Space-Shift Sampling of Graph Signals

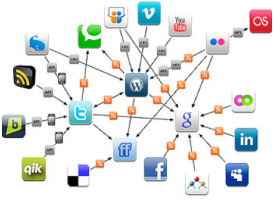
Santiago Segarra, Antonio G. Marques, Geert Leus, and Alejandro Ribeiro

Dept. of Electrical and Systems Engineering
University of Pennsylvania
ssegarra@seas.upenn.edu
http://www.seas.upenn.edu/~ssegarra/

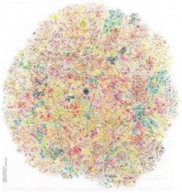
ICASSP, March 23, 2016

Network science analytics

Online social media



Internet

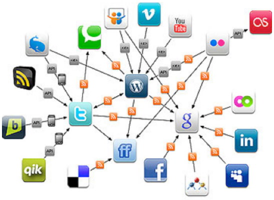


Clean energy and grid analytics

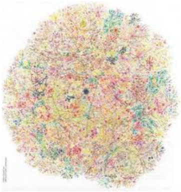
- Desiderata: Process, analyze and learn from network data [Kolaczyk'09]

Network science analytics

Online social media



Internet

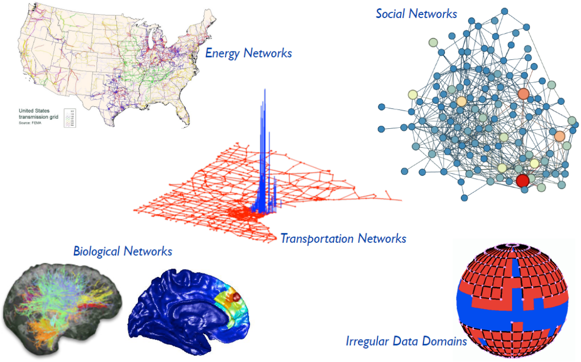


Clean energy and grid analytics

- Desiderata: Process, analyze and learn from network data [Kolaczyk'09]
- Network as graph $G = (\mathcal{V}, \mathcal{E})$: encode pairwise relationships
- Interest here not in G itself, but in data associated with nodes in $\mathcal{V}$

  $\Rightarrow$ The object of study is a graph signal
- Ex: Opinion profile, buffer congestion levels, neural activity, epidemic

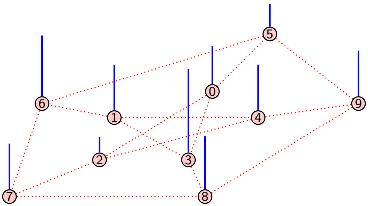
Motivating examples – Graph signals



- Graph SP: broaden classical SP to graph signals [Shuman etal'13]
  - $\Rightarrow$ Our view: GSP well suited to study network processes
- **As.:** Signal properties related to topology of G (e.g., smoothness)
  - $\Rightarrow$ Algorithms that fruitfully leverage this relational structure

Graph signals

- Consider a graph $G(\mathcal{V}, \mathcal{E})$. Graph signals are mappings $x : \mathcal{V} \to \mathbb{R}$
  - $\Rightarrow$ Defined on the vertices of the graph (data tied to nodes)
- May be represented as a vector $\mathbf{x} \in \mathbb{R}^N$
  - $\Rightarrow$ x_n denotes the signal value at the n-th vertex in $\mathcal{V}$
  - $\Rightarrow$ Implicit ordering of vertices

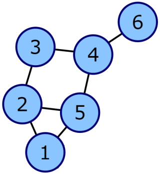


$$\mathbf{x} = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_9 \end{bmatrix} = \begin{bmatrix} 0.6 \\ 0.7 \\ 0.3 \\ \vdots \\ 0.7 \end{bmatrix}$$

Graph-shift operator

- To understand and analyze $\mathbf{x}$, useful to account for G's structure
- Graph G is endowed with a graph-shift operator $\mathbf{S} \in \mathbb{R}^{N \times N}$

