Cornell University

Lecture 20 - Unsupervised GSP Algorithms

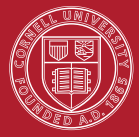
ECE 5260 – ORIE 5735

Anna Scaglione

Cornell Tech

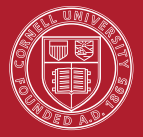
April 20, 2026

Content



Cornell University

Outline



Cornell University

Graph Filters

- Graph Filters are **shift invariant** linear operators, i.e. $\mathbf{S}\mathcal{H}(\mathbf{S}) \equiv \mathcal{H}(\mathbf{S})\mathbf{S}$. To be invariant w.r.t. the GSO a **graph filter** must be a matrix polynomial:

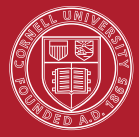
$$\mathcal{H}(\mathbf{S}) \triangleq \sum_{\ell=-\infty}^{+\infty} h_\ell \mathbf{S}^\ell. \tag{1}$$

- Examples: finite order $h_0 + h_1\mathbf{S} + h_2\mathbf{S}^2$, of infinite order filters: $\mathcal{H}(\mathbf{S}) = (\mathbb{I} - \alpha\mathbf{S})^{-1}$, $\mathcal{H}(\mathbf{S}) = e^{\alpha\mathbf{S}}$.
- Graph/network data are **filtered** graph signals, i.e.,

$$\boldsymbol{x} = \mathcal{H}(\mathbf{S})\boldsymbol{s}. \tag{2}$$

- The signal/observation is $\boldsymbol{x}$ while $\boldsymbol{s}$ is viewed as the **excitation**.

Graph Filters



Cornell University

- Graph Filters are **shift invariant** linear operators, i.e. $\mathbf{S}\mathcal{H}(\mathbf{S}) \equiv \mathcal{H}(\mathbf{S})\mathbf{S}$. To be invariant w.r.t. the GSO a **graph filter** must be a matrix polynomial:

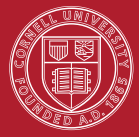
$$\mathcal{H}(\mathbf{S}) \triangleq \sum_{\ell=-\infty}^{+\infty} h_\ell \mathbf{S}^\ell. \tag{1}$$

- Examples: finite order $h_0 + h_1\mathbf{S} + h_2\mathbf{S}^2$, of infinite order filters: $\mathcal{H}(\mathbf{S}) = (\mathbb{I} - \alpha\mathbf{S})^{-1}$, $\mathcal{H}(\mathbf{S}) = e^{\alpha\mathbf{S}}$.
- Graph/network data are **filtered** graph signals, i.e.,

$$x = \mathcal{H}(\mathbf{S})s. \tag{2}$$

- The signal/observation is x while s is viewed as the **excitation**.

Graph Filters



Cornell University

- Graph Filters are **shift invariant** linear operators, i.e. $\mathbf{S}\mathcal{H}(\mathbf{S}) \equiv \mathcal{H}(\mathbf{S})\mathbf{S}$. To be invariant w.r.t. the GSO a **graph filter** must be a matrix polynomial:

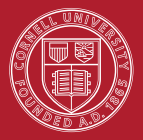
$$\mathcal{H}(\mathbf{S}) \triangleq \sum_{\ell=-\infty}^{+\infty} h_\ell \mathbf{S}^\ell. \tag{1}$$

- Examples: finite order $h_0 + h_1\mathbf{S} + h_2\mathbf{S}^2$, of infinite order filters: $\mathcal{H}(\mathbf{S}) = (\mathbb{I} - \alpha\mathbf{S})^{-1}$, $\mathcal{H}(\mathbf{S}) = e^{\alpha\mathbf{S}}$.
- Graph/network data are **filtered** graph signals, i.e.,

$$\boldsymbol{x} = \mathcal{H}(\mathbf{S})\boldsymbol{s}. \tag{2}$$

- The signal/observation is $\boldsymbol{x}$ while $\boldsymbol{s}$ is viewed as the **excitation**.

Graph Fourier Transform (GFT)



- The Graph Fourier Transform (GFT) basis are the eigen-vectors of the GSO (generalization of the DFT case for the cycle graph).
- $\mathbf{S} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{-1}$. The columns of $\boldsymbol{U}$ are the signals that are invariant w.r.t. the GSO $\mathbf{S}$

$$\tilde{\boldsymbol{x}} = \overbrace{\boldsymbol{U}^{-1}}^{GFT}\boldsymbol{x} \tag{3}$$

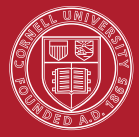
  and the eigenvalues are the graph *frequencies* from low to high: $0 = \lambda_1 \leq \cdots \leq \lambda_N$.

- The transfer function of the GF is:

$$\tilde{\mathcal{H}}(\mathbf{\Lambda}) = \mathbf{diag}(\tilde{\boldsymbol{h}}) \quad \text{where} \quad [\tilde{\boldsymbol{h}}]_i = \tilde{h}_i = \sum_{\ell} h_\ell \lambda_i^\ell$$

node coordinates: $\boldsymbol{x} = \mathcal{H}(\mathbf{S})\boldsymbol{s} \rightarrow$ GFT coordinates $\tilde{\boldsymbol{x}} = \mathrm{diag}(\tilde{\boldsymbol{h}})\tilde{\boldsymbol{s}}$

Graph Fourier Transform (GFT)



- The Graph Fourier Transform (GFT) basis are the eigen-vectors of the GSO (generalization of the DFT case for the cycle graph).
- $\mathbf{S} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{-1}$. The columns of $\boldsymbol{U}$ are the signals that are invariant w.r.t. the GSO $\mathbf{S}$

$$\tilde{\boldsymbol{x}} = \overbrace{\boldsymbol{U}^{-1}}^{GFT}\boldsymbol{x} \tag{3}$$

  and the eigenvalues are the graph *frequencies* from low to high: $0 = \lambda_1 \leq \cdots \leq \lambda_N$.

- The transfer function of the GF is:

$$\tilde{\mathcal{H}}(\mathbf{\Lambda}) = \mathbf{diag}(\tilde{\boldsymbol{h}}) \quad \text{where} \quad [\tilde{\boldsymbol{h}}]_i = \tilde{h}_i = \sum_{\ell} h_\ell \lambda_i^\ell$$

$$\text{node coordinates: } \boldsymbol{x} = \mathcal{H}(\mathbf{S})\boldsymbol{s} \rightarrow \text{ GFT coordinates } \tilde{\boldsymbol{x}} = \text{diag}(\tilde{\boldsymbol{h}})\tilde{\boldsymbol{s}}$$

Graph Fourier Transform (GFT)

- The Graph Fourier Transform (GFT) basis are the eigen-vectors of the GSO (generalization of the DFT case for the cycle graph).
- $\mathbf{S} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{-1}$. The columns of $\boldsymbol{U}$ are the signals that are invariant w.r.t. the GSO $\mathbf{S}$

$$\tilde{\boldsymbol{x}} = \overbrace{\boldsymbol{U}^{-1}}^{GFT}\boldsymbol{x} \tag{3}$$

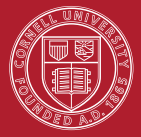
  and the eigenvalues are the graph *frequencies* from low to high: $0 = \lambda_1 \leq \cdots \leq \lambda_N$.

