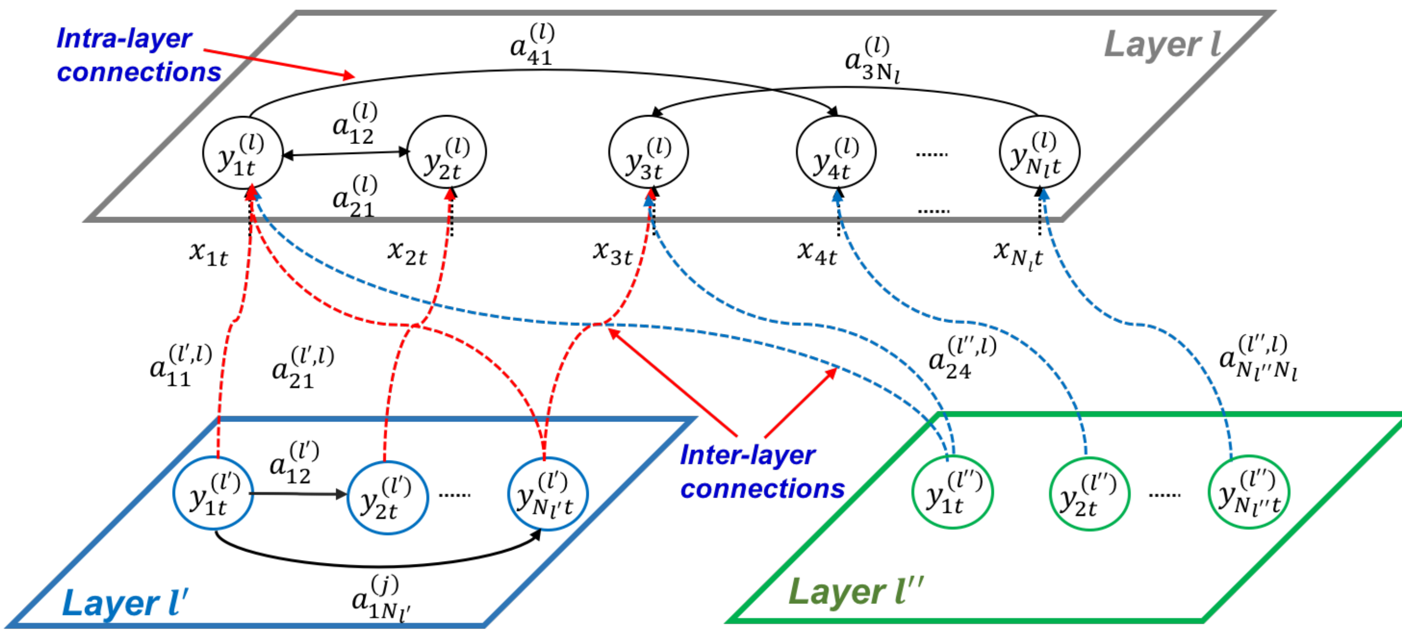
Goal: Given $\{x_{it},\ y_{it}\}$, estimate $\{a_{ij},\ b_{ii}\}$ to capture **directed** dependencies $a_{ij} \neq a_{ji}$

Theorem 1. If $a_{ii} = 0,\ b_{ii} \neq 0,\ \forall i,\ b_{ij} = 0, \forall i \neq j$, then ***A*** and ***B*** identifiable if $\text{kr}_{\mathbf{X}^T} > 2\ \text{max\_deg}\,(\mathcal{G})$, $\mathbf{X} := [\mathbf{x}_1 \ldots \mathbf{x}_T]$

OK if $T > 2N$

J. A. Bazerque, B. Baingana, and G. B. Giannakis, "Identifiability of sparse structural equation models for directed and cyclic networks," in *Proc. of Global Conf. on Signal and Info. Processing*, Austin, TX, Dec. 2013.

From single- to multi-layer SEMs



- Generalizes linear SEM to multi-layer settings
- Layers can model exogenous variables and time snapshots or lags

P. A. Traganitis, Y. Shen, and G. B. Giannakis, "Topology inference for multilayer networks," *Proc. of INFOCOM Workshop on Network Science for Communications*, Atlanta, May 2017.

Multilayer linear SEMs

- Per node

$$y_{it}^{(\ell)} = \sum_{j \neq i} a_{ij}^{(\ell)} y_{jt}^{(\ell)} + \sum_{\ell' \neq \ell} \sum_{j} a_{ij}^{(\ell',\ell)} y_{kt}^{(\ell')} + e_{it}^{(\ell)}$$

Intra-layer term ***Inter-layer term***

- Matrix form

$$\mathbf{Y}^{(\ell)} = \mathbf{Y}^{(\ell)}\mathbf{A}^{(\ell)} + \sum_{\ell' \neq \ell} \mathbf{Y}^{(\ell')}\mathbf{A}^{(\ell',\ell)} + \mathbf{E}^{(\ell)}, \quad \ell = 1, \ldots, L$$

Goal: Given $\{\mathbf{Y}^{(\ell)}\}_{\ell=1}^{L}$ find $\{\mathbf{A}^{(\ell)}\}_{\ell=1}^{L}$ $\{\mathbf{A}^{(\ell',\ell)}\}_{\ell \neq \ell'}^{L}$

- Identifiability

Theorem 2. If $\boldsymbol{Y} := [\boldsymbol{Y}^{(1)} \ldots \boldsymbol{Y}^{(L)}]$, and $\text{kr}_{\boldsymbol{Y}} > 2$ max_deg($\mathcal{G}$), then $\{\mathbf{A}^{(\ell',\ell)}\}_{\ell \neq \ell'}^{L}$, $\{\mathbf{A}^{(\ell)}\}_{\ell=1}^{L}$ identifiable

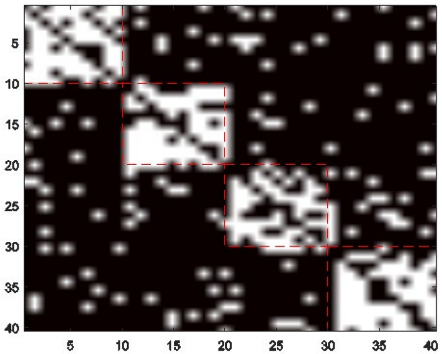
- Estimation via e.g., ordinary or regularized least-squares (LS)

P. A. Traganitis, Y. Shen, and G. B. Giannakis, "Topology inference for multilayer networks," *Proc. of INFOCOM Workshop on Network Science for Communications*, Atlanta, May 2017.

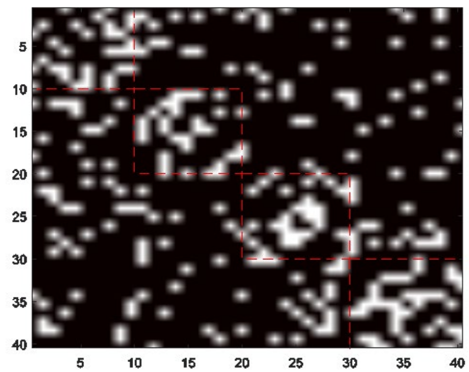
Simulated and real data tests

- **Synthetic network**, *N=40, L = 4* (each layer corresponds to a block diagonal)

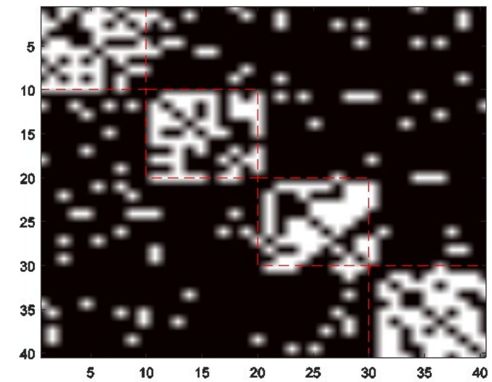
Ground-truth



Single layer SEM

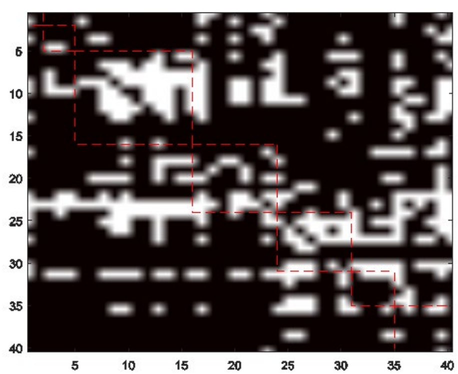


Multilayer SEM

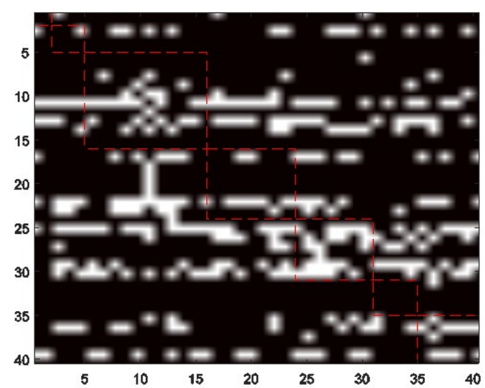


- **US economic sectors,** *N=40* industries, *L = 7* sectors (textiles, automotive …)

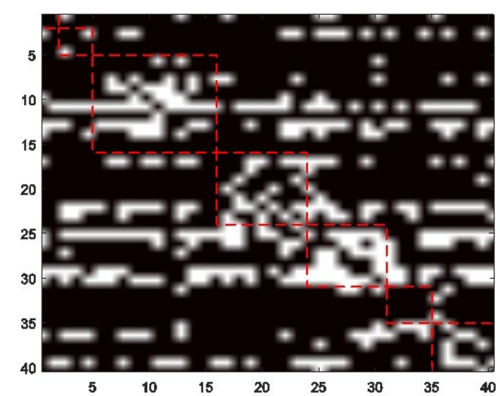
Ground-truth



Single layer SEM



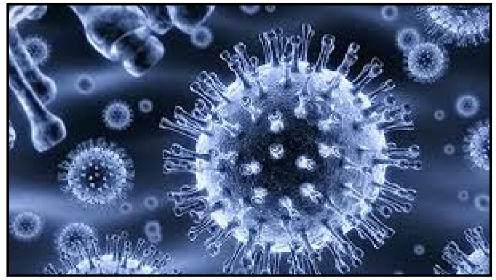
Multilayer SEM



Topology tracking from network cascades

Contagions

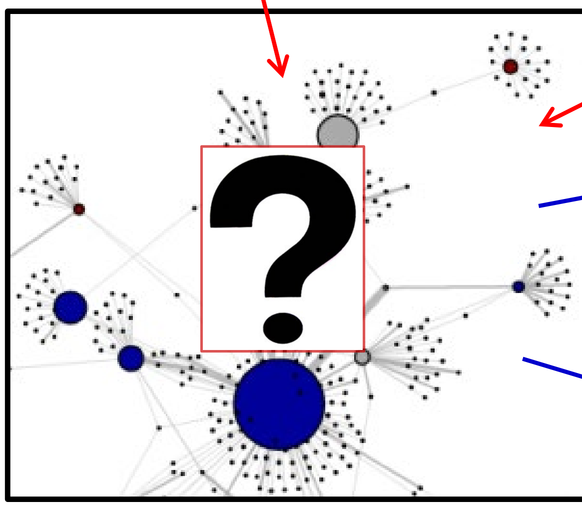
Popular news stories



Infectious diseases

Buying patterns

Propagate in cascades over implicit networks



Network topologies:

Unobservable, dynamic, sparse

Topology inference vital:

Viral advertising, healthcare policy

Desiderata: track unobservable time-varying network topology from cascade traces

B. Baingana, G. Mateos, and G. B. Giannakis, ``Dynamic structural equation models for social network topology inference," *IEEE J. of Selected Topics in Signal Processing*, vol. 8, no. 4, pp. 563-575, Aug. 2014.