  $\Rightarrow S_{ij} = 0$ for $i \neq j$ and $(i, j) \notin \mathcal{E}$ (captures local structure in G)
- $\mathbf{S}$ can take nonzero values in the edges of G or in its diagonal



$$\mathbf{S} = \begin{pmatrix} S_{11} & S_{12} & 0 & 0 & S_{15} & 0 \\ S_{21} & S_{22} & S_{23} & 0 & S_{25} & 0 \\ 0 & S_{23} & S_{33} & S_{34} & 0 & 0 \\ 0 & 0 & S_{43} & S_{44} & S_{45} & S_{46} \\ S_{51} & S_{52} & 0 & S_{54} & S_{55} & 0 \\ 0 & 0 & 0 & S_{64} & 0 & S_{66} \end{pmatrix}$$

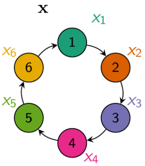
- Ex: Adjacency $\mathbf{A}$, degree $\mathbf{D}$, and Laplacian $\mathbf{L} = \mathbf{D} - \mathbf{A}$ matrices

Relevance of the graph-shift operator

- Q: Why is $\mathbf{S}$ called shift?

Relevance of the graph-shift operator

- Q: Why is $\mathbf{S}$ called shift? A: Resemblance to time shifts



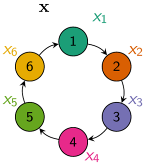
Set $\mathbf{S} = \mathbf{A_{dc}}$

What is $\mathbf{Sx}$?

$$\begin{pmatrix} 0 \\ 0 \\ x_2 \\ x_3 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ x_2 \\ x_3 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Relevance of the graph-shift operator

- Q: Why is **S** called shift? A: Resemblance to time shifts

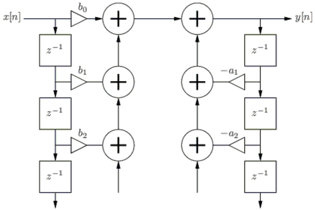


Set $\mathbf{S} = \mathbf{A_{dc}}$

What is $\mathbf{Sx}$?

$$\begin{pmatrix} 0 \\ 0 \\ x_2 \\ x_3 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ x_2 \\ x_3 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

- **S** will be building block for GSP algorithms
  - $\Rightarrow$ Same is true in the time domain (filters and delay)



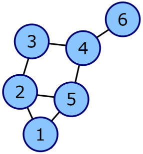
Locality of the graph-shift operator

- $\mathbf{S}$ is a linear operator that can be computed locally at the nodes in $\mathcal{V}$
- Consider the graph signal $\mathbf{y} = \mathbf{S}\mathbf{x}$ and node i's neighborhood $\mathcal{N}_i$

  $\Rightarrow$ Node i can compute y_i if it has access to x_j at $j \in \mathcal{N}_i$

$$y_i = \sum_{j \in \mathcal{N}_i} S_{ij} x_j, \quad i \in \mathcal{V}$$

- Recall $S_{ij} \neq 0$ only if $i = j$ or $(j, i) \in \mathcal{E}$



$$\Rightarrow \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{pmatrix} = \begin{pmatrix} S_{11} & S_{12} & 0 & 0 & S_{15} & 0 \\ S_{21} & S_{22} & S_{23} & 0 & S_{25} & 0 \\ 0 & S_{32} & S_{33} & S_{34} & 0 & 0 \\ 0 & 0 & S_{43} & S_{44} & S_{45} & S_{46} \\ S_{51} & S_{52} & 0 & S_{54} & S_{55} & 0 \\ 0 & 0 & 0 & S_{64} & 0 & S_{66} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{pmatrix}$$

- If $\mathbf{y} = \mathbf{S}^2\mathbf{x} \ \Rightarrow y_i$ found using values x_j within 2 hops

Graph Fourier transform (GFT)