- The transfer function of the GF is:

$$\tilde{\mathcal{H}}(\mathbf{\Lambda}) = \mathbf{diag}(\tilde{\boldsymbol{h}}) \quad \text{where} \quad [\tilde{\boldsymbol{h}}]_i = \tilde{h}_i = \sum_{\ell} h_\ell \lambda_i^\ell$$

node coordinates: $\boldsymbol{x} = \mathcal{H}(\mathbf{S})\boldsymbol{s} \rightarrow$ GFT coordinates $\tilde{\boldsymbol{x}} = \mathrm{diag}(\tilde{\boldsymbol{h}})\tilde{\boldsymbol{s}}$

Graph filters

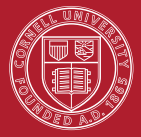


- Thus, graph filters amplify or attenuate different graph frequencies exactly as classical filters do in time-domain DSP.

Transition

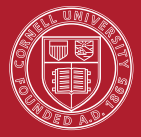
This viewpoint is the basis for modeling graph signals as *low-pass*, *bandlimited*, or approximately *low-rank* in the graph Fourier basis.

Outline



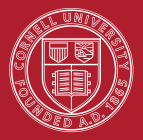
Cornell University

Outline



Cornell University

Low rank data property



Cornell University

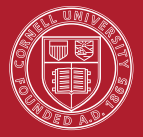
- Let $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T)$. In a number of situations data indexed by graphs exhibit a sample covariance:

$$\hat{\boldsymbol{C}}_x = \frac{1}{T} \sum_{t=1} \boldsymbol{x}_t \boldsymbol{x}_t^\top = \frac{1}{T} \boldsymbol{X} \boldsymbol{X}^\top$$

  that is ill conditioned.

- The GSP point of view is that $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T)$ is potentially low rank because $\boldsymbol{U} \tilde{\boldsymbol{X}} \tilde{\boldsymbol{X}}^\top \boldsymbol{U}^\top = \boldsymbol{U} \text{diag}(\tilde{\boldsymbol{h}}) \tilde{\boldsymbol{C}}_s \text{diag}(\tilde{\boldsymbol{h}}) \boldsymbol{U}^T$
- If the graph filter transfer function is sparse, irrespective of $\tilde{\boldsymbol{C}}_s$ only the non negligible values of $\tilde{h}(\lambda_i)$ will dominate the EVD of $\hat{\boldsymbol{C}}_x$

Low rank data property



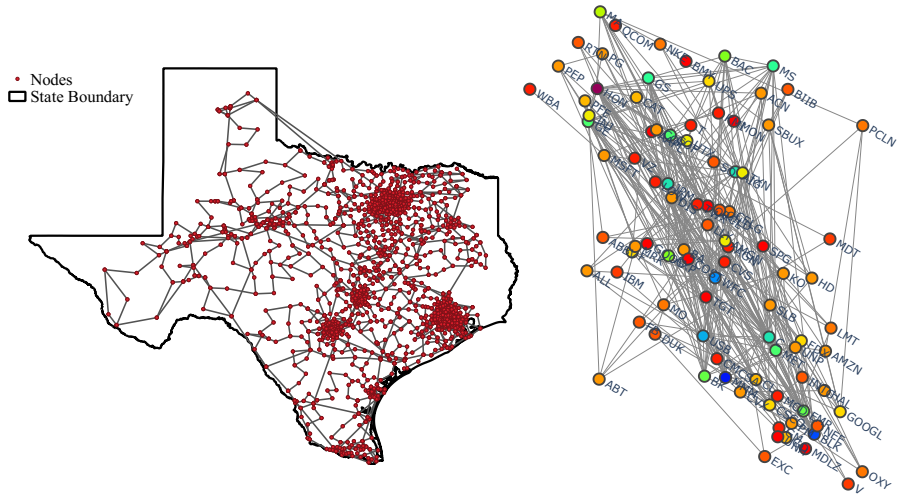
Cornell University

- Let $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T)$. In a number of situations data indexed by graphs exhibit a sample covariance:

$$\hat{\boldsymbol{C}}_x = \frac{1}{T} \sum_{t=1} \boldsymbol{x}_t \boldsymbol{x}_t^\top = \frac{1}{T} \boldsymbol{X} \boldsymbol{X}^\top$$

  that is ill conditioned.
- The GSP point of view is that $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T)$ is potentially low rank because $\boldsymbol{U} \tilde{\boldsymbol{X}} \tilde{\boldsymbol{X}}^\top \boldsymbol{U}^\top = \boldsymbol{U} \text{diag}(\tilde{\boldsymbol{h}}) \tilde{\boldsymbol{C}}_s \text{diag}(\tilde{\boldsymbol{h}}) \boldsymbol{U}^T$
- If the graph filter transfer function is sparse, irrespective of $\tilde{\boldsymbol{C}}_s$ only the non negligible values of $\tilde{h}(\lambda_i)$ will dominate the EVD of $\hat{\boldsymbol{C}}_x$

Examples



Social network **Power network** **Financial network**

Eigenvalues of the covariance matrix of Senate rollcall, voltage phasors and financial stock data. The low-pass nature is sufficient to create the low-rank property.

Orthogonal Projection

- You may recall from algebra the following result

Orthogonal projection

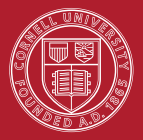
Let $\boldsymbol{U}_p$ be an orthonormal basis of a p-dimensional subspace of $\mathbb{R}^N$. Any vector in $\boldsymbol{U}_p$ can be expressed as $\boldsymbol{U}_p \boldsymbol{v}_p, \ \boldsymbol{v}_p \in \mathbb{R}^p$.
The orthogonal projection $\hat{\boldsymbol{x}}$ of $\boldsymbol{x}$ on the subspace $\boldsymbol{U}_p$, is:

$$\hat{\boldsymbol{x}} = \boldsymbol{U}_p \boldsymbol{U}_p^\top \boldsymbol{x}.$$

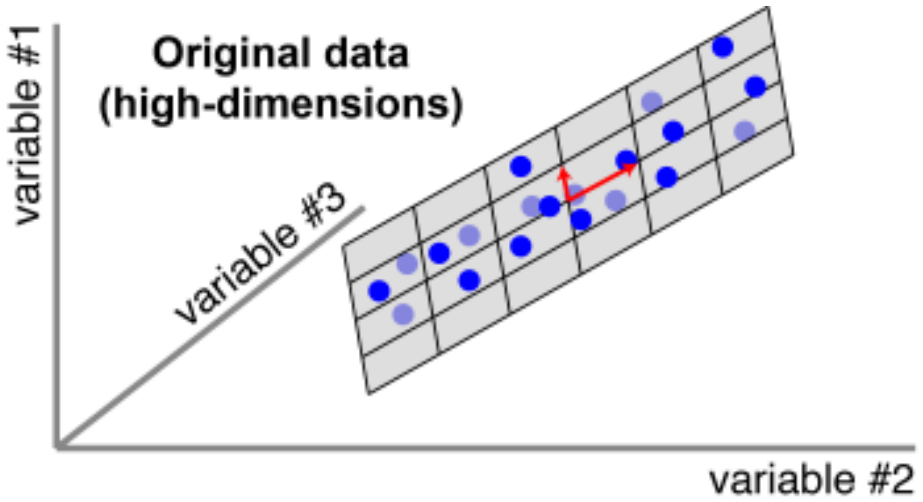
$\hat{\boldsymbol{x}}$ is such that residual error $\hat{\boldsymbol{x}} - \boldsymbol{x}$ has minimum norm, i.e.

