Blind Identification of Invertible Graph Filters with Multiple Sparse Inputs

Chang Ye

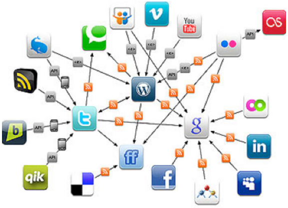
Dept. of ECE and Goergen Institute for Data Science
University of Rochester
cye7@ur.rochester.edu
http://www.ece.rochester.edu/~cye7/

Co-authors: Rasoul Shafipour and Gonzalo Mateos
Acknowledgment: NSF Awards CCF-1750428 and ECCS-1809356

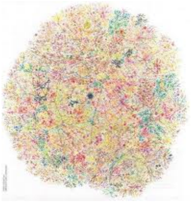
EUSIPCO 2018, Rome, Italy, September 3, 2018

Network Science analytics

Online social media



Internet



Clean energy and grid analytics

- Network as undirected graph $G = (\mathcal{V}, \mathcal{E})$: encode pairwise relationships
- Desiderata: Process, analyze and learn from network data [Kolaczyk'09]

  $\Rightarrow$ Study graph signals, data associated with N nodes in $\mathcal{V}$
- Ex: Opinion profile, buffer congestion levels, neural activity, epidemic

Graph signal processing fundamentals

- Graph signals mappings $x : \mathcal{V} \to \mathbb{R}$, represented as vectors $\mathbf{x} \in \mathbb{R}^N$
  - $\Rightarrow$ **As.:** Signal properties related to topology of G
- To process graph signals $\Rightarrow$ Graph-shift operator $\mathbf{S} \in \mathbb{R}^{N \times N}$
  - $\Rightarrow$ Local $S_{ij} = 0$ for $i \neq j$ and $(i,j) \notin \mathcal{E}$ $\Rightarrow$ Ex: $\mathbf{A}$ or $\mathbf{L} = \mathbf{D} - \mathbf{A}$
  - $\Rightarrow$ Spectrum of symmetric $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$
- Graph Fourier Transform (GFT) for signals: $\tilde{\mathbf{x}} = \mathbf{V}^T\mathbf{x}$
- Graph filters $H : \mathbb{R}^N \to \mathbb{R}^N$ are maps between graph signals
  - $\Rightarrow$ Polynomial in $\mathbf{S}$ with coefficients $\mathbf{h} \in \mathbb{R}^L \Rightarrow \mathbf{H} := \sum_{l=0}^{L-1} h_l \mathbf{S}^l$
  - $\Rightarrow$ Orthogonal frequency operator: $\mathbf{H} = \mathbf{V}\text{diag}(\tilde{\mathbf{h}})\mathbf{V}^T$
  - $\Rightarrow$ Freq. response (GFT for filters): $\tilde{\mathbf{h}} = \mathbf{\Psi}\mathbf{h}$ and $[\mathbf{\Psi}]_{k,l} = {\lambda_k}^{l-1}$

Diffusion processes as graph filter outputs

- **Q:** Upon observing a graph signal $\mathbf{y}$, how was this signal generated?
- Postulate $\mathbf{y}$ is the response of linear diffusion to a sparse input $\mathbf{x}$

$$\mathbf{y} = \alpha_0 \prod_{l=1}^{\infty} (\mathbf{I} - \alpha_l \mathbf{S})\mathbf{x} = \sum_{l=0}^{\infty} \beta_l \mathbf{S}^l \mathbf{x}$$

$\Rightarrow$ Common generative model, e.g., heat diffusion, consensus

- Cayley-Hamilton asserts we can write diffusion as $(L \leq N)$

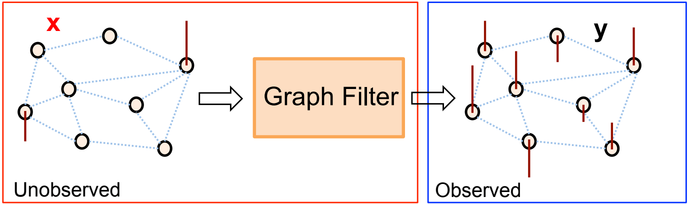
$$\mathbf{y} = \left(\sum_{l=0}^{L-1} h_l \mathbf{S}^l \right) \mathbf{x} := \mathbf{H}\mathbf{x}$$