Linear dynamic SEMs

- **Data:** Infection time of node *i* by contagion *c* during interval *t*

$$y_{ic}^t = \sum_{j \neq i} a_{ij}^t y_{jc}^t + b_{ii}^t x_{ic}^t + e_{ic}^t \qquad \mathbf{Y}_t = \mathbf{A}_t \mathbf{Y}_t + \mathbf{B}_t \mathbf{X}_t + \mathbf{E}_t, \quad t = 1, \ldots, T$$

Goal: Given data $\{\mathbf{Y}_t, \mathbf{X}_t\}$, track topology $\{\mathbf{A}_t\}$ and external influences $\{\mathbf{B}_t\}$

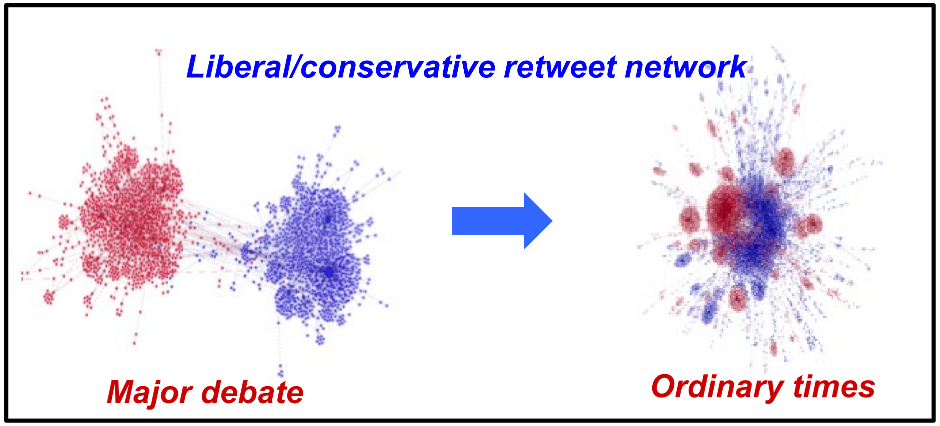
Q: How do network topologies evolve?

- Slow-varying network topologies

  $\mathbf{A}_t$ changes slowly; e.g., Facebook

- Switching among discrete states $\sigma(t) \in \{1, \ldots, S\}$

  $\mathbf{A}_t = \mathbf{A}_{\sigma(t)}$ e.g., *Tweets* during political/sports events



B. Baingana, G. Mateos, and G. B. Giannakis, ``Dynamic structural equation models for social network topology inference,'' *IEEE J. of Selected Topics in Signal Processing*, vol. 8, no. 4, pp. 563-575, Aug. 2014.

Tracking slowly-varying topologies

- Structural spatio-temporal properties
  - Slowly time-varying topology
  - Sparse edge connectivity, $\#\text{edges} = \mathcal{O}(\#\text{nodes})$

- Sparsity-promoting **exponentially-weighted LS estimator (EWLSE)**

$$\{\hat{\mathbf{A}}_t, \hat{\mathbf{B}}_t\} = \arg\min_{\mathbf{A},\mathbf{B}} \; (1/2)\sum_{\tau=1}^{t} \beta^{t-\tau}\|\mathbf{Y}_\tau - \mathbf{A}\mathbf{Y}_\tau - \mathbf{B}\mathbf{X}_\tau\|_F^2 + \lambda_t\|\mathbf{A}\|_1$$

$$\text{s.t.} \;\; a_{ii} = 0, \;\; b_{ij} = 0, \;\; \forall\, i \neq j$$

$$\|\mathbf{A}\|_1 = \sum_{ij} |a_{ij}|$$

  - **Edge sparsity** encouraged by ℓ_1-norm regularization with $\lambda_t > 0$
  - **Tracking** dynamic topologies possible if $\beta < 1$ $(\beta \in (0,1])$

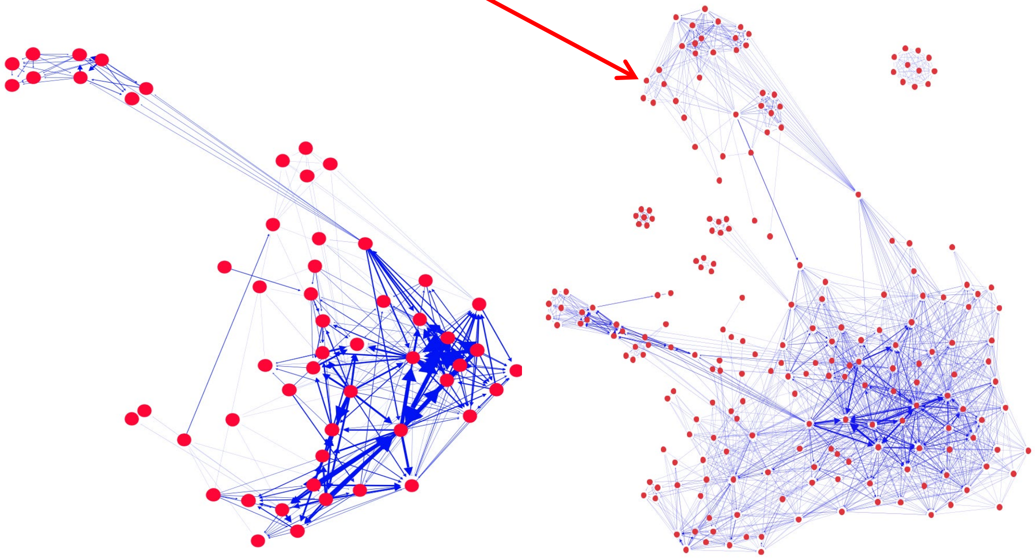
- **Solver:** proximal-splitting optimization methods [Daubechies et al'04]

The rise of Kim Jong-un

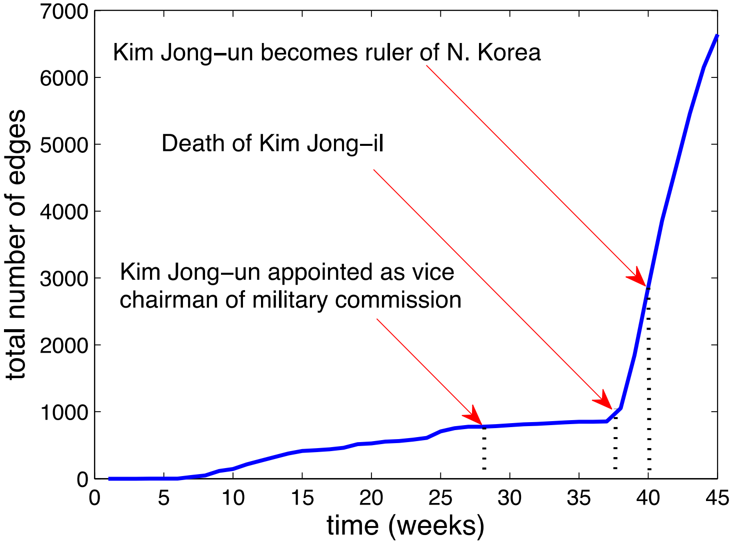
- Web mentions of **"Kim Jong-un"** tracked from Mar.'11 to Feb.'12
- N = **360** websites, C = **466** cascades, T = **45** weeks

Kim Jong-un – Supreme leader of N. Korea

Increased media frenzy following Kim Jong-un's ascent to power in 2011



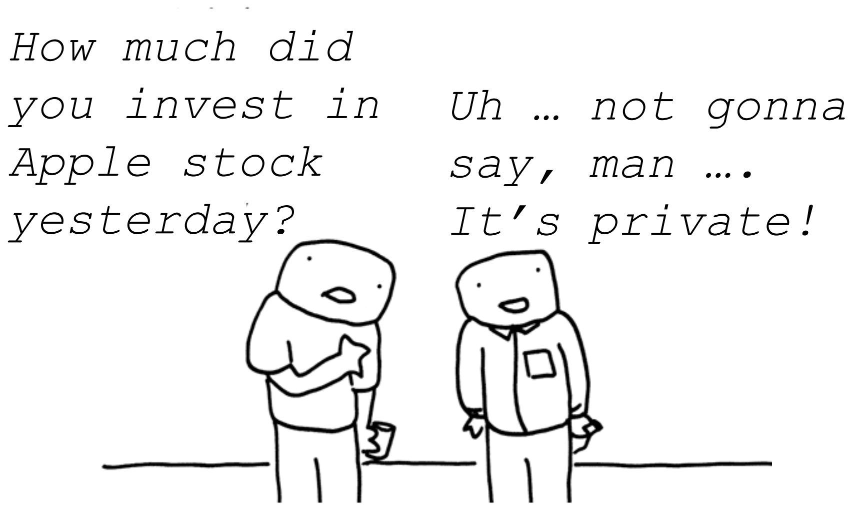
t = 10 weeks ***t = 40 weeks***



Data: SNAP's "Web and blog datasets" http://snap.stanford.edu/infopath/data.html

Identifiability of SEMs with input statistics?