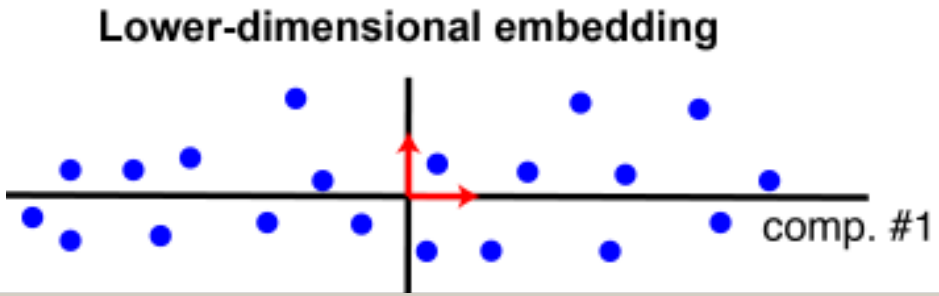
$$\hat{\boldsymbol{x}} = \arg \min_{\boldsymbol{z}: \ \boldsymbol{z} = U_p \boldsymbol{v}_p} \|\boldsymbol{z} - \boldsymbol{x}\|^2 = \boldsymbol{U}_p^\top \boldsymbol{x}$$

Principal Component Analysis (PCA)

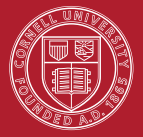


PCA tries to find the best subspace $\boldsymbol{U}_p$ that minimizes the residual. It is a generalization of linear regression: in the latter we try to find the straight line approximating $x(t) = at + b$, in PCA we have multi-dimensional data $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T]$ and try to find an hyperplane





PCA solution



- Let us assume we removed the mean from the data $\boldsymbol{X}$ i.e. $\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_t = 0$.
- As it turns out the solution is as follows:

Solution of batch PCA

The solution of

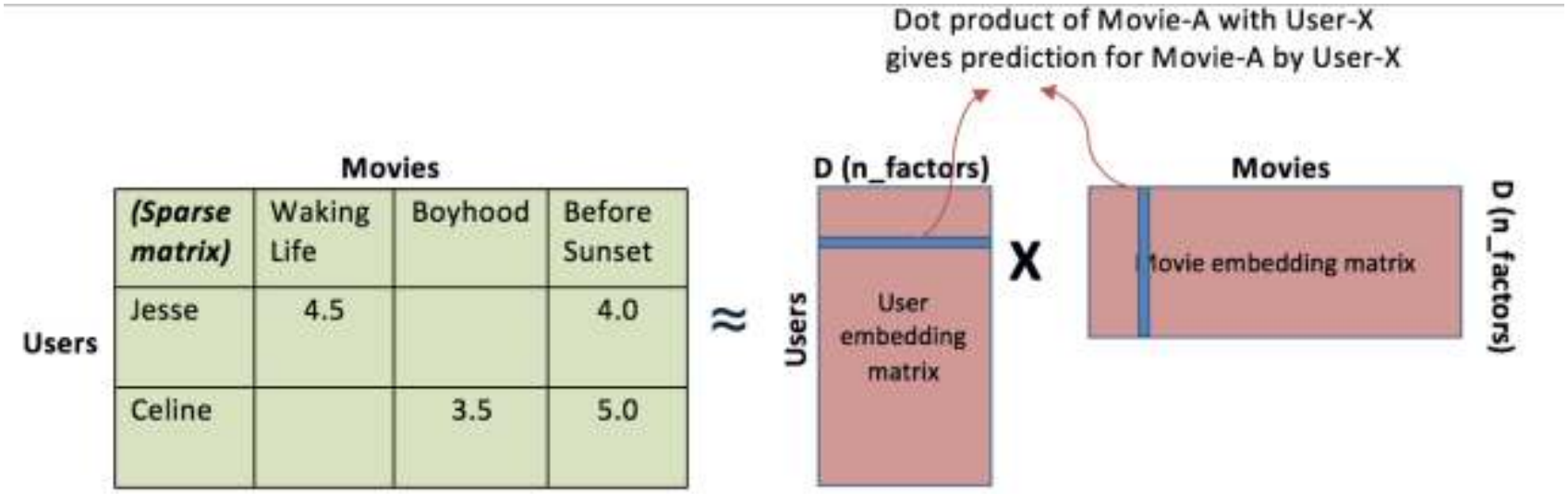
$$\min_{\boldsymbol{U}_p} \sum_{t=1}^{T} \|(\boldsymbol{U}_p\boldsymbol{U}_p^\top - \boldsymbol{I})\boldsymbol{x}_t\|^2 \quad \text{subject to} \quad \boldsymbol{U}_p^\top\boldsymbol{U}_p = \boldsymbol{I}$$

is the matrix of the p principal eigenvectors of $\frac{1}{T}\boldsymbol{X}\boldsymbol{X}^\top = \hat{\boldsymbol{C}}_x$, and the minimum cost is:

$$(\boldsymbol{U}^\star)^H\hat{\boldsymbol{C}}_x\boldsymbol{U}^\star = \text{diag}(\lambda_1, \cdots, \lambda_p).$$

Modeling graph signals' low rank property

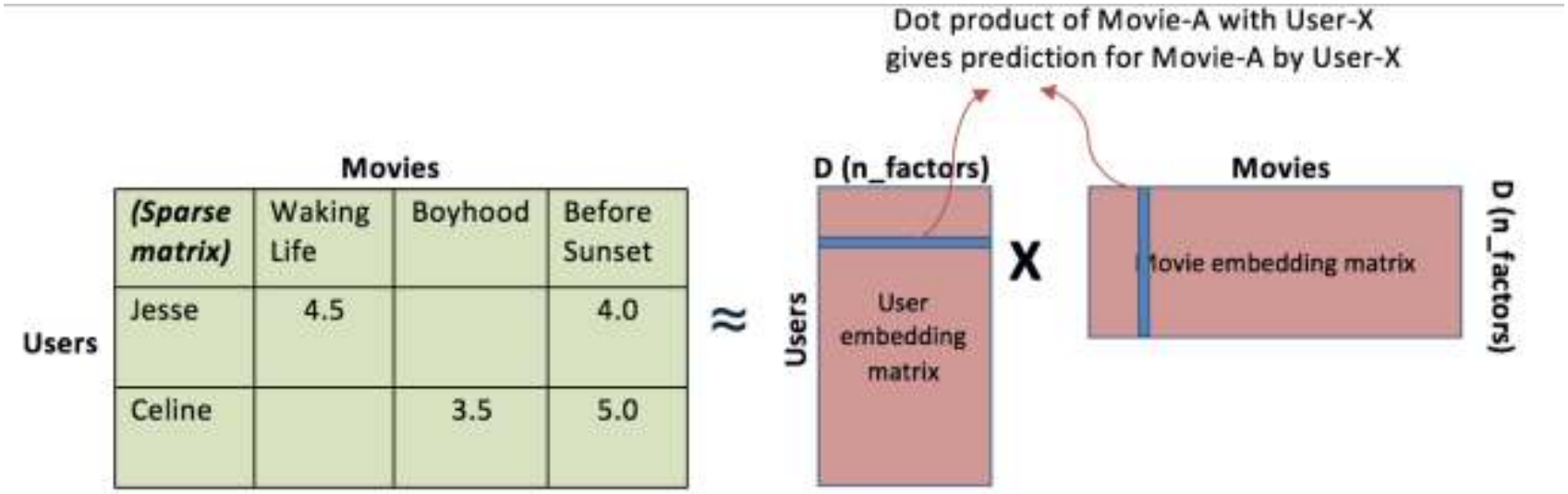
- PCA provides a method for dimensionality reduction and also for interpolation. For instance in collaborative filtering:



- When we model the data structure as $x(t) = \mathcal{H}(\mathrm{S})s$ the low rank structure must emerge from $\mathcal{H}(\mathrm{S})$. This is what we want to discuss next.