- **Model:** Observed network process as output of a graph filter

$\Rightarrow$ View few elements in $\text{supp}(\mathbf{x}) =: \{i : x_i \neq 0\}$ as sources

Motivation and problem statement

- Ex: Global opinion/belief profile formed by spreading a rumor
  - ⇒ What was the rumor? Who started it?
  - ⇒ How do people weigh in peers' opinions to form their own?



- **Problem:** Blind identification of graph filters with sparse inputs
- **Q:** Given $\mathbf{S}$, can we find sparse $\mathbf{x}$ and the filter coeffs. $\mathbf{h}$ from $\mathbf{y} = \mathbf{H}\mathbf{x}$?
  - ⇒ Extends classical blind deconvolution to graphs
  - ⇒ Localization of sources that diffuse on the network

Work in context

- Super-resolution of point sources via convex programming
  - Signals on structured domains (e.g.,time series) [Fernandez-Granda'15]
  - Known diffusion model (low-pass point-spread function)
- Source localization on graphs
  - Maximum-likelihood estimator optimal for trees [Pinto et al'12]
  - Scalable under restrictive dependency assumptions [Feizi el al'16]
  - Non-convex estimators of sparse sources [Pena et al'16], [Hu et al'16]
- Blind identification of graph filters [Segarra et al'17]
  - Matrix lifting can hinder applicability to large graphs
- **Our contribution:** mild requirement of graph filter invertibility

  $\Rightarrow$ Convex formulation amenable to efficient solvers

  $\Rightarrow$ Multi-signal case with arbitrary supports

Blind graph filter identification

- Suppose we observe P output signals $\mathbf{Y} = [\mathbf{y}_1, ..., \mathbf{y}_P] \in \mathbb{R}^{N \times P}$
- Leverage frequency response of graph filters

$$\mathbf{Y} = \mathbf{H}\mathbf{X} \;\Rightarrow\; \mathbf{Y} = \mathbf{V}\text{diag}(\boldsymbol{\Psi}\mathbf{h})\mathbf{V}^T\mathbf{X}$$

$\Rightarrow$ $\mathbf{Y}$ is a bilinear function of the unknowns $\mathbf{h}$ and $\mathbf{X}$

- Ill-posed problem $\Rightarrow$ $L + NP$ unknowns and NP observations

$\Rightarrow$ **As.:** $\mathbf{X}$ has S-sparse columns i.e., $\|\mathbf{X}\|_0 := |\text{supp}(\mathbf{X})| \leq PS$

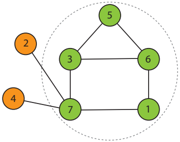
- Blind graph filter identification $\Rightarrow$ Non-convex feasibility problem

$$\text{find } \{\mathbf{h}, \mathbf{X}\}, \quad \text{s. to } \quad \mathbf{Y} = \mathbf{V}\text{diag}\big(\boldsymbol{\Psi}\mathbf{h}\big)\mathbf{V}^T\mathbf{X}, \;\; \|\mathbf{X}\|_0 \leq PS$$

$\Rightarrow$ Identifiability for Bernoulli-Gaussian model on $\mathbf{X}$ [Li et al'17]

Permutation ambiguities

- Beyond scaling, permutation ambiguities can arise with unweighted G



$$\mathbf{v}_4 = \begin{bmatrix} 0 \\ \sqrt{2}/2 \\ 0 \\ -\sqrt{2}/2 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad \mathbf{P} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

- Let $\{\mathbf{X}_0, \tilde{\mathbf{h}}_0\}$ be a solution, i.e., $\mathbf{Y} = \mathbf{V}\text{diag}(\tilde{\mathbf{h}}_0)\mathbf{V}^T\mathbf{X}_0$

