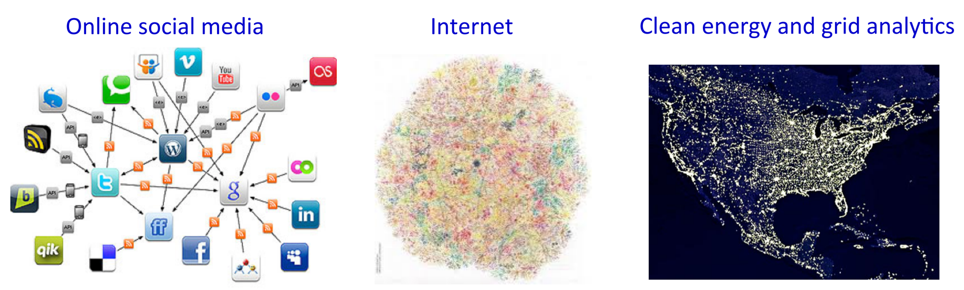
LINEAR NETWORK OPERATORS USING NODE-VARIANT GRAPH FILTERS

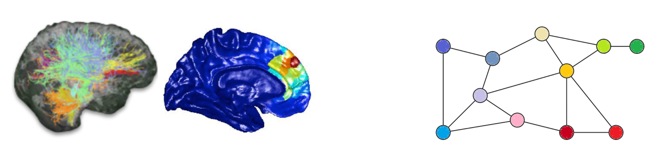
Santiago Segarra*, Antonio G. Marques†, and Alejandro Ribeiro* E-mail: ssegarra@seas.upenn.edu

*Electrical & Systems Engineering, University of Pennsylvania †Signal Theory & Communications, King Juan Carlos University

Graph signal processing - 101



- Desiderata: Process, analyze and learn from network data
- 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

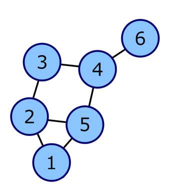


$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_{|\mathcal{V}|} \end{bmatrix} = \begin{bmatrix} 0.6 \\ \vdots \\ 0.7 \end{bmatrix}$$

- Graph SP: extending classical SP results to graph signals
  - $\Rightarrow$ Our view: GSP is well-suited to study network processes
- Filtering, smoothing, prediction, signal synthesis, compression

Graph signals and graph-shift operator

- Graph signals are mappings $x : \mathcal{V} \to \mathbb{R}$
  - $\Rightarrow$ May be represented as a vector $\mathbf{x} \in \mathbb{R}^N$ (with $|\mathcal{V}| = N$)
  - $\Rightarrow$ To understand and analyze $\mathbf{x}$, useful to account for G's structure
- Graph G is endowed with a graph-shift operator $\mathbf{S}$
  - $\Rightarrow$ Matrix $\mathbf{S} \in \mathbb{R}^{N\times N}$ satisfying: $S_{ij} = 0$ for $i \neq j$ and $(i,j) \notin \mathcal{E}$



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

$\mathbf{S}$ captures local structure in G

- Ex: Adjacency $\mathbf{A}$, Degree $\mathbf{D}$ and Laplacian $\mathbf{L}$

Locality of S and frequency-domain representation

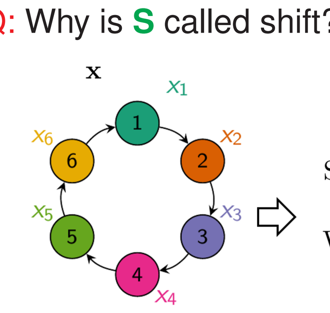
- $\mathbf{S}$ is a local linear operator $\Rightarrow$ If $\mathbf{y} = \mathbf{S}\mathbf{x}$, $y_i = \sum_{j \in \mathcal{N}_i} S_{ij} x_j \Rightarrow$ 1-hop info
- Spectrum of $\mathbf{S}$ useful to analyze $\mathbf{x}$
  - $\Rightarrow$ Consider diagonalizable shifts $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$
- Leverage $\mathbf{S}$ to define graph Fourier transform (GFT) and iGFT

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

- Key message: the two basic elements of GSP are $\mathbf{x}$ and $\mathbf{S}$

Connection with discrete-time signal processing

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



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

What is $\mathbf{Sx}$?