Modeling graph signals' low rank property

- PCA provides a method for dimensionality reduction and also for interpolation. For instance in collaborative filtering:



- When we model the data structure as $\boldsymbol{x}(t) = \mathcal{H}(\mathbf{S})\boldsymbol{s}$ the low rank structure must emerge from $\mathcal{H}(\mathbf{S})$. This is what we want to discuss next.

Modeling graph signals' low rank property

Recall that the GF

$$\mathcal{H}(\mathbf{S}) = \sum_{\ell=1}^{L} h_\ell \mathbf{S}^\ell$$

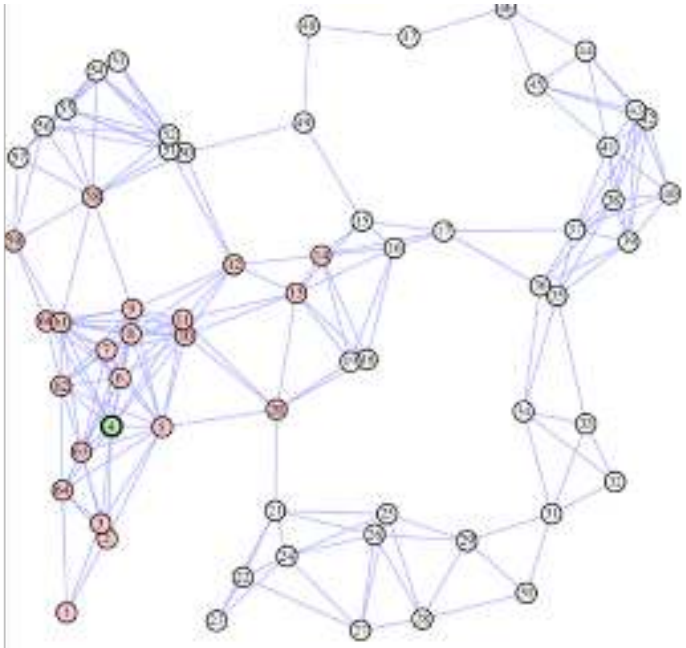
has a GFT response

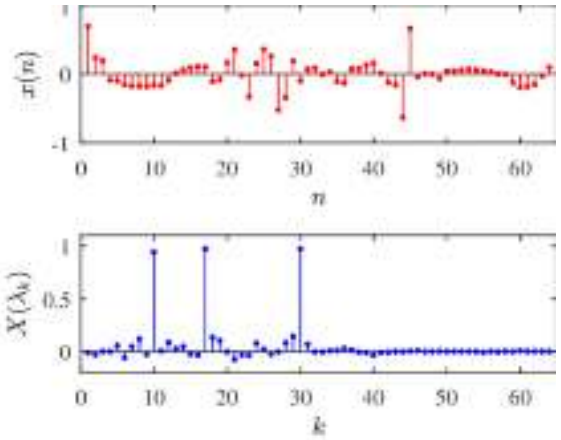
$$\tilde{\boldsymbol{h}}: \quad [\tilde{\boldsymbol{h}}]_i = \tilde{h}_i = \sum_{i=1}^{N} h_k \lambda_i^\ell$$

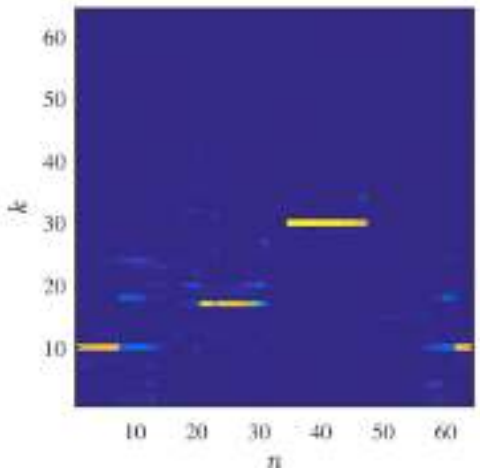
that shapes the GFT spectrum through the multiplicative property:

$$\boldsymbol{x} = \mathcal{H}(\mathbf{S})\boldsymbol{s}$$

$$\Leftrightarrow \quad \tilde{\boldsymbol{x}} = \mathrm{diag}(\tilde{\boldsymbol{h}})\tilde{\boldsymbol{s}}$$







Modeling graph signals' low rank property

- More specifically, where $|\tilde{h}_i| \ll 1$ it is expected that $|\tilde{x}_i| \ll 1$
- As a result, graph filter models can be classified as either low-pass, band-pass, or high-pass, or having a sparse support, depending on its graph frequency response, as we will see next.

GSP point of view

The vector found through PCA are, in fact, GFT components, i.e. eigenvectors of the GSO $\boldsymbol{S}$.

Low-rank property

- Let us assume that we have independent and identically distributed *random experiments*:

$$\boldsymbol{x}_t = \mathcal{H}(\mathbf{S})\, \boldsymbol{s}_t + \boldsymbol{w}_t \quad t = 1, \ldots, T,$$

- Let $\boldsymbol{s}_\ell$, $\boldsymbol{w}_\ell$ are zero-mean white noise with $\mathbb{E}[\boldsymbol{s}_\ell \boldsymbol{s}_\ell^\top] = \boldsymbol{I}$, $\mathbb{E}[\boldsymbol{w}_\ell \boldsymbol{w}_\ell^\top] = \sigma^2 \boldsymbol{I}$. (This is not necessary but makes life easier)
- The first obvious property is that:

$$\hat{C}_x \approx \mathbb{E}[\boldsymbol{x}_t \boldsymbol{x}_t^\top] = \mathcal{H}(\mathbf{S})\mathcal{H}^\top(\mathbf{S}) + \sigma^2 \boldsymbol{I} \equiv \boldsymbol{U}\, \mathrm{diag}(\tilde{\boldsymbol{h}})^2 \boldsymbol{U}^\top + \sigma^2 \boldsymbol{I}.$$

  - **Low-pass graph signals** have the the first $\tilde{h}_1, \ldots, \tilde{h}_k$ which dominate (corresponding to the small $\lambda_1, \ldots, \lambda_k$)
  - **High-pass graph signals** have $\tilde{h}_{N-k}, \ldots, \tilde{h}_N$ which dominate
  - **Sparse spectrum graph signals** have a certain subset $\mathcal{K}$ of the entries of $\tilde{\boldsymbol{h}}$ which dominate.

