Learning graphs from data:

A signal processing perspective

Xiaowen Dong
MIT Media Lab

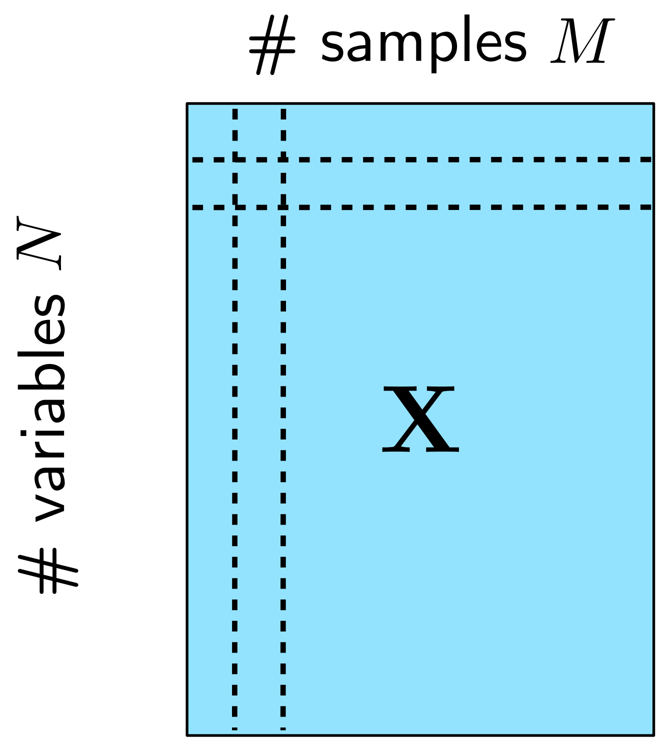
Graph Signal Processing Workshop
Pittsburgh, PA, May 2017

Introduction

- What is the problem of graph learning?

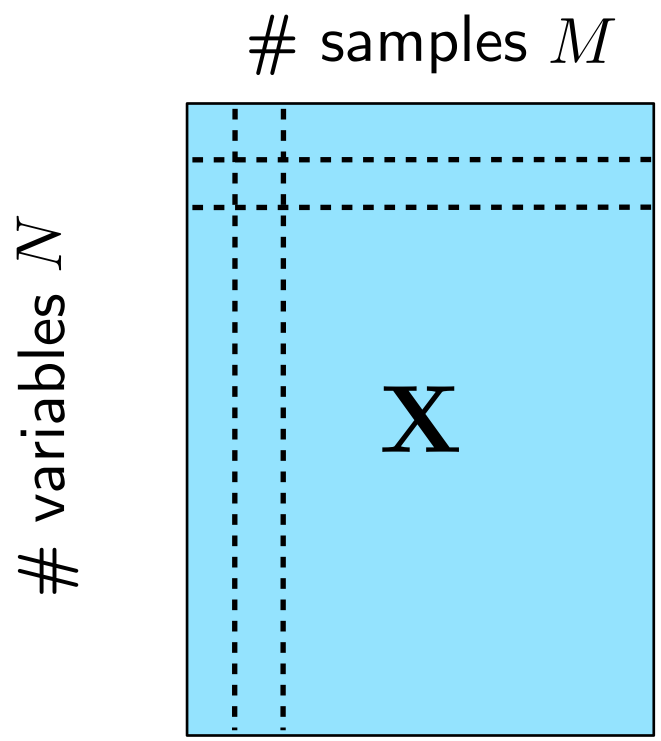
Introduction

- What is the problem of graph learning?
  - Given observations on a number of variables and some prior knowledge (distribution, model, etc)



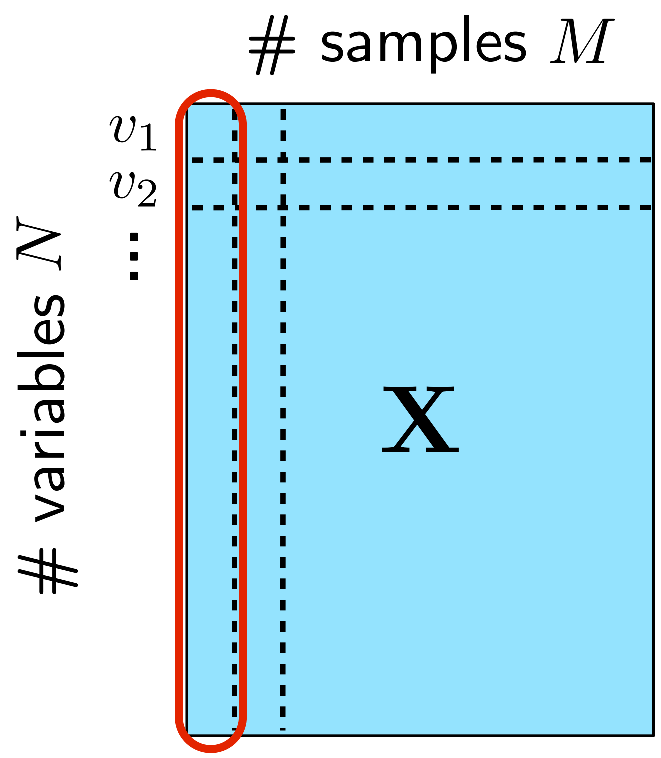
Introduction

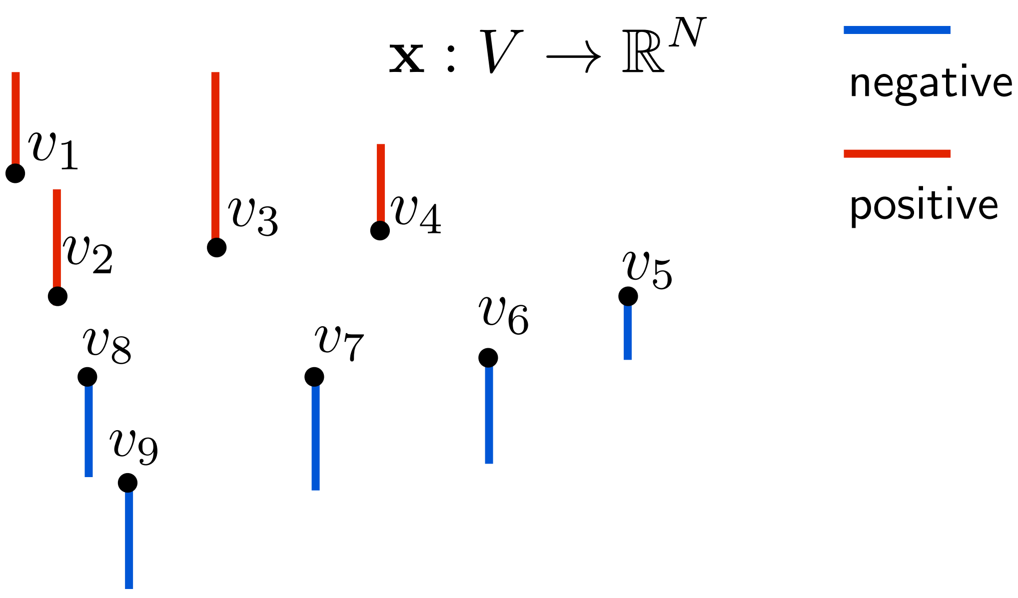
- What is the problem of graph learning?
  - Given observations on a number of variables and some prior knowledge (distribution, model, etc)
  - Build/learn a measure of relations between variables (correlation/covariance, graph topology/operator or equivalent)



Introduction

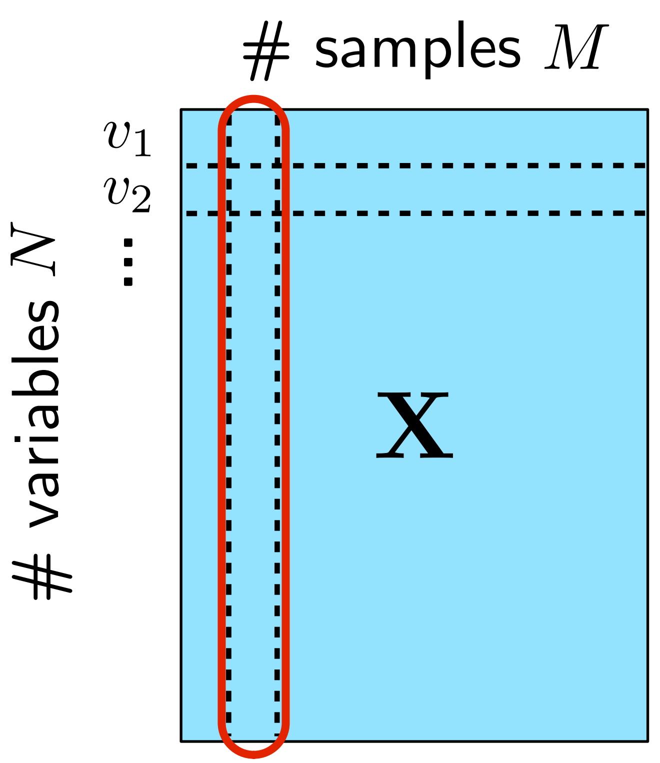
- What is the problem of graph learning?
  - Given observations on a number of variables and some prior knowledge (distribution, model, etc)
  - Build/learn a measure of relations between variables (correlation/covariance, graph topology/operator or equivalent)

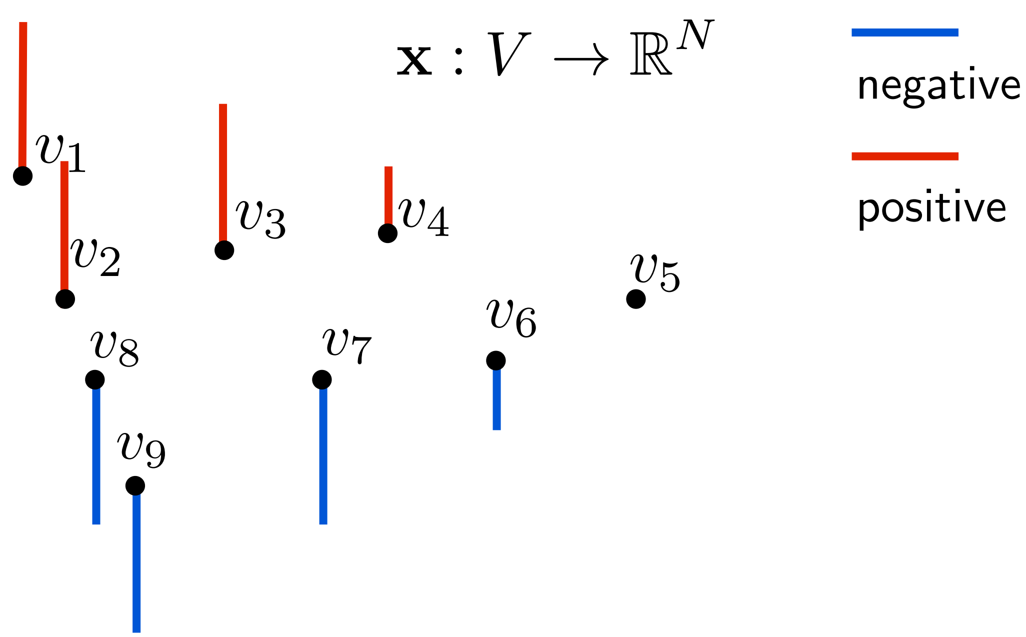




Introduction

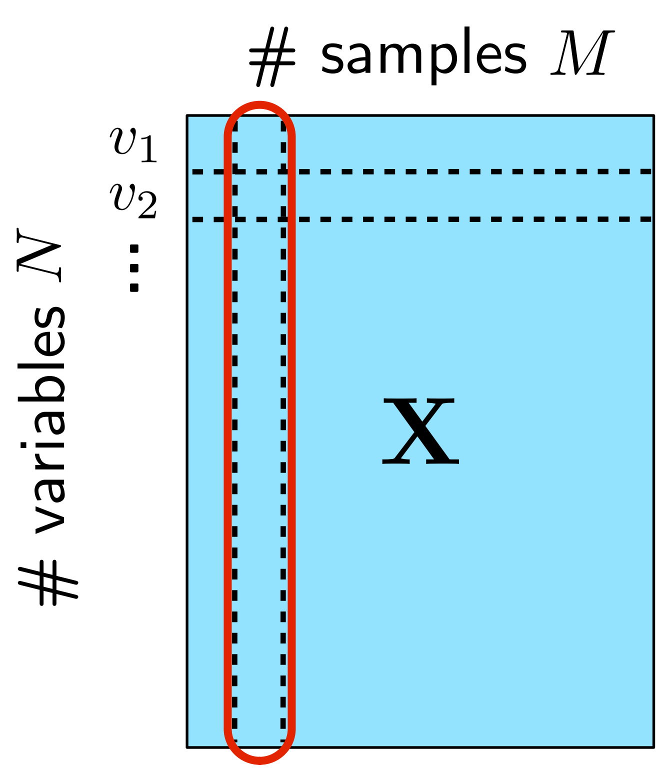
- What is the problem of graph learning?
  - Given observations on a number of variables and some prior knowledge (distribution, model, etc)
  - Build/learn a measure of relations between variables (correlation/covariance, graph topology/operator or equivalent)

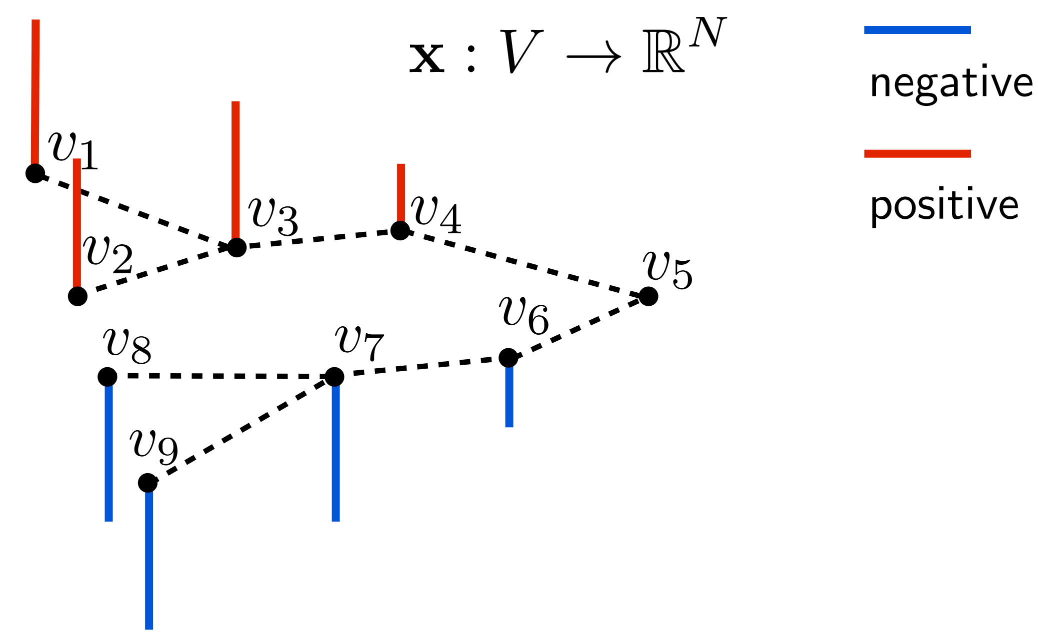




Introduction

- What is the problem of graph learning?
  - Given observations on a number of variables and some prior knowledge (distribution, model, etc)
  - Build/learn a measure of relations between variables (correlation/covariance, graph topology/operator or equivalent)



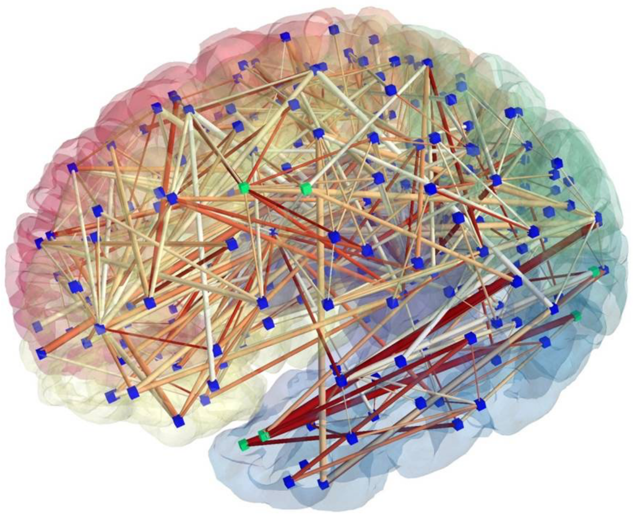


Introduction

- Why is it important?
  - Learning relations between entities benefits numerous application domains

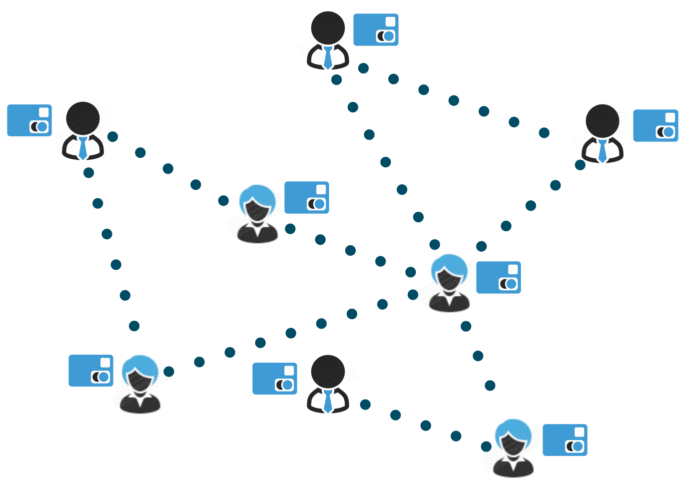
Introduction

- Why is it important?
  - Learning relations between entities benefits numerous application domains



Objective: functional connectivity between brain regions

Input: fMRI recordings in these regions



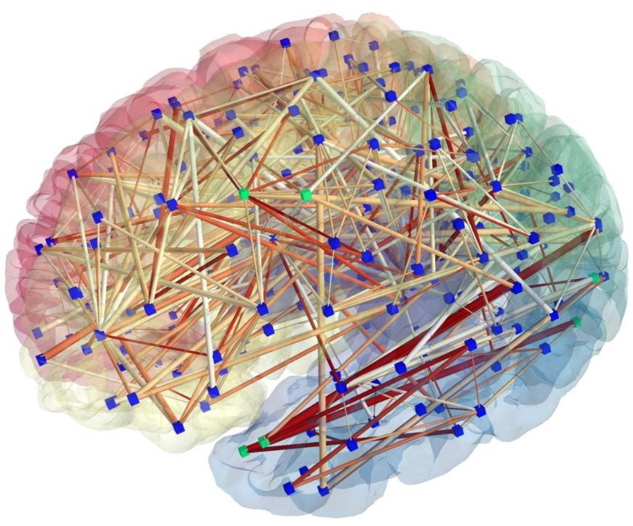
Objective: behavioral similarity/ influence between people

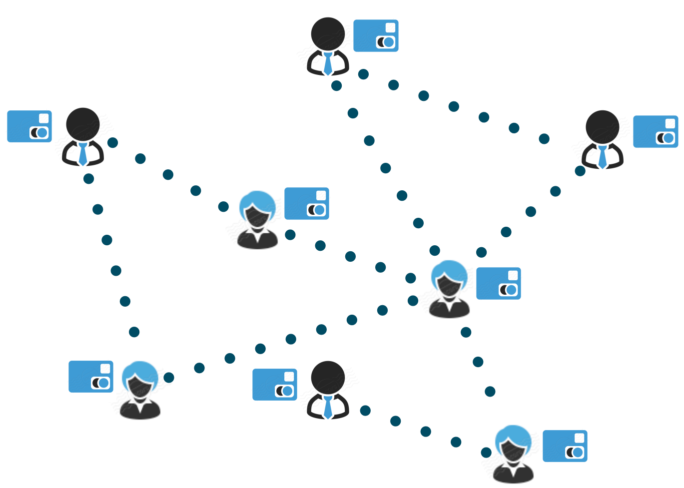
Input: individual history of activities

image credit:
http://blog.myesr.org/mri-reveals-the-human-connectome/
https://www.iconexperience.com

Introduction

- Why is it important?
  - Learning relations between entities benefits numerous application domains
  - The learned relations can help us predict future observations





Objective: functional connectivity between brain regions

Input: fMRI recordings in these regions

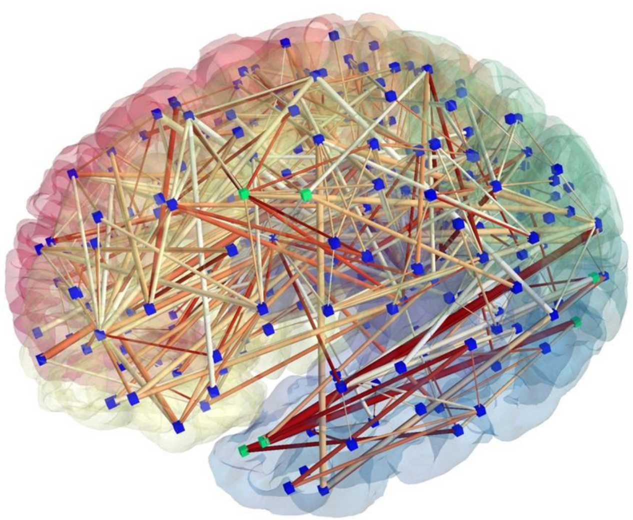
Objective: behavioral similarity/ influence between people

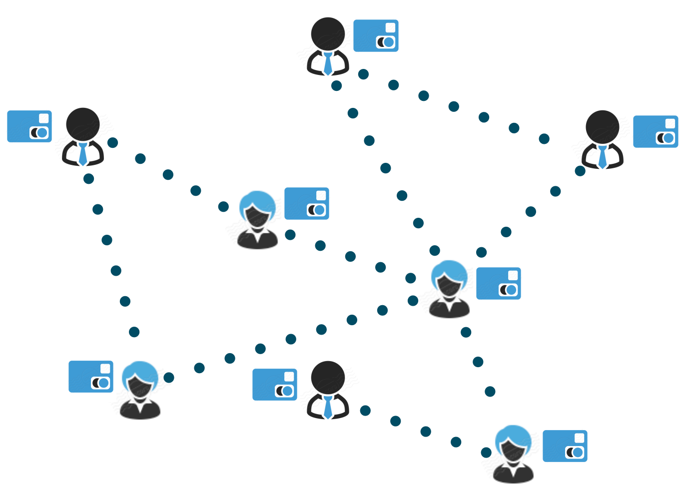
Input: individual history of activities

image credit:
http://blog.myesr.org/mri-reveals-the-human-connectome/
https://www.iconexperience.com

Introduction

- Why is it important?
  - Learning relations between entities benefits numerous application domains
  - The learned relations can help us predict future observations





Objective: functional connectivity between brain regions

Input: fMRI recordings in these regions

Objective: behavioral similarity/ influence between people

Input: individual history of activities

How do we build/learn the graph?