- Limited access to input ***x***
  - Privacy concerns
  - Not explicitly available



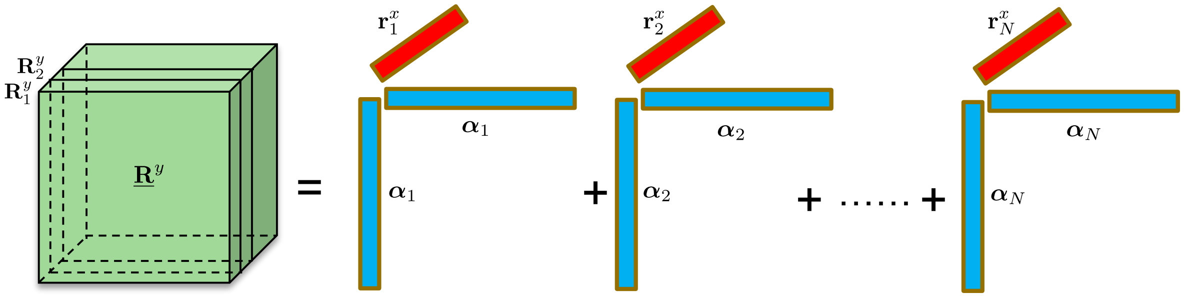
Goal: Given statistics of $\{\mathbf{y}_t, \mathbf{x}_t\}$ identify and track hidden directed network topology

$$\mathbf{y}_t = \mathbf{A}\mathbf{y}_t + \mathbf{B}\mathbf{x}_t \quad \Rightarrow \quad \mathbf{y}_t = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}\mathbf{x}_t = \boldsymbol{\mathcal{A}}\mathbf{x}_t$$

- Covar. over segment *m* $\mathbf{R}_m^y := \mathbb{E}\{\mathbf{y}_t\mathbf{y}_t^\top\}, \quad t \in [\tau_m, \ \tau_{m+1} - 1], m = 1 \ldots M$

Network snapshots as tensor slabs

- Tensor slab $\mathbf{R}_m^y = \boldsymbol{\mathcal{A}}\text{Diag}(\boldsymbol{\rho}_m^x)\boldsymbol{\mathcal{A}}^\top \Rightarrow \underline{\mathbf{R}}^y = \sum_{i=1}^{N} \boldsymbol{\alpha}_i \circ \boldsymbol{\alpha}_i \circ \mathbf{r}_i^x$,



Partially symmetric PARAFAC

$$\underline{\mathbf{R}}^y = (\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{A}}, \mathbf{R}^x)$$

Proposition 1: If $a_{ii} = 0,\ b_{ii} \neq 0,\ \forall i,\ b_{ij} = 0, \forall i \neq j$,
then **A** and **B** are uniquely expressible in terms of $\boldsymbol{\mathcal{A}}$
as $\mathbf{B} = \left(\text{Diag}\left[\boldsymbol{\mathcal{A}}^{-1}\right]\right)^{-1}$ and $\mathbf{A} = \mathbf{I} - \left(\text{Diag}(\boldsymbol{\mathcal{A}}^{-1})\right)^{-1}\boldsymbol{\mathcal{A}}^{-1}$

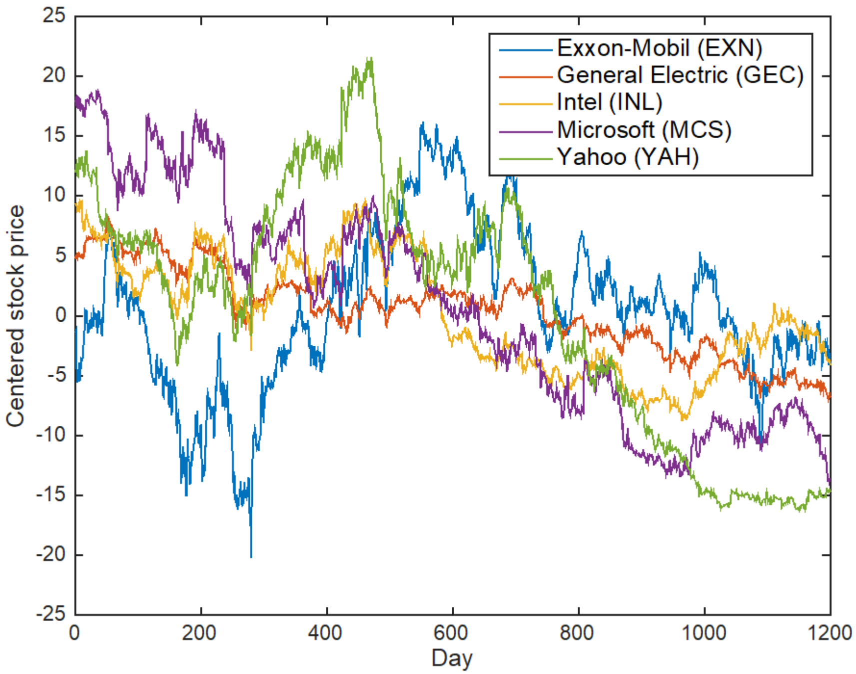
Theorem 2a. If $\text{kr}_{\boldsymbol{R}^x} > 1$, and $\boldsymbol{R}^x$ available, then ***A*** is identifiable.

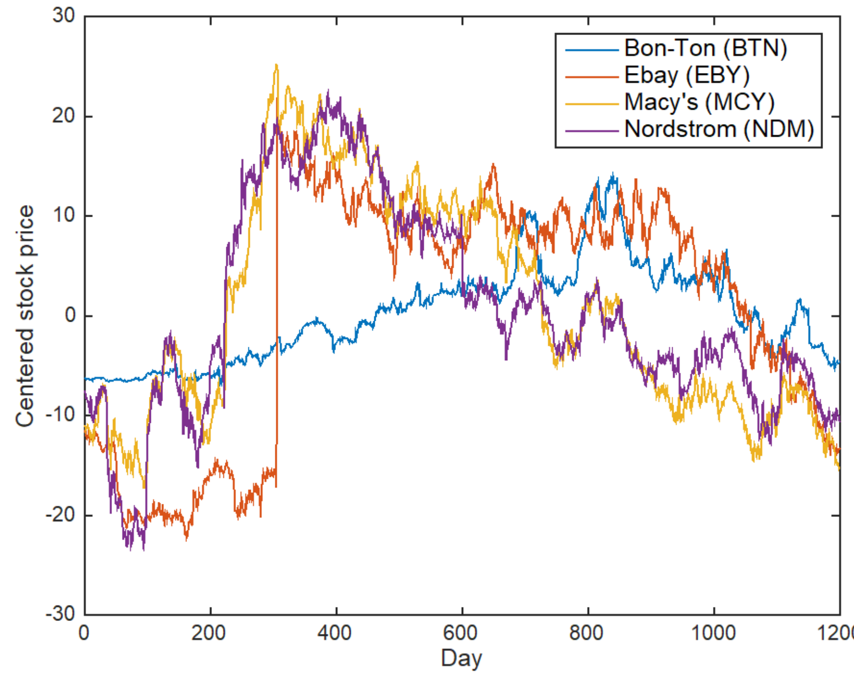
Theorem 2b. If $\text{kr}_{\boldsymbol{R}^x} > 1$, but $\boldsymbol{R}^x$ unknown, ***A*** identifiable within permutations (finite!)

Y. Shen, B. Baingana, and G. B. Giannakis, "Tensor decompositions for identifying directed graph topologies and tracking dynamic networks," *IEEE Transactions on Signal Processing,* vol. 65, no. 14, pp. 3675 - 3687, July 2017.

Real stock networks

- Dec. 23 , 2011 to Sep. 30 , 2016 (1,200 days), M = 12 time segments
- 100 runs each with random initialization

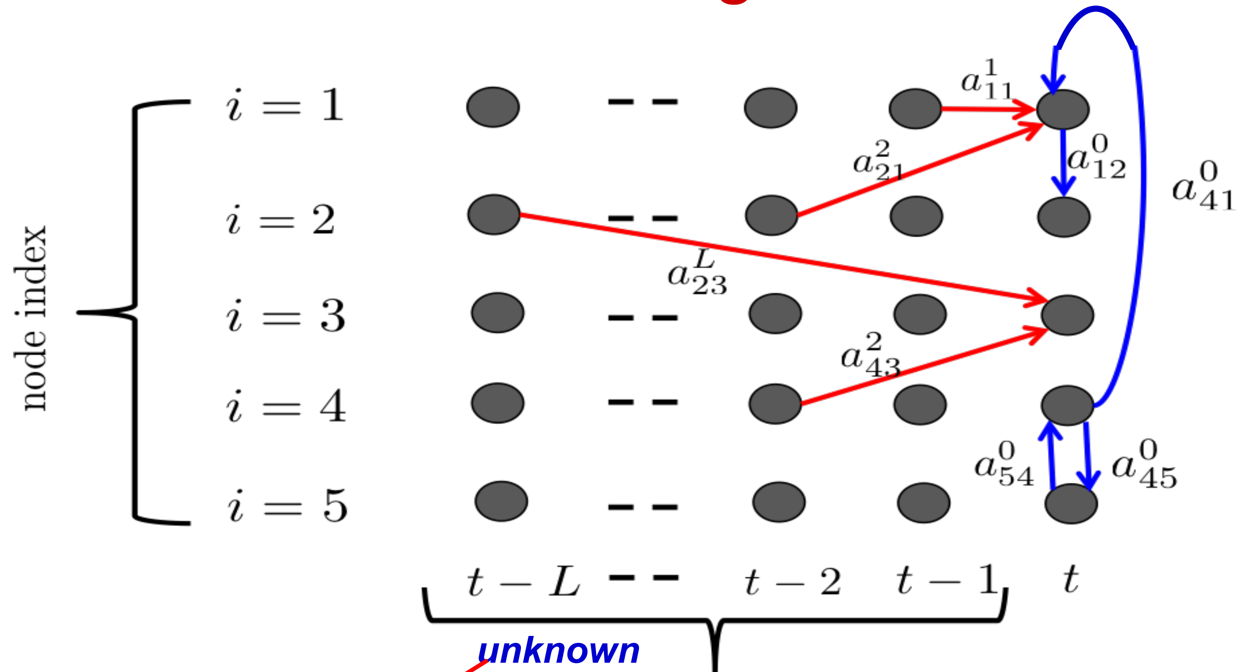




92/100

- Strong connectivity among major technology companies
- Stronger connectivity between Macy's and Nordstrom

Linear structural vector autoregressive models



$$y_{jt} = \sum_{i \neq j} a_{ij}^{0} y_{it} + \sum_{i=1}^{N} \sum_{\ell=1}^{L} a_{ij}^{\ell} y_{j(t-\ell)} + e_{jt}$$

- Endogenous variables here played by lagged exogenous
- Edge weights $\{a_{ij}^{\ell}\}$ capture **directed** causal dependencies
- Edge sparsity ⇨ only a few $\{a_{ij}^{\ell}\}$ are nonzero

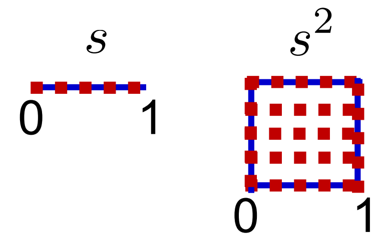
G. Chen, D. R. Glen, Z. S. Saad, J. P. Hamilton, M. E. Thomason, I. H. Gotlib, and R. W. Cox, "Vector autoregression, structural equation modeling, and their synthesis in neuroimaging data analysis'' *Comput. in Biology and Medicine,* pp. 1142–1155, Dec. 2011.