Low-rank property

- Let us assume that we have independent and identically distributed *random experiments*:

$$\boldsymbol{x}_t = \mathcal{H}(\mathbf{S})\, \boldsymbol{s}_t + \boldsymbol{w}_t \quad t = 1, \ldots, T,$$

- Let $\boldsymbol{s}_\ell$, $\boldsymbol{w}_\ell$ are zero-mean white noise with $\mathbb{E}[\boldsymbol{s}_\ell \boldsymbol{s}_\ell^\top] = \boldsymbol{I}$, $\mathbb{E}[\boldsymbol{w}_\ell \boldsymbol{w}_\ell^\top] = \sigma^2 \boldsymbol{I}$. (This is not necessary but makes life easier)
- The first obvious property is that:

$$\hat{C}_x \approx \mathbb{E}[\boldsymbol{x}_t \boldsymbol{x}_t^\top] = \mathcal{H}(\mathbf{S})\mathcal{H}^\top(\mathbf{S}) + \sigma^2 \boldsymbol{I} \equiv \boldsymbol{U}\, \mathrm{diag}(\tilde{\boldsymbol{h}})^2 \boldsymbol{U}^\top + \sigma^2 \boldsymbol{I}.$$

  - **Low-pass graph signals** have the the first $\tilde{h}_1, \ldots, \tilde{h}_k$ which dominate (corresponding to the small $\lambda_1, \ldots, \lambda_k$)
  - **High-pass graph signals** have $\tilde{h}_{N-k}, \ldots, \tilde{h}_N$ which dominate
  - **Sparse spectrum graph signals** have a certain subset $\mathcal{K}$ of the entries of $\tilde{\boldsymbol{h}}$ which dominate.

Low-rank property

- Let us assume that we have independent and identically distributed *random experiments*:

$$\boldsymbol{x}_t = \mathcal{H}(\mathbf{S})\, \boldsymbol{s}_t + \boldsymbol{w}_t \quad t = 1, \ldots, T,$$

- Let $\boldsymbol{s}_\ell$, $\boldsymbol{w}_\ell$ are zero-mean white noise with $\mathbb{E}[\boldsymbol{s}_\ell \boldsymbol{s}_\ell^\top] = \boldsymbol{I}$, $\mathbb{E}[\boldsymbol{w}_\ell \boldsymbol{w}_\ell^\top] = \sigma^2 \boldsymbol{I}$. (This is not necessary but makes life easier)
- The first obvious property is that:

$$\hat{\boldsymbol{C}}_x \approx \mathbb{E}[\boldsymbol{x}_t \boldsymbol{x}_t^\top] = \mathcal{H}(\mathbf{S})\mathcal{H}^\top(\mathbf{S}) + \sigma^2 \boldsymbol{I} \equiv \boldsymbol{U}\text{diag}(\tilde{\boldsymbol{h}})^2 \boldsymbol{U}^\top + \sigma^2 \boldsymbol{I}.$$

  - **Low-pass graph signals** have the the first $\tilde{h}_1, \ldots, \tilde{h}_k$ which dominate (corresponding to the small $\lambda_1, \ldots, \lambda_k$)
  - **High-pass graph signals** have $\tilde{h}_{N-k}, \ldots, \tilde{h}_N$ which dominate
  - **Sparse spectrum graph signals** have a certain subset $\mathcal{K}$ of the entries of $\tilde{\boldsymbol{h}}$ which dominate.

GFT and PCA

- Since PCA yields the dominant eigenvectors of $\hat{\boldsymbol{C}}_x$, then it is clear that those are the same as the Graph Frequencies $\boldsymbol{U}_{\mathcal{K}}$ corresponding to the dominant entries of $\tilde{\boldsymbol{h}}$
- If it is a low-pass signal, then the low GFT frequencies will dominate: they are the *least significant eigenvectors* of the GSO $\mathbf{S}$, i.e. they correspond to $0 = \lambda_1 \leq \lambda_2 \ldots \leq \lambda_k$
  - Recall, the lowest eigenvectors of the GSP are those we use for spectral clustering!

GFT and PCA relation

Since the PCA basis is one to one with GFT components, if we know $\mathbf{S}$ we know what the principal components are and, if we do not know $\mathbf{S}$ we know how they relate to the graph.

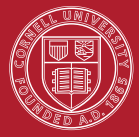
GFT and PCA

- Since PCA yields the dominant eigenvectors of $\hat{\boldsymbol{C}}_x$, then it is clear that those are the same as the Graph Frequencies $\boldsymbol{U}_{\mathcal{K}}$ corresponding to the dominant entries of $\tilde{\boldsymbol{h}}$
- If it is a low-pass signal, then the low GFT frequencies will dominate: they are the *least significant eigenvectors* of the GSO $\mathbf{S}$, i.e. they correspond to $0 = \lambda_1 \leq \lambda_2 \ldots \leq \lambda_k$
  - Recall, the lowest eigenvectors of the GSP are those we use for spectral clustering!

GFT and PCA relation

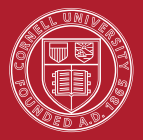
Since the PCA basis is one to one with GFT components, if we know $\mathbf{S}$ we know what the principal components are and, if we do not know $\mathbf{S}$ we know how they relate to the graph.

Outline



Cornell University

Sampling theory from GSP



Cornell University

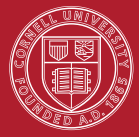
- **Basic intuition**: Assume $\tilde{\boldsymbol{x}}_t$ is dominated by the subset of components $\mathcal{K}$ that are the entries of $[\tilde{\boldsymbol{x}}_t]_{\mathcal{K}}$. Then

$$\hat{\boldsymbol{x}}_t = \mathbf{U}_{\mathcal{K}}[\tilde{\boldsymbol{x}}_t]_{\mathcal{K}} \approx \boldsymbol{x}_t,$$

  where $\mathbf{U}_{\mathcal{K}}$ contains the the subset $\mathcal{K}$ of columns from the GFT basis and $[\tilde{\boldsymbol{x}}_t]_{\mathcal{K}}$ contains the entries of $\tilde{\boldsymbol{x}}_t$ in the set of indexes $\mathcal{K}$.

- In the case of lowpass graph signals $\mathcal{K} = \{1, \ldots, k\}$
- The GSP sampling theorem simply states that, if the graph signal is a linear combinatio of relatively few GFT components, then we can decimate x_t over a subset of entries (samples) $\mathcal{M} \subset \mathcal{V}$ and it is possible to obtain good reconstruction accuracy as long as $|\mathcal{K}| \leq \mathcal{M}$.

Sampling theory from GSP



Cornell University

- **Basic intuition**: Assume $\tilde{\boldsymbol{x}}_t$ is dominated by the subset of components $\mathcal{K}$ that are the entries of $[\tilde{\boldsymbol{x}}_t]_{\mathcal{K}}$. Then

$$\hat{\boldsymbol{x}}_t = \mathbf{U}_{\mathcal{K}}[\tilde{\boldsymbol{x}}_t]_{\mathcal{K}} \approx \boldsymbol{x}_t,$$

  where $\mathbf{U}_{\mathcal{K}}$ contains the the subset $\mathcal{K}$ of columns from the GFT basis and $[\tilde{\boldsymbol{x}}_t]_{\mathcal{K}}$ contains the entries of $\tilde{\boldsymbol{x}}_t$ in the set of indexes $\mathcal{K}$.