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

- When $\mathbf{S} = \mathbf{A}_{dc}$, we have that $\mathbf{V}^{-1}$ is the DFT matrix
- Point of contact between GSP and DSP
- When particularizing GSP to the directed cycle
  - $\Rightarrow$ We recover known result from DSP

Linear (shift-invariant) graph filter

- A graph filter $H : \mathbb{R}^N \to \mathbb{R}^N$ is a map between graph signals
  - $\Rightarrow$ Focus on linear filters $\Rightarrow N \times N$ matrix
- Filter $\mathbf{H}$ is a polynomial in $\mathbf{S}$ of degree $L-1$, with coeff. $\mathbf{h} = [h_0, \ldots, h_{L-1}]^T$

$$\mathbf{H} := h_0\mathbf{S}^0 + h_1\mathbf{S}^1 + \ldots + h_{L-1}\mathbf{S}^{L-1} = \sum_{l=0}^{L-1} h_l \mathbf{S}^l$$

- Key properties: shift invariance and distributed implementation
  - $\Rightarrow$ Satisfies $\mathbf{H}(\mathbf{Sx}) = \mathbf{S}(\mathbf{Hx})$, only L-hop information to form $\mathbf{y} = \mathbf{Hx}$

Frequency response of a graph filter

- Using $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$, filter is $\mathbf{H} = \sum_{l=0}^{L-1} h_l \mathbf{S}^l = \mathbf{V}\left(\sum_{l=0}^{L-1} h_l \mathbf{\Lambda}^l\right)\mathbf{V}^{-1}$
- Since $\mathbf{\Lambda}^l$ are diagonal, the GFT-iGFT can be used to write $\mathbf{y} = \mathbf{Hx}$ as

$$\tilde{\mathbf{y}} = \text{diag}(\tilde{\mathbf{h}})\tilde{\mathbf{x}}$$

  - $\Rightarrow$ Output at frequency k depends only on input at frequency k
- Frequency response of $\mathbf{H}$ is $\tilde{\mathbf{h}} = \mathbf{\Psi}\mathbf{h}$, where $\Psi_{ij} = \lambda_i^{j-1}$ (Vandermonde)
- Note that GFTs for signals ($\tilde{\mathbf{x}} = \mathbf{V}^{-1}\mathbf{x}$) and filters ($\tilde{\mathbf{h}} = \mathbf{\Psi}\mathbf{h}$) are different
  - $\Rightarrow$ If $\mathbf{S} = \mathbf{A}_{dc}$ (periodic signal), both $\mathbf{\Psi}$ and $\mathbf{V}^{-1}$ are equal to the DFT

Motivation and problem formulation

- Design graph filters to implement a given linear transformation
  - $\Rightarrow$ Implementation is distributed by construction
  - $\Rightarrow$ Conditions for perfect and approximate implementation
  - $\Rightarrow$ Leverage results from classical time-invariant systems
- Given a linear transformation $\mathbf{B}$, find the filter coefficients $\mathbf{h}$ s. t.

$$\mathbf{B} = \sum_{l=0}^{L-1} h_l \mathbf{S}^l$$

- Graph-shift operator $\mathbf{S}$ is given
  - $\Rightarrow$ Well-suited for cases where $\mathbf{S}$ is a network process
  - $\Rightarrow$ E.g., diffusion in a social network
  - $\Rightarrow$ Agents exchange information and weigh info observed
  - $\Rightarrow$ Choosing $\mathbf{h}$ $\Rightarrow$ fixing the weights

Conditions for perfect implementation

- Notation: a) D # of distinct $\{\lambda_k\}_{k=1}^N$; b) $\{\gamma_k\}_{k=1}^N$ eigenvalues of $\mathbf{B}$

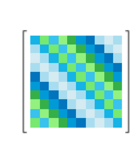
Perfect implementation of linear graph operators

The linear transformation $\mathbf{B}$ can be implemented using a graph filter $\mathbf{H}$ if the three following conditions hold true:
i) Matrices $\mathbf{B}$ and $\mathbf{S}$ are simultaneously diagonalizable.
ii) For all (k_1,k_2) such that $\lambda_{k_1} = \lambda_{k_2}$, it holds that $\gamma_{k_1} = \gamma_{k_2}$.
iii) The degree $L-1$ of $\mathbf{H}$ satisfies $L \geq D$.