From linear to nonlinear SVARMs

$$y_{jt} = \bar{f}_j(\mathbf{y}_{-jt}, \{\mathbf{y}_{t-\ell}\}_{\ell=1}^{L}) + e_{jt}, \quad j = 1, \ldots, N$$

(L+1)N-1 variables

Curse of dimensionality



Idea: Reduce complexity using a generalized additive model

$$\bar{f}_j(\mathbf{y}_{-jt}, \{\mathbf{y}_{t-\ell}\}_{\ell=1}^{L}) = \sum_{i \neq j} \bar{f}_{ij}^{0}(y_{it}) + \sum_{i=1}^{N} \sum_{\ell=1}^{L} \bar{f}_{ij}^{\ell}(y_{i(t-\ell)})$$

(L+1)N-1 **univariate** functions

$$\bar{f}_{ij}^{\ell}(y) := a_{ij}^{\ell} f_{ij}^{\ell}(y) \quad a_{ij}^{\ell} \in \{0, 1\}$$

- Linear SVARM is special case

➢ Draw each univariate function from a reproducing kernel Hilbert space (RKHS)

$$\mathcal{H}_i^{\ell} := \{f_{ij}^{\ell} | f_{ij}^{\ell}(y) = \sum_{t=1}^{\infty} \beta_{ijt}^{\ell} \kappa_i^{\ell}(y, y_{i(t-\ell)})\}$$

$$\{\hat{f}_{ij}^{\ell}\} = \arg \min_{\{f_{ij}^{\ell} \in \mathcal{H}_i^{\ell}\}} \frac{1}{2} \sum_{t=1}^{T} \left[y_{jt} - \sum_{i \neq j} a_{ij}^{0} f_{ij}^{0}(y_{it}) - \sum_{i=1}^{N} \sum_{\ell=1}^{L} a_{ij}^{\ell} f_{ij}^{\ell}(y_{it}) \right]^2 + \lambda \sum_{i=1}^{N} \sum_{\ell=0}^{L} \Omega(\|a_{ij}^{\ell} f_{ij}^{\ell}\|_{\mathcal{H}^{\ell}})$$

Y. Shen, B. Baingana, and G. B. Giannakis, "Nonlinear Structural Vector Autoregressive Models for Inferring Effective Brain Network Connectivity," *IEEE Trans. on Medical Imag.*, 2018 revised; [Online]. Available: https://arxiv.org/abs/1610.06551

Edge sparsity leads to group-sparsity

- Representer theorem [Wahba etal'90] $\hat{f}^{\ell}_{ij}(y) = \sum_{t=1}^{T} \beta^{\ell}_{ijt}\kappa^{\ell}_{i}(y, y_{i(t-\ell)})$
- $\boldsymbol{\alpha}^{\ell}_{ij} := a^{\ell}_{ij}\boldsymbol{\beta}^{\ell}_{ij}$, $\boldsymbol{\beta}^{\ell}_{ij} := [\beta^{\ell}_{ij1}, \ldots, \beta^{\ell}_{ijT}]^{\top}$, $[\mathbf{K}^{\ell}_{i}]_{t,\tau} = \kappa^{\ell}_{i}(y_{it}, y_{i(\tau-\ell)})$

$$\{\hat{\boldsymbol{\alpha}}^{\ell}_{ij}\} = \underset{\hat{\boldsymbol{\alpha}}^{0}_{ii}=\mathbf{0}, \{\boldsymbol{\alpha}^{\ell}_{ij}\}}{\arg\min} \ \frac{1}{2}\left\|\mathbf{Y} - \sum_{l=1}^{L} \bar{\mathbf{K}}^{\ell}\mathbf{W}^{\ell}_{\alpha}\right\|_{F}^{2} + \lambda \sum_{\ell=0}^{L}\sum_{j=1}^{N}\sum_{i=1}^{N} \sqrt{(\boldsymbol{\alpha}^{\ell}_{ij})^{\top}\mathbf{K}^{\ell}_{i}\boldsymbol{\alpha}^{\ell}_{ij}}$$

- $\mathbf{Y} := [\mathbf{y}_1, \ldots, \mathbf{y}_N] \in \mathbb{R}^{T\times N}$ $\quad \bar{\mathbf{K}}^{\ell} := [\mathbf{K}^{\ell}_{1} \ldots \mathbf{K}^{\ell}_{N}]$
- Edge sparsity ⇨ **group sparsity** of $\mathbf{W}^{\ell}_{\alpha}$

$$\mathbf{W}^{\ell}_{\alpha} := \begin{bmatrix} \boldsymbol{\alpha}^{\ell}_{11} & \cdots & \boldsymbol{\alpha}^{\ell}_{1N} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\alpha}^{\ell}_{N1} & \cdots & \boldsymbol{\alpha}^{\ell}_{NN} \end{bmatrix}$$

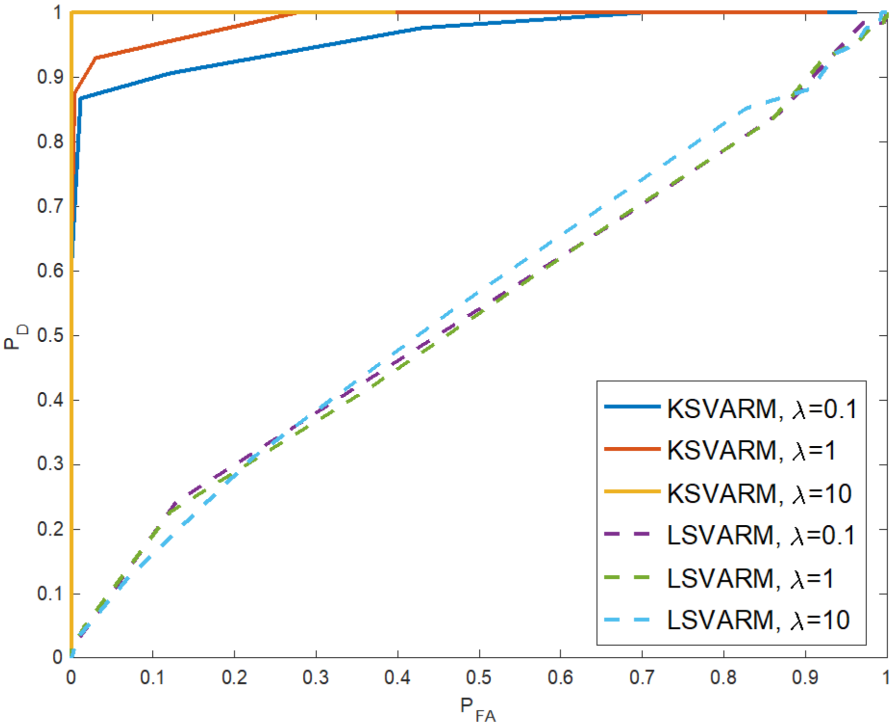
Bottomline: Nonzero $\{\boldsymbol{\alpha}^{\ell}_{ij}\}$ reveal edges; ADMM solver

- Multi-kernels can choose optimal kernel combination from a prescribed dictionary of kernels

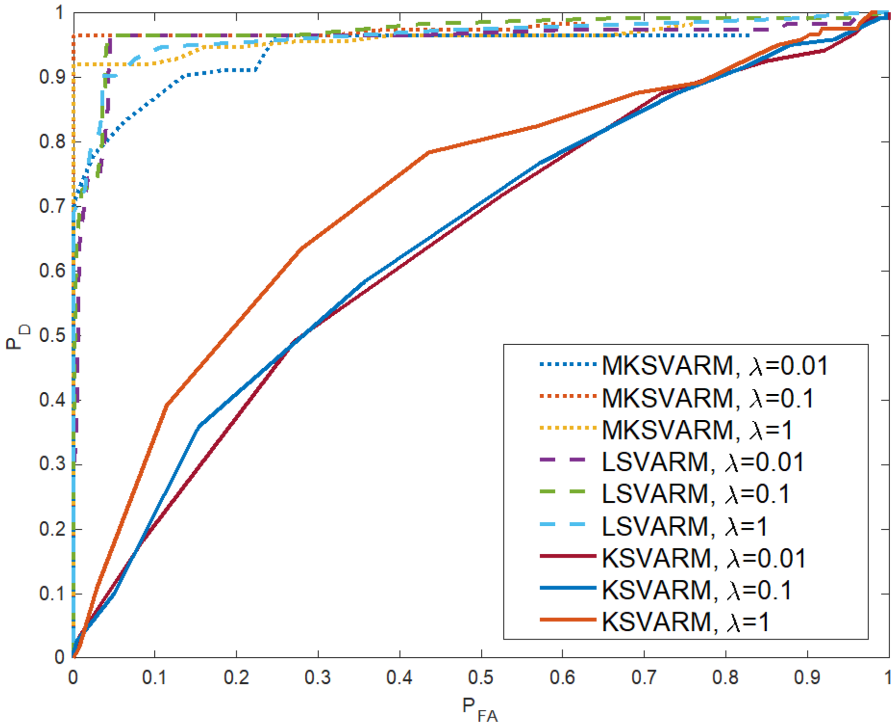
Simulated test

- Synthetic graph via Erdős–Rényi model, *N*=20, *T=40*
  - $p = 0.3 \quad \boldsymbol{\alpha}_{ij}^{\ell} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
  - $\sigma_e = 0.1$