image credit:
http://blog.myesr.org/mri-reveals-the-human-connectome/
https://www.iconexperience.com

Outline

- A (partial) historic overview
- A signal processing perspective
  - GSP idea for graph learning
  - Three signal/graph models
- Perspective

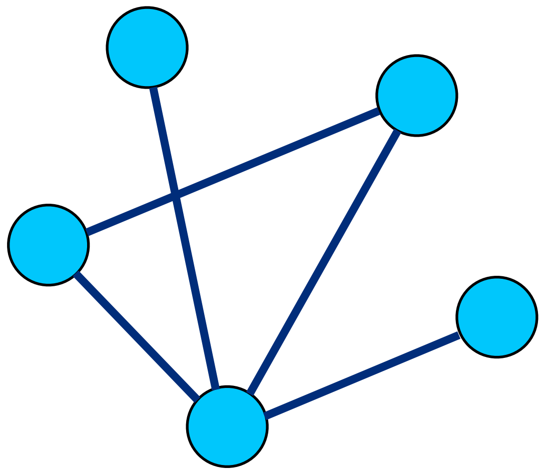
A (partial) historical overview

- Simple and intuitive methods
  - Sample correlation
  - Similarity function (e.g., Gaussian RBF)

A (partial) historical overview

- Simple and intuitive methods
  - Sample correlation
  - Similarity function (e.g., Gaussian RBF)

- Learning graphical models

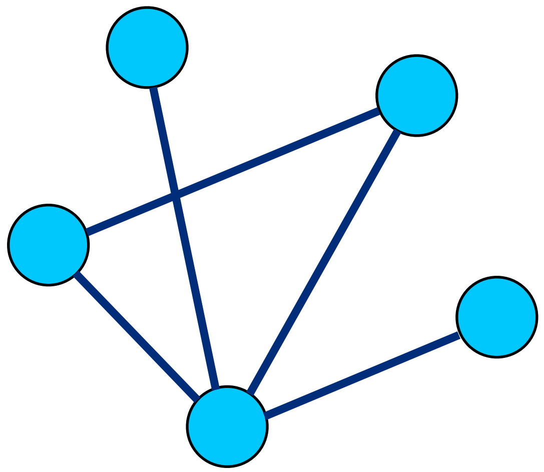


Undirected graphical models:
Markov random fields (MRF)

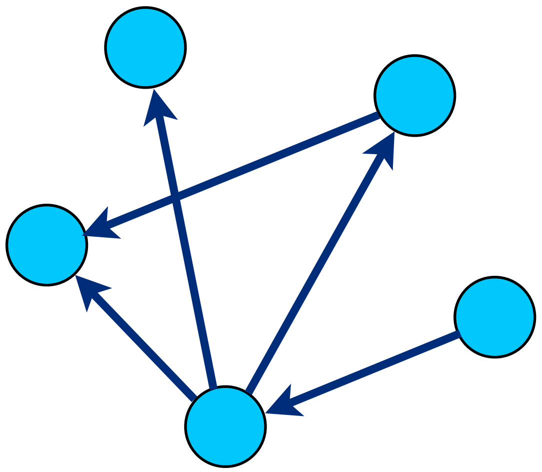
A (partial) historical overview

- Simple and intuitive methods
  - Sample correlation
  - Similarity function (e.g., Gaussian RBF)

- Learning graphical models



Undirected graphical models:
Markov random fields (MRF)

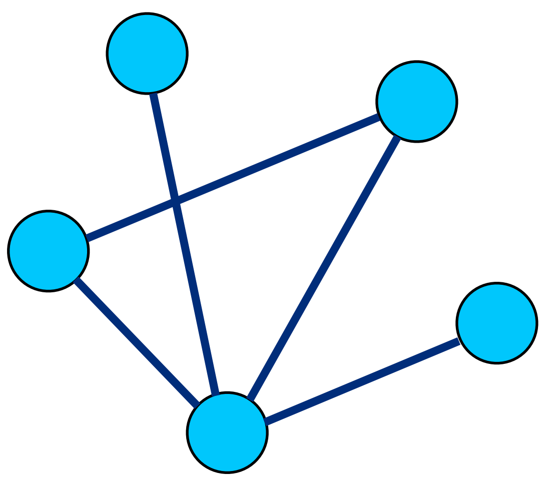


Directed graphical models:
Bayesian networks (BN)

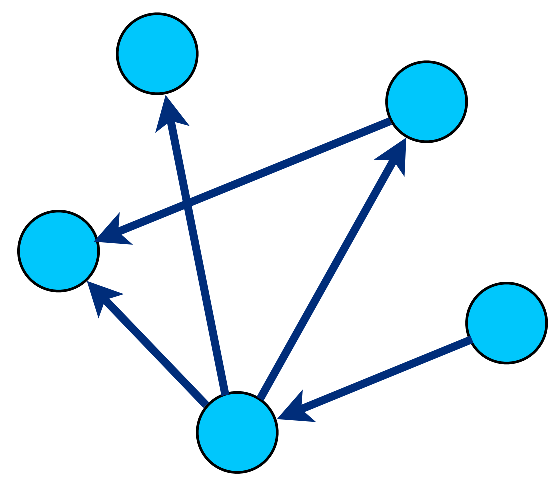
A (partial) historical overview

- Simple and intuitive methods
  - Sample correlation
  - Similarity function (e.g., Gaussian RBF)

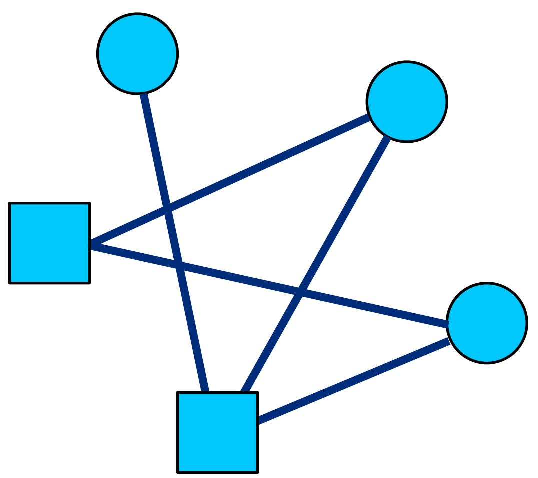
- Learning graphical models



Undirected graphical models: Markov random fields (MRF)



Directed graphical models: Bayesian networks (BN)

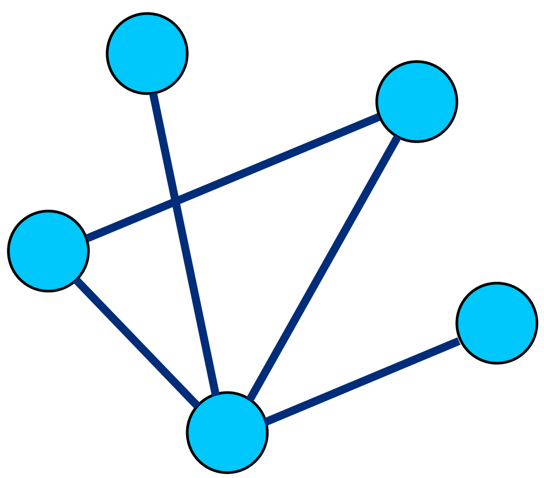


Factor graphs

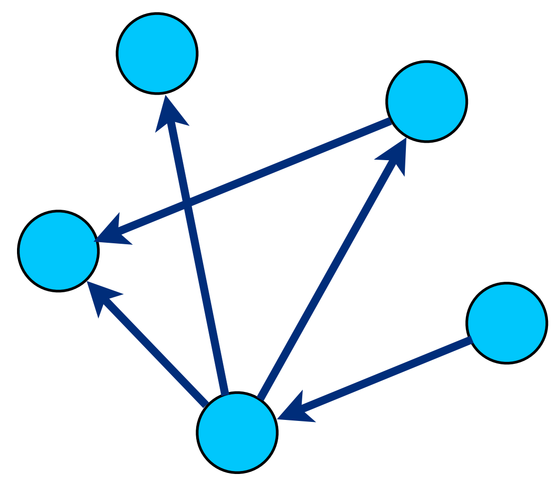
A (partial) historical overview

- Simple and intuitive methods
  - Sample correlation
  - Similarity function (e.g., Gaussian RBF)

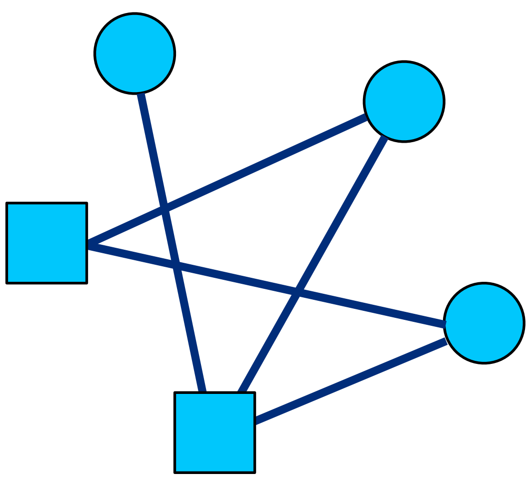
- Learning graphical models



Undirected graphical models:
Markov random fields (MRF)



Directed graphical models:
Bayesian networks (BN)



Factor graphs

A (partial) historical overview

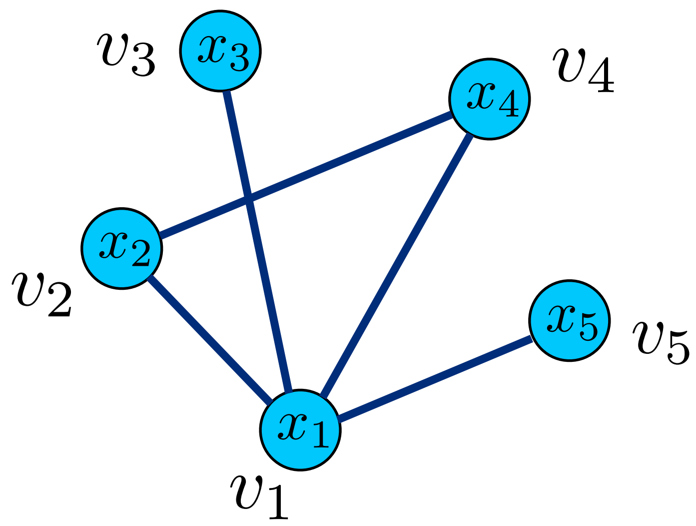
- Learning pairwise MRF

A (partial) historical overview

- Learning pairwise MRF

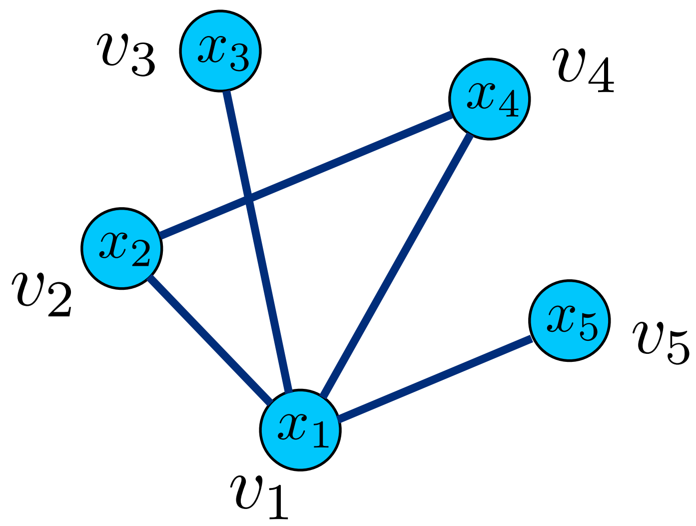
conditional independence:

$$(i, j) \notin E \Leftrightarrow x_i \perp x_j \mid \mathbf{x} \setminus \{x_i, x_j\}$$



A (partial) historical overview

- Learning pairwise MRF



conditional independence:

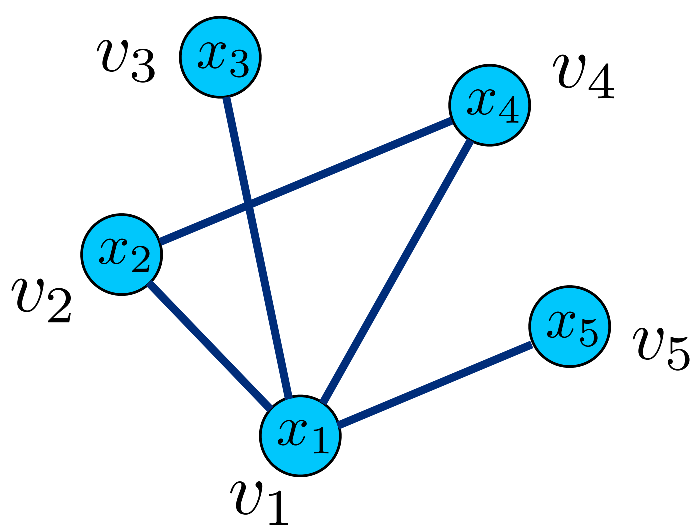
$$(i,j) \notin E \Leftrightarrow x_i \perp x_j \mid \mathbf{x} \setminus \{x_i, x_j\}$$

probability parameterized by $\boldsymbol{\Theta}$:

$$P(\mathbf{x}|\boldsymbol{\Theta}) = \frac{1}{Z(\boldsymbol{\Theta})} \exp\Big(\sum_{i \in V} \theta_{ii} x_i^2 + \sum_{(i,j) \in E} \theta_{ij} x_i x_j\Big)$$

A (partial) historical overview

- Learning pairwise MRF



conditional independence:

$$(i,j) \notin E \Leftrightarrow x_i \perp x_j \mid \mathbf{x} \setminus \{x_i, x_j\}$$

probability parameterized by $\mathbf{\Theta}$:

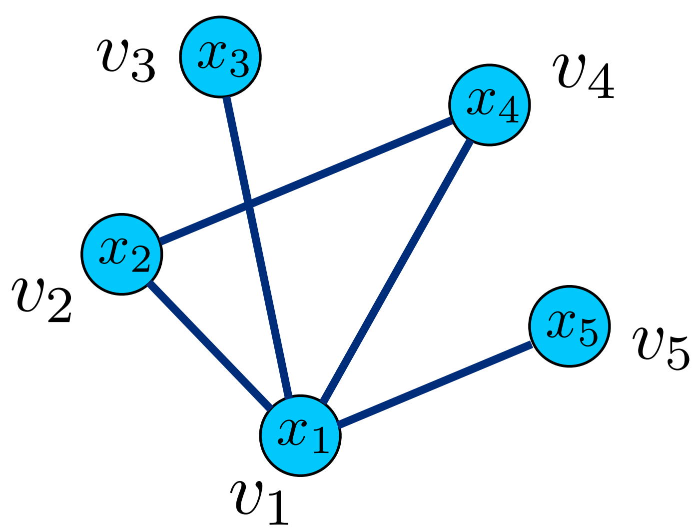
$$P(\mathbf{x}|\mathbf{\Theta}) = \frac{1}{Z(\mathbf{\Theta})} \exp\Big(\sum_{i \in V} \theta_{ii} x_i^2 + \sum_{(i,j) \in E} \theta_{ij} x_i x_j \Big)$$

Gaussian graphical models with precision $\mathbf{\Theta}$:

$$P(\mathbf{x}|\mathbf{\Theta}) = \frac{|\mathbf{\Theta}|^{1/2}}{(2\pi)^{N/2}} \exp\Big(-\frac{1}{2} \mathbf{x}^T \mathbf{\Theta} \mathbf{x} \Big)$$

A (partial) historical overview

- Learning pairwise MRF



conditional independence:

$$(i,j) \notin E \Leftrightarrow x_i \perp x_j \mid \mathbf{x} \setminus \{x_i, x_j\}$$

probability parameterized by $\boldsymbol{\Theta}$:

$$P(\mathbf{x}|\boldsymbol{\Theta}) = \frac{1}{Z(\boldsymbol{\Theta})} \exp\big(\sum_{i \in V} \theta_{ii} x_i^2 + \sum_{(i,j) \in E} \theta_{ij} x_i x_j\big)$$

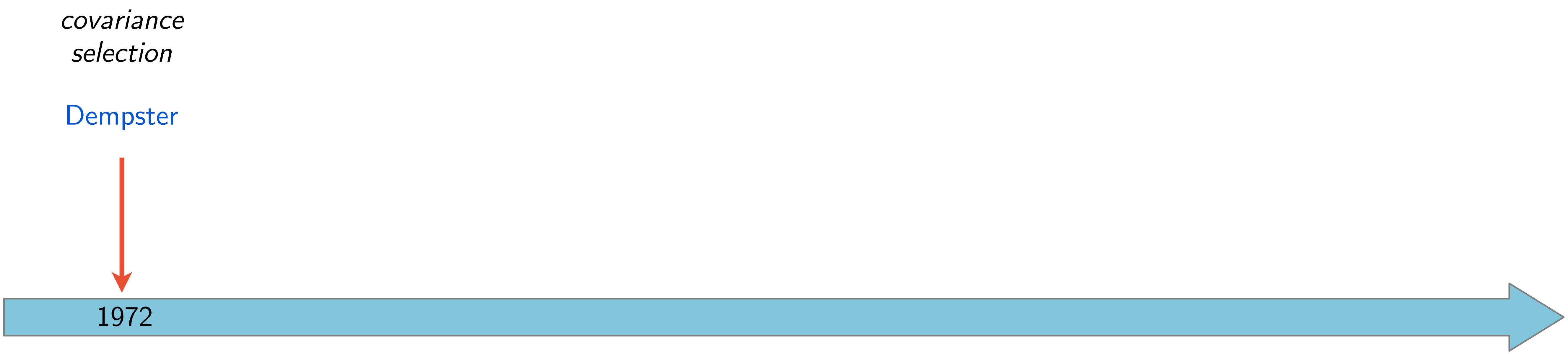
Gaussian graphical models with precision $\boldsymbol{\Theta}$:

$$P(\mathbf{x}|\boldsymbol{\Theta}) = \frac{|\boldsymbol{\Theta}|^{1/2}}{(2\pi)^{N/2}} \exp\big(-\frac{1}{2} \mathbf{x}^T \boldsymbol{\Theta} \mathbf{x}\big)$$

Learning a sparse $\boldsymbol{\Theta}$:

- **interactions are mostly local**
- **computationally more tractable**

A (partial) historical overview



Prune the smallest elements in precision (inverse covariance) matrix

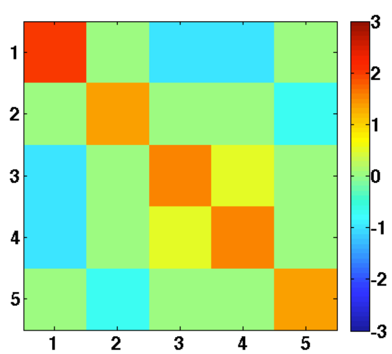
A (partial) historical overview

covariance selection

Dempster

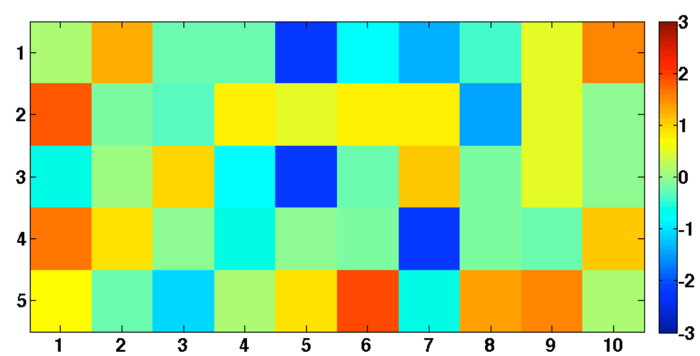
1972

Prune the smallest elements in precision (inverse covariance) matrix



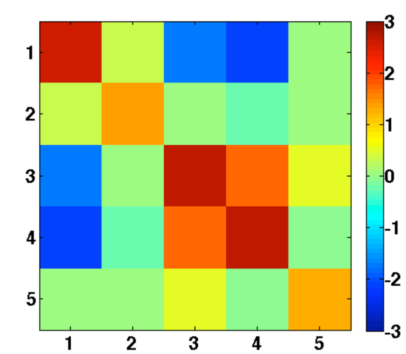
$\boldsymbol{\Theta}$

groundtruth precision



$\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Theta})$

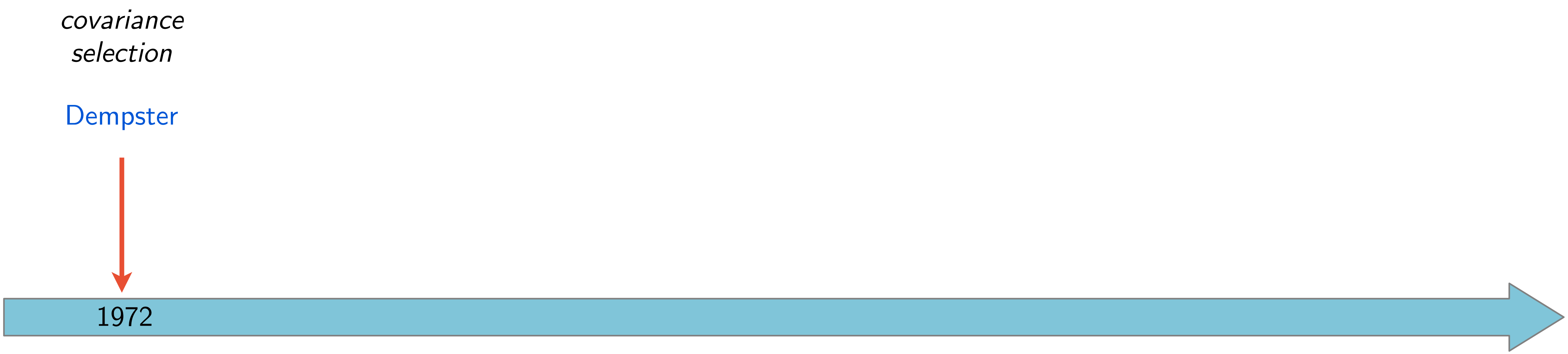
data matrix



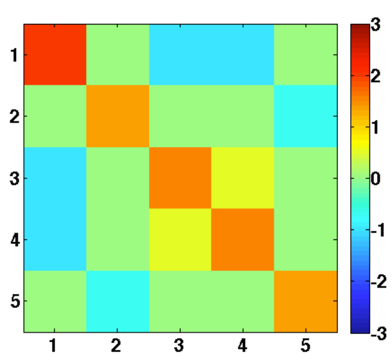
$\mathbf{S}^{-1}$

inverse of sample covariance

A (partial) historical overview

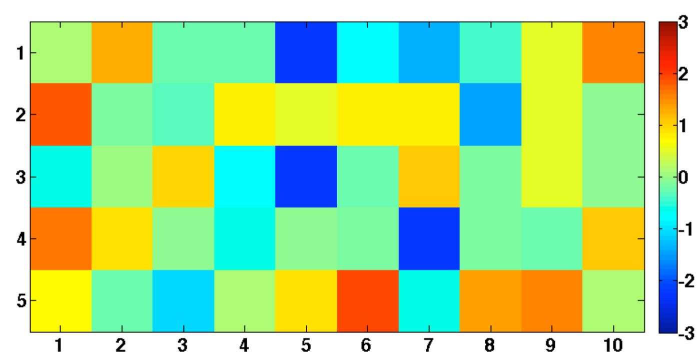


Prune the smallest elements in precision (inverse covariance) matrix



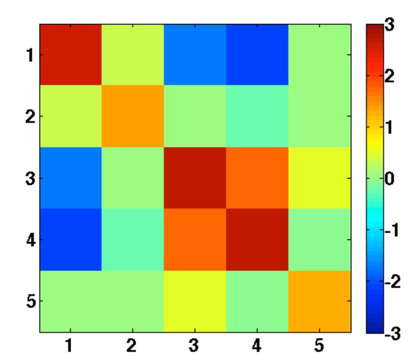
$\mathbf{\Theta}$

groundtruth precision



$\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Theta})$

data matrix

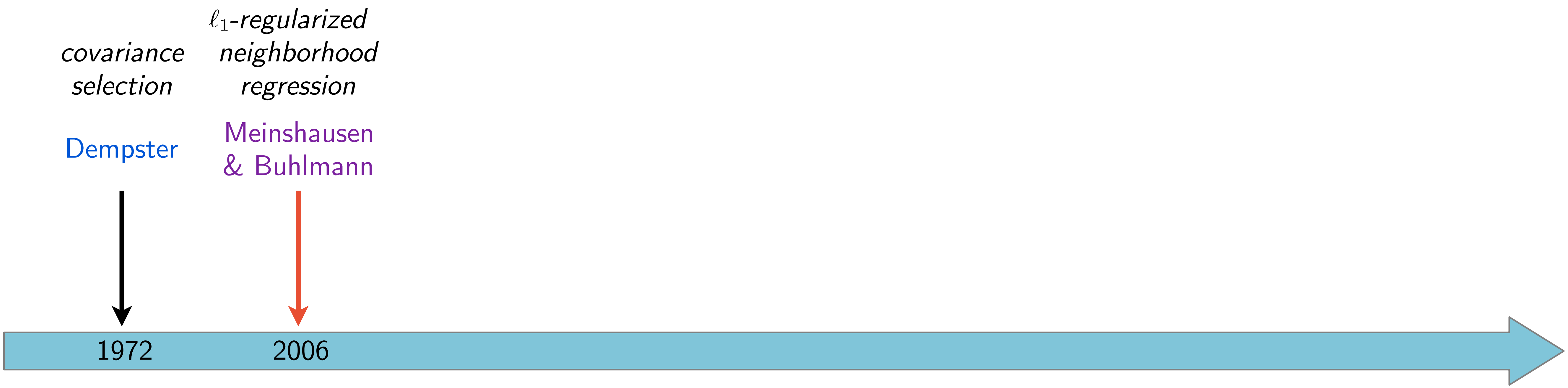


$\mathbf{S}^{-1}$

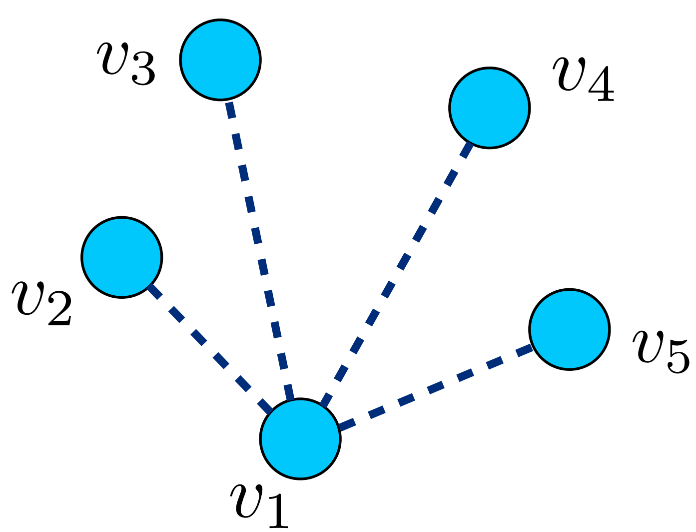
inverse of sample covariance

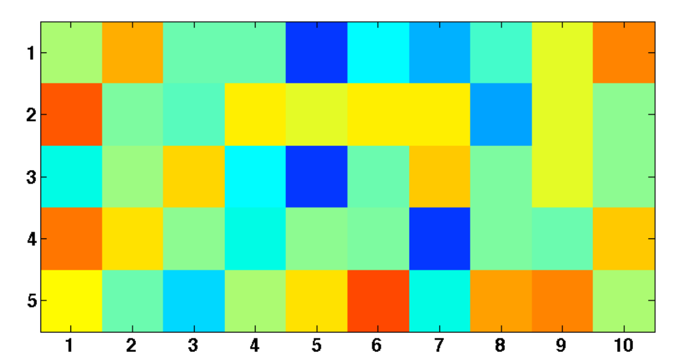
Not applicable when sample covariance is not invertible!

A (partial) historical overview

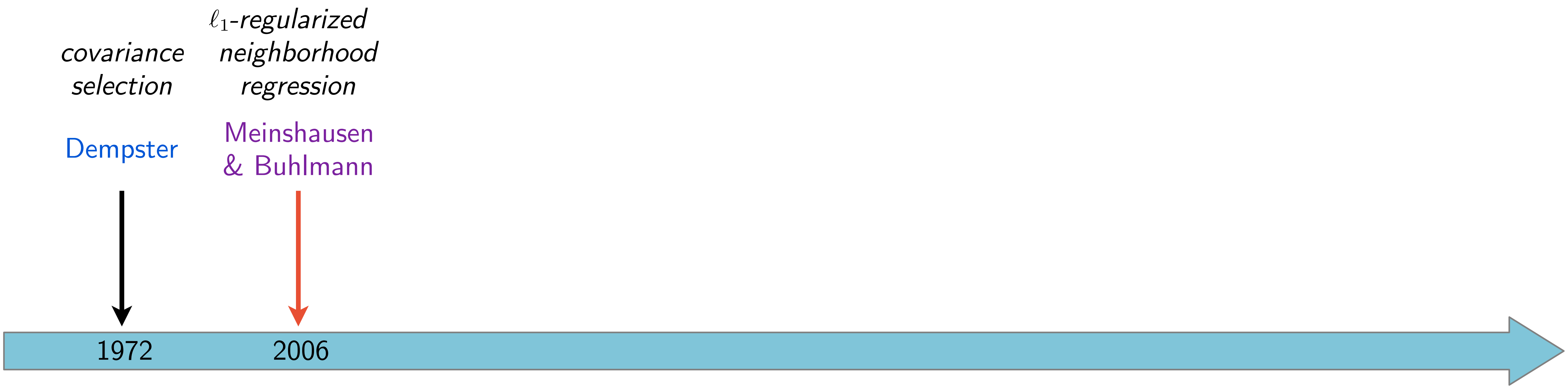


Learning a graph = learning neighborhood of each node

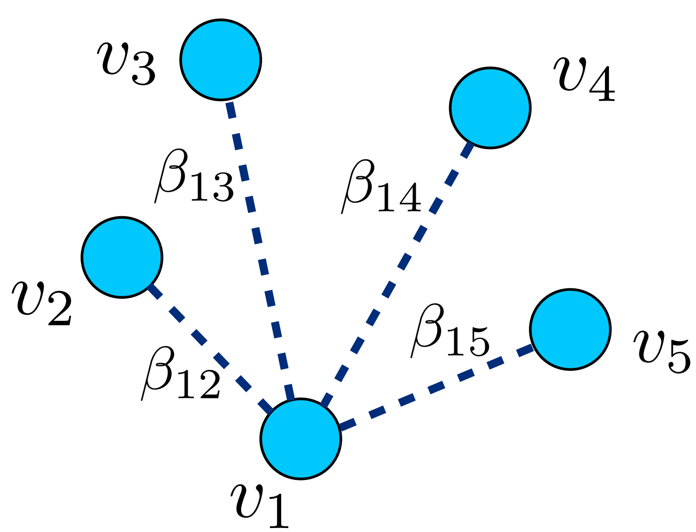


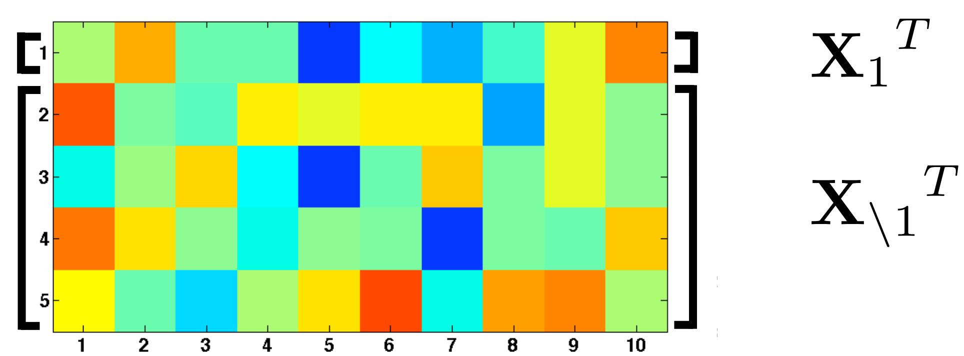


A (partial) historical overview

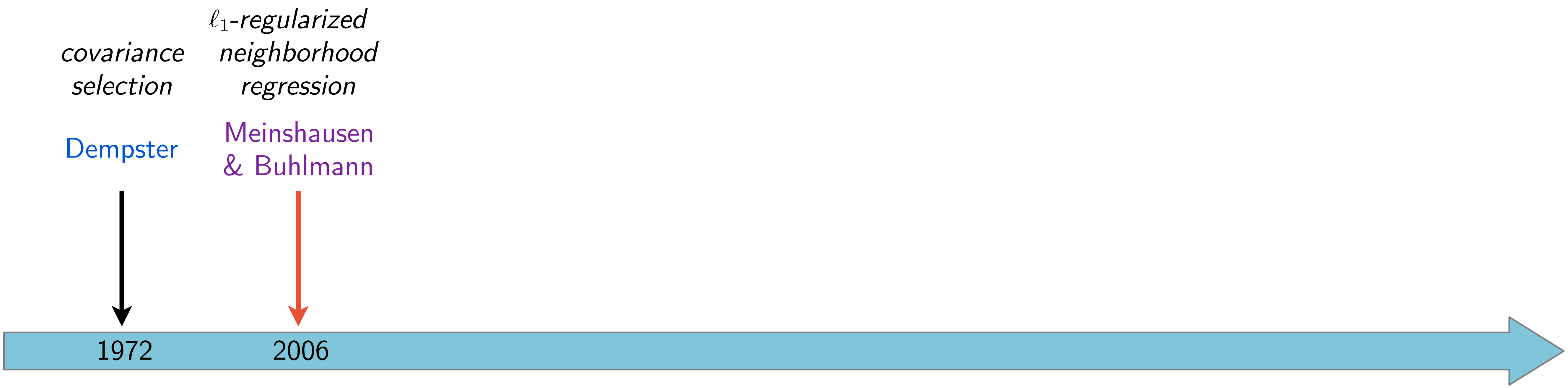


Learning a graph = learning neighborhood of each node

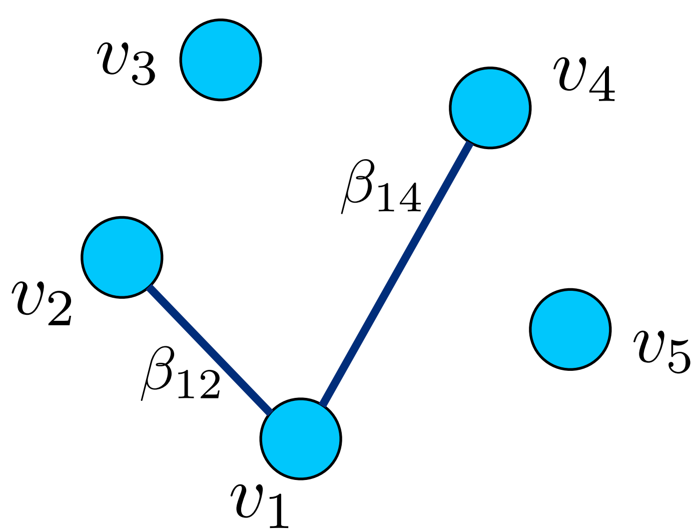


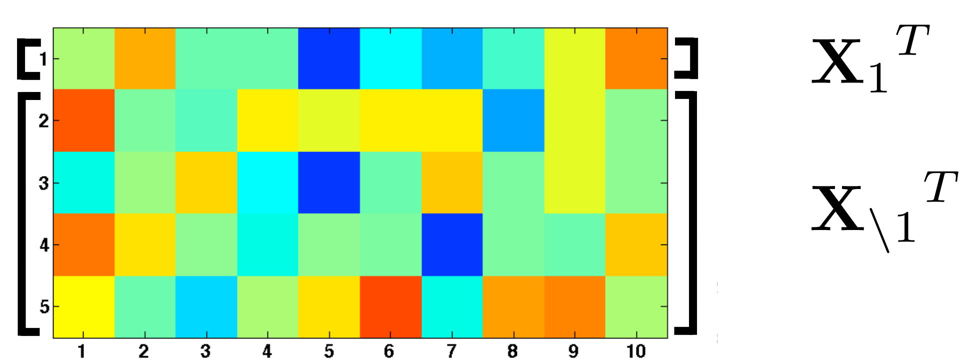


A (partial) historical overview



Learning a graph = learning neighborhood of each node

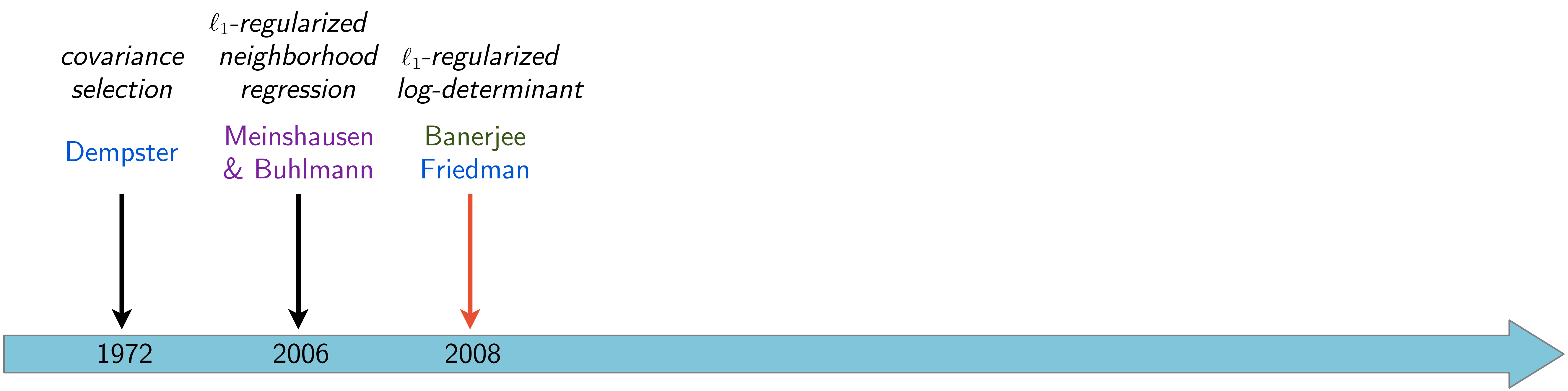




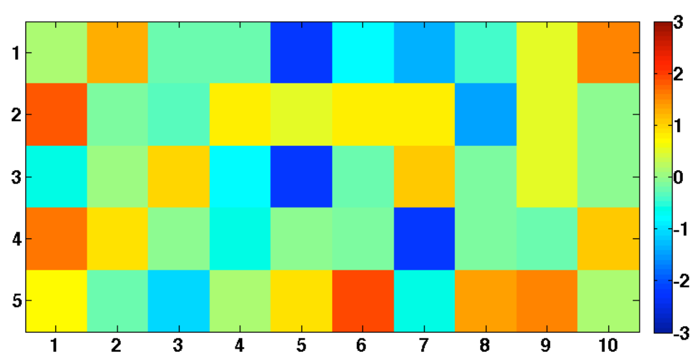
LASSO regression:

$$\min_{\boldsymbol{\beta}_1} ||\mathbf{X}_1 - \mathbf{X}_{\setminus 1}\boldsymbol{\beta}_1||^2 + \lambda||\boldsymbol{\beta}_1||_1$$

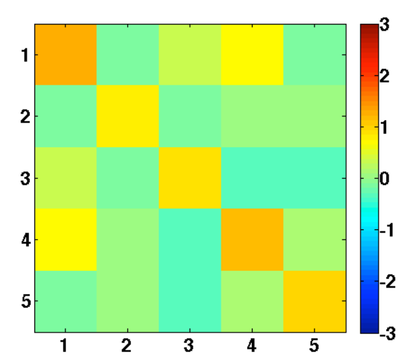
A (partial) historical overview



Estimation of sparse precision matrix

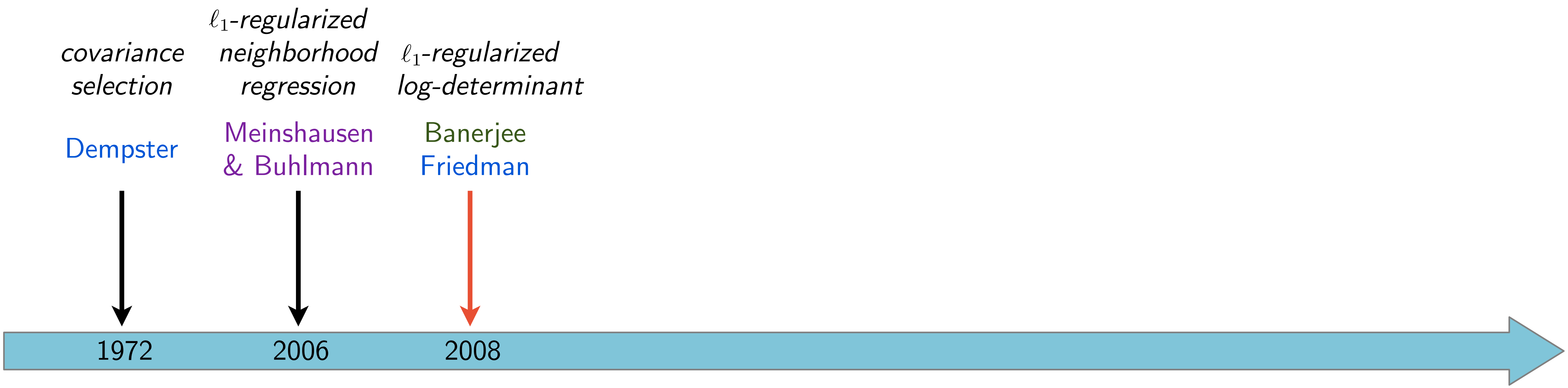


$\mathbf{X}$

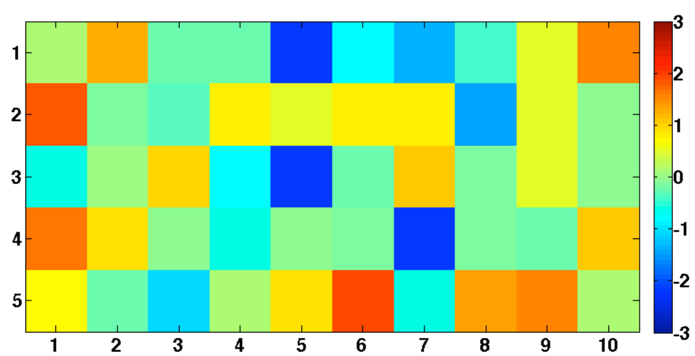


$\mathbf{S}$

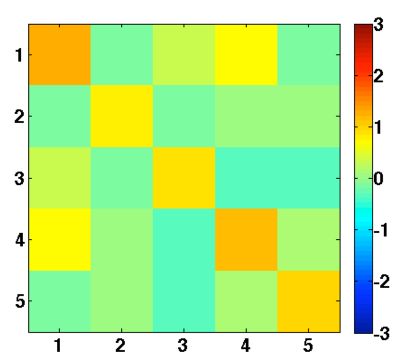
A (partial) historical overview



Estimation of sparse precision matrix



$\mathbf{X}$

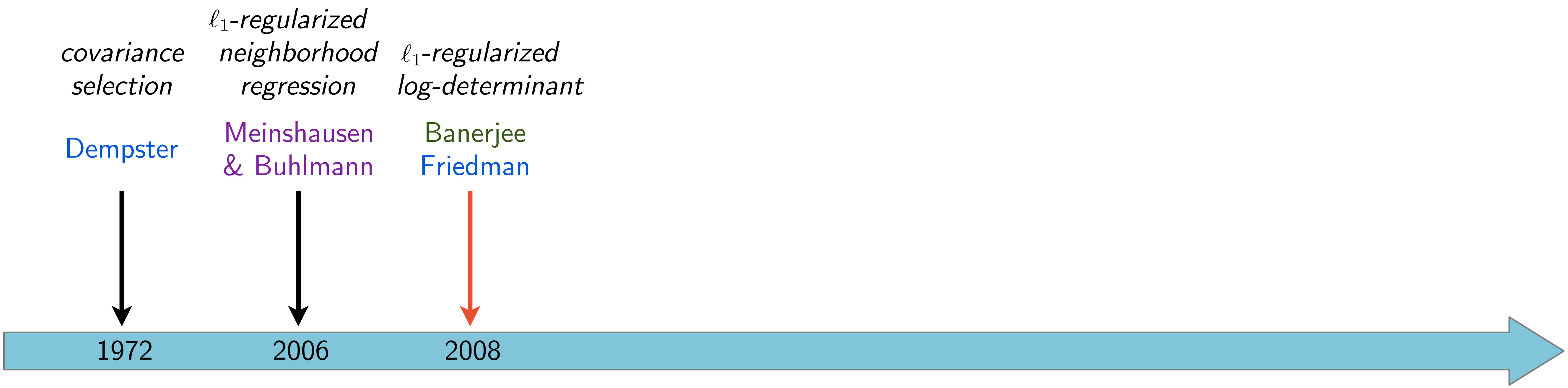


$\mathbf{S}$

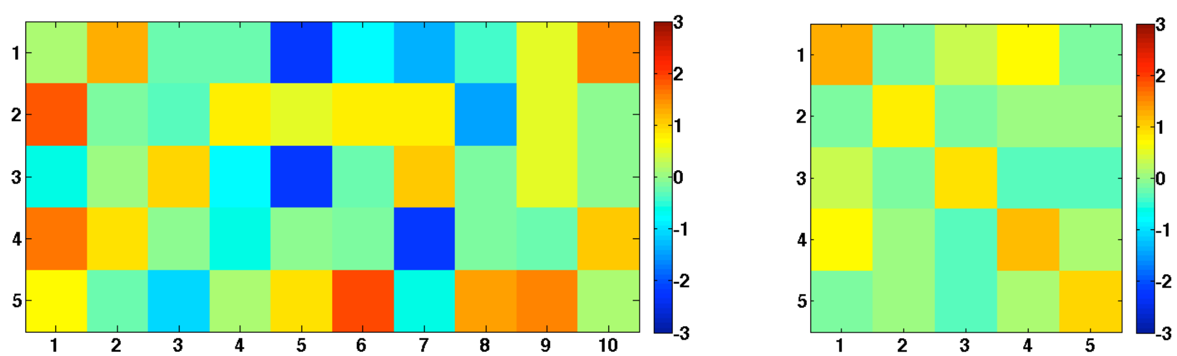
graphical LASSO maximizes likelihood of precision matrix $\boldsymbol{\Theta}$:

$$|\boldsymbol{\Theta}|^{M/2}\exp\Big(-\sum_{m=1}^{M}\frac{1}{2}\mathbf{X}(m)^T\boldsymbol{\Theta}\mathbf{X}(m)\Big)$$