- In the case of lowpass graph signals $\mathcal{K} = \{1, \ldots, k\}$

- The GSP sampling theorem simply states that, if the graph signal is a linear combinatio of relatively few GFT components, then we can decimate x_t over a subset of entries (samples) $\mathcal{M} \subset \mathcal{V}$ and it is possible to obtain good reconstruction accuracy as long as $|\mathcal{K}| \leq \mathcal{M}$.

Sampling theory from GSP

Cornell University

- **Basic intuition**: Assume $\tilde{\boldsymbol{x}}_t$ is dominated by the subset of components $\mathcal{K}$ that are the entries of $[\tilde{\boldsymbol{x}}_t]_{\mathcal{K}}$. Then

$$\hat{\boldsymbol{x}}_t = \mathbf{U}_{\mathcal{K}}[\tilde{\boldsymbol{x}}_t]_{\mathcal{K}} \approx \boldsymbol{x}_t,$$

  where $\mathbf{U}_{\mathcal{K}}$ contains the the subset $\mathcal{K}$ of columns from the GFT basis and $[\tilde{\boldsymbol{x}}_t]_{\mathcal{K}}$ contains the entries of $\tilde{\boldsymbol{x}}_t$ in the set of indexes $\mathcal{K}$.

- In the case of lowpass graph signals $\mathcal{K} = \{1, \ldots, k\}$
- The GSP sampling theorem simply states that, if the graph signal is a linear combinatio of relatively few GFT components, then we can decimate $\boldsymbol{x}_t$ over a subset of entries (samples) $\mathcal{M} \subset \mathcal{V}$ and it is possible to obtain good reconstruction accuracy as long as $|\mathcal{K}| \leq \mathcal{M}$.

Sampling theory from GSP

Cornell University

Suppose $\boldsymbol{x}_t \approx \hat{\boldsymbol{x}}_t = \mathbf{U}_{\mathcal{K}}[\tilde{\boldsymbol{x}}_t]_{\mathcal{K}} = \mathbf{U}_{\mathcal{K}}\mathbf{U}_{\mathcal{K}}^{\top}\boldsymbol{x}_t$. Note that the GFT of $\hat{\boldsymbol{x}}$ is zero for all graph frequencies $i \notin \mathcal{K}$.

Vertex-limiting (sampling) operator

Measurements from nodes $i \in \mathcal{M}$ $[\boldsymbol{x}_t]_{\mathcal{M}} = \boldsymbol{P}_{\mathcal{M}}^{\top}\boldsymbol{x}_t$

Vertex-limiting operator: $\boldsymbol{\mathcal{D}}_{\mathcal{M}} = \boldsymbol{P}_{\mathcal{M}}\boldsymbol{P}_{\mathcal{M}}^{\top}$

- Note: $[\boldsymbol{x}_t]_{\mathcal{M}} = \boldsymbol{P}_{\mathcal{M}}^{\top}\boldsymbol{x}_t \approx \boldsymbol{P}_{\mathcal{M}}^{\top}\boldsymbol{U}_{\mathcal{K}}[\tilde{\boldsymbol{x}}_t]_{\mathcal{K}} \quad \rightarrow \quad [\tilde{\boldsymbol{x}}_t]_{\mathcal{K}} = \left(\boldsymbol{P}_{\mathcal{M}}^{\top}\boldsymbol{U}_k\right)^{\dagger}[\boldsymbol{x}_t]_{\mathcal{M}}$
- The reconstruction formula is:

$$\hat{\boldsymbol{x}}_t = \boldsymbol{U}_{\mathcal{K}}\left(\boldsymbol{P}_{\mathcal{M}}^{\top}\boldsymbol{U}_{\mathcal{K}}\right)^{\dagger}[\boldsymbol{x}_t]_{\mathcal{M}} \tag{4}$$

$|\mathcal{M}| \geq |\mathcal{K}|$ is a necessary condition and the optimal sampling pattern choose rows of $\mathbf{U}_{\mathcal{K}}$ so as to maximize the smallest singular value, $\sigma_{\min}\left(\boldsymbol{\mathcal{D}}_{\mathcal{M}}\mathbf{U}_{\mathcal{K}}\right)$.

Interpolation w.o. GSP: Matrix completion

- Let $\boldsymbol{X}^{\text{samp}}$ be a matrix that contains the set of entries of the vectors $\boldsymbol{x}_t$ that are known $\left\{[\boldsymbol{x}_1]_{\mathcal{M}_1}, \ldots, [\boldsymbol{x}_T]_{\mathcal{M}_T}\right\}$ and zeros when they are missing.
- Assume we know tend to be well approximated in a subspace of dimension $p > n$, even though it is not exactly p-dimensional

Matrix Completion Problem

Let $\|\boldsymbol{X}\|^\star$ denote the nuclear norm, i.e. the sum of the magnitudes of the singular values of $\boldsymbol{X}$

$$\min_{\boldsymbol{X} \in \mathbb{R}^{n \times m}} \|\mathcal{M}(\boldsymbol{X}) - \boldsymbol{X}^{\text{samp}}\|_F^2 + \gamma \left\{ \|\boldsymbol{X}\|^\star + \sum_{t=2}^{m} \|\boldsymbol{x}_t - \boldsymbol{x}_{t-1}\|_2^2 \right\}.$$

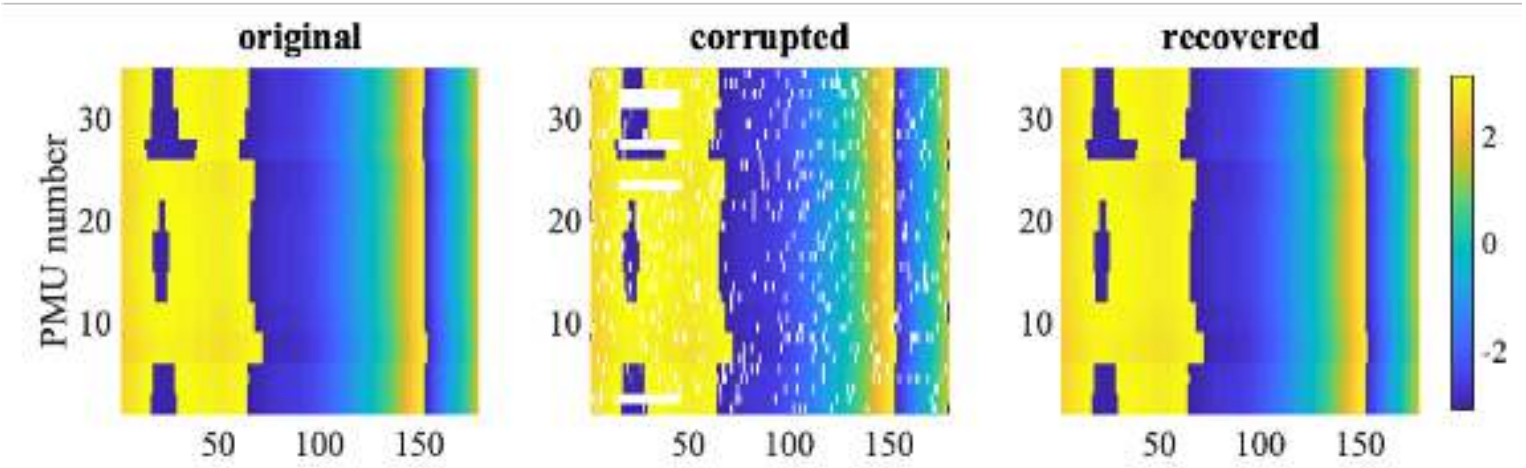
- The regularization term $\{\|\boldsymbol{X}\|^\star + \sum_{t=2}^{m} \|\boldsymbol{x}_t - \boldsymbol{x}_{t-1}\|_2^2\}$ tends to promote having just a few dominant singular values