Polynomial SVARM

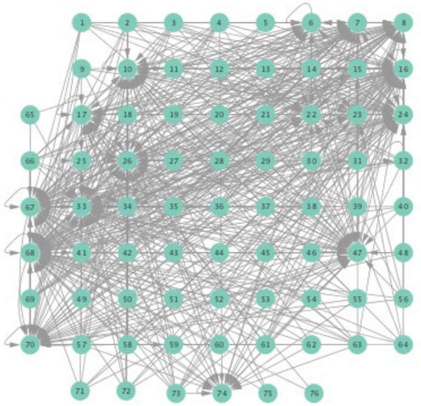


Linear SVARM

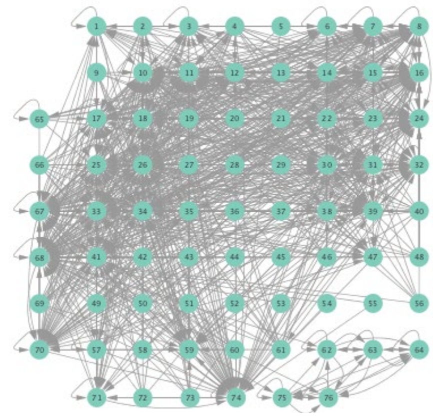


Brain is densely networked

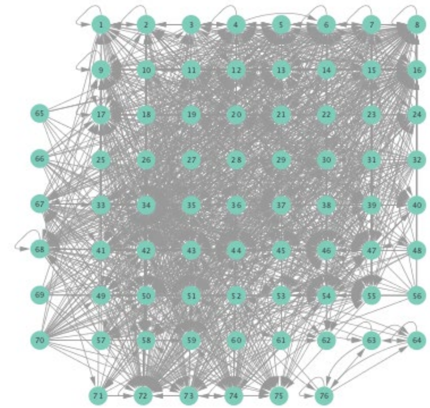
- **Data:** electrocorticography (ECoG) data for epilepsy [Kramer et al' 08]
  - **Y**: ECoG data samples
  - *N = 76* electrodes, *T = 200 samples, L=1*



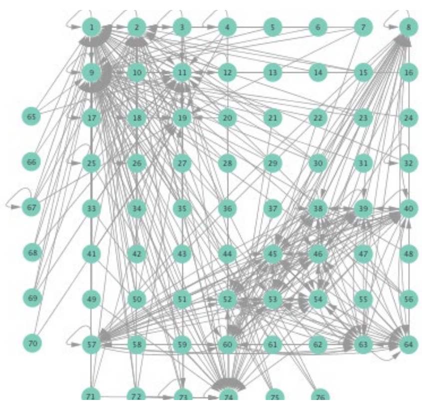
Linear SVARM (Preictal)



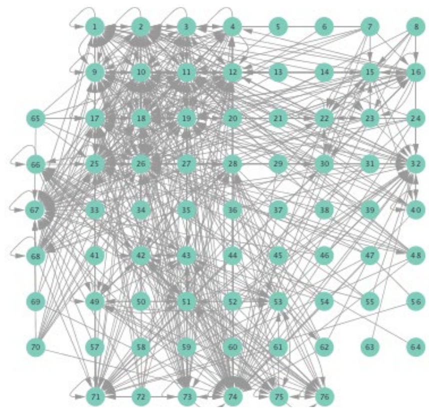
Kernel SVARM (preictal)



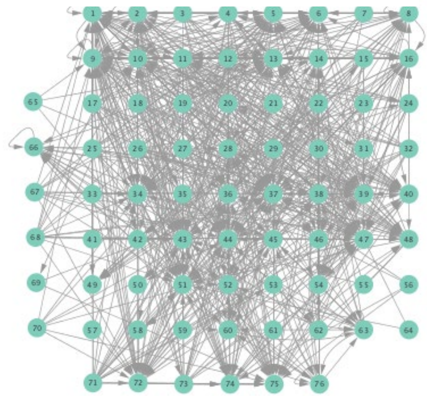
MKL SVARM (preictal)



Linear SVARM (Ictal)



Kernel SVARM (preictal)



MKL SVARM (ictal)

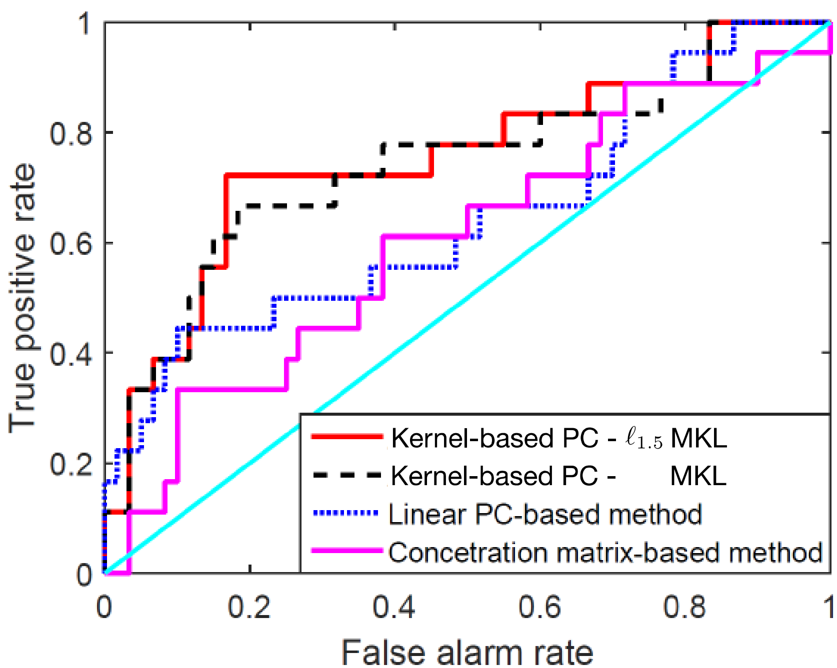
- Diffusion of information is inhibited after the onset of an epileptic seizure

M. A. Kramer, E. D. Kolaczyk, and H. E. Kirsch, "Emergent network topology at seizure onset in humans," Epilepsy Research, vol. 79, no. 2, pp. 173–186, May 2008.

Identifying connectivity of meshed power grids

- Grid of *N*=14 buses; nodal vectors are voltage angle time courses
  - IEEE-14 bus benchmark; voltage angles obtained using MATPOWER

- Real load data from 2012 Global Energy Forecasting Competition

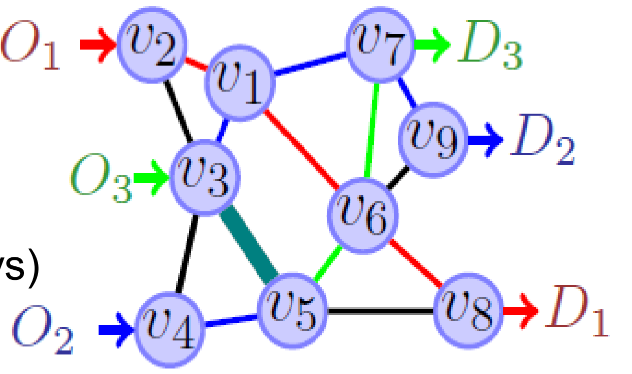


- Nonlinear predictors improve ID of mesh grid connectivity

L. Zhang, G. Wang, and G. B. Giannakis, "Going Beyond Linear Dependencies to Unveil Connectivity of Meshed Grids," *Proc. of CAMSAP*, Curacao, Dutch Antilles, Dec. 10-13, 2017.

Interpolating and extrapolating over networks

- Undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with $|\mathcal{V}| = N$



- Process y_{nt} per node n, timeslot t
  - Only measure $M < N$ nodes (e.g., link counts, delays)

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

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

$$\mathbf{z}_t := [z_{1t}, \ldots, z_{Nt}]^\top, \mathbf{y}_t := [y_{1t}, \ldots, y_{Mt}]^\top$$

  - Rows of $\mathbf{M}_t$ selected from $\mathbf{I}_{N \times N}$

Goal: Impute misses and predict $\mathbf{y}_t$ from selected node observations $\mathbf{z}_t$

Motivating application: Estimate network delays

P. A. Forero, K. Rajawat, and G. B. Giannakis, "Prediction of partially observed dynamical processes over networks via dictionary learning," *IEEE Trans. Signal Processing*, vol. 62, no. 13, pp. 3305-3320, 2014.

Graph-regularized dictionary learning approach

- Generalizes low-rank based matrix completion that cannot predict!

- Adopt and learn basis and expansion coefficients $(\mathbf{y}_t = \mathbf{B}\mathbf{s}_t)$

Smoothness over graph

$$\arg\min_{\mathbf{S},\mathbf{B}:\{\|\mathbf{b}_q\|_2\leq 1\}_{q=1}^{Q}} \sum_{t=1}^{T} \left[\|\mathbf{z}_t - \mathbf{M}_t\mathbf{B}\mathbf{s}_t\|_2^2 + \lambda_s\|\mathbf{s}_t\|_1 + \lambda_w \mathbf{s}_t^\top \mathbf{B}^\top \mathbf{L}\mathbf{B}\mathbf{s}_t\right]$$

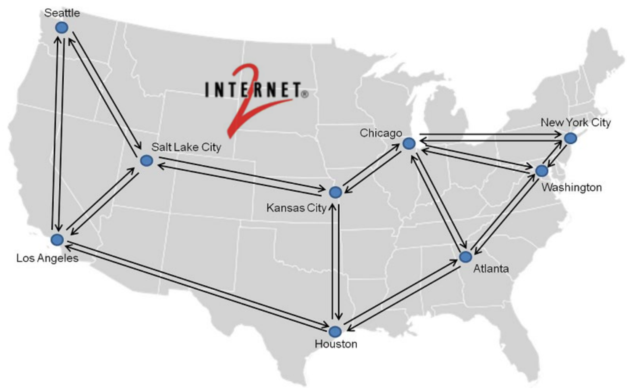
Dictionary: $N \times Q$ ***Sparse coefficients*** ***Graph Laplacian***

- With adjacency matrix ***A***, graph Laplacian $\mathbf{L} := \mathrm{Diag}(\mathbf{A}\mathbf{1}_N) - \mathbf{A}$

$$\mathbf{s}_t^\top \mathbf{B}^\top \mathbf{L}\mathbf{B}\mathbf{s}_t = (1/2)\sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}(y_{it} - y_{jt})^2$$