A (partial) historical overview



Estimation of sparse precision matrix

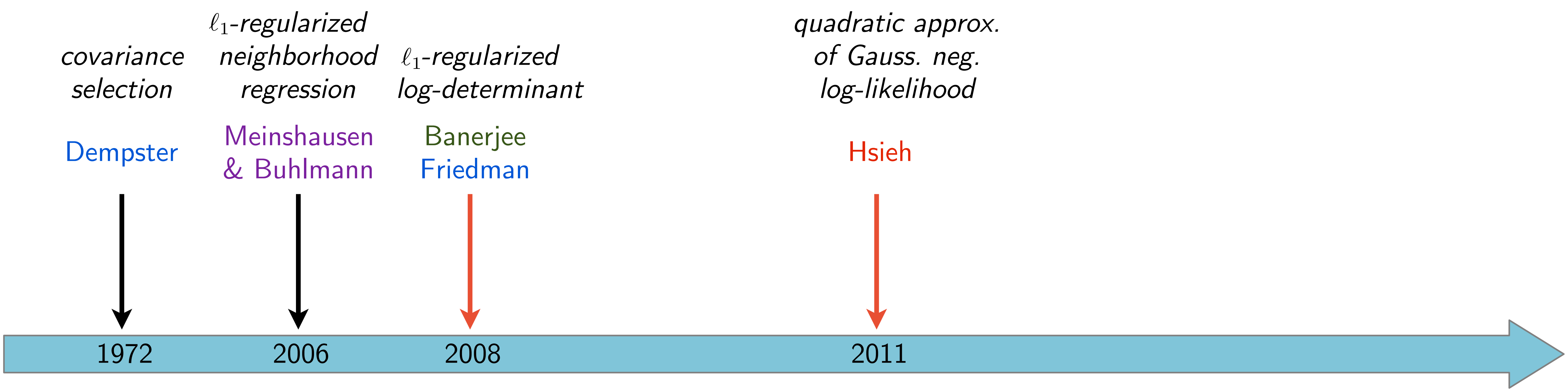


$\mathbf{X}$ $\quad$ $\mathbf{S}$

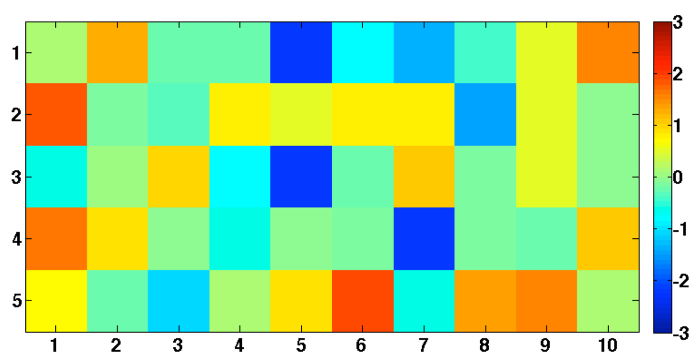
graphical LASSO maximizes likelihood of precision matrix $\mathbf{\Theta}$:

$$\max_{\mathbf{\Theta}} \underbrace{\log \det \mathbf{\Theta} - \mathrm{tr}(\mathbf{S}\mathbf{\Theta})}_{\text{log-likelihood function}} - \rho ||\mathbf{\Theta}||_1$$

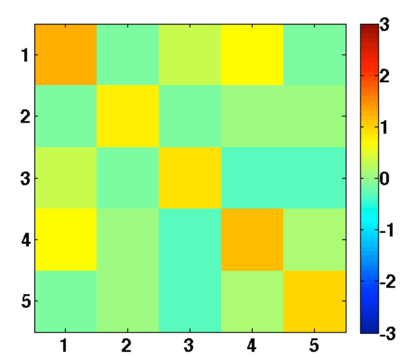
A (partial) historical overview



Estimation of sparse precision matrix



$\mathbf{X}$

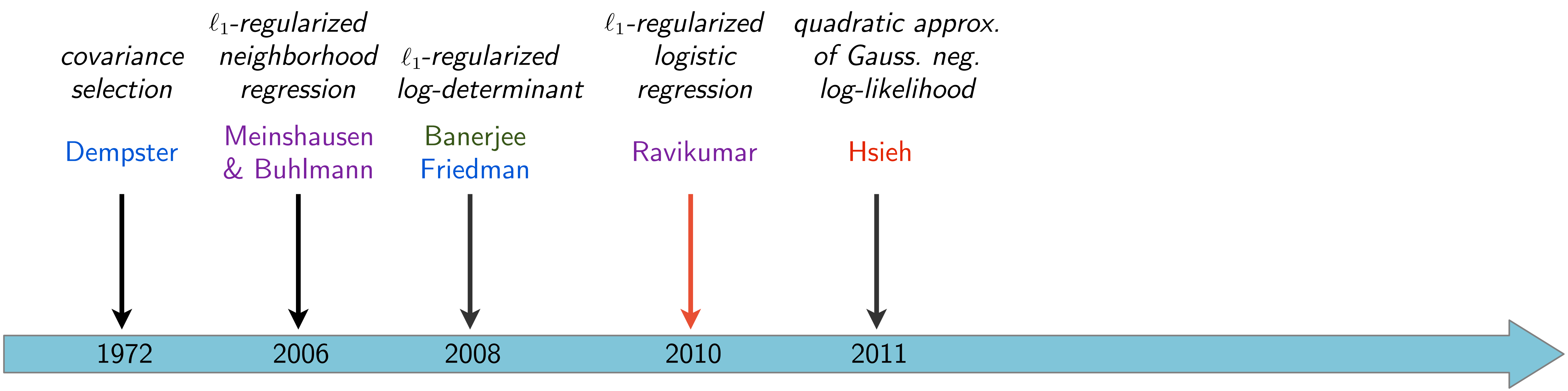


$\mathbf{S}$

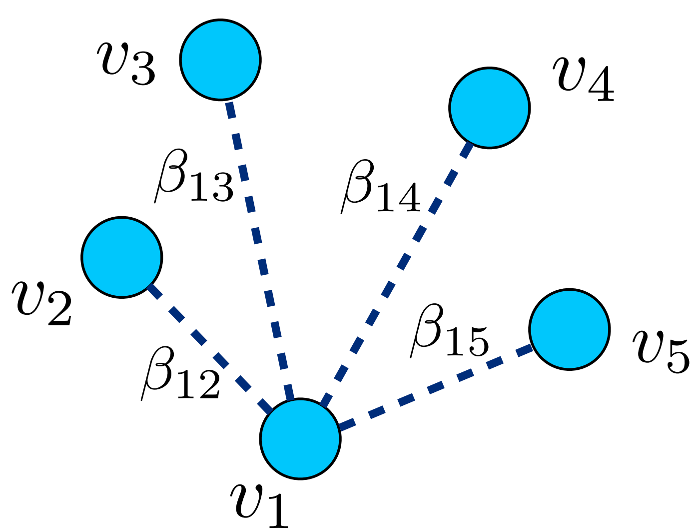
graphical LASSO maximizes likelihood of precision matrix $\boldsymbol{\Theta}$:

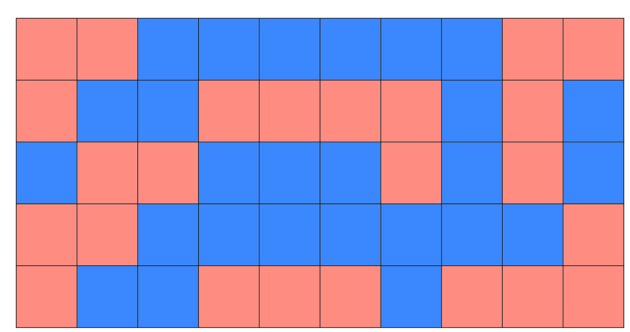
$$\max_{\boldsymbol{\Theta}} \underbrace{\log \det \boldsymbol{\Theta} - \mathrm{tr}(\mathbf{S}\boldsymbol{\Theta})}_{\text{log-likelihood function}} - \rho ||\boldsymbol{\Theta}||_1$$

A (partial) historical overview

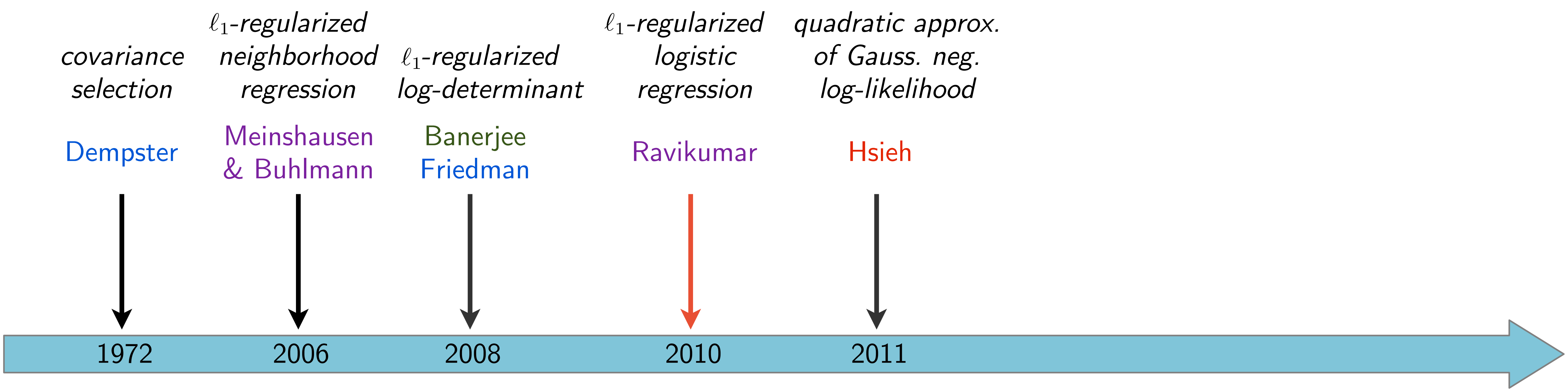


Neighborhood learning for discrete variables

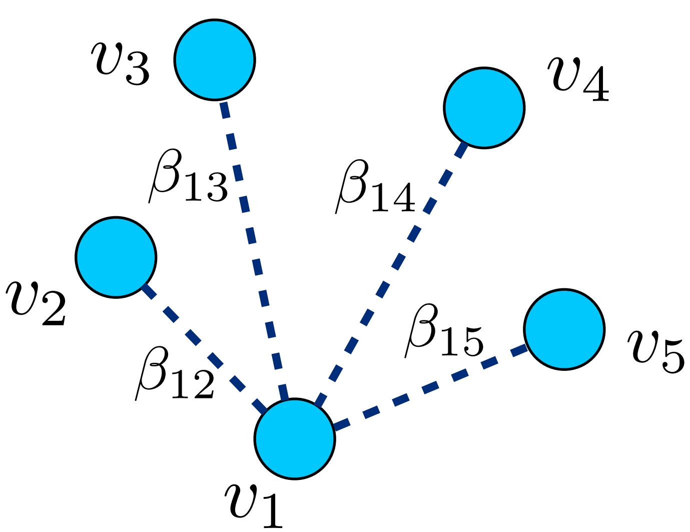


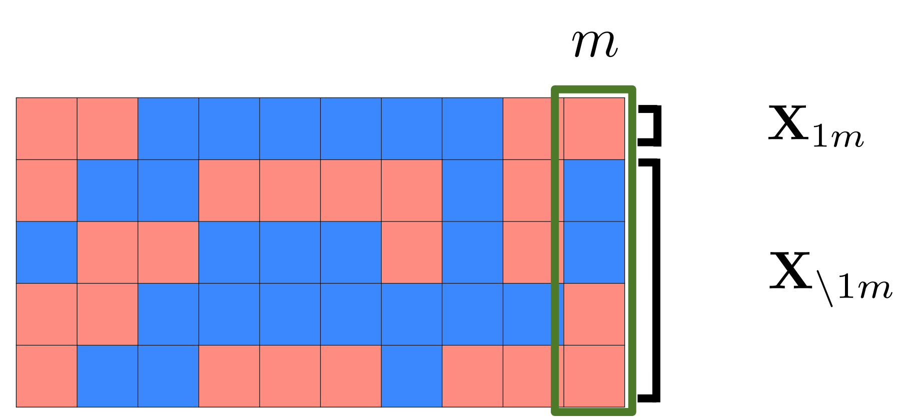


A (partial) historical overview

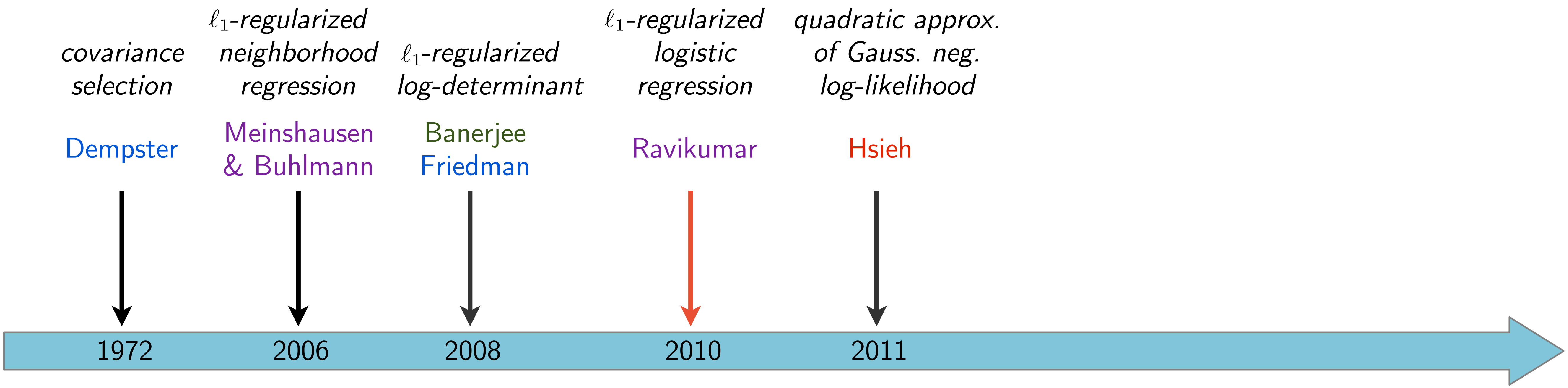


Neighborhood learning for discrete variables

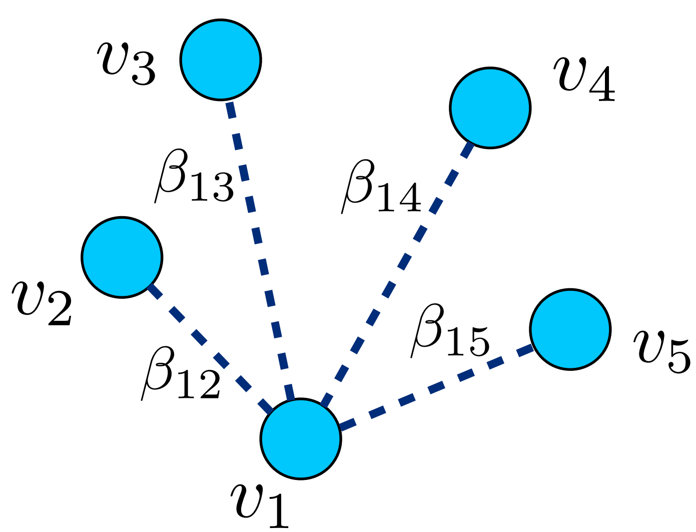


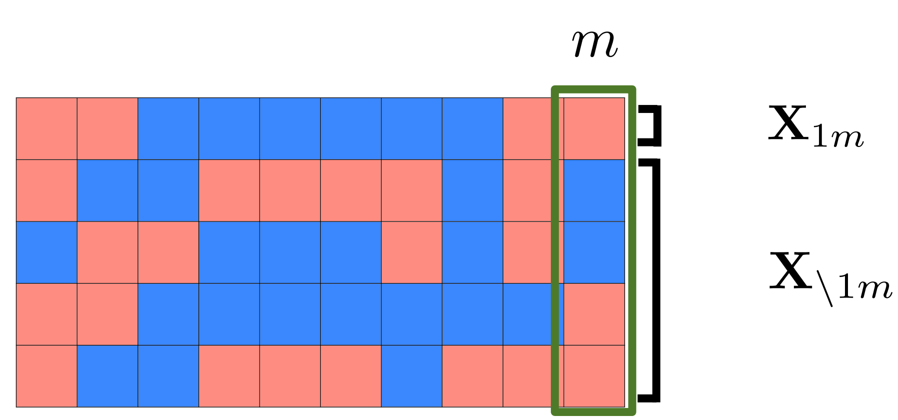


A (partial) historical overview



Neighborhood learning for discrete variables





regularized logistic regression:

$$\max_{\boldsymbol{\beta}_1} \log \underbrace{P_\beta(\mathbf{X}_{1m}|\mathbf{X}_{\backslash 1m})}_{\text{logistic function}} - \lambda||\boldsymbol{\beta}_1||_1$$

A (partial) historical overview

- Simple and intuitive methods
  - Sample correlation
  - Similarity function (e.g., Gaussian RBF)
- Learning graphical models
  - Classical learning approaches lead to both positive/negative relations
  - What about learning a graph topology with non-negative weights?

A (partial) historical overview

- Simple and intuitive methods
  - Sample correlation
  - Similarity function (e.g., Gaussian RBF)
- Learning graphical models
  - Classical learning approaches lead to both positive/negative relations
  - What about learning a graph topology with non-negative weights?
- Learning topologies with non-negative weights
  - M-matrices (sym., p.d., non-pos. off-diag.) have been used as precision, leading to attractive GMRF (Slawski and Hein 2015)

A (partial) historical overview

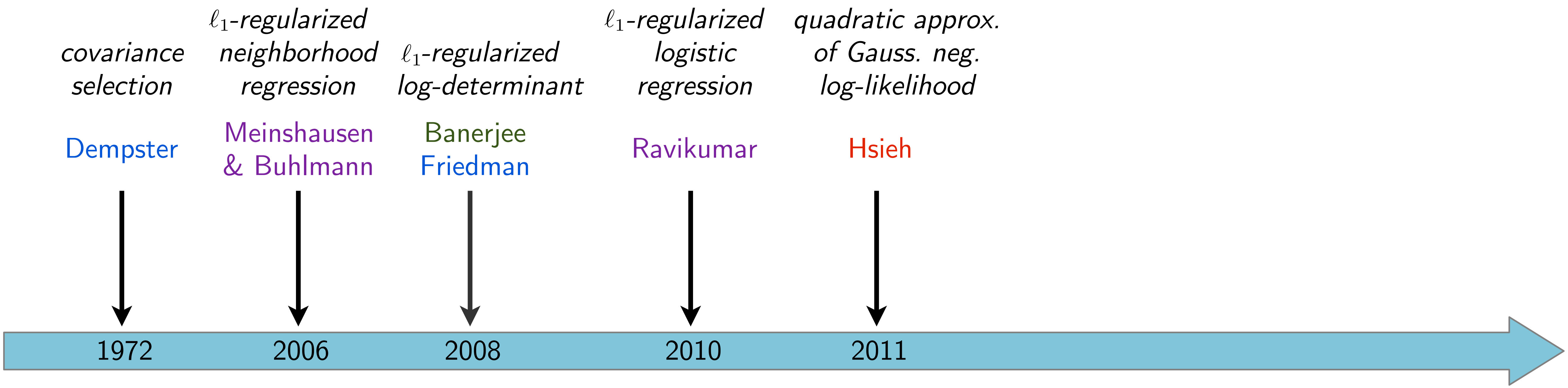
- Simple and intuitive methods
  - Sample correlation
  - Similarity function (e.g., Gaussian RBF)
- Learning graphical models
  - Classical learning approaches lead to both positive/negative relations
  - What about learning a graph topology with non-negative weights?
- Learning topologies with non-negative weights
  - M-matrices (sym., p.d., non-pos. off-diag.) have been used as precision, leading to attractive GMRF (Slawski and Hein 2015)
  - The combinatorial graph Laplacian $\mathbf{L} = \mathbf{Deg} - \mathbf{W}$ belongs to M-matrices and is equivalent to graph topology

A (partial) historical overview

- Simple and intuitive methods
  - Sample correlation
  - Similarity function (e.g., Gaussian RBF)
- Learning graphical models
  - Classical learning approaches lead to both positive/negative relations
  - What about learning a graph topology with non-negative weights?
- Learning topologies with non-negative weights
  - M-matrices (sym., p.d., non-pos. off-diag.) have been used as precision, leading to attractive GMRF (Slawski and Hein 2015)
  - The combinatorial graph Laplacian $\mathbf{L} = \mathbf{Deg} - \mathbf{W}$ belongs to M-matrices and is equivalent to graph topology

From arbitrary precision matrix to graph Laplacian!