- **As.:** $\mathbf{S}$ related to generation (description) of the signals of interest
  $\Rightarrow$ Spectrum of $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$ will be especially useful to analyze $\mathbf{x}$
- The Graph Fourier Transform (GFT) of $\mathbf{x}$ is defined as

$$\tilde{\mathbf{x}} = \mathbf{V}^{-1}\mathbf{x}$$

- While the inverse GFT (iGFT) of $\tilde{\mathbf{x}}$ is defined as

$$\mathbf{x} = \mathbf{V}\tilde{\mathbf{x}}$$

  $\Rightarrow$ Eigenvectors $\mathbf{V} = [\mathbf{v}_1, ..., \mathbf{v}_N]$ are the frequency basis (atoms)
- Ex: For the directed cycle (temporal signal) $\Rightarrow$ GFT $\equiv$ DFT
  $\Rightarrow$ DFT matrix diagonalizes circulant matrices like $\mathbf{S} = \mathbf{A}_{dc}$

Sampling of graph signals

- Sampling is a cornerstone inverse problem in classical SP
  - $\Rightarrow$ How to find $\mathbf{x} \in \mathbb{R}^N$ using $P < N$ observations?

Sampling of graph signals

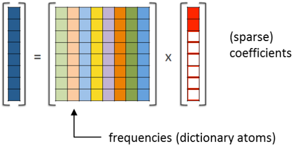
- Sampling is a cornerstone inverse problem in classical SP
  - $\Rightarrow$ How to find $\mathbf{x} \in \mathbb{R}^N$ using $P < N$ observations?
- Our focus on bandlimited signals, but other models possible

$\Rightarrow$ $\tilde{\mathbf{x}} = \mathbf{V}^{-1}\mathbf{x}$ sparse

$\Rightarrow$ $\mathbf{x} = \sum_{k \in \mathcal{K}} \tilde{x}_k \mathbf{v}_k$, with $|\mathcal{K}| = K < N$

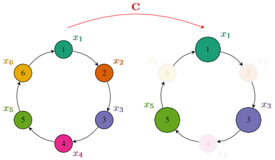
$\Rightarrow$ $\mathbf{S}$ involved in generation of $\mathbf{x}$

$\Rightarrow$ Agnostic to the particular form of $\mathbf{S}$

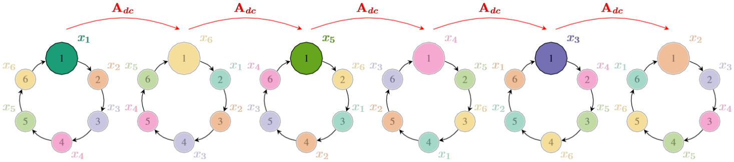


- Two sampling schemes were introduced in the literature
  - $\Rightarrow$ Selection [Anis14, Chen15, Tsitsvero15, Puy15, Wang15]
  - $\Rightarrow$ Aggregation [Marques15, Segarra15]
- We combine both to create a hybrid scheme $\Rightarrow$ Space-shift sampling

- There are two ways of interpreting sampling of time signals
- We can either freeze the signal and sample values at different times



- We can fix a point (present) and sample the evolution of the signal

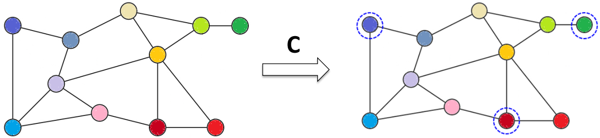


- Both strategies coincide for time signals but not for general graphs
  - $\Rightarrow$ Give rise to selection and aggregation sampling

Selection sampling

- Intuitive extension of sampling to graph signals
  - $\Rightarrow$ Select a subset of the nodes and observe the signal value
  - $\Rightarrow$ Let $\mathbf{C} \in \{0,1\}^{P \times N}$ be a selection matrix (P rows of $\mathbf{I}_N$)

$$\bar{\mathbf{x}} = \mathbf{C}\mathbf{x}$$



- Goal: recover $\mathbf{x}$ based on $\bar{\mathbf{x}}$
  - $\Rightarrow$ Assume that the support of $\mathcal{K}$ is known (w.l.o.g. $\mathcal{K} = \{k\}_{k=1}^{K}$)
  - $\Rightarrow$ Since $\tilde{x}_k = 0$ for $k > K$, define $\tilde{\mathbf{x}}_K$ as (with $\mathbf{E}_K := [\mathbf{e}_1, ..., \mathbf{e}_K]$)