Smoothness and interpolation with GSP

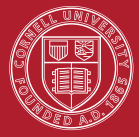
- For Graph-Temporal processes that are Low-pass using the GSO yields better results,:

$$\min_{\boldsymbol{X}\in\mathbb{R}^{n\times m}} \|\mathcal{M}(\boldsymbol{X})-\boldsymbol{X}^{\text{samp}}\|_F^2+\gamma\left\{\sum_{t=1}^{T}\boldsymbol{x}_t^\top\mathbf{S}\boldsymbol{x}_t+\sum_{t=2}^{m}\|\boldsymbol{x}_t-\boldsymbol{x}_{t-1}\|_2^2\right\}.$$

- The regularization $\sum_{t=1}^{T}\boldsymbol{x}_t^\top\mathbf{S}\boldsymbol{x}_t$ seeks to reduce spatial variations main benefit of this approach is that it can recover entirely missing rows of data, unlike matrix completion

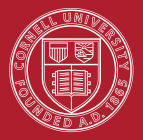


Outline



Cornell University

Smoothness and Graph Learning



Cornell University

- For low-pass graph signals with independent samples, in general we can formulate the following problem. Graph-Topology learning formulation:

$$\begin{aligned} \min_{\boldsymbol{z}_t, t=1,\dots,t, \mathbf{S}} \quad & \frac{1}{m} \sum_{t=1}^{T} \left(\frac{1}{\sigma^2} \|\boldsymbol{z}_t - \boldsymbol{x}_t\|_2^2 + \mu \boldsymbol{z}_t^\top \mathbf{S} \boldsymbol{z}_t \right) \\ \text{s.t.} \quad & S_{ji} = S_{ij} \leq 0, \ \forall \ i \neq j, \ \mathbf{S1} = \mathbf{0}, \end{aligned}$$

  finding the Laplacian that best smoothens the signal ($\boldsymbol{z}_t$ is a smooth signal close to the observation $\boldsymbol{x}_t$ which may be noisy).

- Note that if the previous iterpolation problem is solved w.r.t. both $\boldsymbol{X}$ and $\mathbf{S}$ we obtain a similar formulation, where also temporal smoothness is accounted for

$$\begin{aligned} \min_{\boldsymbol{Z} \in \mathbb{R}^{n \times m}, \mathbf{S}} \quad & \|\mathcal{M}(\boldsymbol{Z}) - \boldsymbol{X}^{\text{samp}}\|_F^2 + \gamma \left(\sum_{t=1}^{T} \boldsymbol{z}_t^\top \mathbf{S} \boldsymbol{z}_t + \sum_{t=2}^{m} \|\boldsymbol{z}_t - \boldsymbol{z}_{t-1}\|_2^2 \right) \\ \text{s.t.} \quad & S_{ji} = S_{ij} \leq 0, \ \forall \ i \neq j, \ \mathbf{S1} = \mathbf{0}, \end{aligned}$$

Smoothness and Graph Learning

- For low-pass graph signals with independent samples, in general we can formulate the following problem. Graph-Topology learning formulation:

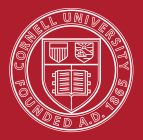
$$\begin{aligned} \min_{\boldsymbol{z}_t, t=1,\dots,t, \mathbf{S}} \quad & \frac{1}{m} \sum_{t=1}^{T} \left(\frac{1}{\sigma^2} \|\boldsymbol{z}_t - \boldsymbol{x}_t\|_2^2 + \mu \boldsymbol{z}_t^\top \mathbf{S} \boldsymbol{z}_t \right) \\ \text{s.t.} \quad & S_{ji} = S_{ij} \leq 0, \ \forall \ i \neq j, \ \mathbf{S1} = \mathbf{0}, \end{aligned}$$

  finding the Laplacian that best smoothens the signal ($\boldsymbol{z}_t$ is a smooth signal close to the observation $\boldsymbol{x}_t$ which may be noisy).

- Note that if the previous iterpolation problem is solved w.r.t. both $\boldsymbol{X}$ and $\mathbf{S}$ we obtain a similar formulation, where also temporal smoothness is accounted for

$$\begin{aligned} \min_{\boldsymbol{Z} \in \mathbb{R}^{n \times m}, \mathbf{S}} \ & \|\mathcal{M}(\boldsymbol{Z}) - \boldsymbol{X}^{\text{samp}}\|_F^2 + \gamma \left(\sum_{t=1}^{T} \boldsymbol{z}_t^\top \mathbf{S} \boldsymbol{z}_t + \sum_{t=2}^{m} \|\boldsymbol{z}_t - \boldsymbol{z}_{t-1}\|_2^2 \right) \\ \text{s.t.} \quad & S_{ji} = S_{ij} \leq 0, \ \forall \ i \neq j, \ \mathbf{S1} = \mathbf{0}, \end{aligned}$$

Blind community detection (CD)



Cornell University

- We know that the principal components of $\boldsymbol{C}_y$ will be dominated by a rank-k component spanned by $\boldsymbol{U}_k$ which are the least significant eigenvectors of the Laplacian

Blind CD algorithm

Upon observing m low-pass graph signals:

(i) find the top-k $\widehat{\boldsymbol{U}}_k \in \mathbb{R}^{n \times k}$ of $\widehat{\boldsymbol{C}}_y = \frac{1}{m} \sum_{\ell=1}^{m} \boldsymbol{y}_\ell \boldsymbol{y}_\ell^\top$;
(ii) apply k-means on the rows of $\widehat{\boldsymbol{U}}_k$.