A (partial) historical overview



$$\max_{\mathbf{\Theta}} \log \det \mathbf{\Theta} - \mathrm{tr}(\mathbf{S}\mathbf{\Theta}) - \rho||\mathbf{\Theta}||_1$$

graph Laplacian **L** can be the precision,
BUT it is singular

A (partial) historical overview

covariance selection
Dempster
1972

ℓ_1-regularized neighborhood regression
Meinshausen & Buhlmann
2006

ℓ_1-regularized log-determinant
Banerjee
Friedman
2008

ℓ_1-regularized logistic regression
Ravikumar
2010

quadratic approx. of Gauss. neg. log-likelihood
Hsieh
2011

Lake

ℓ_1-regularized log-determinant on generalized ***L***

$$\max_{\boldsymbol{\Theta}} \log \det \boldsymbol{\Theta} - \mathrm{tr}(\mathbf{S}\boldsymbol{\Theta}) - \rho ||\boldsymbol{\Theta}||_1$$

$$\text{s.t. } \boldsymbol{\Theta} = \mathbf{L} + \frac{1}{\sigma^2}\mathbf{I}$$

graph Laplacian **L** can be the precision,
BUT it is singular

A (partial) historical overview

1972	2006	2008	2010	2011
covariance selection	*ℓ_1-regularized neighborhood regression*	*ℓ_1-regularized log-determinant*	*ℓ_1-regularized logistic regression*	*quadratic approx. of Gauss. neg. log-likelihood*
Dempster	Meinshausen & Buhlmann	Banerjee Friedman	Ravikumar	Hsieh

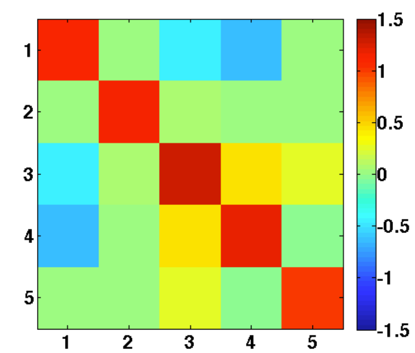
$$\max_{\boldsymbol{\Theta}} \log \det \boldsymbol{\Theta} - \mathrm{tr}(\mathbf{S}\boldsymbol{\Theta}) - \rho ||\boldsymbol{\Theta}||_1$$

$$\text{s.t. } \boldsymbol{\Theta} = \mathbf{L} + \frac{1}{\sigma^2}\mathbf{I}$$

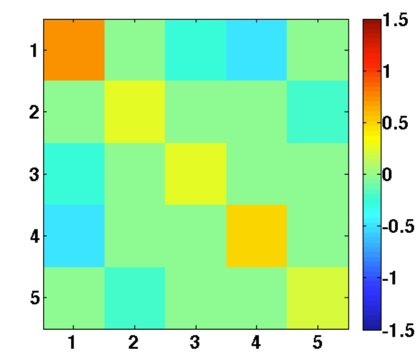
graph Laplacian **L** can be the precision, BUT it is singular

Lake (2010)

*ℓ_1-regularized log-determinant on generalized **L***



precision by graphical LASSO



Laplacian by Lake et al.

A (partial) historical overview

1972	*covariance selection*	Dempster
2006	*ℓ_1-regularized neighborhood regression*	Meinshausen & Buhlmann
2008	*ℓ_1-regularized log-determinant*	Banerjee Friedman
2010	*ℓ_1-regularized logistic regression*	Ravikumar
2011	*quadratic approx. of Gauss. neg. log-likelihood*	Hsieh

$$\max_{\mathbf{\Theta}} \log \det \mathbf{\Theta} - \mathrm{tr}(\mathbf{S}\mathbf{\Theta}) - \rho||\mathbf{\Theta}||_1$$

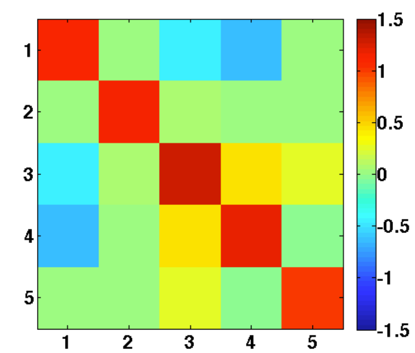
$$\text{s.t. } \mathbf{\Theta} = \mathbf{L} + \frac{1}{\sigma^2}\mathbf{I}$$

graph Laplacian **L** can be the precision, BUT it is singular

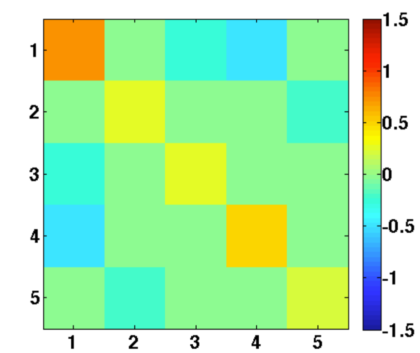
Lake

*ℓ_1-regularized log-determinant on generalized **L***

Slawski and Hein (2015)

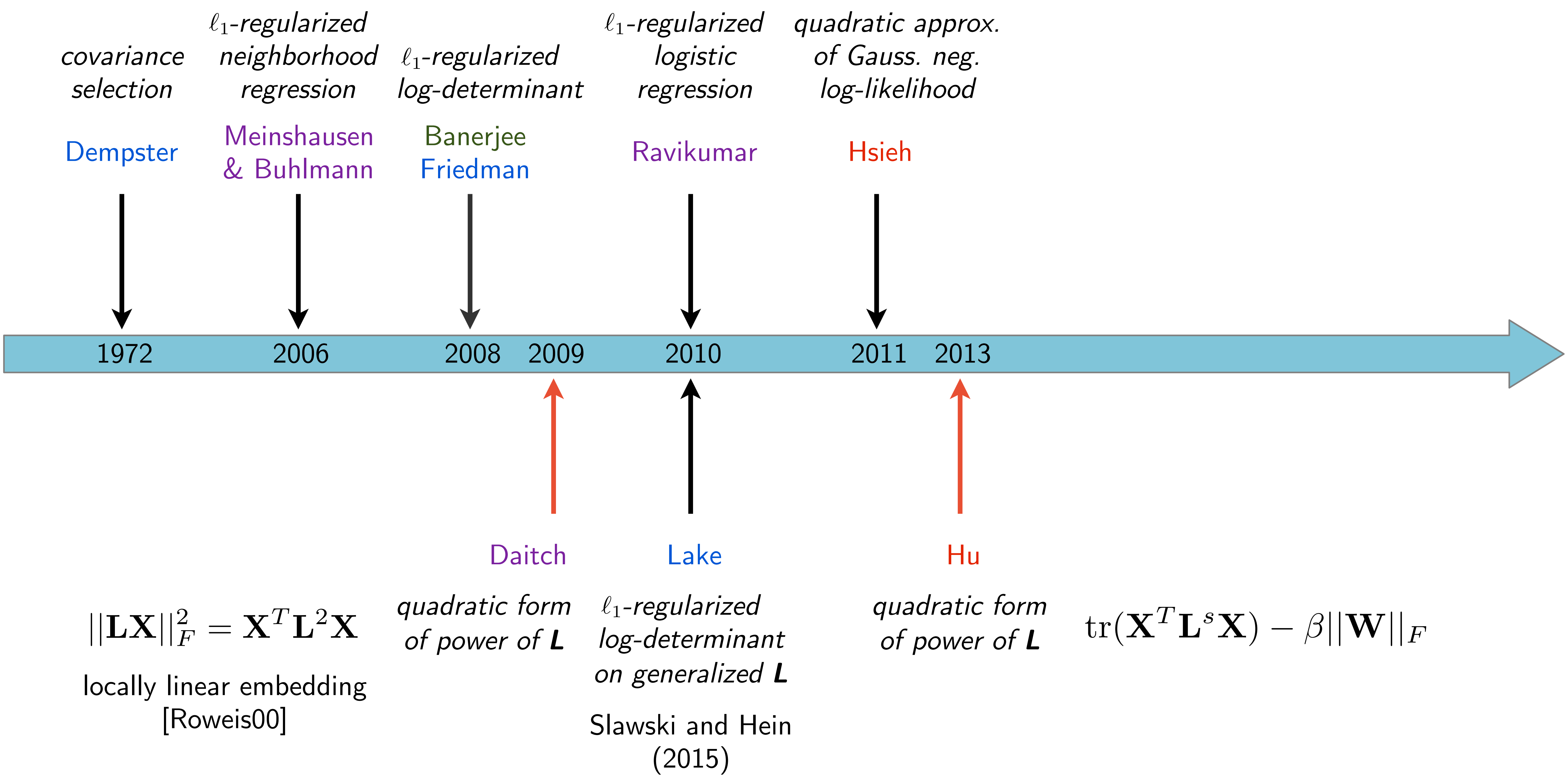


precision by graphical LASSO

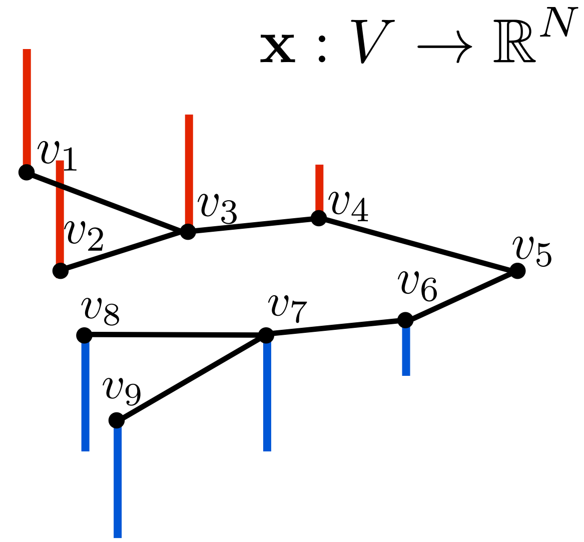


Laplacian by Lake et al.

A (partial) historical overview

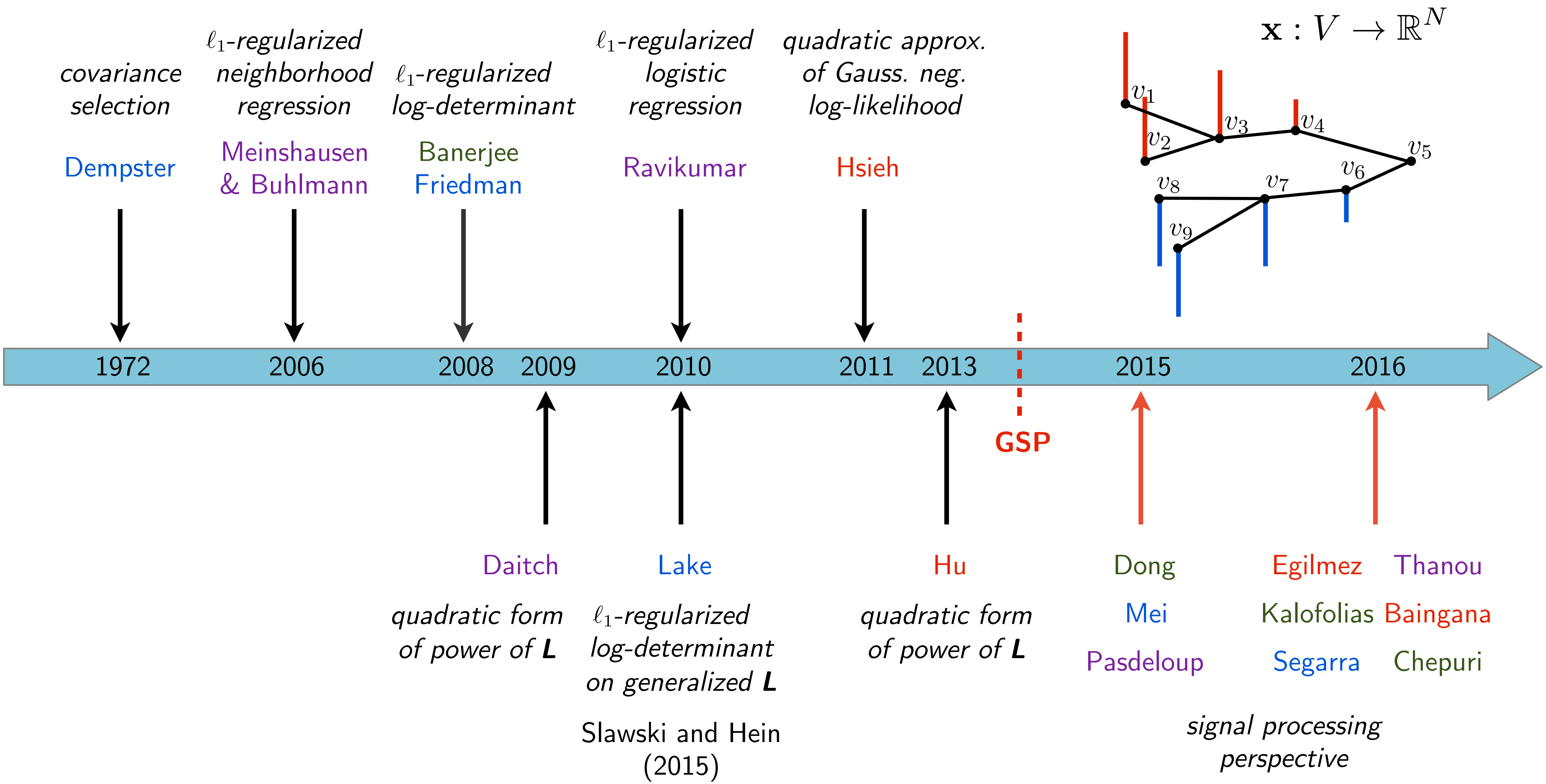


A (partial) historical overview



- 1972 — Dempster: *covariance selection*
- 2006 — Meinshausen & Buhlmann: *ℓ_1-regularized neighborhood regression*
- 2008 — Banerjee, Friedman: *ℓ_1-regularized log-determinant*
- 2009 — Daitch: *quadratic form of power of **L***
- 2010 — Ravikumar: *ℓ_1-regularized logistic regression*
- 2010 — Lake: *ℓ_1-regularized log-determinant on generalized **L***
  Slawski and Hein (2015)
- 2011 — Hsieh: *quadratic approx. of Gauss. neg. log-likelihood*
- 2013 — Hu: *quadratic form of power of **L***
- GSP

A (partial) historical overview



A signal processing perspective

- Existing approaches have limitations
  - Simple correlation or similarity functions are not enough
  - Most classical methods for learning graphical models do not directly lead to topologies with non-negative weights
  - There is no strong emphasis on signal/graph interaction with spectral/frequency-domain interpretation

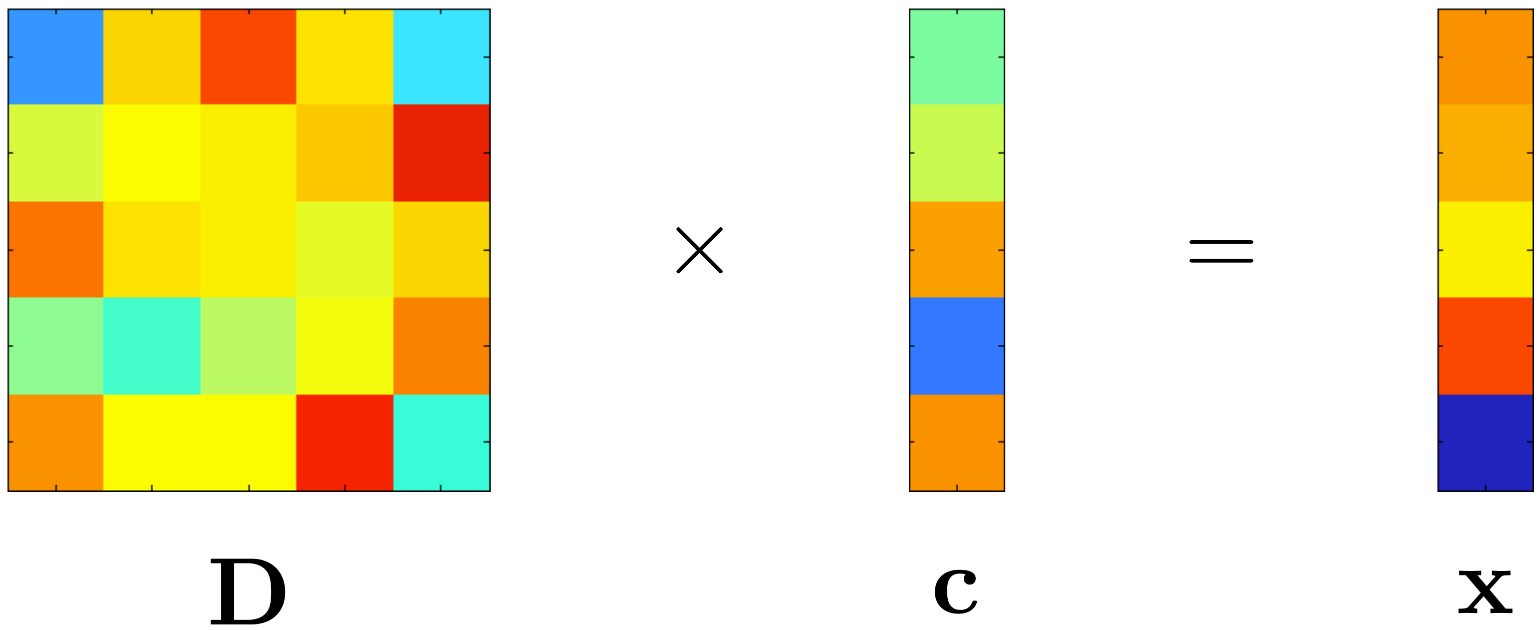
A signal processing perspective

- Existing approaches have limitations
  - Simple correlation or similarity functions are not enough
  - Most classical methods for learning graphical models do not directly lead to topologies with non-negative weights
  - There is no strong emphasis on signal/graph interaction with spectral/frequency-domain interpretation

- Opportunity and challenge for graph signal processing
  - GSP tools such as frequency-analysis and filtering can contribute to the graph learning problem
  - Filtering-based approaches can provide generative models for signals with complex non-Gaussian behavior

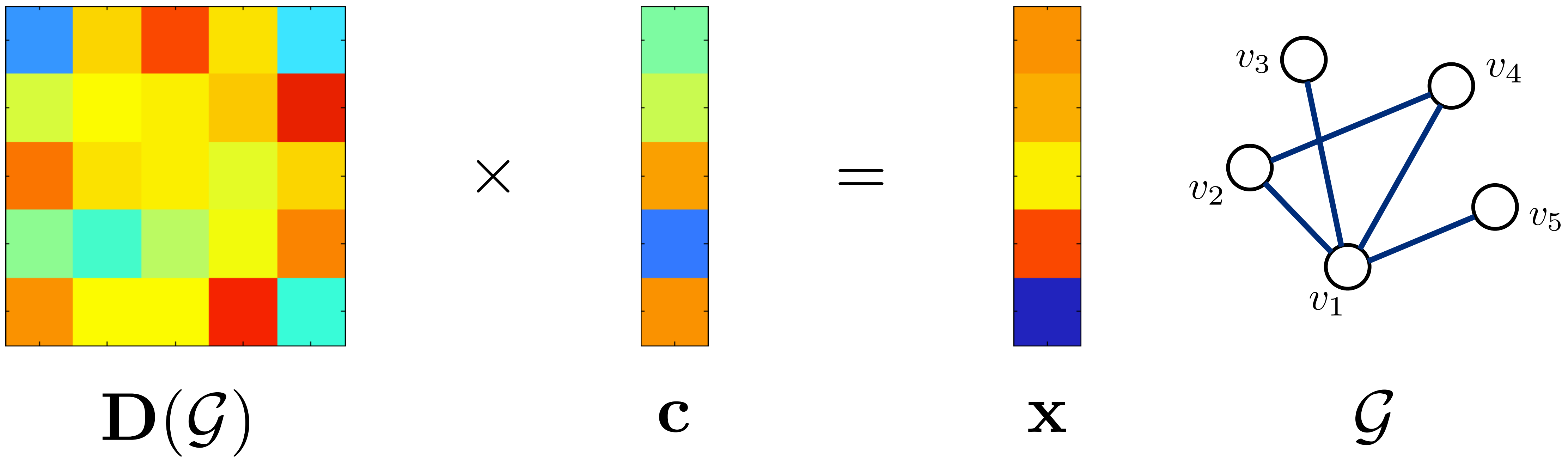
A signal processing perspective

- Signal processing is about $\mathbf{D}\ \mathbf{c} = \mathbf{x}$



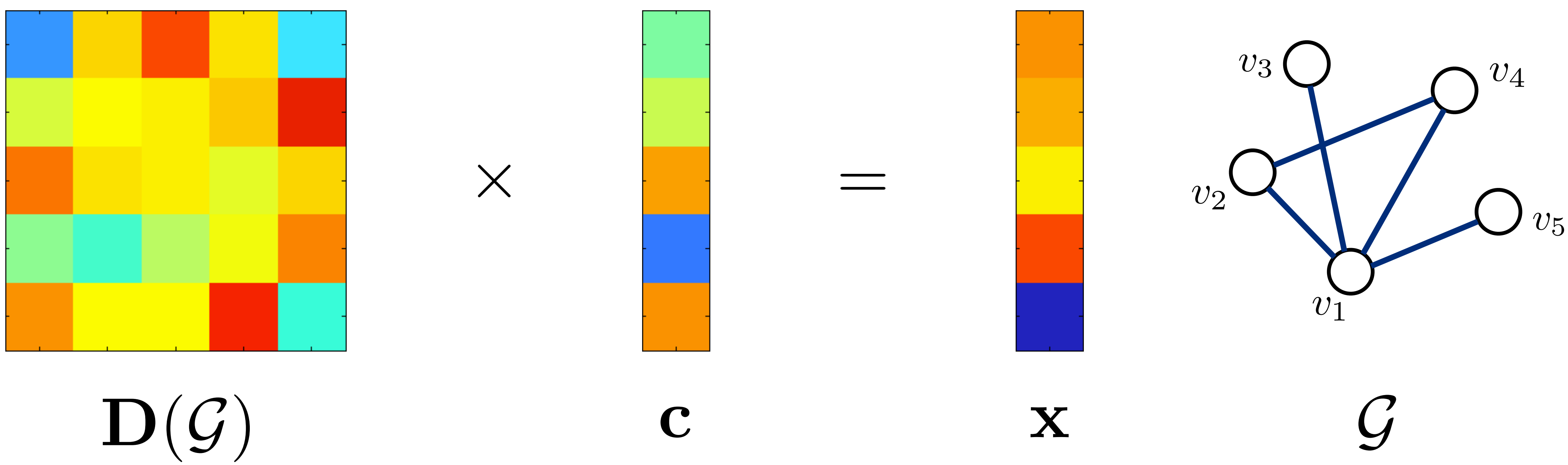
A signal processing perspective

- Graph signal processing is about $\mathbf{D}(\mathbf{G})\ \mathbf{c} = \mathbf{x}$



A signal processing perspective

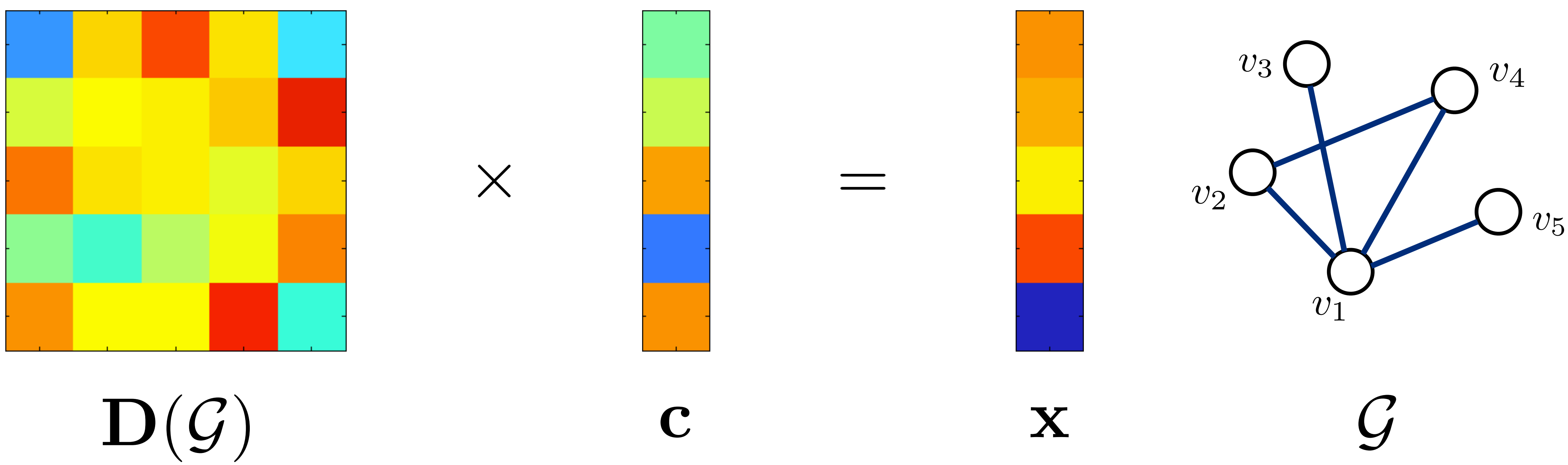
- Forward: Given **G** and **x**, design **D** to study **c**