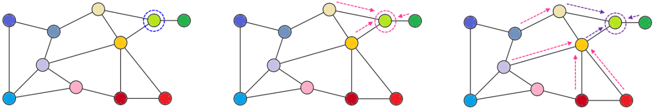
$$\tilde{\mathbf{x}}_K := [\tilde{x}_1, ..., \tilde{x}_K]^T = \mathbf{E}_K^T \tilde{\mathbf{x}}$$

- Use $\bar{\mathbf{x}}$ to find $\tilde{\mathbf{x}}_K$, and then recover $\mathbf{x}$ as

$$\mathbf{x} = \mathbf{V}_K \tilde{\mathbf{x}}_K = \mathbf{V}_K (\mathbf{C}\mathbf{V}_K)^{-1} \bar{\mathbf{x}}$$

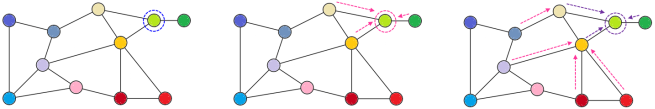
Aggregation sampling

- Idea: incorporating $\mathbf{S}$ to the sampling procedure
- Consider shifted (aggregated) signals $\mathbf{y}^{(l)} = \mathbf{S}^l\mathbf{x}$
  $\Rightarrow$ $\mathbf{y}^{(l)} = \mathbf{S}\mathbf{y}^{(l-1)}$ $\Rightarrow$ they can be found sequentially
  $\Rightarrow$ $S_{ij} = 0$ if $i \notin \mathcal{N}_j$ $\Rightarrow$ Only local exchanges are required
- Form signal $\mathbf{y}_i = [y_i^{(0)}, y_i^{(1)}, ..., y_i^{(N-1)}]^T$



Aggregation sampling

- Idea: incorporating $\mathbf{S}$ to the sampling procedure
- Consider shifted (aggregated) signals $\mathbf{y}^{(l)} = \mathbf{S}^l\mathbf{x}$
  $\Rightarrow \mathbf{y}^{(l)} = \mathbf{S}\mathbf{y}^{(l-1)} \Rightarrow$ they can be found sequentially
  $\Rightarrow S_{ij} = 0$ if $i \notin \mathcal{N}_j \Rightarrow$ Only local exchanges are required
- Form signal $\mathbf{y}_i = [y_i^{(0)}, y_i^{(1)}, ..., y_i^{(N-1)}]^T$



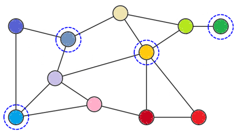
- Sampled signal $\bar{\mathbf{y}}_i = \mathbf{C}\mathbf{y}_i \;\Rightarrow \bar{\mathbf{y}}_i$ can be obtained locally by node i
- Goal: recover $\mathbf{x}$ based on $\bar{\mathbf{y}}_i \Rightarrow$ Find $\tilde{\mathbf{x}}_K$ and recover $\mathbf{x}$ as $\mathbf{x} = \mathbf{V}_K\tilde{\mathbf{x}}_K$
- Define $\bar{\mathbf{u}}_i := \mathbf{V}_K^T\mathbf{e}_i$ and the Vandermonde matrix $\boldsymbol{\Psi}$ s.t. $\Psi_{kl} = \lambda_k^{l-1}$

$$\mathbf{x} = \mathbf{V}_K\tilde{\mathbf{x}}_K = \mathbf{V}_K \mathrm{diag}^{-1}(\bar{\mathbf{u}}_i)(\mathbf{C}\boldsymbol{\Psi}^T\mathbf{E}_K)^{-1}\bar{\mathbf{y}}_i$$

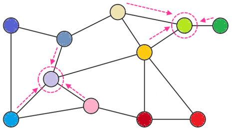
Space-shift sampling

- Hybrid scheme combining selection and aggregation sampling
  - $\Rightarrow$ Selection $\Rightarrow$ sampling the dimension of nodes
  - $\Rightarrow$ Aggregation $\Rightarrow$ sampling the dimension of shift applications
  - $\Rightarrow$ Space-shift $\Rightarrow$ sampling the 2D space spanned by the above

