Online Graph Learning under Smoothness Priors

Seyed Saman Saboksayr

Dept. of Electrical and Computer Engineering

University of Rochester

Email: ssaboksa@ur.rochester.edu

Co-authors: Gonzalo Mateos and Mujdat Cetin

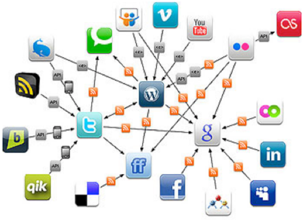
Acknowledgement: NSF awards CCF-1750428, CCF-1934962 and ECCS-1809356

EUSIPCO, August 23-27, 2021

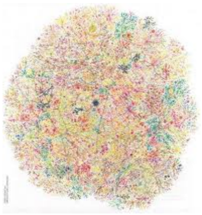
Network science analytics

UNIVERSITY of ROCHESTER

Online social media



Internet



Clean energy and grid analytics

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

  $\Rightarrow$ Use $\mathcal{G}$ to study graph signals, data associated with nodes in $\mathcal{V}$
- Ex: Opinion profile, buffer congestion levels, functional brain connectivity
- **Q:** What about streaming data from (possibly) dynamic networks?

Topology inference: Motivation and context

UNIVERSITY of ROCHESTER

- Network topology inference from nodal observations [Kolaczyk'09]
  - Partial correlations and conditional dependence [Dempster'74]
  - Sparsity [Friedman et al'07] and consistency [Meinshausen-Buhlmann'06]
- Key in neuroscience [Kassam et al'13] and financial market analytics [Palomar et al'20]
- Noteworthy GSP-based approaches
  - Graphical models [Egilmez et al'16], [Rabbat'17], [Kumar et al'19], ...
  - Smooth signals [Dong et al'15], [Kalofolias'16], [Sardellitti et al'17], ...
  - Streaming data [Shafipour et al'18], [Natali et al'20], [Saboksayr et al'21], ...
- **Our contribution:** Graph learning from streaming signals

$\Rightarrow$ Topology inference via convergent online proximal gradient (PG) iterations

UNIVERSITY of ROCHESTER

Setup

- Sparse network $\mathcal{G}$ with unknown adjacency matrix $\mathbf{W}$ (or $\mathbf{W}_t$ in dynamic setting)
- Observe streaming smooth signals $\{\mathbf{x}_t\}_{t=1}^{T}$ defined on $\mathcal{G}$
- Total variation (smoothness measure) of signal $\mathbf{x}$ with respect to $\mathbf{L} = \mathbf{D} - \mathbf{W}$

$$\mathrm{TV}(\mathbf{x}) := \mathbf{x}^{\top}\mathbf{L}\mathbf{x} = \frac{1}{2}\sum_{i \neq j} W_{ij}\,(x_i - x_j)^2$$

Problem statement

Given a set $\mathcal{X} = \{\mathbf{x}_t\}_{t=1}^{T}$ of graph signal observations acquired at time t, learn an undirected graph $\mathcal{G}(\mathcal{V}, \mathcal{E}, \mathbf{W})$ such that the observations in $\mathcal{X}$ are smooth on $\mathcal{G}$.

- Given $\mathcal{X}$ one can form the data matrix $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_T] \in \mathbb{R}^{N \times T}$

⇒ A link between smoothness and sparsity (considering $\mathbf{Z} \in \mathbb{R}_+^{N \times N}$ as Euclidean-distance matrix)

$$\sum_{t=1}^{T} \mathrm{TV}(\mathbf{x}_t) = \mathrm{trace}(\mathbf{X}^\top \mathbf{L} \mathbf{X}) = \frac{1}{2} \|\mathbf{W} \circ \mathbf{Z}\|_1$$