Fourier/wavelet atoms	graph Fourier/ wavelet coefficient	[Coifman06,Narang09,Hammond11, Shuman13,Sandryhaila13]
trained dictionary atoms	graph dictionary coefficient	[Zhang12,Thanou14]

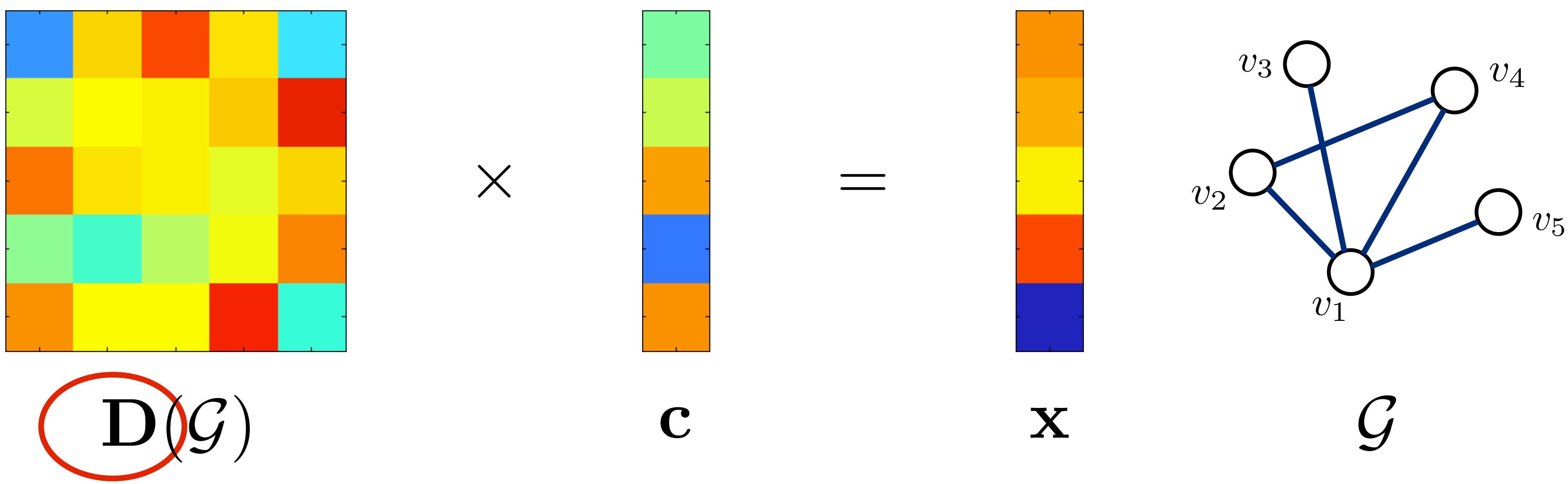
A signal processing perspective

- Backward (graph learning): Given **x**, design **D** and **c** to infer **G**



A signal processing perspective

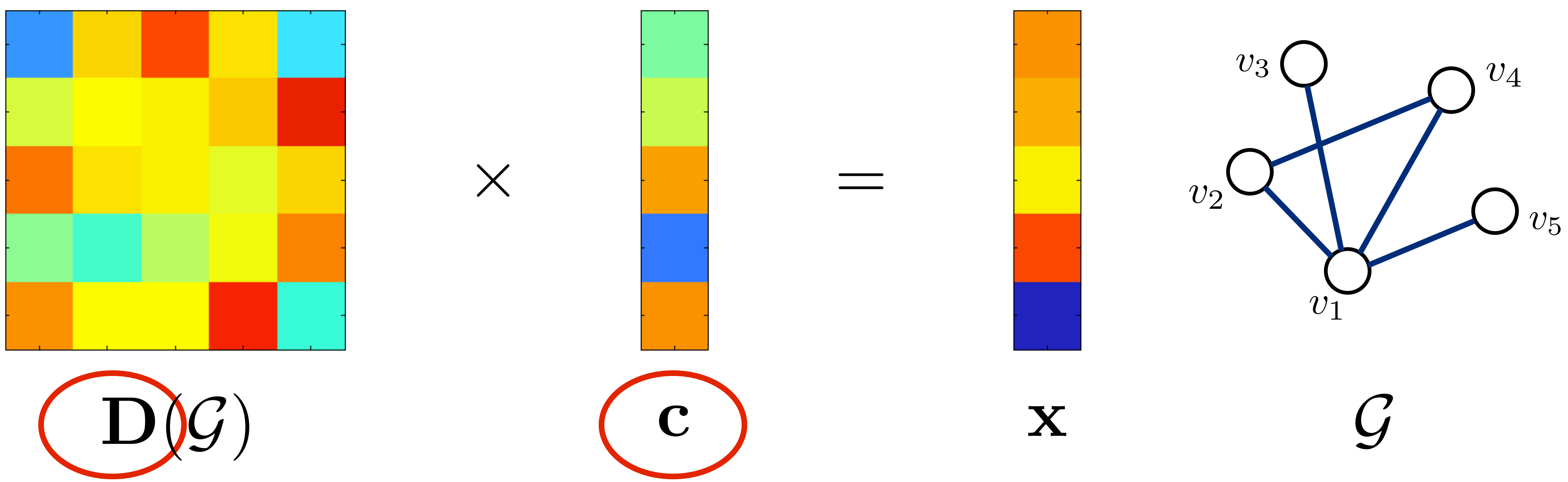
- Backward (graph learning): Given **x**, design **D** and **c** to infer **G**



- The key is a signal/graph model behind **D**
- Designed around graph operators (adjacency/Laplacian matrices, shift operators)

A signal processing perspective

- Backward (graph learning): Given **x**, design **D** and **c** to infer **G**



- The key is a signal/graph model behind **D**
- Designed around graph operators (adjacency/Laplacian matrices, shift operators)
- Choice of/assumption on **c** often determines signal characteristics

Model 1: Global smoothness

- Signal values vary smoothly between all pairs of nodes that are connected
- Example: Temperature of different locations in a flat geographical region
- Usually quantified by the Laplacian quadratic form:

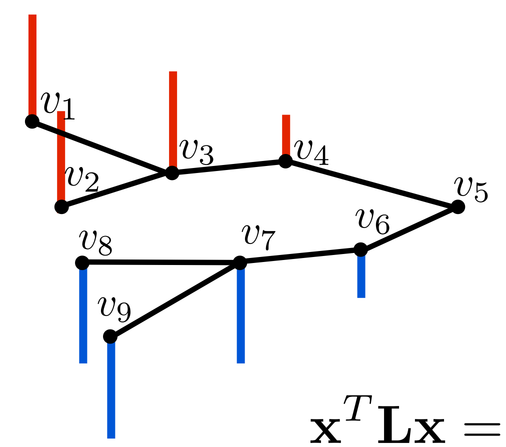
$$\mathbf{x}^T\mathbf{L}\mathbf{x} = \frac{1}{2}\sum_{i,j}\mathbf{W}_{ij}\left(\mathbf{x}(i) - \mathbf{x}(j)\right)^2$$

Model 1: Global smoothness

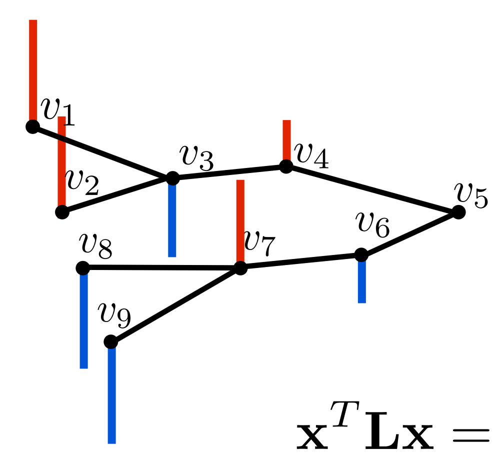
- Signal values vary smoothly between all pairs of nodes that are connected
- Example: Temperature of different locations in a flat geographical region
- Usually quantified by the Laplacian quadratic form:

$$\mathbf{x}^T\mathbf{L}\mathbf{x} = \frac{1}{2}\sum_{i,j}\mathbf{W}_{ij}\left(\mathbf{x}(i) - \mathbf{x}(j)\right)^2$$

$\mathbf{x} : V \to \mathbb{R}^N$



$\mathbf{x}^T\mathbf{L}\mathbf{x} = 1$

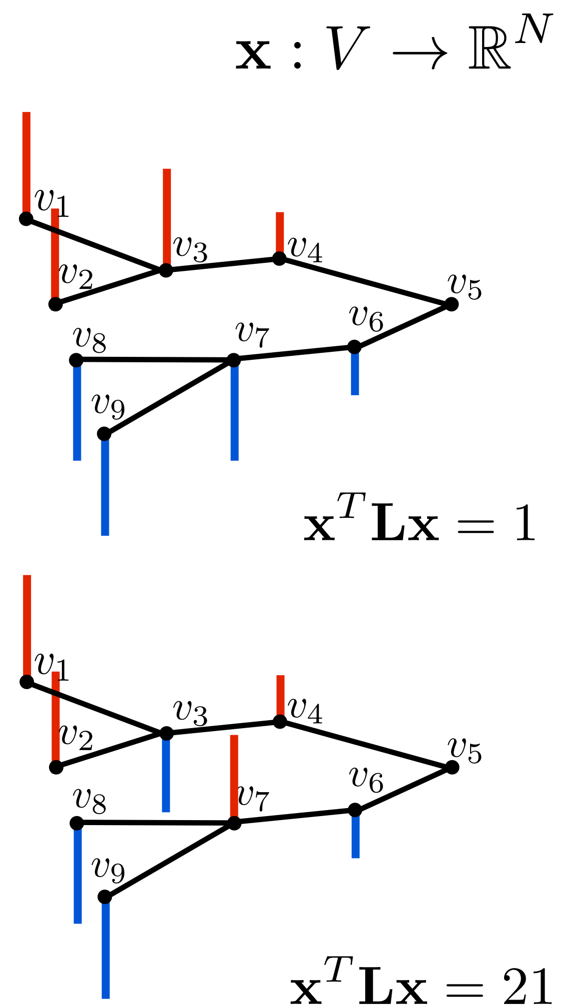


$\mathbf{x}^T\mathbf{L}\mathbf{x} = 21$

Model 1: Global smoothness

- Signal values vary smoothly between all pairs of nodes that are connected
- Example: Temperature of different locations in a flat geographical region
- Usually quantified by the Laplacian quadratic form:

$$\mathbf{x}^T\mathbf{L}\mathbf{x} = \frac{1}{2}\sum_{i,j}\mathbf{W}_{ij}\left(\mathbf{x}(i) - \mathbf{x}(j)\right)^2$$



Similar to previous approaches:

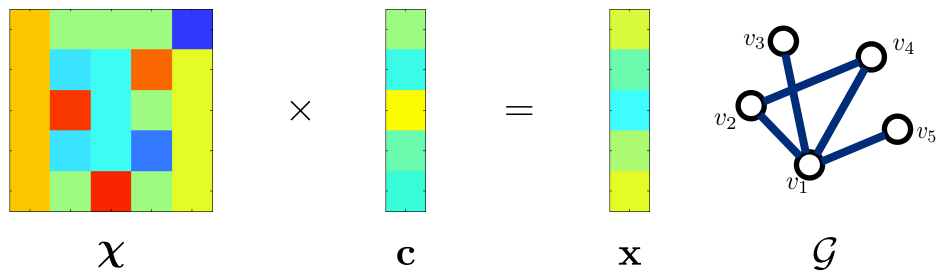
Lake (2010): $\max_{\mathbf{\Theta}=\mathbf{L}+\frac{1}{\sigma^2}\mathbf{I}} \log\det\mathbf{\Theta} - \frac{1}{M}\mathrm{tr}(\mathbf{X}\mathbf{X}^T\mathbf{\Theta}) - \rho||\mathbf{\Theta}||_1$

Daitch (2009): $\min_{\mathbf{L}} \mathbf{X}^T\mathbf{L}^2\mathbf{X}$

Hu (2013): $\min_{\mathbf{L}} \mathrm{tr}(\mathbf{X}^T\mathbf{L}^s\mathbf{X}) - \beta||\mathbf{W}||_F$

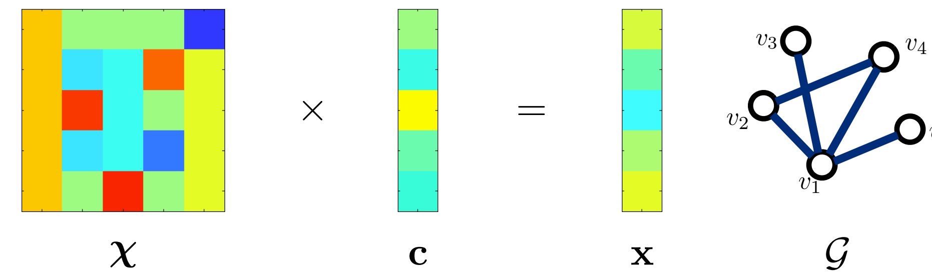
Model 1: Global smoothness

- Dong et al. (2015) & Kalofolias (2016)
  - $\mathbf{D}(\mathcal{G}) = \boldsymbol{\chi}$ (eigenvector matrix of **L**)
  - Gaussian assumption on $\mathbf{c}$: $\mathbf{c} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda})$



Model 1: Global smoothness

- Dong et al. (2015) & Kalofolias (2016)



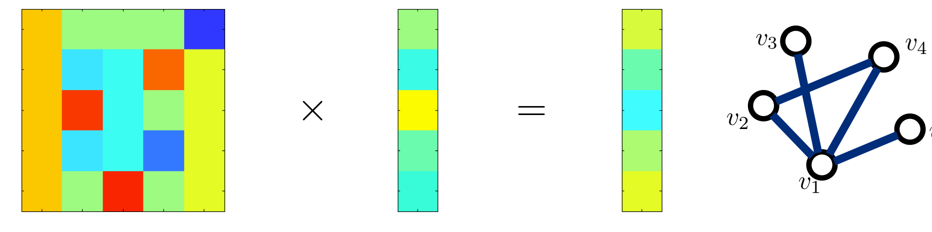
  - $\mathbf{D}(\mathcal{G}) = \boldsymbol{\chi}$ (eigenvector matrix of **L**)
  - Gaussian assumption on **c**: $\mathbf{c} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda})$
  - Maximum a posterior (MAP) estimation on **c** leads to minimization of Laplacian quadratic form:

$$\min_{c} ||\mathbf{x} - \boldsymbol{\chi}\mathbf{c}||_2^2 - \log P_c(\mathbf{c})$$

$$\min_{\mathbf{L},\mathbf{Y}} \underbrace{||\mathbf{X} - \mathbf{Y}||_F^2}_{\text{data fidelity}} + \alpha \underbrace{\mathrm{tr}(\mathbf{Y}^T\mathbf{L}\mathbf{Y})}_{\text{smoothness on } \mathbf{Y}} + \beta \underbrace{||\mathbf{L}||_F^2}_{\text{regularization}}$$

Model 1: Global smoothness

- Dong et al. (2015) & Kalofolias (2016)



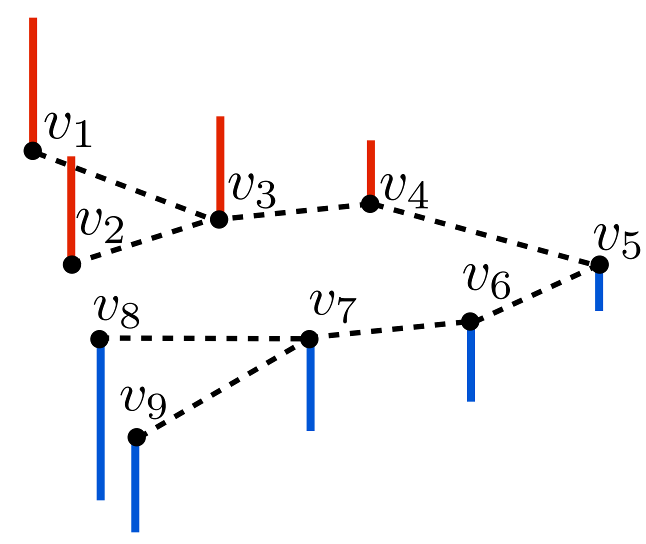
  - $\mathbf{D}(\mathcal{G}) = \chi$ (eigenvector matrix of $\mathbf{L}$)
  - Gaussian assumption on $\mathbf{c}$: $\mathbf{c} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Lambda})$
  - Maximum a posterior (MAP) estimation on $\mathbf{c}$ leads to minimization of Laplacian quadratic form:

$$\min_{c} ||\mathbf{x} - \chi\mathbf{c}||_2^2 - \log P_c(\mathbf{c})$$

$$\Downarrow$$

$$\min_{\mathbf{L},\mathbf{Y}} ||\mathbf{X} - \mathbf{Y}||_F^2 + \alpha\, \mathrm{tr}(\mathbf{Y}^T\mathbf{L}\mathbf{Y}) + \beta ||\mathbf{L}||_F^2$$

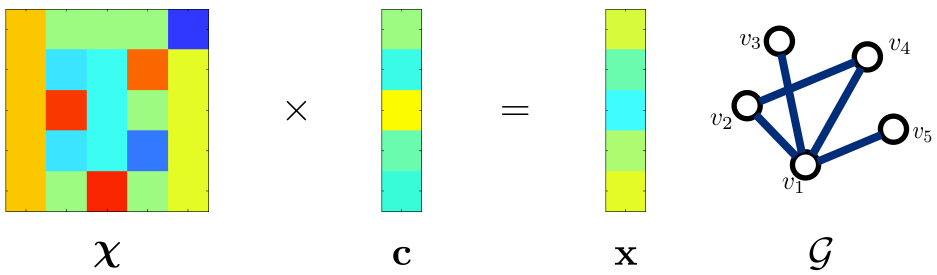
data fidelity smoothness on $\mathbf{Y}$ regularization



$\mathbf{X}$

Model 1: Global smoothness

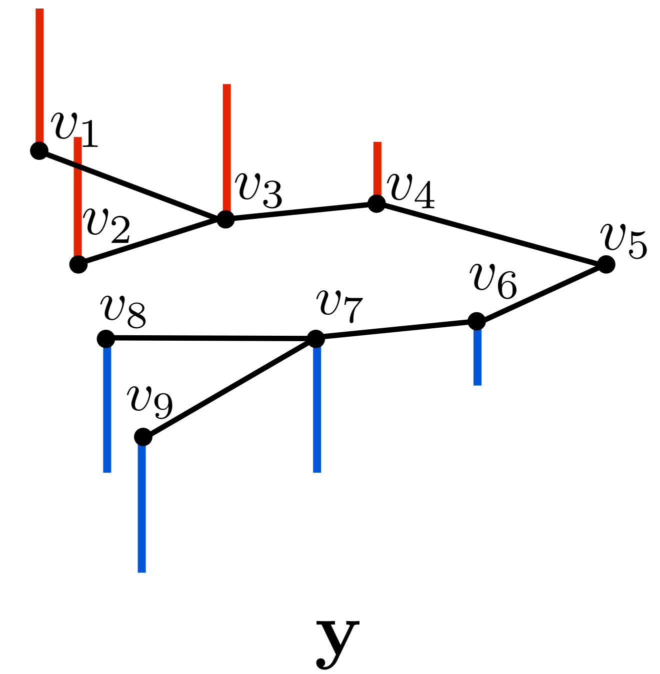
- Dong et al. (2015) & Kalofolias (2016)
  - $\mathbf{D}(\mathcal{G}) = \boldsymbol{\chi}$ (eigenvector matrix of **L**)
  - Gaussian assumption on **c**: $\mathbf{c} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Lambda})$
  - Maximum a posterior (MAP) estimation on **c** leads to minimization of Laplacian quadratic form:



$$\min_{c} ||\mathbf{x} - \boldsymbol{\chi}\mathbf{c}||_2^2 - \log P_c(\mathbf{c})$$

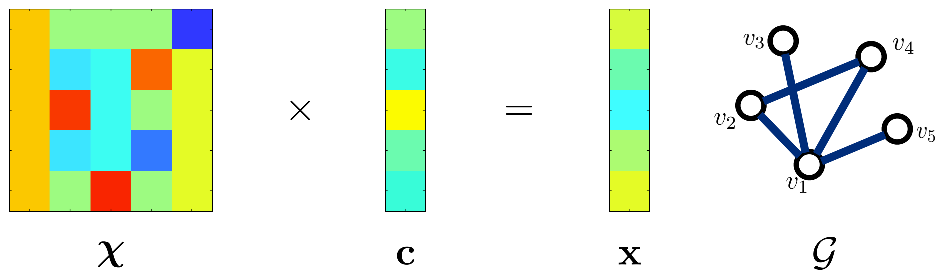
$$\min_{\mathbf{L},\mathbf{Y}} ||\mathbf{X} - \mathbf{Y}||_F^2 + \alpha\, \mathrm{tr}(\mathbf{Y}^T\mathbf{L}\mathbf{Y}) + \beta ||\mathbf{L}||_F^2$$

data fidelity — smoothness on **Y** — regularization



Model 1: Global smoothness

- Dong et al. (2015) & Kalofolias (2016)

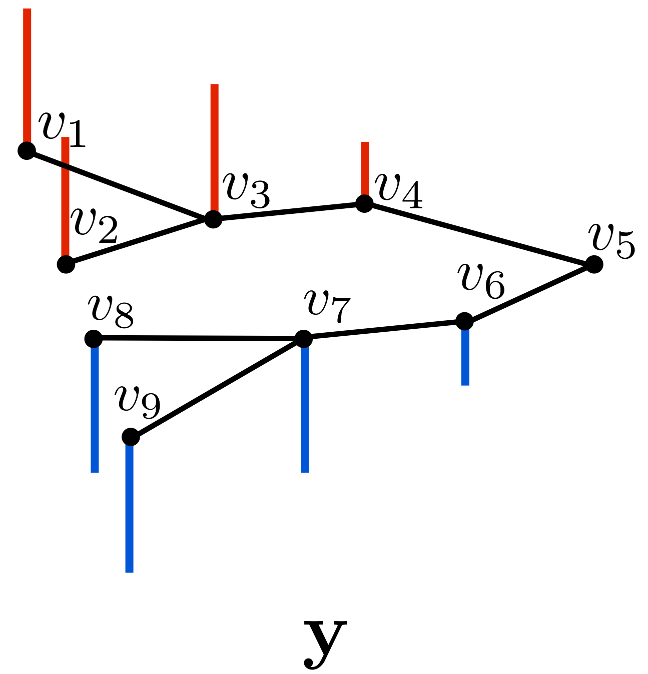


  - $\mathbf{D}(\mathcal{G}) = \boldsymbol{\chi}$ (eigenvector matrix of **L**)
  - Gaussian assumption on **c**: $\mathbf{c} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Lambda})$
  - Maximum a posterior (MAP) estimation on **c** leads to minimization of Laplacian quadratic form:

$$\min_{c} ||\mathbf{x} - \boldsymbol{\chi}\mathbf{c}||_2^2 - \log P_c(\mathbf{c})$$

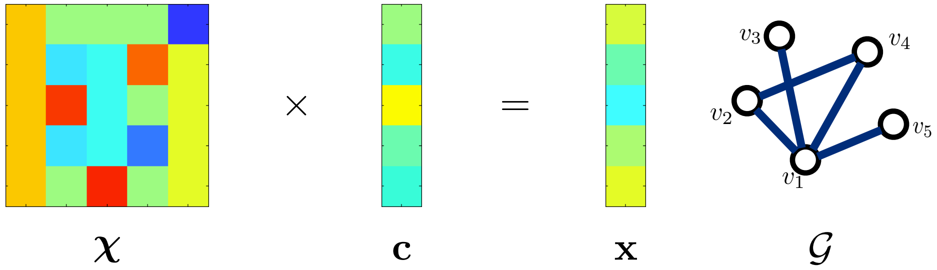
$$\min_{\mathbf{L},\mathbf{Y}} ||\mathbf{X} - \mathbf{Y}||_F^2 + \alpha \, \mathrm{tr}(\mathbf{Y}^T\mathbf{L}\mathbf{Y}) + \beta ||\mathbf{L}||_F^2$$

data fidelity smoothness on **Y** regularization



Learning enforces signal property (global smoothness)!