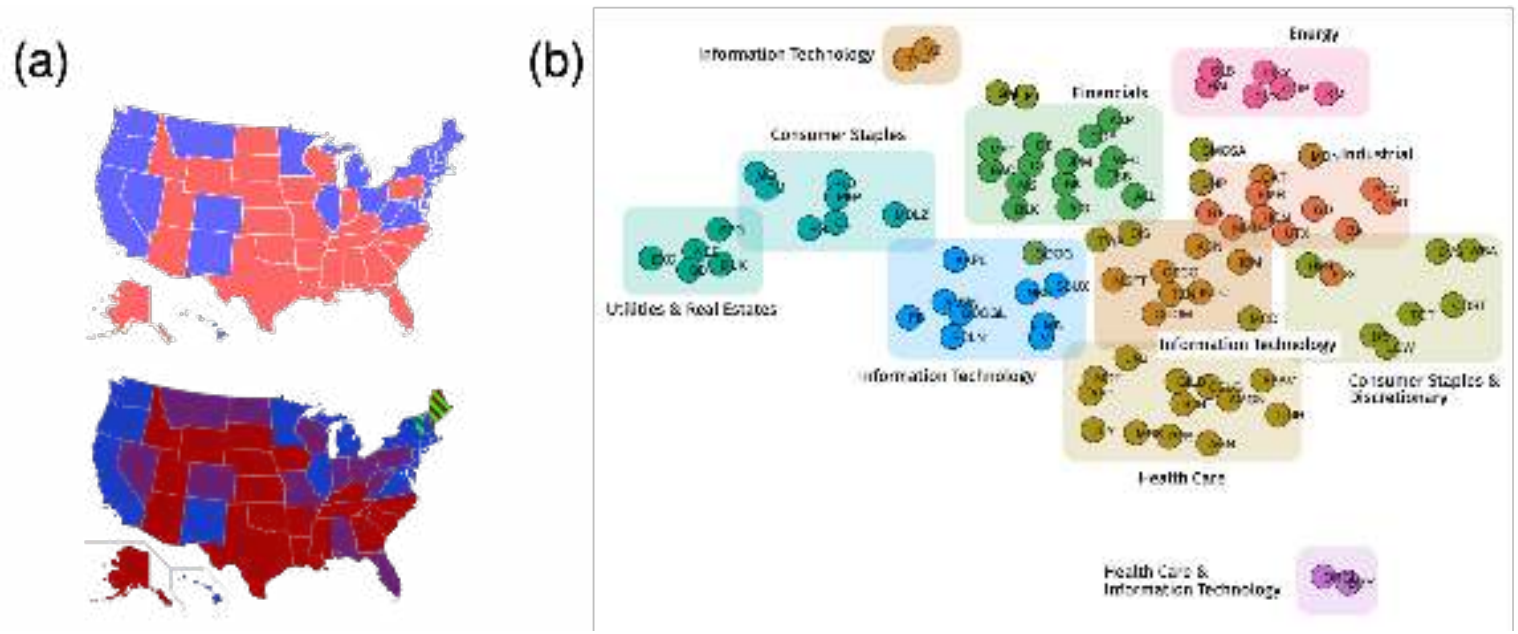
- If we denote the detected communities as $\widehat{\mathcal{N}}_1, \ldots, \widehat{\mathcal{N}}_k$, then

$$F(\widehat{\mathcal{N}}_1, \ldots, \widehat{\mathcal{N}}_k; \boldsymbol{U}_k) - F^\star = \mathcal{O}(\eta_k + \sigma + m^{-1/2}).$$

- In other words, the BlindCD approaches the performance of SC if the graph filter is k-low-pass with $\eta_k \ll 1$, the observation noise σ and the observation noise σ is small.

Blind CD in Opinion Dynamics

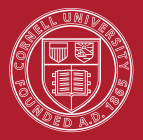
Cornell University



(a) US Senate from the 115th US Congress (2017-2019) $m = 502$ voting rounds - grouping senators from each state.

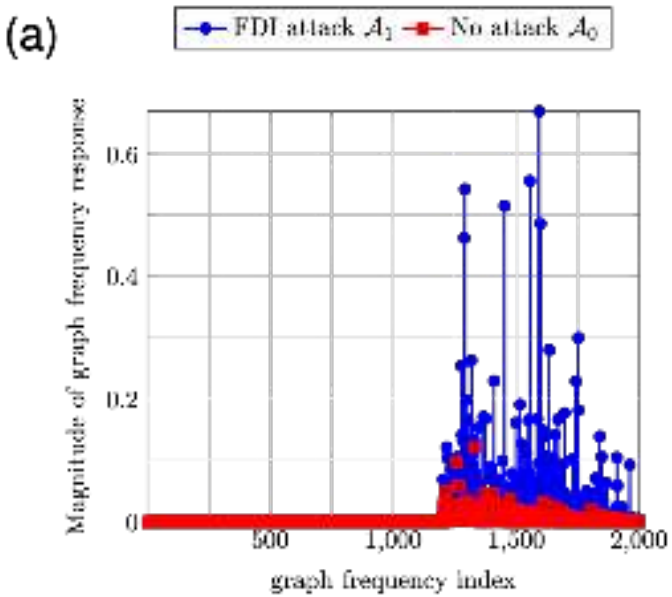
(b) Daily return data of stocks in the S&P100 index from Feb. 2013.

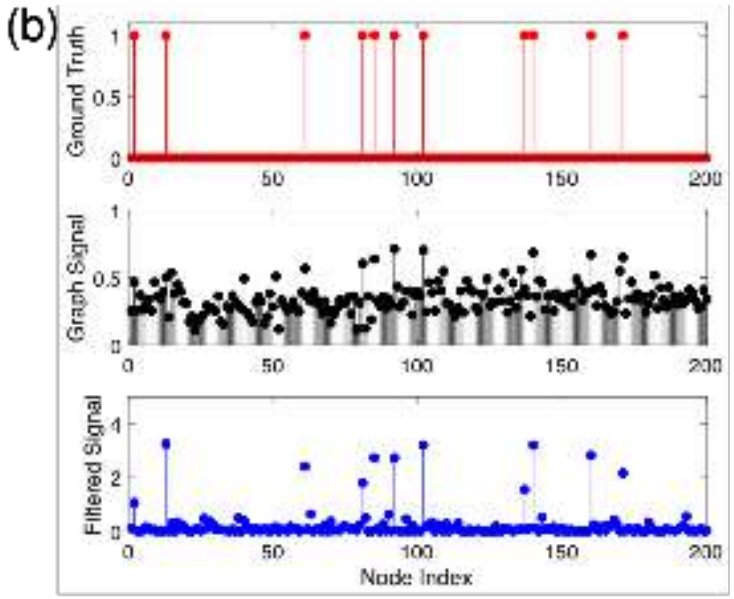
Anomaly detection with Low-pass GSP



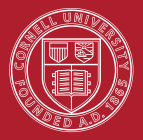
- We can cast the anomaly detection as a canonical hypothesis test:
  - Null Hypothesis $\mathcal{A}_0$: The high frequency components of the graph signal are small
  - Positive Hypothesis $\mathcal{A}_1$: The high frequency graph signal components are large due to a sudden change or inconsistent data

$$\boldsymbol{x}_\ell = \begin{cases} \mathcal{H}(\mathbf{S})\boldsymbol{s}_\ell, & \text{under } \mathcal{A}_0, \\ \mathcal{H}(\mathbf{S})\boldsymbol{s}_\ell + \boldsymbol{w}_\ell & \text{under } \mathcal{A}_1, \end{cases} \quad \Rightarrow \quad \Gamma_\ell = \|\mathcal{H}_{\mathsf{HPF}}(\mathbf{S})\boldsymbol{x}_\ell\| \underset{\mathcal{A}_1}{\overset{\mathcal{A}_0}{\lessgtr}} \delta.$$





Conclusion



- We introduced a class of unsupervised algorithms that used data to make inferences and are based on the GSP framework.
  - Least square reconstruction of graph signals from sparse nodal measurements and optimum sampling design
  - Matrix completion versus regularized reconstruction of low-pass graph temporal signals
  - Joint network inference and data reconstruction
  - Community detection from graph signals
  - Anomaly detection
- Next we discuss how to perform supervised inference with Graph Convolutional Neural Networks