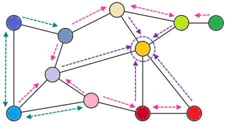
Selection: 4 nodes, 1 sample



Space-shift: 2 nodes, 2 samples



Aggregat.: 1 node, 4 samples



- Define the matrix $\mathbf{Y} := [\mathbf{y}^{(0)}, \ldots, \mathbf{y}^{(N-1)}] = [\mathbf{x}, \mathbf{S}\mathbf{x}, \ldots, \mathbf{S}^{N-1}\mathbf{x}]$
  - $\Rightarrow$ Selection samples the first column of $\mathbf{Y}$
  - $\Rightarrow$ Aggregation samples the i-th row of $\mathbf{Y}$
  - $\Rightarrow$ Space-shift samples the whole matrix $\mathbf{Y}$

Space-shift sampling: Recovery

- Define the matrix $\bar{\boldsymbol{\Upsilon}} := [\text{diag}(\bar{\mathbf{u}}_1), \ldots, \text{diag}(\bar{\mathbf{u}}_N)]^T$ and $\gamma := \mathrm{vec}(\mathbf{Y}^T)$
- Let $\mathbf{C} \in \{0,1\}^{K \times N^2}$ be a selection matrix $\Rightarrow \bar{\gamma} = \mathbf{C}\gamma$

Recovery of space-shift sampling

Signal $\mathbf{x}$ can be recovered from K space-shift samples as

$$\mathbf{x} = \mathbf{V}_K \tilde{\mathbf{x}}_K = \mathbf{V}_K \left(\mathbf{C}(\mathbf{I} \otimes (\boldsymbol{\Psi} \mathbf{E}_K)) \bar{\boldsymbol{\Upsilon}} \right)^{-1} \bar{\gamma}$$

provided that the inverse exists.

- If $\mathbf{C}(\mathbf{I} \otimes (\boldsymbol{\Psi} \mathbf{E}_K)) \bar{\boldsymbol{\Upsilon}}$ is not invertible $\Rightarrow$ additional samples required
- In general, invertibility is not easy to check a priori $\Rightarrow$ Selection
- For some forms of $\mathbf{C}$, invertibility can be ensured $\Rightarrow$ Aggregation

Space-shift sampling: Discussion

Appealing features of Space-shift Sampling

- Natural scheme when $\mathbf{S}$ encodes an underlying network dynamics
- Appropriate for inference based on a few access nodes
- Includes cases where node observes neighboring signal values
- Consistent with sampling in DSP
- Recovery error is reduced by combining selection and aggregation

Extensions

- Sampling in the presence of noise
  - $\Rightarrow$ Design of optimal sampling schemes
  - $\Rightarrow$ Aggregating nodes and $\mathbf{C}$ play a key role in minimizing error
- Unknown frequency support $\Rightarrow$ Sparse recovery

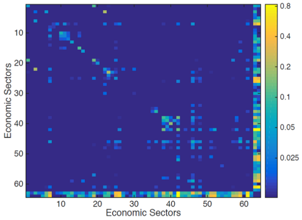
Joint recovery and support identification

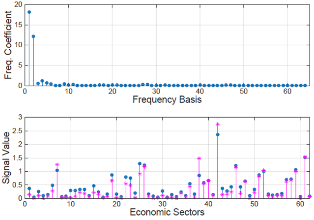
- We have assumed the first K frequencies of $\mathbf{x}$ to be active
- A more challenging problem $\Rightarrow$ Frequency support $\mathcal{K}$ is unknown
- Defining $\boldsymbol{\Upsilon} := [\mathrm{diag}(\mathbf{u}_1), \ldots, \mathrm{diag}(\mathbf{u}_N)]^T$, reformulate the problem

$$\tilde{\mathbf{x}}^* = \arg\min_{\tilde{\mathbf{x}}} \quad ||\tilde{\mathbf{x}}||_0 \qquad \text{s.t.} \quad \bar{\gamma} = \mathbf{C}\big(\mathbf{I} \otimes \boldsymbol{\Psi}\big)\boldsymbol{\Upsilon}\tilde{\mathbf{x}}$$