- Framework for learning graphs under a smoothness prior [Kalofolias'16]

$$\min_{\mathbf{W}} \|\mathbf{W} \circ \mathbf{Z}\|_1 + g(\mathbf{W}) \quad \text{s. t.} \quad \mathrm{diag}(\mathbf{W}) = \mathbf{0},\ W_{ij} = W_{ji} \geq 0,\ i \neq j$$

⇒ Convex objective function $g(\mathbf{W})$ encodes assumptions about the network $\mathcal{G}$

⇒ Amenable to the proximal gradient method, $\mathcal{O}(N^2)$ complexity per iteration

- **Q:** Online estimation from streaming data $\{\mathbf{x}_1, \ldots, \mathbf{x}_t, \mathbf{x}_{t+1}, \ldots\}$

  $\Rightarrow$ At time t solve the time-varying composite optimization

$$\mathbf{w}_t^\star \in \underset{\mathbf{w}}{\operatorname{argmin}}\, F_t(\mathbf{w}) := \underbrace{\mathbb{I}\{\mathbf{w} \succeq \mathbf{0}\} + 2\mathbf{w}^\top \mathbf{z}_{1:t}}_{h_t(\mathbf{w})} \underbrace{-\alpha \mathbf{1}^\top \log(\mathbf{S}\mathbf{w}) + \beta 2\|\mathbf{w}\|^2}_{g(\mathbf{w})}.$$

- **Step 1:** Recursively update the Euclidean-distance vector via exponential moving average

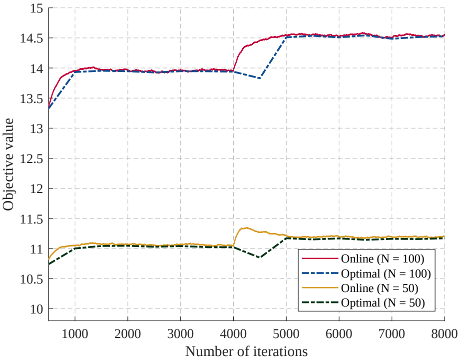
$$\bar{\mathbf{z}}_t = (1-\gamma)\bar{\mathbf{z}}_{t-1} + \gamma \mathbf{z}_t$$

- **Step 2:** Run a single iteration of the PG algorithm [Madden et al'19]

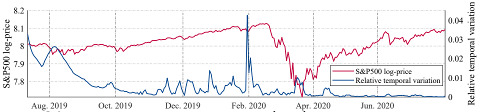
$$\mathbf{w}_{t+1} = \mathbf{prox}_{\mu_t h_t}(\mathbf{w}_t - \mu_t \nabla g(\mathbf{w}_t))$$

  $\Rightarrow$ Memory footprint and computational complexity does not grow with t

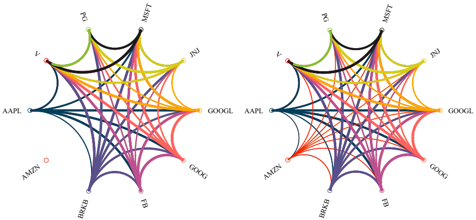
  $\Rightarrow$ Guaranteed to converge within a neighborhood of the optimal time-varying batch solution



Mean of objective value as a function of acquired time samples (iteration). Indicates that the proposed method can effectively track its offline counterpart.



The S&P500 log-price per day (red). The daily relative temporal variation of the learned graphs (blue).



Jan 10, 2020 *Mar 9, 2020*

The estimated network of the market over two different days.

More information

- **Our paper:** S. S. Saboksayr, G. Mateos, and M. Cetin, "Online graph learning under smoothness priors" in European Signal Process. Conf. (EUSIPCO), Dublin, Ireland, 2021.

- **Extended journal paper:** S. S. Saboksayr, G. Mateos, and M. Cetin, "Online discriminative graph learning from multi-class smooth signals,"Signal Processing, vol. 186,p. 108101, 2021.

- Session: Signal Processing over Graphs and Networks (Wednesday, 25 August, 13:30 - 16:30 IST)