  $\Rightarrow$ Define unit-norm $\mathbf{u}^{(i,j)} \in \mathbb{R}^N$, with $u_i^{(i,j)} = -u_j^{(i,j)} = \frac{1}{\sqrt{2}}$

- If $\mathbf{v}_k = \mathbf{u}^{(i,j)}$, then $\exists \{\mathbf{X}_1, \tilde{\mathbf{h}}_1\}$ such that $\mathbf{Y} = \mathbf{V}\text{diag}(\tilde{\mathbf{h}}_1)\mathbf{V}^T\mathbf{X}_1$

$$\mathbf{X}_1 := \mathbf{P}\mathbf{X}_0, \quad \tilde{\mathbf{h}}_1 := \text{diag}(\mathbf{p})\tilde{\mathbf{h}}_0$$

$$\mathbf{P} := \mathbf{I} - 2\mathbf{u}^{(i,j)}(\mathbf{u}^{(i,j)})^T = \mathbf{V}\text{diag}(\mathbf{p})\mathbf{V}^T$$

$\Rightarrow$ Compare with cyclic-shift ambiguity for discrete-time signals

Inverse filter and convex relaxation

- Inverse filter $\mathbf{G} = \mathbf{H}^{-1}$ is also a graph filter on G [Sandryhaila-Moura'13]
  - $\Rightarrow$ Requires $\tilde{h}_i = \sum_{l=0}^{L-1} h_l \lambda_i^l \neq 0$, for all $i = 1, \ldots, N$
  - $\Rightarrow$ Inverse-filter coefficients $\mathbf{g} \in \mathbb{R}^N$, frequency response $\tilde{\mathbf{g}} = \mathbf{\Psi g}$
- Recast as linear inverse problem [Wang-Chi'16]

$$\min_{\{\tilde{\mathbf{g}}, \mathbf{X}\}} \|\mathbf{X}\|_0, \quad \text{s. to} \quad \mathbf{X} = \mathbf{V}\text{diag}(\tilde{\mathbf{g}})\mathbf{V}^T\mathbf{Y}, \quad \mathbf{X} \neq \mathbf{0}$$

- Still NP hard. Relax! and minimize convex $\|\mathbf{X}\|_1$

$$\hat{\tilde{\mathbf{g}}} = \operatorname*{argmin}_{\tilde{\mathbf{g}}} \|(\mathbf{Y}^T\mathbf{V} \odot \mathbf{V})\tilde{\mathbf{g}}\|_1, \quad \text{s. to} \quad \mathbf{1}^T\tilde{\mathbf{g}} = 1$$

  - $\Rightarrow$ Constraint fixes the scale and avoids all-zero solution
  - $\Rightarrow$ ℓ_1-synthesis problem, efficient solvers available

Recovery guarantee for ℓ_1 relaxation

- Let $\{\mathbf{X}_0, \tilde{\mathbf{g}}_0\}$ be the solution, i.e., $\mathbf{X}_0 = \mathbf{V}\text{diag}(\tilde{\mathbf{g}}_0)\mathbf{V}^T\mathbf{Y}$
  $\Rightarrow \mathcal{I}$ indexes the support of $\text{vec}(\mathbf{X}_0)$, complement is $\mathcal{I}^c$
- Define $\mathbf{Z} := \mathbf{Y}^T\mathbf{V} \odot \mathbf{V} \in \mathbb{R}^{NP\times N}$
  $\Rightarrow \mathbf{Z}_{\mathcal{S}}$ is the submatrix of $\mathbf{Z}$ with rows indexed by $\mathcal{S} \subset \{1, ..., NP\}$.