- i) $\Rightarrow$ frequency basis of $\mathbf{B}$ and $\mathbf{S}$ the same $\Rightarrow$ necessary
- ii) $\Rightarrow$ two equal freqs. in $\mathbf{S}$ must be equal in $\mathbf{B}$ $\Rightarrow$ necessary

- Restrictive conditions but not impossible
  - $\Rightarrow$ Consensus $\mathbf{B}_{\text{con}} = \mathbf{1}\mathbf{1}^T$ favors i) and ii) because it is rank-one

In time:
i) $\Leftrightarrow$ $\mathbf{B}$ circulant

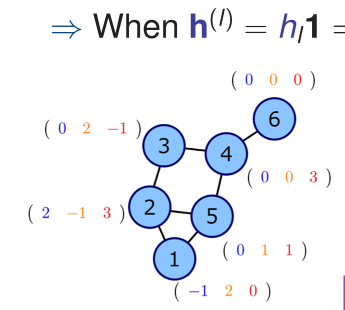


Node-variant graph filters: definition

- A generalization of graph filters:

$$\mathbf{H}_{\text{nv}} := \sum_{l=0}^{L-1} \text{diag}(\mathbf{h}^{(l)})\mathbf{S}^l$$

  - $\Rightarrow$ When $\mathbf{h}^{(l)} = h_l\mathbf{1} \Rightarrow$ regular (node-invariant) filter



$$\mathbf{x}^{(0)} = \mathbf{x} = \begin{pmatrix} -1 \\ 2 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad \mathbf{x}^{(1)} = \mathbf{S}\mathbf{x}^{(0)} = \begin{pmatrix} 2 \\ -1 \\ 2 \\ 0 \\ 1 \\ 0 \end{pmatrix} \quad \mathbf{x}^{(2)} = \mathbf{S}\mathbf{x}^{(1)} = \begin{pmatrix} 0 \\ 3 \\ -1 \\ 3 \\ 1 \\ 0 \end{pmatrix}$$

$$\boxed{\mathbf{y} = h^{(0)}\mathbf{x}^{(0)} + h^{(1)}\mathbf{x}^{(1)} + h^{(2)}\mathbf{x}^{(2)}} \qquad \boxed{y_i = h_i^{(0)}x_i^{(0)} + h_i^{(1)}x_i^{(1)} + h_i^{(2)}x_i^{(2)}}$$

- In general, when $\mathbf{H}_{\text{nv}}$ is applied to a signal $\mathbf{x}$
  - $\Rightarrow$ Each node applies different weights to the shifted signals $\mathbf{S}^l\mathbf{x}$
  - $\Rightarrow$ More flexible and still distributed, not shift-invariant

Node-variant graph filters: frequency response

- Collect the coefficients of node i in $\mathbf{h}_i$, such that $[\mathbf{h}_i]_l = [\mathbf{h}^{(l)}]_i$
- Focus on the filter output at node i, $\mathbf{e}_i^T\mathbf{H}_{\text{nv}}\mathbf{x}$, and define $\mathbf{u}_i := \mathbf{V}^T\mathbf{e}_i$

$$\boldsymbol{\eta}_i^T = \mathbf{e}_i^T\mathbf{H}_{\text{nv}} = \sum_{l=0}^{L-1} [\mathbf{h}_i]_l \mathbf{e}_i^T\mathbf{V}\mathbf{\Lambda}^l\mathbf{V}^{-1} = \mathbf{u}_i^T\text{diag}(\mathbf{\Psi}\mathbf{h}_i)\mathbf{V}^{-1}$$

