Identifying weak network activation patterns

Aarti Singh

Network Measurement and Mapping Conference (NMMC) June 8, 2009
Weak network activations

- Onset of malicious activity or congestion in the Internet
- Sensor network monitoring incipient contamination
- fMRI signals recorded in brain network
- Expression levels in gene networks
- Anomalous patterns arising in social networks
“Weak” activation patterns

Observation at node $i$:

$$y_i \sim \mathcal{N}(\mu x_i, \sigma^2) \quad i = 1, \ldots, p$$

Weak in strength: $\mu < \sigma \sqrt{2 \log p}$

If $\mu > \sigma \sqrt{2 \log p}$, signal can be detected with max order statistic

$$\max_{i=1,\ldots,p} |y_i| < \sigma \sqrt{2 \log p} \quad \text{in the absence of signal}$$
“Weak” activation patterns

Observation at node $i$

$y_i \sim \mathcal{N}(\mu x_i, \sigma^2)$ \quad $i = 1, \ldots, p$

Binary pattern 0/1

Signal strength

Weak in extent: $\|x\|_0 < \sqrt{p}$

If $\|x\|_0 > \sqrt{p}$, signal can be detected with averaging statistic

$$\frac{1}{\sqrt{p}} \sum_{i=1}^{p} y_i \sim \mathcal{N} \left( \frac{\|x\|_0}{\sqrt{p}} \mu, \sigma^2 \right)$$
“Weak” activation patterns

Observation at node $i$

$$y_i \sim \mathcal{N}(\mu x_i, \sigma^2) \quad i = 1, \ldots, p$$

Weak in strength: signal $\mu < \sigma \sqrt{2 \log p}$

Weak in extent: # active nodes $\|x\|_0 < \sqrt{p}$

Invisible in local node signatures, as well as network-wide coarse signature
Detect/localize Weak patterns

Detection: \( H_0 : y \sim \mathcal{N}(0, \sigma^2 I) \)
\( H_1 : y \sim \mathcal{N}(\mu x, \sigma^2 I) \)

Localization: find \( \{i : x_i = 1\} \)

Threshold of detectability/localizability: (Ingster, Jin-Donoho’03, Abramovich et al ’01)

\[ \# \text{ active nodes} \sim p^{1-\alpha}, \alpha \in (1/2, 1) \]

\[ \mu^* = \sigma \sqrt{2\eta_\alpha \log p} \]

Subtle adaptive testing procedures:
Detection - Higher Criticism statistic (Jin-Donoho’03)
Localization - False Discovery Rate statistic (Abramovich-Benjamini-Donoho’01)
Network activation patterns

Graph $G = (V, E)$

- How can we exploit the (possibly non-local) dependencies between node measurements to boost performance?
- Method must be adaptive to network behavior and/or structure.
- How do node interactions effect thresholds of detectability/localizability?
Modeling network activation patterns

Hidden Multi-scale Ising model:  Prior on activation patterns

\[ \mathbf{x} = \{z_i\}_{i=\text{Leaf node}} \in \{0, 1\}^p \quad \mathbf{z} \in \{0, 1\}^{|V|} \]

\[ p(\mathbf{z}) = \frac{1}{Z} \exp \left( \beta \sum_{(i,j) \in E} [z_i z_j + (1 - z_i)(1 - z_j)] \right) \]

strength of interaction  # edge agreements
Detect/localize **Weak network patterns**

Tree $T = (V, E)$

Roadmap:

- Learn dependencies from multiple snapshots $(y_1^{(i)}, y_2^{(i)}, \ldots, y_p^{(i)})_{i=1}^n$ of noisy network data

- Sparsifying transform for network data (Capture network signal energy in few large coefficients = boost signal-to-noise ratio)

- Solve detection/localization problems in transform domain
Ising model correlation structure

Correlation, \( r_{ij} = \mathbb{E}[x_i x_j] \propto \tanh(\beta)^{d(i,j)} \) (Falk’75)
I. Hierarchical Correlation clustering

Tree $T = (V, E)$

Input: Clusters $\{1, 2, \ldots, p\}$ Correlations $\{r_{ij}\}$

Repeat: Until one cluster remains

Choose cluster pair with highest correlation

Merge pair to form a new cluster $\hat{i}$

Update correlation of new cluster with remaining clusters $r_{\hat{i}j} = \min_{k \in i} r_{kj}$

Given true correlations, provably recovers correct clustering (Duffield’02)
Sample complexity of clustering

Using empirical correlations, the hierarchical clustering algorithm can recover correct clustering with probability $> 1 - \delta$ using

$$
n \geq \frac{\log(p/\delta)}{(\epsilon(\beta) - \sigma^2)^2} - \Omega(\log p)
$$

i.i.d noiseless snapshots of network data $(y_1^{(i)}, y_2^{(i)}, \ldots, y_p^{(i)})_{i=1}^n$, where

$$
\epsilon(\beta) = \min_{\epsilon \in \{1, \ldots, L\}} (\tau_\epsilon - \tau_{\epsilon-1}) = (\tanh \beta)^{2L} (1 - (\tanh \beta)^2)
$$
II. Construct sparsifying transform

