Inferring Network Structure from Indirect Observations

Mike Rabbat

McGill University, Montréal, Canada

GSP Workshop, Philadelphia, May 2016
Motivation

- Need “the graph(s)” to do graph signal processing
- Fundamental problem: Given signals, recover the graph
The Agenda

- Historical overview
- Case study: Inferring network structure from cascades
- Open directions
Combinatorial vs. Statistical Approaches

Combinatorial Graph Reconstruction

• Let $G = (V, E)$ be a graph
• A card $G_i = V \setminus v_i$ is a vertex-deleted subgraph
• The deck is the set of all cards

$$\text{Deck}(G) = \{G_i = V \setminus v_i : v_i \in V\}$$

• Does $\text{Deck}(G)$ uniquely determine $G$?
• How many cards are necessary?
Statistical Approaches – Historical Overview

• Covariance selection (Dempster ’72)
  • Estimate $\Sigma = \mathbb{E} [xx^\top]$ from $x_1, x_2, \ldots, x_m \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \Sigma)$
  • Explore sparsity of $\Sigma^{-1}$
  • $\Sigma^{-1}_{i,j} = 0$ encodes conditional independence relationships

• Theoretical and computational advances
  • Meinshausen and Bühlmann (2006)
  • Ravikumar, Wainwright, and Lafferty (2010)
  • Friedman, Hastie, and Tibshirani (2007)
  • Banerjee, El Ghaoui, and d’Aspremont (2008)
  • Marjanovic and Hero (2014)
  • Main tools: Convex optimization + sparsity/high-d statistics
• Learning probabilistic graphical models
  • Beyond Gaussian/Ising
  • Survey by Heckerman (1995)
  • Main tools: Search over discrete space of graphs
Learning causal networks

- Edge $u \rightarrow v$ iff $\neg v \implies \neg u$
- Interventions: Pearl (2000)
- Predictive: Granger (1969)

Other network representations

- Partial correlations
- Structural equation models
- ...
Application-Specific Approaches

Network tomography

### Session: Thursday P.M.

<table>
<thead>
<tr>
<th>Time</th>
<th>Session</th>
</tr>
</thead>
<tbody>
<tr>
<td>4:00</td>
<td>6:00 Graph topology identification</td>
</tr>
<tr>
<td>4:00</td>
<td><strong>Bastien Pasdeloup, Vincent Gripon, Dominique Pastor, Grégoire Mercier and Michael Rabbat.</strong> Characterizing graphs from diffused signals</td>
</tr>
<tr>
<td>4:20</td>
<td><strong>Benjamin Miller.</strong> Recent Advances in Subgraph Detection</td>
</tr>
<tr>
<td>4:40</td>
<td><strong>Santiago Segarra, Antonio Marques, Gonzalo Mateos and Alejandro Ribeiro.</strong> Network Topology Identification from Spectral Templates</td>
</tr>
<tr>
<td>5:00</td>
<td><strong>Eduardo Pavez, Antonio Ortega and Hilmi E. Egilmez.</strong> Generalized Laplacian Estimation for Graph Signal Processing</td>
</tr>
<tr>
<td>5:20</td>
<td><strong>Dorina Thanou, Xiaowen Dong and Pascal Frossard.</strong> Learning Graph Laplacians: A Signal Representation Perspective</td>
</tr>
<tr>
<td>5:40</td>
<td><strong>Brian Baingana and Georgios Giannakis.</strong> Tracking switching network topologies from propagating graph signals</td>
</tr>
</tbody>
</table>
Relevant Questions

**Identifiability:** Do the observations (asymptotically) uniquely identify $G$?

**Consistency:** Does $\hat{G}$ (asymptotically) converge to $G$?

**Statistical complexity:** How many samples are required for $\hat{G} = G$ w.h.p.?

**Computational complexity:** To compute $\hat{G}$ from observations?
Case Study

Observations of Cascades

Observe multiple cascades $t^1, t^2, \ldots, t^m$ where $t^c_u \in [0, T^c] \cup \{\infty\}$