- The output of the filter at node i, $\boldsymbol{\eta}_i^T\mathbf{x}$ is the inner product of
  - $\Rightarrow$ $\mathbf{V}^{-1}\mathbf{x}$ $\Rightarrow$ the frequency representation of the input, and
  - $\Rightarrow$ $\mathbf{u}_i$ $\Rightarrow$ how strongly the frequencies are expressed by node i
  - $\Rightarrow$ Modulated by $\mathbf{\Psi}\mathbf{H}_i$ $\Rightarrow$ Frequency response associated to i

Perfect reconstruction with node-variant filters

- Node-variant filters can implement a larger class of transformations
  - $\Rightarrow$ Pick $\mathbf{h}^{(l)}$ for $l = 0, \cdots, L-1$ so that

$$\mathbf{B} = \sum_{l=0}^{L-1} \text{diag}(\mathbf{h}^{(l)})\mathbf{S}^l$$

- Defining $\mathbf{b}_i := \mathbf{B}^T\mathbf{e}_i$ and $\bar{\mathbf{b}}_i := \mathbf{V}^T\mathbf{b}_i$

Perfect implementation using node-variant filters

The linear transformation $\mathbf{B}$ can be implemented using the node-variant filter $\mathbf{H}_{\text{nv}}$ if the three following conditions hold for all i:
i) $[\bar{\mathbf{b}}_i]_k = 0$ for those k such that $[\mathbf{u}_i]_k = 0$.
ii) For all (k_1,k_2) s. t. $\lambda_{k_1} = \lambda_{k_2}$, it holds $[\bar{\mathbf{b}}_i]_{k_1}/[\mathbf{u}_i]_{k_1} = [\bar{\mathbf{b}}_i]_{k_2}/[\mathbf{u}_i]_{k_2}$.
iii) The degree $L-1$ of $\mathbf{H}_{\text{nv}}$ satisfies $L \geq D$.

- Less restrictive that for node-invariant filters
  - $\Rightarrow$ Not surprising since we have additional freedom in the design
- Some shift operators $\mathbf{S}$ can implement any transformation $\mathbf{B}$

Corollary

Any linear transformation $\mathbf{B}$ can be implemented using a node-varying filter of degree $N-1$ if the graph-shift operator $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$ satisfies:
i) all the entries of $\mathbf{V}$ are non-zero.
ii) all the eigenvalues $\{\lambda_k\}_{k=1}^N$ are distinct.

- Under these conditions, unique set of optimal coefficients $\{\mathbf{h}_i^*\}_{i=1}^N$

$$\mathbf{h}_i^* = \mathbf{\Psi}^{-1}\text{diag}(\mathbf{u}_i)^{-1}\mathbf{V}^T\mathbf{b}_i$$

Approximate implementation: No prior knowledge

- When perfect reconstruction is infeasible
  - $\Rightarrow$ Minimize a pre-specified error metric, design $\mathbf{H}_{\text{nv}}$ to resemble $\mathbf{B}$
  - $\Rightarrow$ We look at minimizing $\|\mathbf{H}_{\text{nv}} - \mathbf{B}\|_{\text{F}}$ and $\|\mathbf{H}_{\text{nv}} - \mathbf{B}\|_2$
- Define $\mathbf{\Phi}_i := (\mathbf{V}^{-1})^T\text{diag}(\mathbf{u}_i)\mathbf{\Psi}$ and $\tilde{\mathbf{U}} := [\text{diag}(\mathbf{u}_1), \ldots, \text{diag}(\mathbf{u}_N)]^T$
- Collect the desired coefficient vectors in $\mathbf{\Gamma} = [\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_N]$