Unbalanced orthonormal Haar basis on clusters

Merge two cluster pairs \((c_1, c_2)\)

\[
b(j) = \frac{\sqrt{|c_1||c_2|}}{\sqrt{|c_1| + |c_2|}} \left[ \frac{1}{|c_2|} \mathbf{1}_{j \in c_2} - \frac{1}{|c_1|} \mathbf{1}_{j \in c_1} \right] \quad j = 1, \ldots, p
\]

Basis \(B = \{b\} \cup \frac{1}{\sqrt{p}} \mathbf{1}\)
II. Construct sparsifying transform

Unbalanced orthonormal Haar basis on clusters

Merge two cluster pairs \((c_1, c_2)\)

\[
b(j) = \frac{\sqrt{|c_1| |c_2|}}{\sqrt{|c_1| + |c_2|}} \left[ \frac{1}{|c_2|} \mathbf{1}_{j \in c_2} - \frac{1}{|c_1|} \mathbf{1}_{j \in c_1} \right] \quad j = 1, \ldots, p
\]
Transform domain sparsity

Consider a pattern $\mathbf{x}$ distributed according to the hidden multi-scale Ising model defined on a tree with depth $L$ and maximum degree $d$.

Then with probability $\geq 1 - \delta$, the number of non-zero Haar basis coefficients are bounded by

$$\#\{\mathbf{b}^T \mathbf{x} > 0\} \leq c_\delta d L^2 p e^{-2\beta}$$

where $c_\delta > 0$ is a constant.

Higher $\beta$ $\implies$ correlated network activity $\implies$ more sparsity

Efficient “compression” of network data

Boost signal-to-noise ratio by adaptive multi-scale aggregation
Basis for Graphs – related work

- **Graph wavelets (Crovella,Kolacyzk’03)**
  - highly overcomplete basis (#nodes x #hops)

- **Diffusion wavelets (Coifman,Maggioni’06)**
  - multi-scale and approximation properties not well understood

- **Treelets (Lee,Nadler,Wasserman’08)**
  - not a sparsifying transform if nodes have different variances

- **Balanced Haar on a dendogram (Murtagh’05)**
  - not orthogonal, basis vectors not constant on merged sub-clusters

\[
\begin{bmatrix}
  x_1, x_2, x_3 \\
  x_4
\end{bmatrix} \quad \begin{bmatrix}
  \frac{1}{8} & \frac{1}{8} & \frac{1}{4} & -\frac{1}{2}
\end{bmatrix}
\]

\[
\begin{bmatrix}
  \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & -\frac{3}{2}
\end{bmatrix}
\]
III. Identifying weak network patterns

- Project network measurements $\mathbf{y}$ onto basis $\mathbf{B} : \mathbf{B}^T \mathbf{y}$

**Detection:**
- $H_0 : \mathbf{y} \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$
- $H_1 : \mathbf{y} \sim \mathcal{N}(\mu \mathbf{x}, \sigma^2 \mathbf{I})$
- # active nodes $\sim p^{1-\alpha}$

**Localization:**
- find $\{i : x_i = 1\}$
- recover $\mathbf{x}$

**Transform domain**
- $H_0 : \mathbf{B}^T \mathbf{y} \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$
- $H_1 : \mathbf{B}^T \mathbf{y} \sim \mathcal{N}(\mu \mathbf{B}^T \mathbf{x}, \sigma^2 \mathbf{I})$
- # non-zero coeffs $\sim pe^{-2\beta}$

- **ONB**

- **Identify weak network patterns in transform domain**

  - Detection - Higher Criticism statistic (Jin-Donoho’03)
  - Localization - False Discovery Rate statistic (Abramovich-Benjamini-Donoho’01)
Identifying weak network patterns

- How do node interactions effect thresholds of detectability/localizability?

Threshold of detectability/localizability

- canonical domain: $\mu^* = \sigma \sqrt{2\eta_\alpha \log p}$
- transform domain: $\mu^* |b^T x| = \sigma \sqrt{2\eta_\beta \log p}$

Effective signal strength

Basis coeff: $\frac{1}{\sqrt{p}} 1^T x$ (averaging statistic)

$|b^T x| \propto$ Extent of correlated activity, $1/#$ changepoints

$x$  

\begin{align*}
\mathbf{b} & = \begin{bmatrix}
1 & 1 & 1 & 1 & -1 & -1 & -1 & -1
\end{bmatrix}^T \\
& = \begin{bmatrix}
1 & 1 & -1 & -1 & 0 & 0 & 0 & 0
\end{bmatrix}^T \\
|b^T x| & = \frac{4}{\sqrt{8}} = \sqrt{2} \\
& = \frac{2}{\sqrt{4}} = 1
\end{align*}
Identifying weak network patterns

- How do node interactions effect thresholds of detectability/localizability?