Each cascade is a graph signal (of exposure times)
Continuous-time diffusion model
Gomez-Rodriguez et al. (2011)

1. Cascade begins at seed node $s \overset{i.i.d.}{\sim} \mathbb{P}(s)$
2. Subsequent transmission over edge $j \to i$ governed by density

$$f(t_i|t_j; \alpha_{ji}) = f(t_i - t_j; \alpha_{ji})$$
Examples

- Exponential

\[ f(\tau; \alpha_{ji}) = \alpha_{ji} e^{-\alpha_{ji} \tau} \mathbb{1} \{ \tau \geq 0 \} \]

- Power-law \((\delta > 0)\)

\[ f(\tau; \alpha_{ji}) = \frac{\alpha_{ji}}{\delta} \left( \frac{\tau}{\delta} \right)^{-1 - \alpha_{ji}} \mathbb{1} \{ \tau \geq \delta \} \]

- Rayleigh

\[ f(\tau; \alpha_{ji}) = \alpha_{ji} \tau e^{-\alpha_{ji} \tau^2 / 2} \mathbb{1} \{ \tau \geq 0 \} \]
Continuous-time diffusion model
Gomez-Rodriguez et al. (2011)

1. Cascade begins at seed node $s \overset{i.i.d.}{\sim} \mathbb{P}(s)$
2. Subsequent transmission over edge $j \rightarrow i$ governed by density

$$f(t_i|t_j; \alpha_{ji}) = f(t_i - t_j; \alpha_{ji})$$

Also define survival function

$$s(t_i|t_j; \alpha_{ji}) = 1 - \int_{t_j}^{t_i} f(t - t_i; \alpha_{ji}) dt$$

and hazard function

$$h(t_i|t_j; \alpha_{ji}) = \frac{f(t_i|t_j; \alpha_{ji})}{s(t_i|t_j; \alpha_{ji})}$$
Likelihood of infected nodes

Likelihood that a particular node $j$ is the parent of $i$,

$$f(t_i|t_j; \alpha_{ji}) \prod_{k \neq j: t_k < t_i} s(t_i|t_k; \alpha_{ki})$$

Marginalizing over all nodes infected before $i$,

$$\sum_{j: t_j < t_i} f(t_i|t_j; \alpha_{ji}) \prod_{k \neq j: t_k < t_i} s(t_i|t_k; \alpha_{ki})$$

$$= \prod_{k: t_k < t_i} s(t_i|t_k; \alpha_{ki}) \sum_{j: t_j < t_i} h(t_i|t_j; \alpha_{ji})$$
Likelihood of a Cascade

Network structure is encoded in matrix $A = [\alpha_{ji}]$

For a cascade $t = (t_1, \ldots, t_N)$ with $T = \max\{t_i : t_i < \infty\}$

$$f(t; A) = \prod_{i:t_i \leq T} \prod_{u:t_u > T} s(T|t_i; \alpha_{iu})$$
$$\times \prod_{k:t_k < t_i} s(t_i|t_k; \alpha_{ki}) \sum_{j:t_j < t_i} h(t_i|t_j; \alpha_{ji})$$

Maximum likelihood: Given $t^1, \ldots, t^m$, find $A$ to

minimize $-\sum_{c=1}^{m} \log f(t^c; A)$
subject to $\alpha_{ji} \geq 0, \ i, j = 1, \ldots, N, i \neq j$

Note: Problem decouples into independent per-node sub-problems
For node $i$ optimize over $\alpha_i = \{\alpha_{ji} : j = 1, \ldots, N, j \neq i\}$.

minimize $\ell_i(\alpha_i) \overset{\text{def}}{=} -\sum_{c=1}^{m} g_i(t^c; \alpha_i)$
subject to $\alpha_{ji} \geq 0, \quad j = 1, \ldots, N$

where (e.g., for exponential delays)

$$g_i(t; \alpha_i) = \begin{cases} 
\log \left( \sum_{j : t_j < t_i} \alpha_{ji} \right) - \sum_{j : t_j < t_i} \alpha_{ji} (t_i - t_j) & \text{if } t_i \leq T \\
- \sum_{j : t_j \leq T} \alpha_{ji} (T - t_j) & \text{if } t_i > T
\end{cases}$$