Approximate implementation with no knowledge of x

The optimal filter coefficients defined as $\{\mathbf{h}_{i,\text{F}}^*\}_{i=1}^N := \text{argmin}_{\{\mathbf{h}_i\}_{i=1}^N} \|\mathbf{H}_{\text{nv}} - \mathbf{B}\|_{\text{F}}$ and $\{\mathbf{h}_{i,2}^*\}_{i=1}^N := \text{argmin}_{\{\mathbf{h}_i\}_{i=1}^N} \|\mathbf{H}_{\text{nv}} - \mathbf{B}\|_2$ are, respectively, given by

$$\mathbf{h}_{i,\text{F}}^* = \mathbf{\Phi}_i^\dagger\mathbf{b}_i \overset{\star}{=} (\mathbf{\Phi}_i^T\mathbf{\Phi}_i)^{-1}\mathbf{\Phi}_i^T\mathbf{b}_i,$$

for all i, where $\overset{\star}{=}$ holds if $\mathbf{\Phi}_i$ has full column rank, and

$$\{\mathbf{\Gamma}_2^*, s^*\} = \underset{\{\mathbf{\Gamma}, s\}}{\text{argmin}} \quad s$$
$$\text{s. to} \begin{bmatrix} s\mathbf{I} & (\mathbf{I} \odot \mathbf{\Psi}\mathbf{\Gamma})^T\tilde{\mathbf{U}}\mathbf{V}^{-1} - \mathbf{B} \\ ((\mathbf{I} \odot \mathbf{\Psi}\mathbf{\Gamma})^T\tilde{\mathbf{U}}\mathbf{V}^{-1} - \mathbf{B})^T & s\mathbf{I} \end{bmatrix} \succeq 0.$$

- For $\|\cdot\|_F$, optimal coefficients result from a least-squares problem
- There is no closed form for the $\|\cdot\|_2$ case
  - $\Rightarrow$ Efficiently obtain it by solving a semi-definite program

Approximate implementation: Prior knowledge

- Incorporate prior knowledge of the input signal $\mathbf{x}$ into the design
- Goal is to minimize a metric of the error vector $\mathbf{d} := \mathbf{H}_{\text{nv}}\mathbf{x} - \mathbf{B}\mathbf{x}$
- When $\mathbf{x}$ is zero-mean with covariance $\mathbf{R}_\mathbf{x}$, the error covariance is

$$\mathbf{R}_\mathbf{d} := \mathbb{E}[\mathbf{d}\mathbf{d}^T] = (\mathbf{H}_{\text{nv}} - \mathbf{B})\mathbf{R}_\mathbf{x}(\mathbf{H}_{\text{nv}} - \mathbf{B})^T$$

- Pick $\mathbf{h}_i$ to minimize some metric of the error covariance $\mathbf{R}_\mathbf{d}$
  - $\Rightarrow$ $\text{trace}(\mathbf{R}_\mathbf{d})$ $\Rightarrow$ Mean squared error
  - $\Rightarrow$ $\lambda_{\text{max}}(\mathbf{R}_\mathbf{d})$ $\Rightarrow$ worst-case error

Approximate implementation with statistical knowledge of x

The optimal filter coefficients defined as $\{\mathbf{h}_{i,\text{Tr}}^*\}_{i=1}^N := \text{argmin}_{\{\mathbf{h}_i\}_{i=1}^N} \text{trace}(\mathbf{R}_\mathbf{d})$ and $\{\mathbf{h}_{i,\lambda}^*\}_{i=1}^N := \text{argmin}_{\{\mathbf{h}_i\}_{i=1}^N} \lambda_{\text{max}}(\mathbf{R}_\mathbf{d})$ are, respectively, given by

$$\mathbf{h}_{i,\text{Tr}}^* = (\mathbf{R}_\mathbf{x}^{1/2}\mathbf{\Phi}_i)^\dagger\mathbf{R}_\mathbf{x}^{1/2}\mathbf{b}_i \overset{\star}{=} (\mathbf{\Phi}_i^T\mathbf{R}_\mathbf{x}\mathbf{\Phi}_i)^{-1}\mathbf{\Phi}_i^T\mathbf{R}_\mathbf{x}\mathbf{b}_i,$$