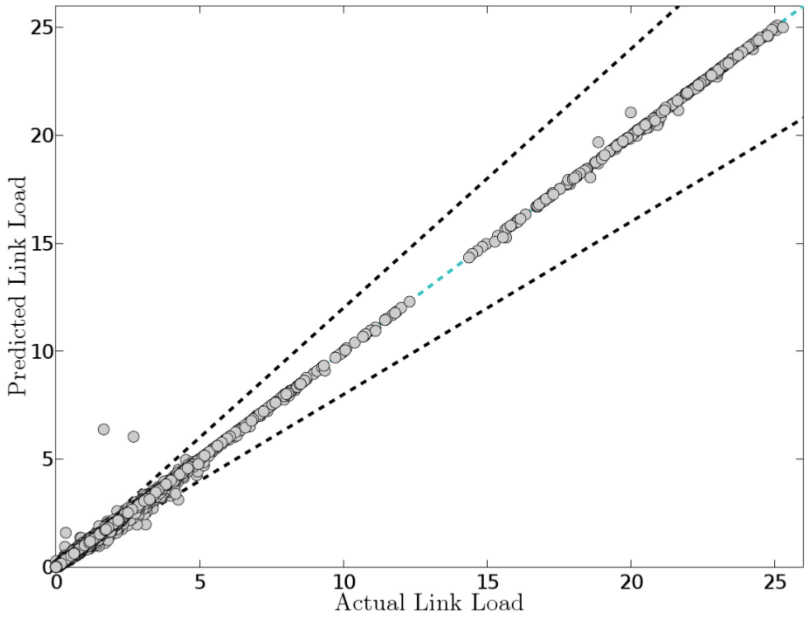
promotes smoothness

Test case: Internet2

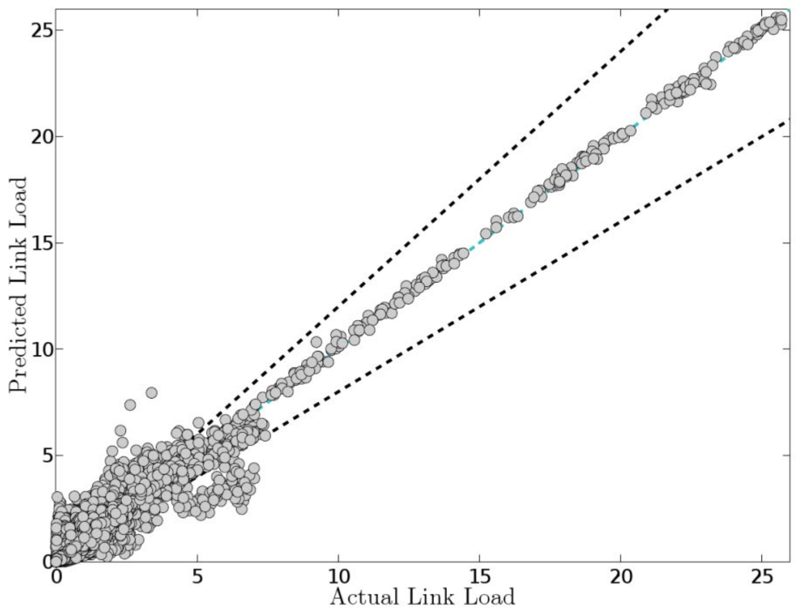


- Link count measurements: *L=54, T=2,000* (other features possible, e.g., delays)

Training phase – 30 links measured



Operational phase – 30 links predicted

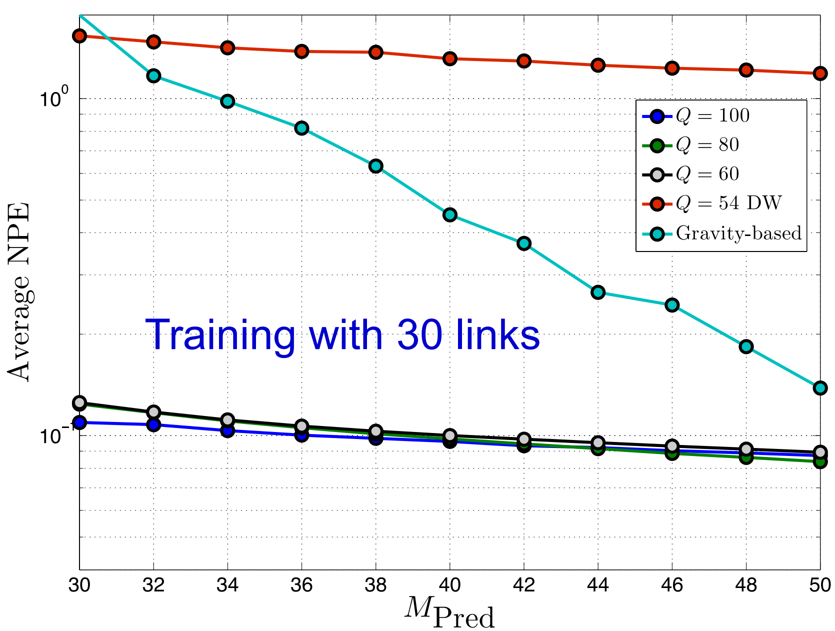


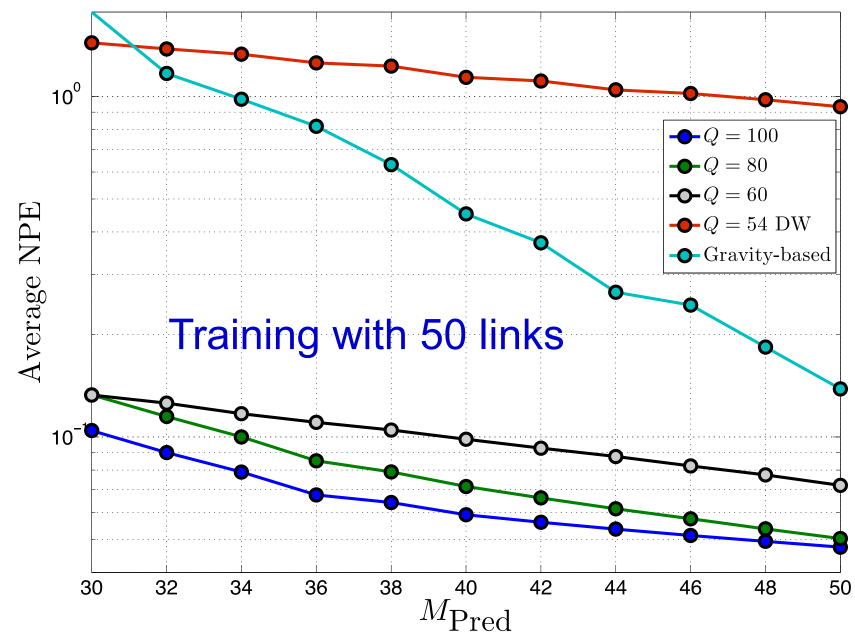
- Prediction improves as link load increases

Performance comparisons

- Normalized prediction error: $\mathrm{NPE} := \frac{1}{Lt_0} \sum_{\tau=1}^{t_0} \|\mathbf{y}_\tau - \hat{\mathbf{y}}_\tau\|_2^2$
  - Q = number of columns of ***B***; t_0=*2,000*

- Gravity-based [Zhang et al'05]; Diffusion wavelets [Coifman-Maggioni'07]





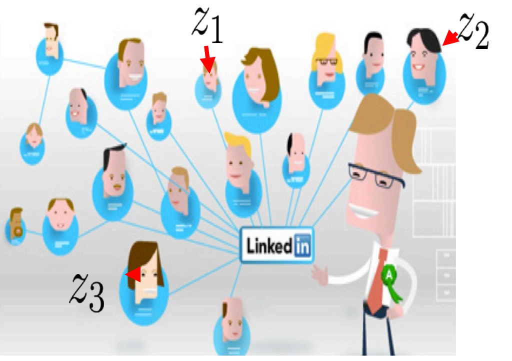
- Graph-regularized DL with semi-supervised predictor outperforms alternatives

Graph-adaptive kernel-based interpolation

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

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

$M < N$



Goal: Given $\mathbf{z}_t$, $\mathbf{M}_t$, and $\mathcal{G}_t$, estimate $\mathbf{y}_t$

- **RKHS model**: $\mathbf{y}_t \in \mathcal{H}_{\mathbf{K}_t}$ iff $\mathbf{y}_t = \mathbf{K}_t \boldsymbol{\alpha}_t, \ \boldsymbol{\alpha}_t \in \mathbb{R}^N$
  - Graph-dependent symmetric $\mathbf{K}_t \geq \mathbf{0}$
  - **Ex.** Laplacian $(\mathbf{L}_t)$ family $\mathbf{K}_t := r^{-1}(\mathbf{L}_t)$

- Kernel ridge regression (KRR) [Smola-Kondor '03]

$$\hat{\mathbf{y}}_t = \arg\min_{\mathbf{y}} \ \frac{1}{M} \|\mathbf{z}_t - \mathbf{M}_t \mathbf{y}\|_2^2 + \mu \|\mathbf{y}\|_{\mathbf{K}_t}^2 \qquad \|\mathbf{y}\|_{\mathbf{K}_t}^2 := \mathbf{y}^T \mathbf{K}_t^{-1} \mathbf{y}$$

$$= \mathbf{K}_t \mathbf{M}_t^T (\mathbf{M}_t \mathbf{K}_t \mathbf{M}_t^T + \mu M \mathbf{I}_M)^{-1} \mathbf{z}_t$$

D. Romero, M. Ma, and G. B. Giannakis, "Kernel-based reconstruction of graph signals,“ *IEEE Trans. on Signal Processing*, vol. 65, pp. 764-778, February 2017.