- Identifiable when $\mathbf{C}(\mathbf{I} \otimes \boldsymbol{\Psi})\boldsymbol{\Upsilon}$ is full spark and has at least $2K$ rows
- For some $\mathbf{C}$, full-spark can be assessed by inspecting $\{\lambda_i\}_{i=1}^N$ and $\mathbf{V}$
- Computationally, the ℓ_0 norm renders the optimization non-convex
  - $\Rightarrow$ Convexify it by replacing the ℓ_0 with an ℓ_1 norm
- Recoverability based on the coherence and the RIP of $\mathbf{C}(\mathbf{I} \otimes \boldsymbol{\Psi})\boldsymbol{\Upsilon}$
- With noise, the constraint can be replaced by $\|\bar{\gamma} - \mathbf{C}\big(\mathbf{I} \otimes \boldsymbol{\Psi}\big)\boldsymbol{\Upsilon}\tilde{\mathbf{x}}\|_2^2 < \epsilon$

- 62 economic sectors in USA + 2 synthetic sectors
  - $\Rightarrow$ Graph: average flows of production in 2007-2010, $\mathbf{S} = \mathbf{A}$
  - $\Rightarrow$ Signal $\mathbf{x}$: Production of sectors in 2011 (approx. bandlimited)



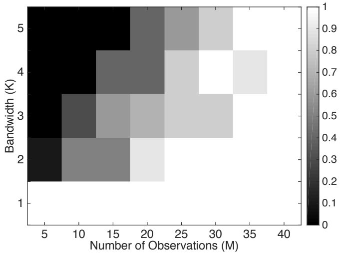


- Comparable minimum errors
- Median errors reduced via space-shift sampling

Sampling strategy				Error Min.	Error Median
$[\mathbf{x}]_i$	$[\mathbf{x}]_j$	$[\mathbf{x}]_k$	$[\mathbf{x}]_l$	.0039	4.2
$[\mathbf{x}]_i$	$[\mathbf{Sx}]_i$	$[\mathbf{S}^2\mathbf{x}]_i$	$[\mathbf{S}^3\mathbf{x}]_i$	.0035	.019
$[\mathbf{Sx}]_i$	$[\mathbf{Sx}]_j$	$[\mathbf{Sx}]_k$	$[\mathbf{Sx}]_l$	.0035	.030
$[\mathbf{S}^2\mathbf{x}]_i$	$[\mathbf{S}^2\mathbf{x}]_j$	$[\mathbf{S}^2\mathbf{x}]_k$	$[\mathbf{S}^2\mathbf{x}]_l$	.0035	.0055
$[\mathbf{x}]_i$	$[\mathbf{Sx}]_i$	$[\mathbf{x}]_j$	$[\mathbf{Sx}]_j$	.0035	.039

Sampling the US economy: Support identification

- Signals of bandwidth $K \in \{1, 2, \ldots, 5\}$ on the economic network
  - $\Rightarrow$ Value of K known but not the specific support
- Nr. of observations M, i.e. rows of $\mathbf{C}$, where $M \in \{5, 10, \ldots, 40\}$
  - $\Rightarrow$ Chosen among values in original signal $\mathbf{x}$ and first shift $\mathbf{Sx}$
- Solve iterative randomized version of convex relaxation



- For large M and small K
  - $\Rightarrow$ perfect recovery
- Gradual detriment for more adverse configurations

Conclusion

- Presented basic building blocks of GSP
  - $\Rightarrow$ Graph signal $\mathbf{x}$, graph-shift operator $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$, GFT $\mathbf{V}^{-1}$
- Discussed differences between selection and aggregation sampling
- Selection and aggregation can be combined in space-shift sampling
  - $\Rightarrow$ All of them reduce to traditional sampling in DSP
- Natural scheme for network processes
  - $\Rightarrow$ Appropriate for inference with few access nodes
- Conditions for perfect recovery and joint support identification
- Illustrated concepts via the U.S. economic network