for all i, where $\overset{\star}{=}$ holds if $\mathbf{R}_\mathbf{x}^{1/2}\mathbf{\Phi}_i$ has full column rank, and

$$\{\mathbf{\Gamma}_\lambda^*, s^*\} = \underset{\{\mathbf{\Gamma}, s\}}{\text{argmin}} \quad s$$
$$\text{s. to} \begin{bmatrix} s\mathbf{I} & (\mathbf{I} \odot \mathbf{\Psi}\mathbf{\Gamma})^T\tilde{\mathbf{U}}\mathbf{V}^{-1} - \mathbf{B} \\ ((\mathbf{I} \odot \mathbf{\Psi}\mathbf{\Gamma})^T\tilde{\mathbf{U}}\mathbf{V}^{-1} - \mathbf{B})^T & s\mathbf{R}_\mathbf{x}^{-1} \end{bmatrix} \succeq 0.$$

- No prior knowledge scenarios equivalent to the cases $\mathbf{R}_\mathbf{x} = \mathbf{I}$
  - $\Rightarrow$ $\|\mathbf{H}_{\text{nv}} - \mathbf{B}\|_{\text{F}} \equiv \text{trace}(\mathbf{R}_\mathbf{d})$ and $\|\mathbf{H}_{\text{nv}} - \mathbf{B}\|_2 \equiv \lambda_{\text{max}}(\mathbf{R}_\mathbf{d})$
- Additional assumptions can be incorporated into the designs

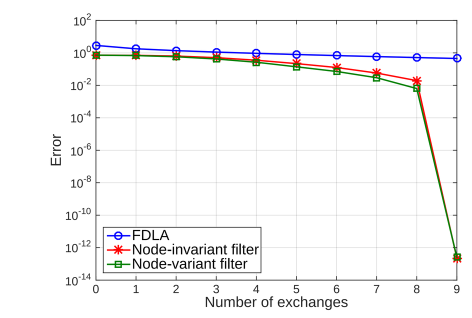
Finite-time consensus

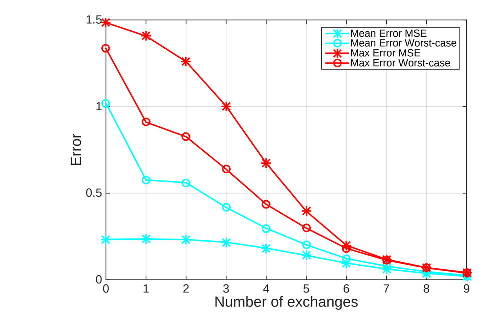
- Local implementation of the consensus operator $\mathbf{B}_{\text{con}} = \mathbf{1}\mathbf{1}^T/N$

Corollary

The operator $\mathbf{B}_{\text{con}}$ can be written as a filter $\sum_{l=0}^{N-1} c_l\mathbf{S}^l$ for some $\mathbf{S}$ associated with an undirected graph $\mathcal{G}$ if and only if $\mathcal{G}$ is connected.

- If the weights in $\mathbf{S}$ can be selected (e.g., choose $\mathbf{S} = \mathbf{L}$)
  - $\Rightarrow$ consensus is achieved in finite time for connected graphs
- Key: $\mathbf{B}_{\text{con}}$ is low-rank (repeated eigenvalues well-suited for approx.)
- We compare the performance of three methods: 1) Asymptotic fastest distributed linear averaging (FDLA), 2) Graph filter approx., 3) Node-variant graph filter approx.
- Compare worst-case and mean error design (50 nodes)