Proposition: $\hat{\tilde{\mathbf{g}}} = \tilde{\mathbf{g}}_0$ if the two following conditions are satisfied
1) $\text{rank}(\mathbf{Z}_{\mathcal{I}^c}) = N - 1$; and
2) There exists $\mathbf{f} \in \mathbb{R}^{NP}$ such that $\mathbf{Z}^T\mathbf{f} = \gamma\mathbf{1}$, for some $\gamma \neq 0$ and

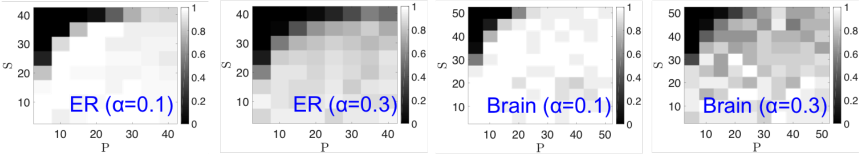
$$\mathbf{f}_{\mathcal{I}} = \text{sign}(\mathbf{Z}_{\mathcal{I}}\tilde{\mathbf{g}}_0) \text{ and } \|\mathbf{f}_{\mathcal{I}^c}\|_\infty < 1$$

- Cond. 1) ensures uniqueness of solution $\hat{\tilde{\mathbf{g}}}$
- Cond. 2) guarantees existence of a dual certificate $\mathbf{f}$ for ℓ_0 optimality

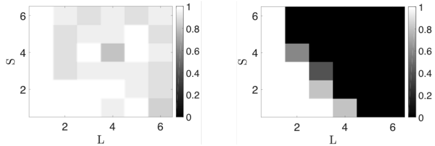
Simulation setup

- Consider undirected graphs with $\mathbf{S} = \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}$
  - $\Rightarrow$ Erdős-Rényi (ER) graphs with $N = 50$ and edge prob. $p = 0.3$
  - $\Rightarrow$ Structural brain network with $N = 66$ [Hagmann et al'08]
- $\mathbf{X}_0$ adheres to a Bernoulli-Gaussian model. Vary P and S
- Filter $\mathbf{h}_0 = (\mathbf{e}_1 + \alpha\mathbf{b})/\|\mathbf{e}_1 + \alpha\mathbf{b}\|_1$ as in [Wang-Chi'16]
  - $\Rightarrow$ $\mathbf{e}_1 = [1, 0, ..., 0]^T \in \mathbb{R}^L$ and $\mathbf{b} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
  - $\Rightarrow$ Recovery performance increases while $\alpha \geq 0$ decreases
- Observation matrix $\rightarrow$ $\mathbf{Y} = \mathbf{V}\text{diag}(\mathbf{\Psi}\mathbf{h}_0)\mathbf{V}^T\mathbf{X}_0$
- **Figure of merit:** Relative recovery error $e_X = \|\hat{\mathbf{X}} - \mathbf{X}_0\|/\|\mathbf{X}_0\|$
  - $\Rightarrow$ Successful recovery $e_X < 0.01$. Show rates over 20 realizations

Recovery performance



- Successful recovery over most of the (S, P) plane
  - $\Rightarrow$ Using multiple signals aids recovery
  - $\Rightarrow$ Performance improves with smaller α



- Brain graph ($\alpha = 0.5$). Proposed (left) and [Segarra et al'17] (right)
  - $\Rightarrow$ Performance of matrix lifting approach degrades faster with L

Concluding summary

- Blind identification of graph filters with multiple sparse inputs
  - ⇒ Extends blind deconvolution of space/time signals to graphs
  - ⇒ Key: model diffusion process as output of graph filter
- Invertible graph filter assumption
  - ⇒ From a bilinear to a linear inverse problem
  - ⇒ Devoid of matrix lifting → Scales better to large graphs
  - ⇒ Encouraging performance for random and real-world graphs
- Ongoing work
  - ⇒ Exact recovery under the Bernoulli-Gaussian model
  - ⇒ Stable recovery from noisy and sampled observations
- Envisioned application domains
  - (a) Localize sources of epileptic seizure
  - (b) Event-driven information cascades and "fake-news" detection
  - (c) Trace "patient zero" for an epidemic outbreak