Spatio-temporal processes on graphs

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

- **Superimposed state model** for e.g., packet delays, stock values, temperature, …

$$\mathbf{y}_t = \mathbf{y}_t^{(\nu)} + \mathbf{y}_t^{(\chi)}, \quad \mathbf{y}_t^{(\chi)} = \mathbf{A}_{t,t-1}\mathbf{y}_{t-1}^{(\chi)} + \boldsymbol{\eta}_t, \quad t = 1, 2, \ldots$$

- Spatial $\mathbf{y}_t^{(\nu)}$ temporally uncorrelated ('fast' dynamics across slots)
- Spatio-temporal $\mathbf{y}_t^{(\chi)}$ VARM ('slow' dynamics; trend)

- **Space-time kriging ridge regression (KRR)**

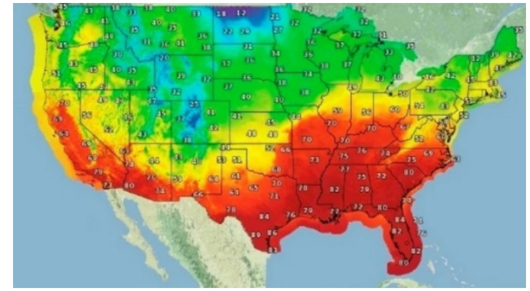
$$\arg\min_{\{\mathbf{y}_{t'}^{(\chi)},\mathbf{y}_{t'}^{(\nu)}\}_{t'=1}^{t}} \sum_{t'=1}^{t} \frac{1}{M_{t'}} \|\mathbf{z}_{t'} - \mathbf{M}_{t'}\mathbf{y}_{t'}^{(\chi)} - \mathbf{M}_{t'}\mathbf{y}_{t'}^{(\nu)}\|_2^2 + \mu_1 \sum_{t'=1}^{t} \|\mathbf{y}_{t'}^{(\chi)} - \mathbf{A}_{t',t'-1}\mathbf{y}_{t'-1}^{(\chi)}\|_{\mathbf{K}_{t'}^{(\eta)}}^2$$

$$+\mu_2 \sum_{t'=1}^{t} \|\mathbf{y}_{t'}^{(\nu)}\|_{\mathbf{K}_{t'}^{(\nu)}}^2$$

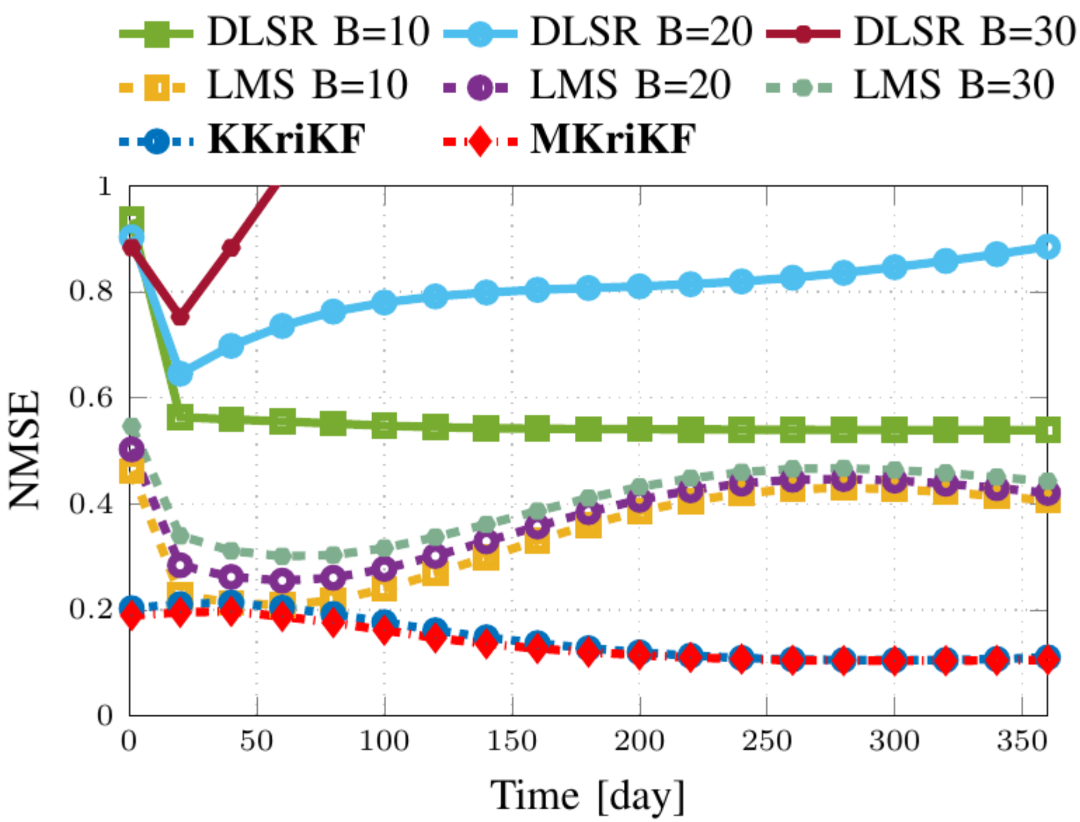
Result. KeKriKF produces the sequence of filtered $\{\hat{\mathbf{y}}_{t'|t'}^{(\chi)}, \hat{\mathbf{y}}_{t'|t'}^{(\nu)}\}_{t'=1}^{t}$

K. Rajawat, E. Dall'Anese, and G. B. Giannakis, "Dynamic network delay cartography," *IEEE Transactions on Information Theory*, vol. 60, no. 05, pp. 2910-2920, May 2014.

Temperature reconstruction

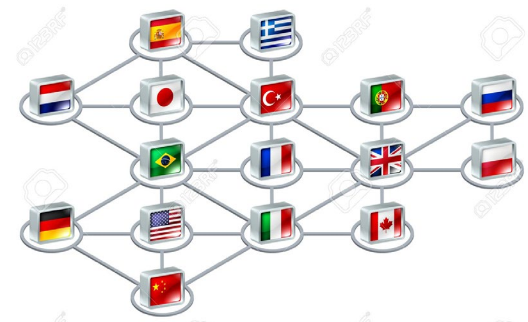


- Temperature sensor network *N=109*
  - Compare reconstruction NMSE per day

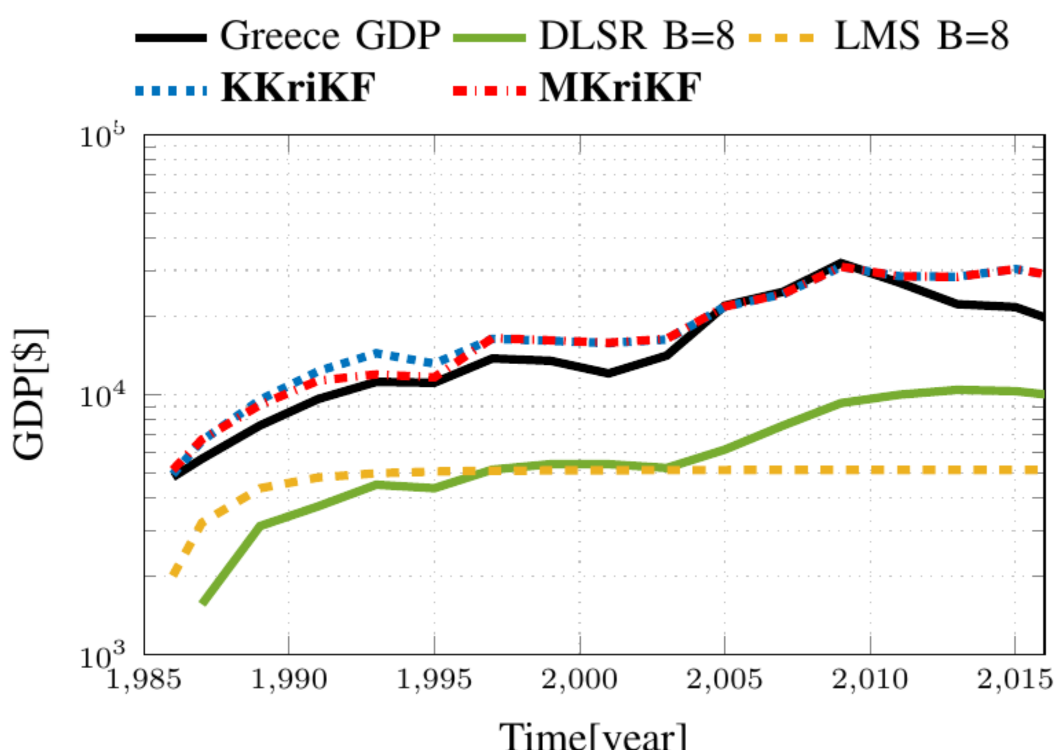


V. N. Ioannidis, D. Romero, and G. B. Giannakis, ``Inference of Spatio-Temporal Functions over Graphs via Multi-Kernel Kriged Kalman Filtering,'' *IEEE Trans. on Signal Processing*, pp. 3228-39, June 2018.

GDP prediction



- Financial network between *N=127* countries
  - Track gross domestic product (GDP) at an unobserved country



V. N. Ioannidis, D. Romero, and G. B. Giannakis, "Inference of spatio-temporal processes over dynamic graphs via kernel kriged Kalman filters," *Proceedings of EUSIPCO*, Kos Island, Greece, Aug. 28-Sept. 3, 2017.

Joint ID of topologies and signals on graphs

Q: What if topology unknown and just a subset of data available due to privacy/large-scale ?