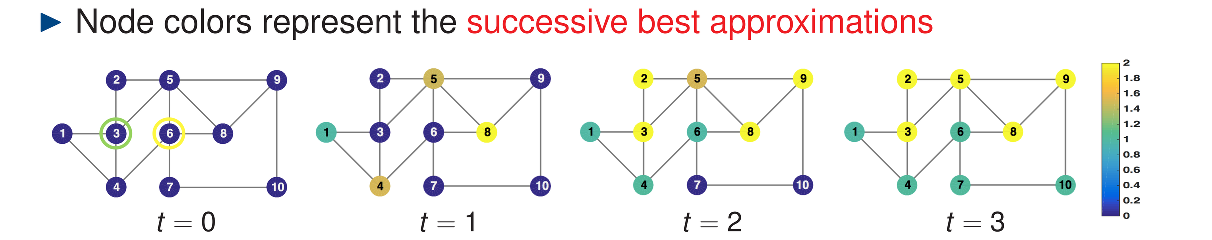
Analog network coding (ANC)

- In ANC we focus on the transmission from sources to sinks
  - $\Rightarrow$ Leverage graph filters to design ANC schemes
  - $\Rightarrow$ Subset of the nodes in a graph
  - $\Rightarrow$ Main results still hold for the modified framework

- Consider a 10-node undirected graph
- Nodes 3 and 6 are sources
- Node 3 transmits to 1, 4, 6, 7, and 10
- Node 6 transmits to the remaining ones

$$\mathbf{B}_{\mathcal{SR}} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}^T$$

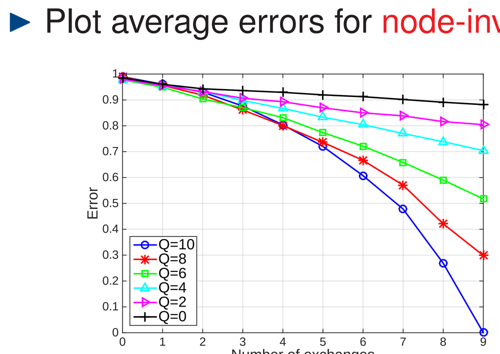
- Denote by g and w the signals injected by nodes 3 and 6, respectively
- Node colors represent the successive best approximations

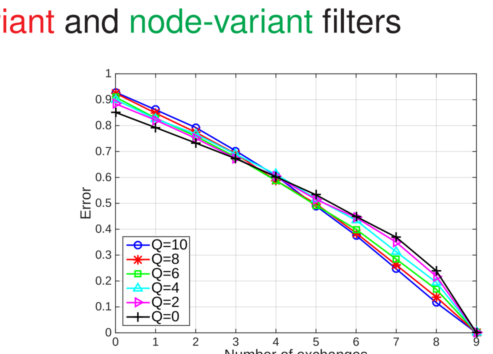


node	recover signal	observed signal $\mathbf{z}^{(t)}$ $t=0$	$t=1$	$t=2$	$t=3$	Optimal coeff.
1	g	0	g	$g+w$	$5g+2w$	[0, 1]
3	w	g	0	$4g+2w$	$4g+3w$	[-2, 0, 0.5]
10	g	0	0	w	$g+2w$	[0, 0, -2, 1]

Spectrum mismatch

- Assume that $\mathbf{B}$ and $\mathbf{S}$ share Q eigenvectors
  - $\Rightarrow$ Higher Q, closer to being simultaneously diagonalizable
- Generate 1000 Erdős-Rényi graphs (N=10)
- Plot average errors for node-invariant and node-variant filters





- Performance of node-invariant filter is sensitive to changes in Q
- Node-variant filters are robust to the spectral differences of $\mathbf{B}$ and $\mathbf{S}$

References

- D. I. Shuman et al., "The emerging field of signal processing on graphs," Signal Process. Mag., 2013.
- A. Sandryhaila and J. Moura, "Discrete signal processing on graphs," IEEE TSP, 2013.
- D. I. Shuman, P. Vandergheynst, and P. Frossard, "Chebyshev polynomial approximation for distributed signal processing," DCOSS, 2011.
- A. Sandryhaila, S. Kar, J. Moura, "Finite-time distributed consensus through graph filters," ICASSP, 2014.
- S. Barbarossa, G. Scutari, and T. Battisti, "Distributed signal subspace projection algorithms with maximum convergence rate for sensor networks with topological constraints," ICASSP, 2009.
- S. Segarra et al., "Distributed linear network operators using graph filters," arXiv, 2015.