Threshold of detectability/localizability

**canonical domain:** \( \mu^* = \sigma \sqrt{2\eta_{\alpha} \log p} \)

**transform domain:** \( \mu^* | b^T x | = \sigma \sqrt{2\eta_{\beta} \log p} \)

Effective signal strength

Basis coeff: \( \frac{1}{\sqrt{p}} 1^T x \) (averaging statistic)

\[ |b^T x| \propto \text{Extent of correlated activity, } 1/\#\text{change points} \]

\[ x \quad \bullet \quad \bullet \quad \bullet \quad \bullet \quad \bullet \quad \bullet \quad \bullet \quad \bullet \quad \bullet \]

\[ b \quad \frac{1}{\sqrt{8}} \begin{bmatrix} 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \end{bmatrix} \]

\[ b^T x = \frac{1}{\sqrt{4}} \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & -1 & -1 \end{bmatrix} \]

\[ |b^T x| = \frac{4}{\sqrt{8}} = \sqrt{2} \]

\[ \frac{1}{\sqrt{4}} \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & -1 & -1 \end{bmatrix} \]

\[ \frac{2}{\sqrt{4}} = 1 \]
Initial Experiments

Multi-scale Ising model with $p = 1296$ leafs, degree $d = 6$, depth $L = 4$

FDR performance

- Transform basis
- Canonical basis

Noise bound

$\sigma \sqrt{2 \log p} = 0.38$

Signal strength, $\mu$

Correlated locations

$\mu$
Preliminary results:

iPlane dataset (active traceroutes sent from Planetlab monitors)
Hierarchical clustering on node similarities
> 80% nodes merged with same AS or parent AS (as verified with BGP data)
Collaborators

- Robert Calderbank (Princeton)
- Rob Nowak (Wisconsin)
- Jennifer Rexford, David Shue (Princeton)
- Ingrid Daubechies, Eugene Brevdo (Princeton)

asimg@math.princeton.edu
http://pacm.princeton.edu/~asimgh
Backup Slides
Modeling network activation patterns

Prior work:

Linear topology (Single Interval of activation)

“Near-Optimal detection of geometric objects by fast multi-scale methods” (Arias-Castro, Donoho, Huo’03)

Lattice and tree topology (Path of activation)

“Searching for trail of evidence in a maze” (Arias-Castro, Candes, Helgason, Zeitouni, ’07)
Modeling network activation patterns

Probabilistic graphical model: $\text{Graph } G = (V, E)$

$$p(x) \propto \exp \left( 2 \sum_{(i,j) \in E} \beta_{ij} [x_i x_j + (1 - x_i)(1 - x_j)] \right)$$

Prior on activation patterns

Ernst Ising

Coupling of electron spins
Transform domain sparsity

Consider a pattern \( \mathbf{x} \) distributed according to the hidden multi-scale Ising model defined on a tree with depth \( L \) and maximum degree \( d \).

Then with probability \( > 1 - \delta \), the number of non-zero Haar basis coefficients are bounded by

\[
\#\{ \mathbf{b}^T \mathbf{x} > 0 \} \leq c_\delta dL^2 p e^{-2\beta}
\]

where \( c_\delta > 0 \) is a constant.

Proof sketch: \( \mathbf{b}^T \mathbf{1} = 0 \)

\[
\#\{ \mathbf{b}^T \mathbf{x} > 0 \} \leq dL \# \text{changepoints}
\leq c_\delta dL |E| e^{-2\beta} \quad (\text{w.p.} > 1 - \delta)
\leq c_\delta dL^2 p e^{-2\beta} \quad \text{Relative Chernoff bound}
\]

\#changepoints \sim \text{Binomial}(|E|, \frac{1}{1 + e^{2\beta}})
III. Identifying weak network patterns

- Project network measurements $\mathbf{y}$ onto basis $\mathbf{B} = \{\mathbf{b}\} \cup \frac{1}{\sqrt{p}} \mathbf{1} : \mathbf{B}^T \mathbf{y}$

Detection:

- $H_0 : \mathbf{y} \sim \mathcal{N}(0, \sigma^2 \mathbb{I})$
- $H_1 : \mathbf{y} \sim \mathcal{N}(\mu \mathbf{x}, \sigma^2 \mathbb{I})$

- $\# \text{ active nodes} \sim p^{1-\alpha}$

ONB

Transform domain:

- $H_0 : \mathbf{B}^T \mathbf{y} \sim \mathcal{N}(0, \sigma^2 \mathbb{I})$
- $H_1 : \mathbf{B}^T \mathbf{y} \sim \mathcal{N}(\mu \mathbf{B}^T \mathbf{x}, \sigma^2 \mathbb{I})$

- $\# \text{ non-zero coeffs} \sim c_5 d L^2 p e^{-2\beta}$
  - $\sim p e^{-2\beta}$

Unbalanced trees (large $L$) $\implies$ less sparsity

Assume: $L \sim (\log p)^c$, $c > 0$
Initial Experiments

Multi-scale Ising model with $p = 1296$ leaves, degree $d = 6$, depth $L = 4$

Noise bound

$$\sigma \sqrt{2 \log p} = 0.38$$