- Linear SEM: $\mathbf{y}_l = \mathbf{A}\mathbf{y}_l + \boldsymbol{\eta}_l$
- Measurements: $\mathbf{z}_l = \underbrace{\mathbf{M}_l}_{M_l \times N} \mathbf{y}_l + \boldsymbol{\epsilon}_l$ $\quad l = 1, \ldots, L$

Goal: Given $\{\mathbf{z}_l, \mathbf{M}_l\}_{l=1}^L$, identify $\mathbf{A}$ and $\{\mathbf{y}_l\}_{l=1}^L$

- Joint inference of signals and (directed) graphs (**JISG**)

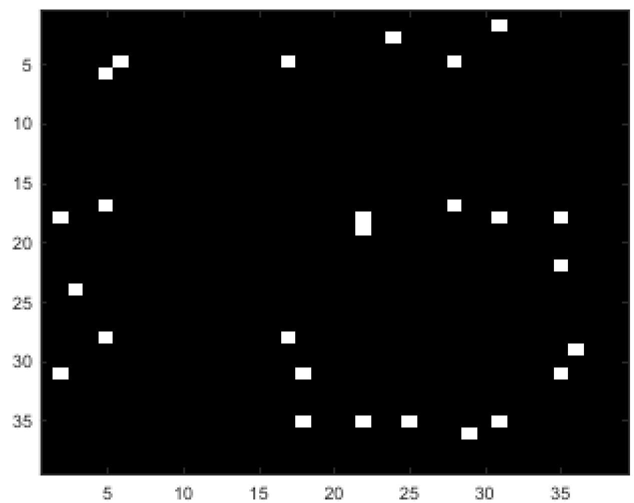
$$\min_{\mathbf{A} \in \mathcal{A}, \{\mathbf{y}_l\}_{l=1}^L} \sum_{l=1}^{L} \|\mathbf{y}_l - \mathbf{A}\mathbf{y}_l\|_2^2 + \sum_{l=1}^{L} \|\mathbf{z}_l - \mathbf{M}_l\mathbf{y}_l\|_2^2 + \lambda_1 \|\mathbf{A}\|_1 + \lambda_2 \|\mathbf{A}\|_F^2$$

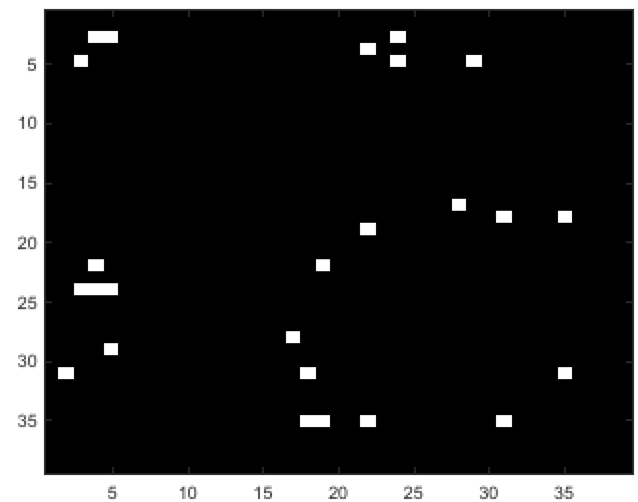
- BCD/ADMM solver: Guaranteed convergence at reduced complexity (separable per *l*)
- Generalizable to nonlinear SEM; multi-layer; and dynamic signals and graphs

V. N. **I**oannidis, Y. **S**hen, G. B. **G**iannakis, “Semi-blind inference of topologies and signals over graphs,” *Proc. of IEEE Data Science Workshop*, Lausanne, Switzerland, June 2018.

Testing JISG on gene-regulatory networks

- *N*=39 immune-related genes; *L*=69 unrelated individuals; *y:* gene expression level
  - SEM oracle observes all genes *M*=39 (left); **JISG** with *M*=31 (right) $M_l = M, \ \forall l$

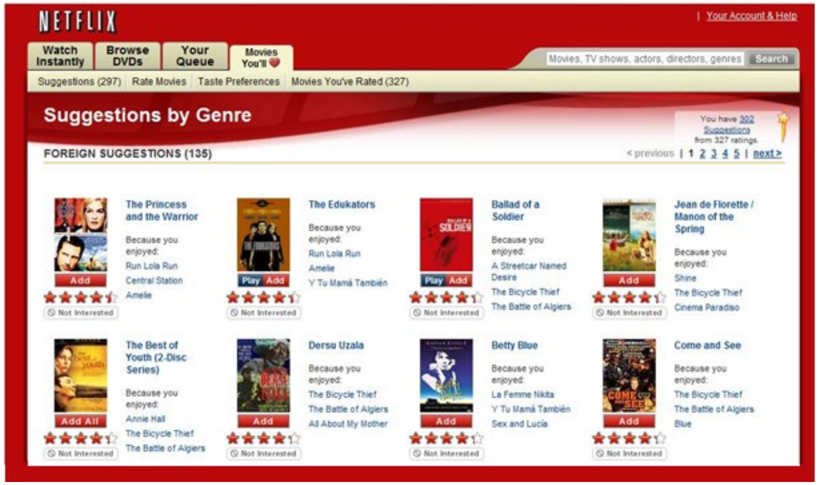


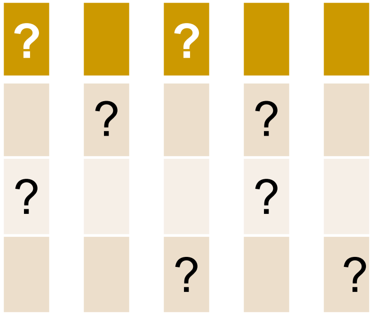


  - NMSE for $\{\hat{\mathbf{y}}_l\}_{l=1}^{L}$ = 0.017
  - JISG-based recovery similar to that of the oracle

V. N. **I**oannidis, Y. **S**hen, G. B. **G**iannakis, "Semi-blind inference of topologies and signals over graphs," *Proc. of IEEE Data Science Workshop*, Lausanne, Switzerland, June 2018.

Top-*N* recommender systems and SLIM





Goal: Given subset of user-item ratings, rank \`*N*-best' candidates of unavailable ratings

- Sparse linear model (SLIM) of ratings: SEM followed by interpolation and ranking

Topology ID via SEM:

$$\min_{\{a_{ii'}\}} \|\mathbf{r}_i - \sum_{i'} a_{ii'} \mathbf{r}_{i'}\|_2^2 + \lambda \sum_{i'} |a_{ii'}|$$

$$\text{s.t. } a_{ii} = 0, \quad a_{ii'} \geq 0 \ \ \forall i'.$$

Interpolate and rank:

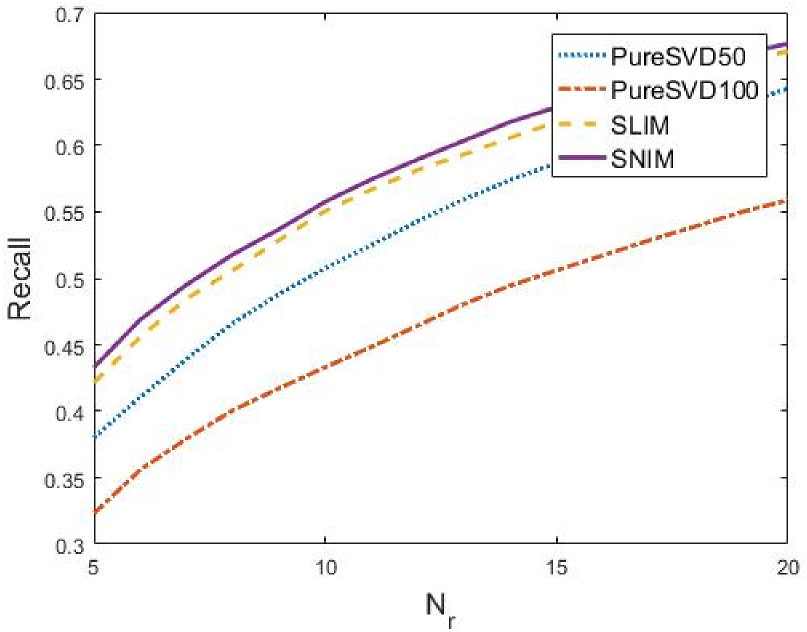
$$\hat{r}_{ui} = \sum_{i'} a_{ii'} r_{ui'}$$

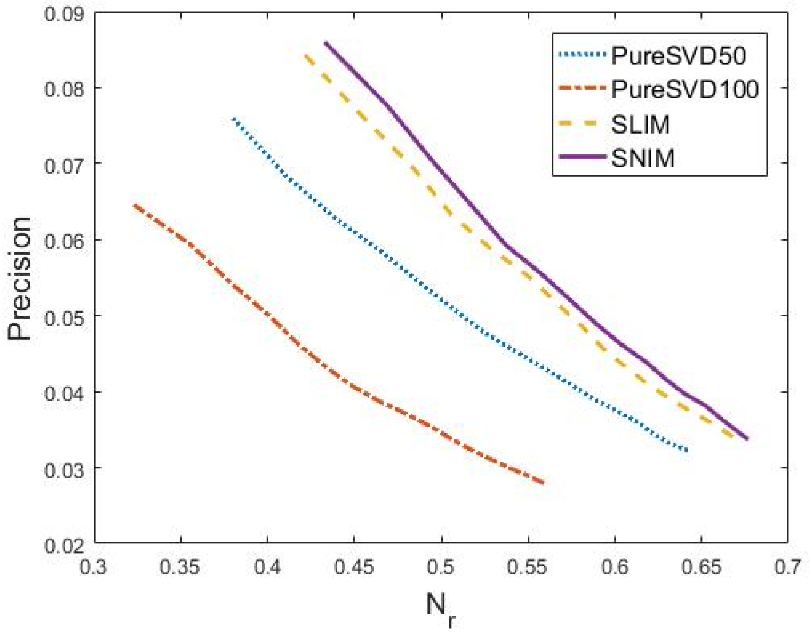
Our idea: Instead of SLIM, employ sparse nonlinear SVARM with *L=0* (SNIM)

X. Ning and G. Karypis, "SLIM: Sparse linear methods for top-n recommender systems," *Proc. of Intl. Conf. on Data Mining*, Vacouver, Canada, Dec. 2011, pp. 497–506.

Movielens dataset

- 3,706 users; 6,040 movies; and 1M ratings
  - **Training set**: 97% randomly sampled ratings; **Probe set**: All 5 star ratings in testing set
  - # hits: number of ratings in probe set that also appear in the recommendation list





$$\text{recall}(N_r) = \#\text{hits}/\#\text{probe}$$

$$\text{precision}(N_r) = \text{recall}(N_r)/N_r$$

  - Here SNIM outperforms SLIM by a slim margin

G. B. Giannakis, Y. Shen, and G. V. Karanikolas "Nonlinear and dynamical models for learning connectivity and processes over graphs," *Proceedings of the IEEE,* submitted September 2017 (invited).

Current research and outlook

- **Topology identification** – a “deterministic” RKHS-based approach
  - Directed and linear multi-layer graphs are allowed with identifiability guarantees
  - Nonlinear dependencies accommodated through multi-kernel regression
  - Slow-varying and switching dynamics can be afforded

- **Learning of processes on graphs**
  - Interpolation and extrapolation from partially-observed processes on graphs
  - Topology can be known or unknown
  - Kernel Kriged Kalman Filtering for inference of dynamic processes

- Ongoing research and future directions
  - Graph-adaptive dimensionality reduction/manifold learning
  - Tracking and identifiability of nonlinear and dynamic topologies
  - RKHS-principled multi-kernel learning vis-à-vis DNNs

Thank you!

G. B. Giannakis, Y. Shen, and G. V. Karanikolas, “Topology Identification and Learning over Graphs: Accounting for Nonlinearities and Dynamics,” *Proceedings of the IEEE,* vol. 106, pp. 787-807, May 2018.