Note: Problem is convex in $\alpha_i$ for all models mentioned before
Identifiability and Consistency

Are per-node problems identifiable? \( (\ell_i(\alpha) = \ell_i(\alpha^*) \iff \alpha = \alpha^*) \)

The Hessian \( Q(\alpha_i) = \nabla^2 \ell_i(\alpha_i) \) has the form

\[
Q(\alpha_i) = X(\alpha_i)[X(\alpha_i)]^T
\]

where \( X(\alpha_i) \in \mathbb{R}^{N \times m} \) with entries

\[
[X(\alpha_i)]_{jc} = \begin{cases} 
(\sum_{k: t_k^c < t_i^c} \alpha_{ki})^{-1} & \text{if } t_j^c < t_i^c \\
0 & \text{if } t_j^c > t_i^c.
\end{cases}
\]

Suppose \( P(s) > 0 \) for all \( s = 1, \ldots, N \).

Then, as \( m \rightarrow \infty \), \( X(\alpha_i) \) is full row-rank.

\( \implies \) \( Q(\alpha_i) \) is positive definite

\( \implies \) If \( \alpha \neq \alpha^* \) then \( \ell_i(\alpha) > \ell_i(\alpha^*) \).

Consistency also follows since the problem is strictly convex.
Finite Sample Recovery

When can we recover $G$ (w.h.p.) from a finite number cascades?

- Are some networks more difficult to recover than others?
- What kind of cascades are needed for recovery?

Examine “population” log-likelihood $\mathbb{E} [\ell_i(\alpha)] = \mathbb{E} [g_i(t; \alpha)]$. 
Let $Q^* \overset{\text{def}}{=} \nabla^2 \mathbb{E} \left[ g_i(t; \alpha) \right] |_{\alpha = \alpha^*}$.
Let $P$ denote the true set of parents.

**Dependency Condition:** There exist constants $C_{\text{min}}, C_{\text{max}} > 0$ such that

$$C_{\text{min}} < \lambda_{\text{min}}(Q^*_PP) < \lambda_{\text{max}}(Q^*_PP) < C_{\text{max}}.$$

**Incoherence Condition:** There exists a constant $\epsilon \in (0, 1]$ such that

$$\|Q^*_{PP}(Q^*_PP)^{-1}\|_{\infty} \leq 1 - \epsilon.$$
Conditions depend on $A = [\alpha_{ji}]$ and $P(s)$
Finite-Sample Recovery Results

Choose $\hat{\alpha}_i$ to solve the regularized problem

\[
\text{minimize } \ell_i(\alpha_i) + \lambda \|\alpha_i\|_1 \\
\text{subject to } \alpha_{ji} \geq 0, \quad j = 1, \ldots, N
\]

Theorem (Gomez-Rodriguez et al. (2016)): Suppose

\[
\lambda \geq C_1 \frac{2 - \epsilon}{\epsilon} \sqrt{\frac{\log N}{N}}.
\]

If $m > C_2 |P|^3 \log N$ then with probability at least $1 - 2 \exp(C_3 \lambda^2 m)$,

1. $\hat{\alpha}_i$ is unique, and
2. $\hat{\alpha}_i = \alpha^*_i$. 

Summary

Identifiability,
Consistency,
Statistical complexity, and
Computational complexity

Open questions for this particular model:
  • Noise/uncertainty in observations
  • Allow for / model unobserved incubation periods
  • Track time-varying $\alpha_{ji}$'s
Directions For Topology Identification in GSP

- Inferring topologies such that observed signals are smooth
  - Generative models (diffusions, Gaussian)
  - Discriminative
- Inference with confidences (or $p(G|X)$)
- Understand how errors in $\hat{G}$ impact subsequent GSP operations (filtering, sampling)
- Infer topology characteristics

michael.rabbat@mcgill.ca