Model 1: Global smoothness



- Egilmez et al. (2016)

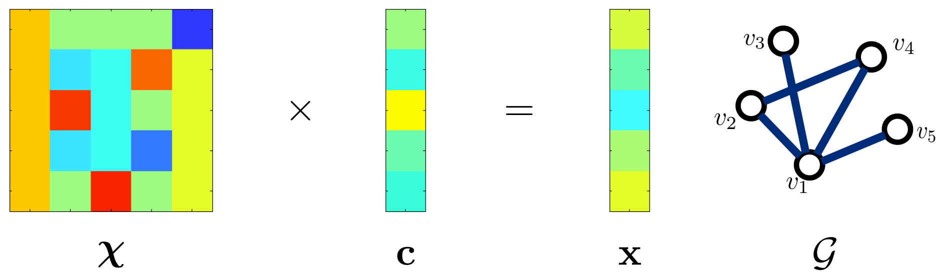
$$\min_{\boldsymbol{\Theta}} \operatorname{tr}(\boldsymbol{\Theta}\mathbf{K}) - \log \det \boldsymbol{\Theta} \quad \text{s.t.} \quad \mathbf{K} = \mathbf{S} - \frac{\alpha}{2}(\mathbf{1}\mathbf{1}^T - \mathbf{I})$$

  - Solve for $\boldsymbol{\Theta}$ as three different graph Laplacian matrices:

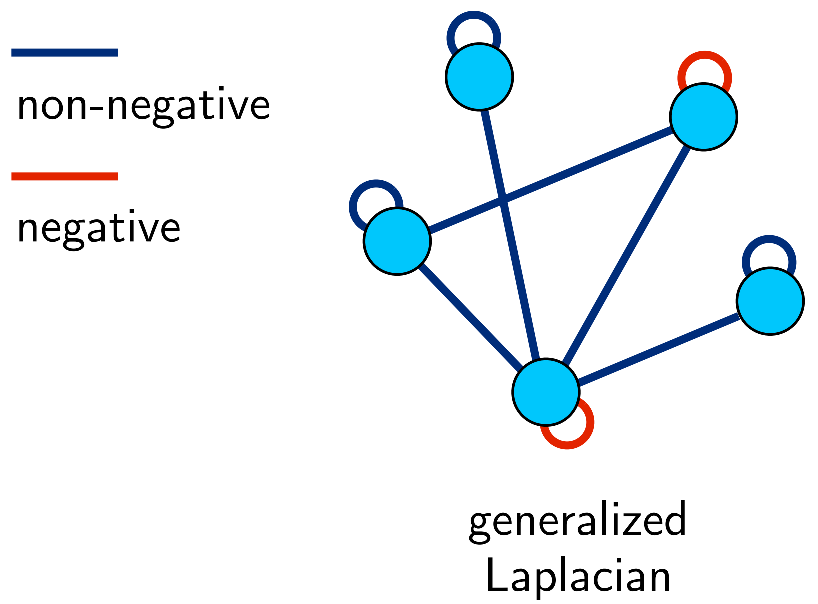
Model 1: Global smoothness

- Egilmez et al. (2016)

$$\min_{\boldsymbol{\Theta}} \operatorname{tr}(\boldsymbol{\Theta}\mathbf{K}) - \log \det \boldsymbol{\Theta} \quad \text{s.t.} \quad \mathbf{K} = \mathbf{S} - \frac{\alpha}{2}(\mathbf{1}\mathbf{1}^T - \mathbf{I})$$



  - Solve for $\boldsymbol{\Theta}$ as three different graph Laplacian matrices:

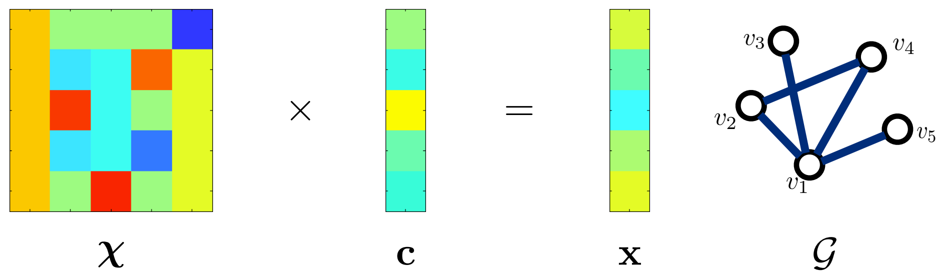


$$\boldsymbol{\Theta} = \mathbf{L} + \mathbf{V}$$
$$= \mathbf{Deg} - \mathbf{W} + \mathbf{V}$$

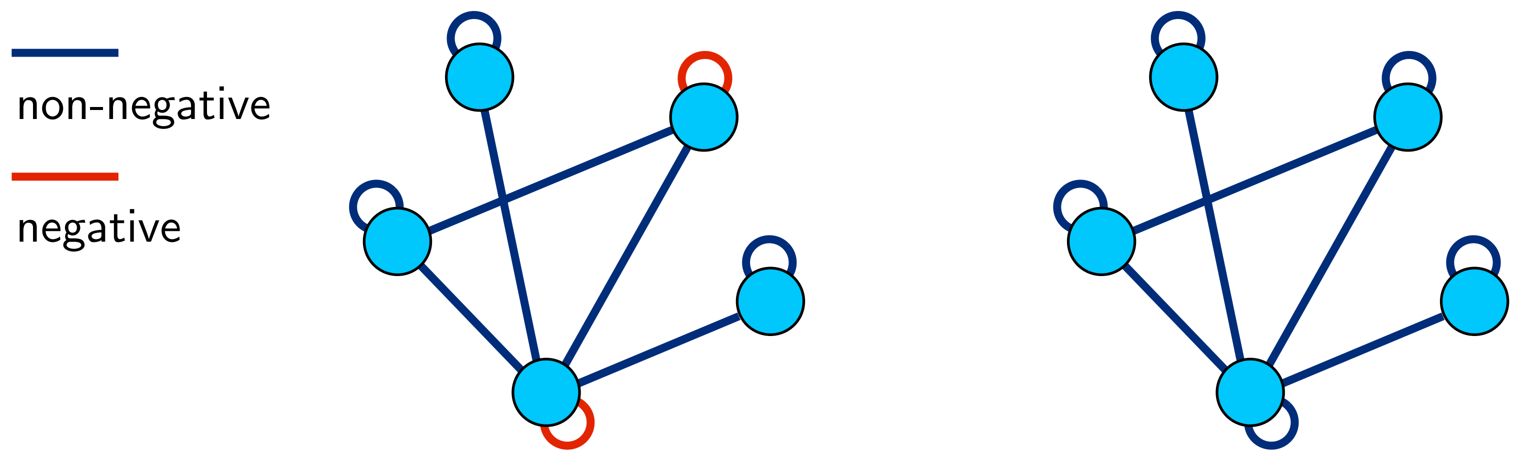
Model 1: Global smoothness

- Egilmez et al. (2016)

$$\min_{\boldsymbol{\Theta}} \operatorname{tr}(\boldsymbol{\Theta}\mathbf{K}) - \log \det \boldsymbol{\Theta} \quad \text{s.t.} \quad \mathbf{K} = \mathbf{S} - \frac{\alpha}{2}(\mathbf{1}\mathbf{1}^T - \mathbf{I})$$



- Solve for $\boldsymbol{\Theta}$ as three different graph Laplacian matrices:



generalized Laplacian

$$\boldsymbol{\Theta} = \mathbf{L} + \mathbf{V} = \mathbf{Deg} - \mathbf{W} + \mathbf{V}$$

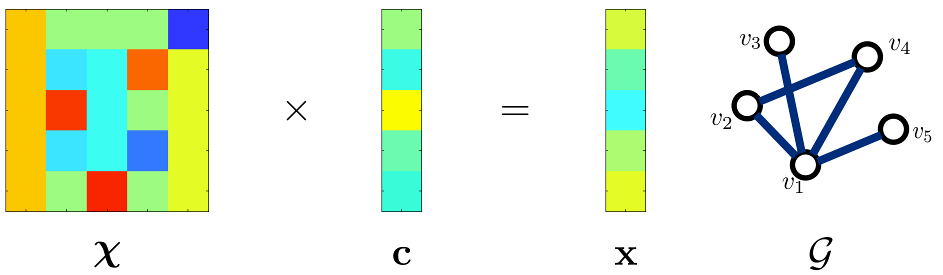
diagonally dominant generalized Laplacian

$$\boldsymbol{\Theta} = \mathbf{L} + \mathbf{V} = \mathbf{Deg} - \mathbf{W} + \mathbf{V} \ (\mathbf{V} \geq \mathbf{0})$$

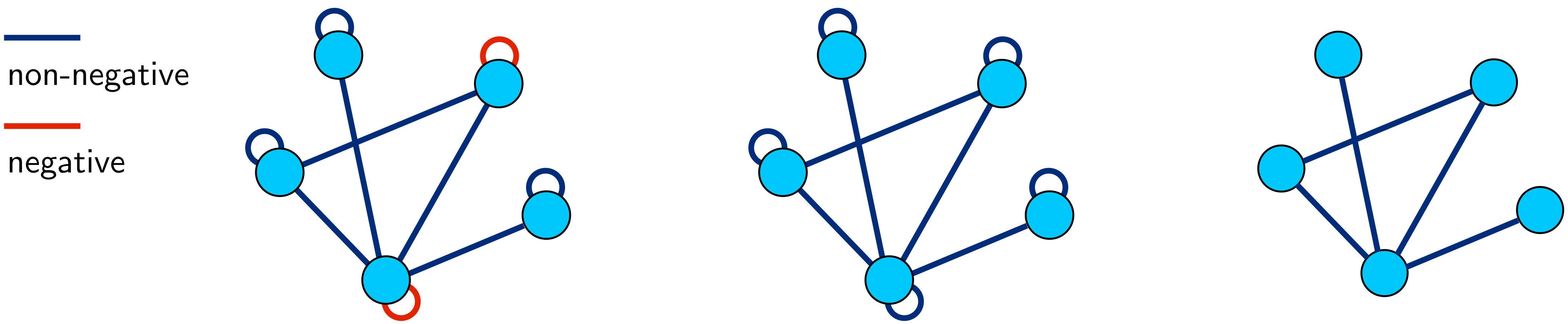
Model 1: Global smoothness

- Egilmez et al. (2016)

$$\min_{\boldsymbol{\Theta}} \operatorname{tr}(\boldsymbol{\Theta}\mathbf{K}) - \log \det \boldsymbol{\Theta} \quad \text{s.t.} \quad \mathbf{K} = \mathbf{S} - \frac{\alpha}{2}(\mathbf{1}\mathbf{1}^T - \mathbf{I})$$



  - Solve for $\boldsymbol{\Theta}$ as three different graph Laplacian matrices:



generalized Laplacian

$$\boldsymbol{\Theta} = \mathbf{L} + \mathbf{V} = \mathbf{Deg} - \mathbf{W} + \mathbf{V}$$

diagonally dominant generalized Laplacian

$$\boldsymbol{\Theta} = \mathbf{L} + \mathbf{V} = \mathbf{Deg} - \mathbf{W} + \mathbf{V} \ (\mathbf{V} \geq \mathbf{0})$$

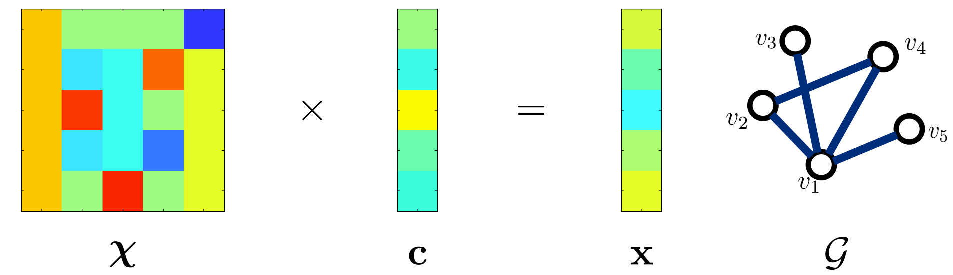
combinatorial Laplacian

$$\boldsymbol{\Theta} = \mathbf{L} = \mathbf{Deg} - \mathbf{W}$$

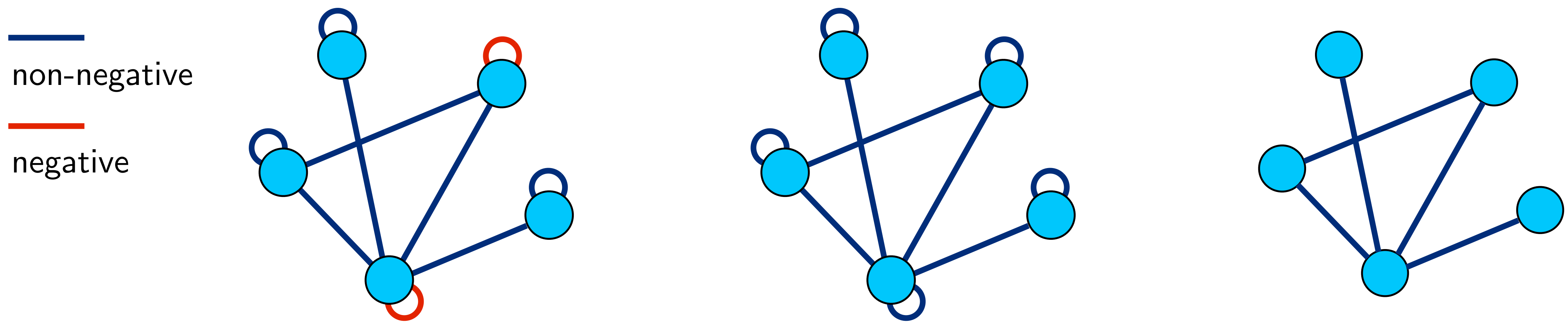
Model 1: Global smoothness

- Egilmez et al. (2016)

$$\min_{\boldsymbol{\Theta}} \operatorname{tr}(\boldsymbol{\Theta}\mathbf{K}) - \log\det\boldsymbol{\Theta} \quad \text{s.t.} \quad \mathbf{K} = \mathbf{S} - \frac{\alpha}{2}(\mathbf{1}\mathbf{1}^T - \mathbf{I})$$



  - Solve for $\boldsymbol{\Theta}$ as three different graph Laplacian matrices:



generalized Laplacian | diagonally dominant generalized Laplacian | combinatorial Laplacian

$$\boldsymbol{\Theta} = \mathbf{L} + \mathbf{V} = \mathbf{Deg} - \mathbf{W} + \mathbf{V}$$

$$\boldsymbol{\Theta} = \mathbf{L} + \mathbf{V} = \mathbf{Deg} - \mathbf{W} + \mathbf{V} \ (\mathbf{V} \geq \mathbf{0})$$

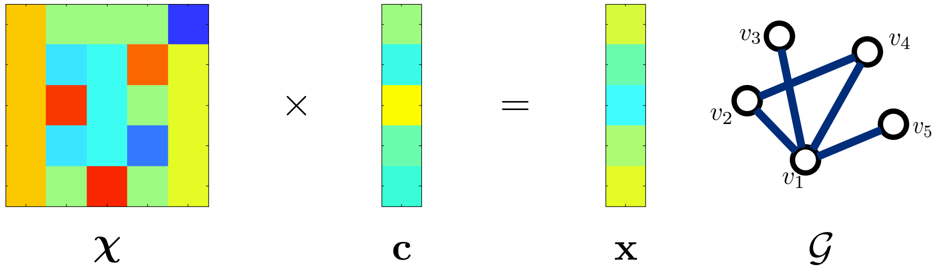
$$\boldsymbol{\Theta} = \mathbf{L} = \mathbf{Deg} - \mathbf{W}$$

Generalizes graphical LASSO and Lake

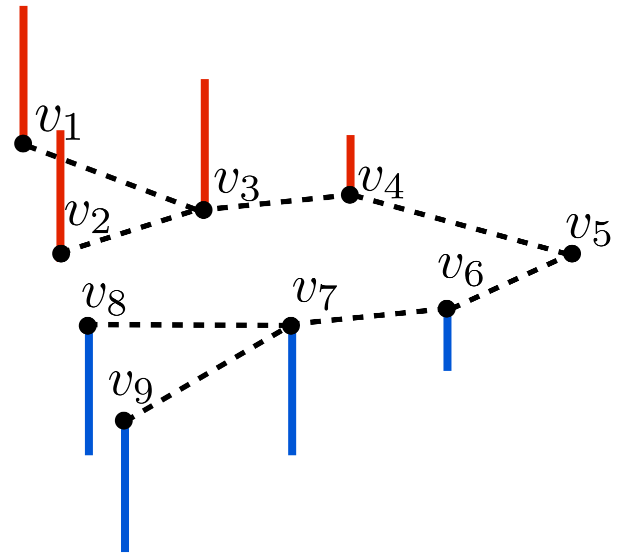
Adding priors on edge weights leads to interpretation of MAP estimation

Model 1: Global smoothness

- Chepuri et al. (2016)

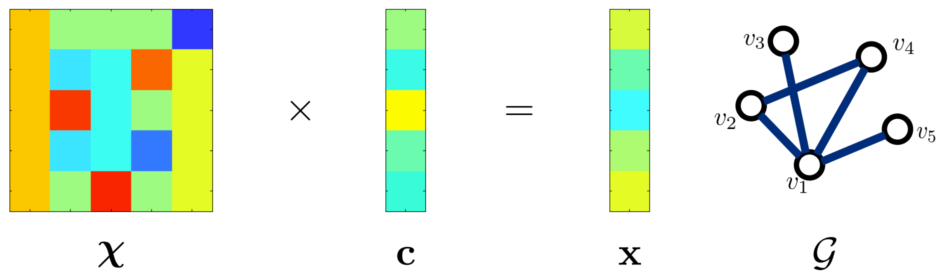


  - An edge selection mechanism based on the same smoothness measure:

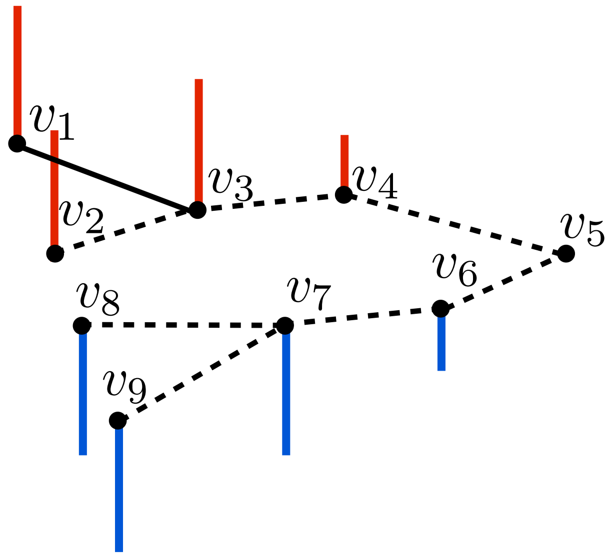


Model 1: Global smoothness

- Chepuri et al. (2016)

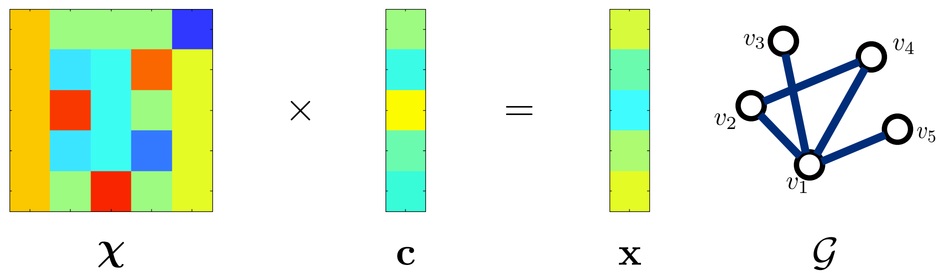


  - An edge selection mechanism based on the same smoothness measure:

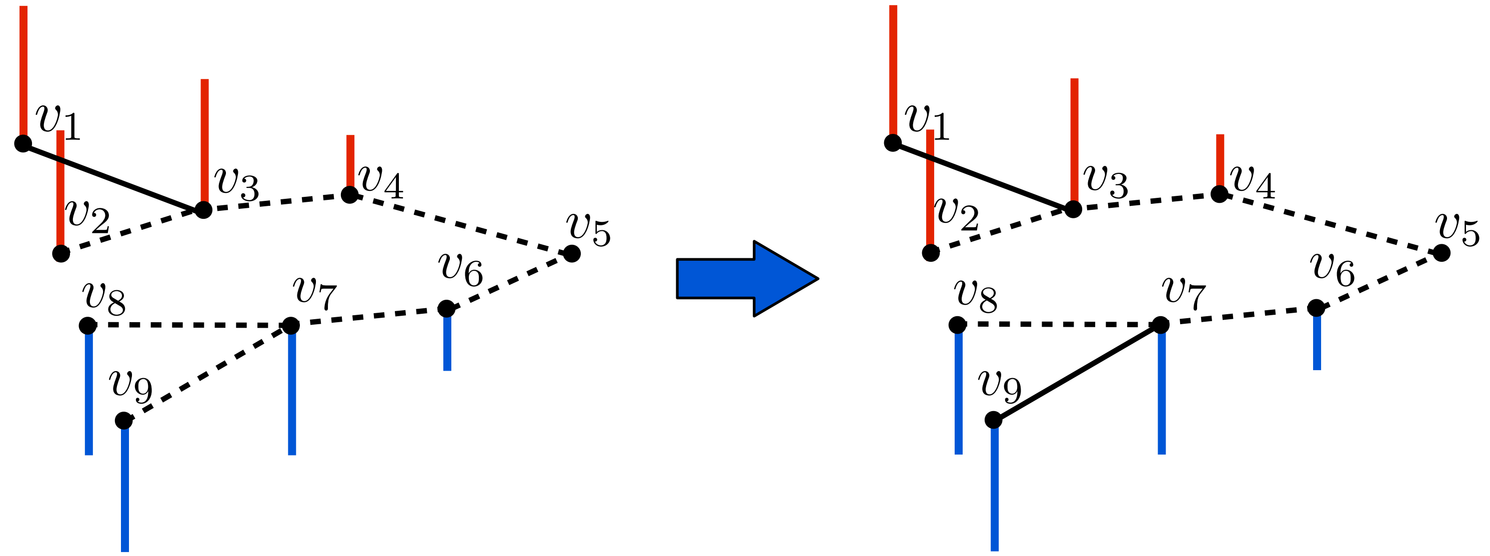


Model 1: Global smoothness

- Chepuri et al. (2016)

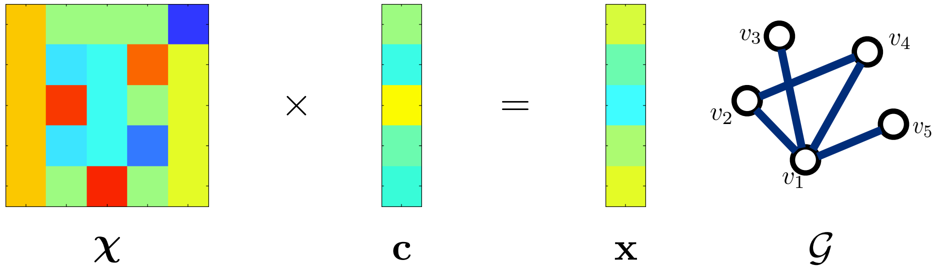


  - An edge selection mechanism based on the same smoothness measure:

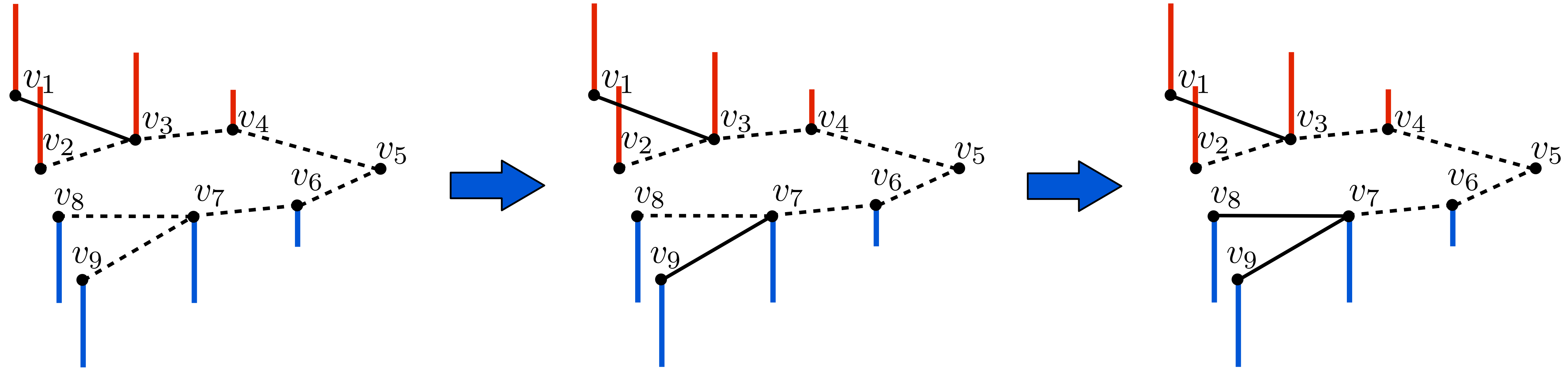


Model 1: Global smoothness

- Chepuri et al. (2016)

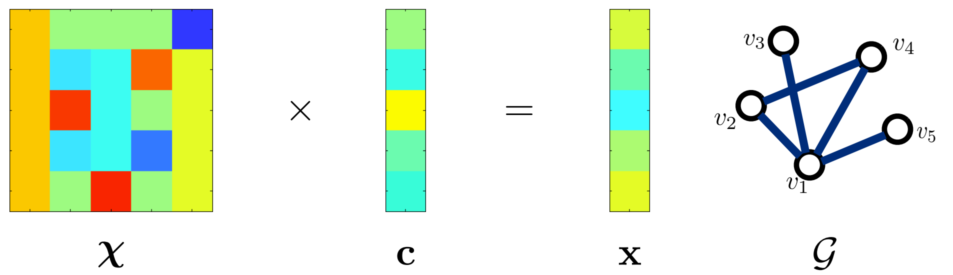


  - An edge selection mechanism based on the same smoothness measure:

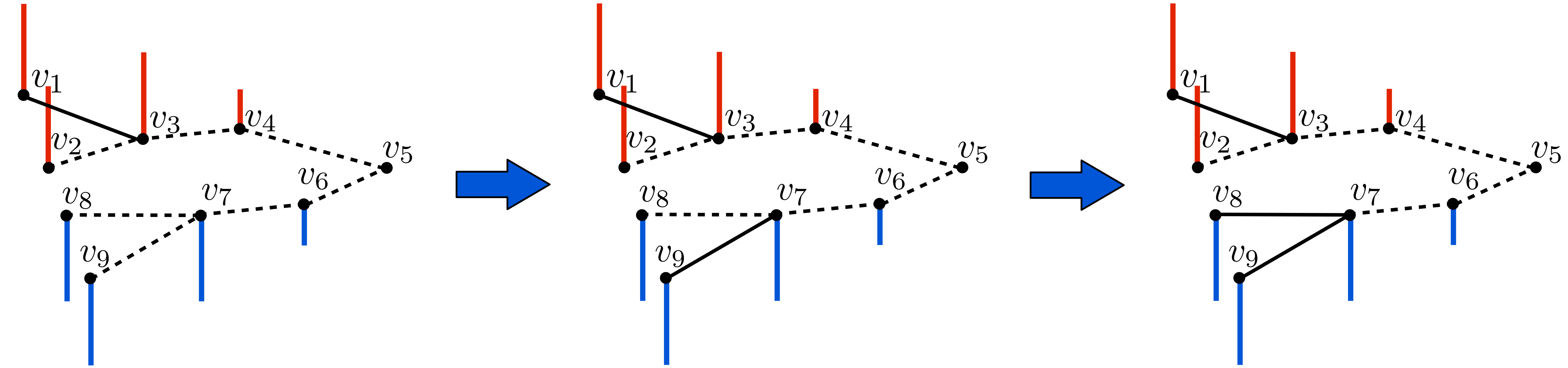


Model 1: Global smoothness

- Chepuri et al. (2016)



  - An edge selection mechanism based on the same smoothness measure:



Similar in spirit to Dempster

Good for learning unweighted graph

Explicit edge-handler is desirable in some applications

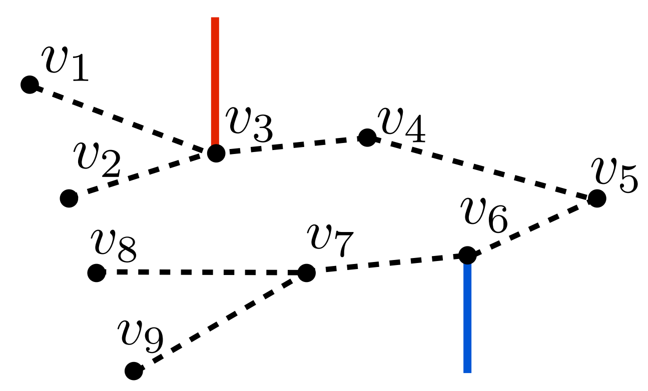
Model 2: Diffusion process

- Signals are outcome of some diffusion processes on the graph (more of local smoothness than global one!)

- Example: Movement of people/vehicles in transportation network

Model 2: Diffusion process

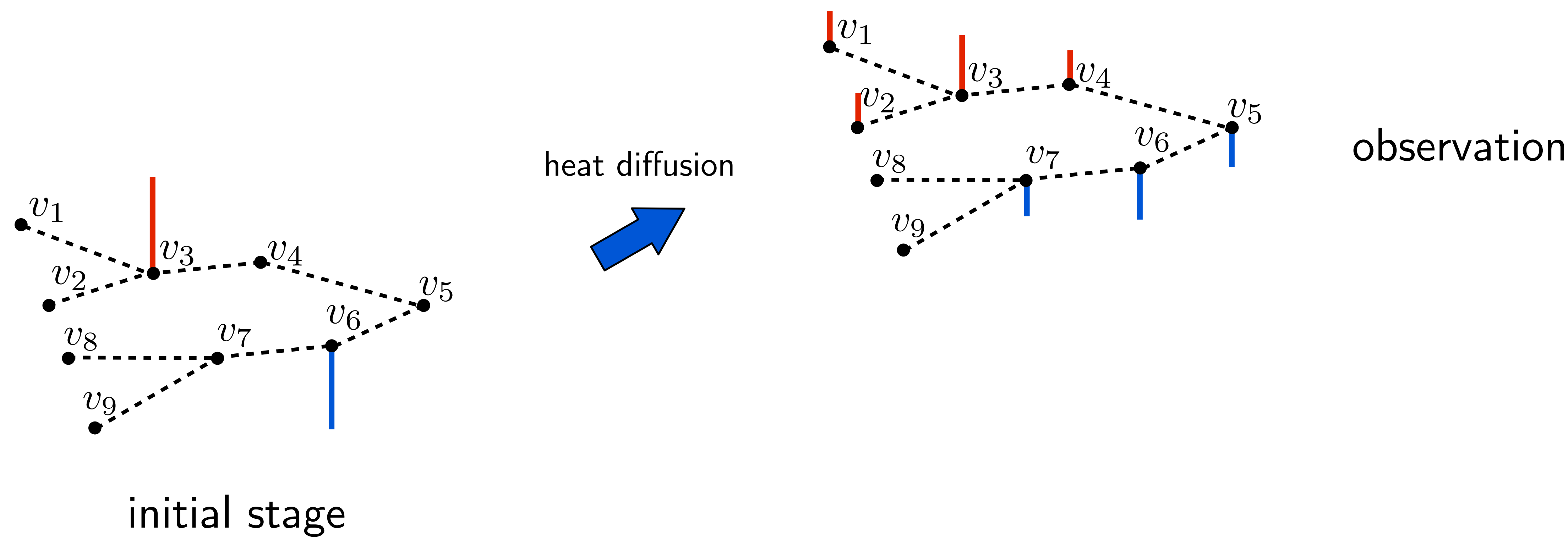
- Signals are outcome of some diffusion processes on the graph (more of local smoothness than global one!)
- Example: Movement of people/vehicles in transportation network
- Characterized by diffusion operators



initial stage

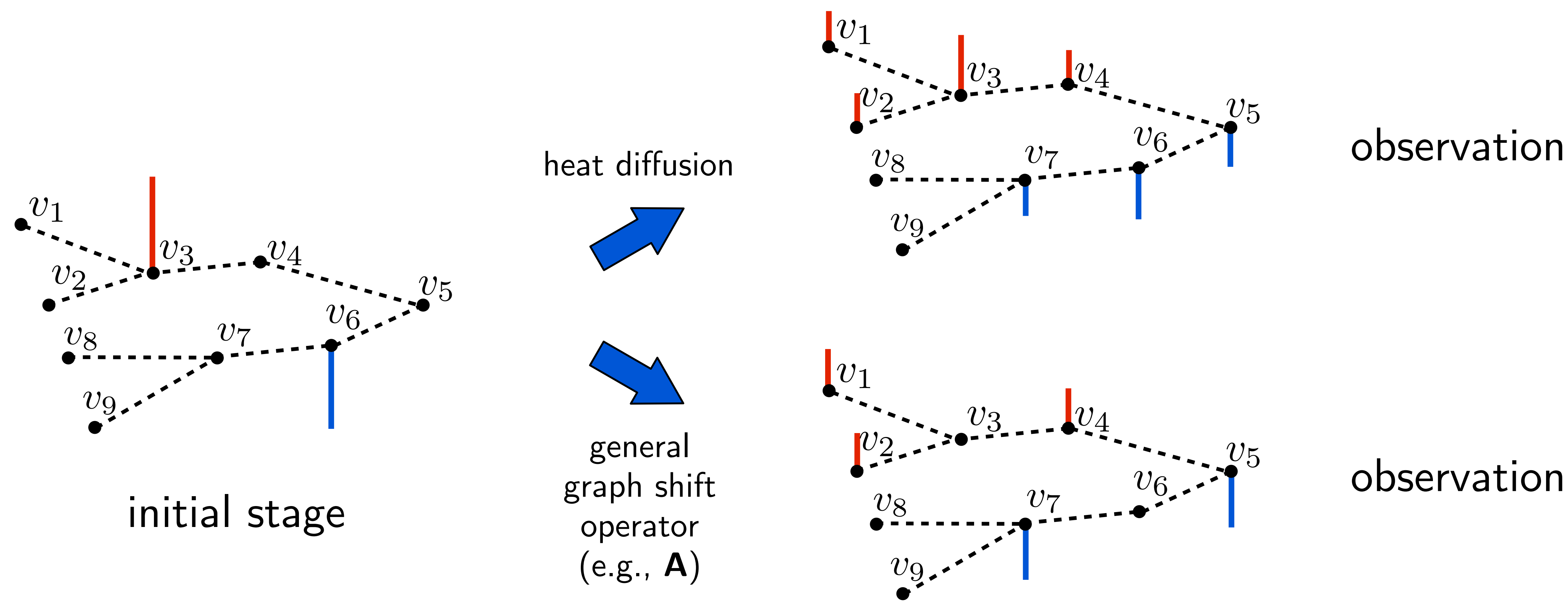
Model 2: Diffusion process

- Signals are outcome of some diffusion processes on the graph (more of local smoothness than global one!)
- Example: Movement of people/vehicles in transportation network
- Characterized by diffusion operators



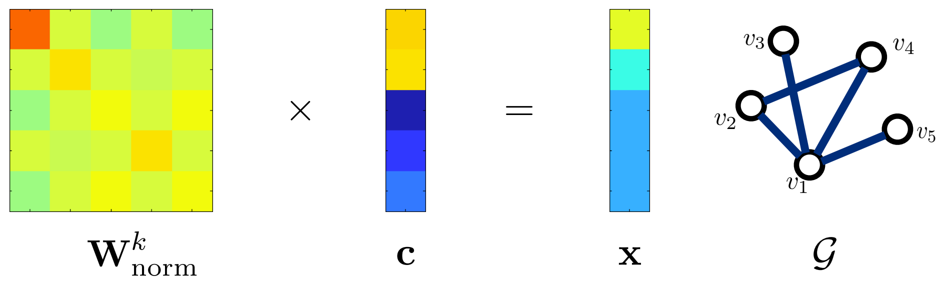
Model 2: Diffusion process

- Signals are outcome of some diffusion processes on the graph (more of local smoothness than global one!)
- Example: Movement of people/vehicles in transportation network
- Characterized by diffusion operators



Model 2: Diffusion process

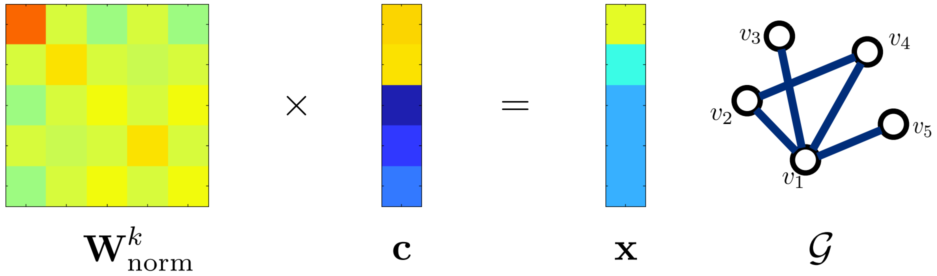
- Pasdeloup et al. (2015, 2016)



  - $\mathbf{D}(\mathcal{G}) = \mathbf{T}^{\mathbf{k}(m)} = \mathbf{W}^{\mathbf{k}(m)}_{\mathrm{norm}}$
  - $\{\mathbf{c_m}\}$ are i.i.d. samples with independent entries

Model 2: Diffusion process

- Pasdeloup et al. (2015, 2016)

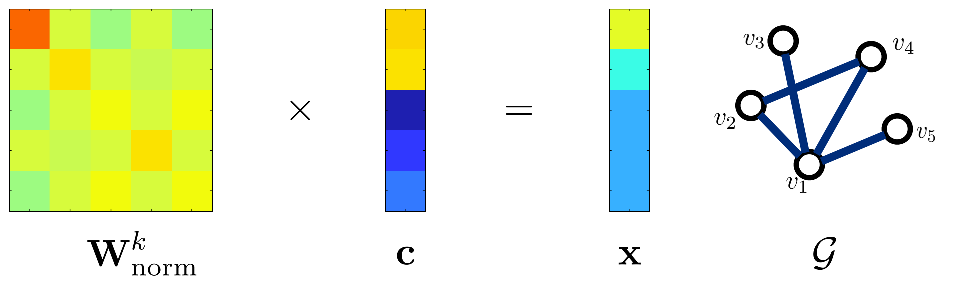


  - $\mathbf{D}(\mathcal{G}) = \mathbf{T}^{\mathbf{k}(m)} = \mathbf{W}_{\mathrm{norm}}^{\mathbf{k}(m)}$
  - $\{\mathbf{c_m}\}$ are i.i.d. samples with independent entries
  - Two-step approach:
    - Estimate eigenvector matrix from sample covariance (if covariance unknown):

$$\mathbf{\Sigma} = \mathbb{E}\Big[\sum_{m=1}^{M} \mathbf{X}(m)\mathbf{X}(m)^T\Big] = \sum_{m=1}^{M} \mathbf{W}_{\mathrm{norm}}^{2\mathbf{k}(m)} \quad \text{(polynomial of } \mathbf{W}_{\mathrm{norm}}\text{)}$$

Model 2: Diffusion process

- Pasdeloup et al. (2015, 2016)

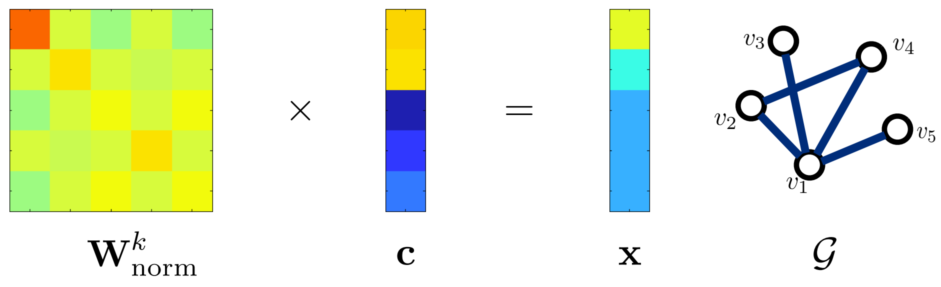


  - $\mathbf{D}(\mathcal{G}) = \mathbf{T}^{\mathbf{k}(m)} = \mathbf{W}^{\mathbf{k}(m)}_{\text{norm}}$
  - $\{\mathbf{c_m}\}$ are i.i.d. samples with independent entries
  - Two-step approach:
    - Estimate eigenvector matrix from sample covariance (if covariance unknown):

$$\mathbf{\Sigma} = \mathbb{E}\Big[\sum_{m=1}^{M} \mathbf{X}(m)\mathbf{X}(m)^T\Big] = \sum_{m=1}^{M} \mathbf{W}^{2\mathbf{k}(m)}_{\text{norm}}$$ (polynomial of $\mathbf{W}_{\text{norm}}$)

    - Optimize for eigenvalues given constraints of $\mathbf{W}_{\text{norm}}$ (mainly non-negativity of off-diagonal of $\mathbf{W}_{\text{norm}}$ and eigenvalue range) and some priors (e.g., sparsity)

Model 2: Diffusion process



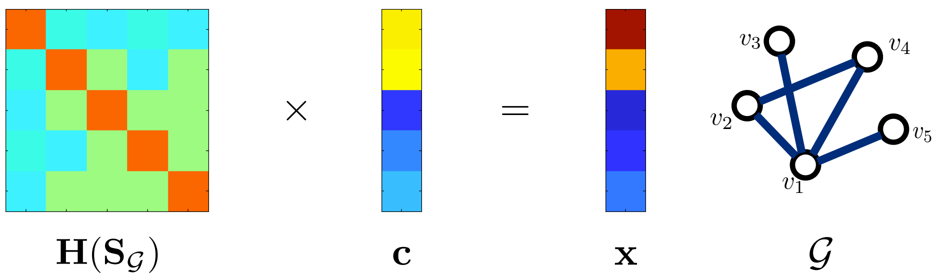
- Pasdeloup et al. (2015, 2016)
  - $\mathbf{D}(\mathcal{G}) = \mathbf{T}^{\mathbf{k}(m)} = \mathbf{W}_{\mathrm{norm}}^{\mathbf{k}(m)}$
  - $\{\mathbf{c_m}\}$ are i.i.d. samples with independent entries
  - Two-step approach:
    - Estimate eigenvector matrix from sample covariance (if covariance unknown):

$$\mathbf{\Sigma} = \mathbb{E}\Big[\sum_{m=1}^{M} \mathbf{X}(m)\mathbf{X}(m)^T\Big] = \sum_{m=1}^{M} \mathbf{W}_{\mathrm{norm}}^{2\mathbf{k}(m)} \quad \text{(polynomial of } \mathbf{W}_{\mathrm{norm}}\text{)}$$

    - Optimize for eigenvalues given constraints of $\mathbf{W}_{\mathrm{norm}}$ (mainly non-negativity of off-diagonal of $\mathbf{W}_{\mathrm{norm}}$ and eigenvalue range) and some priors (e.g., sparsity)

More a "graph-centric" learning framework: Cost function on graph components instead of signals

Model 2: Diffusion process

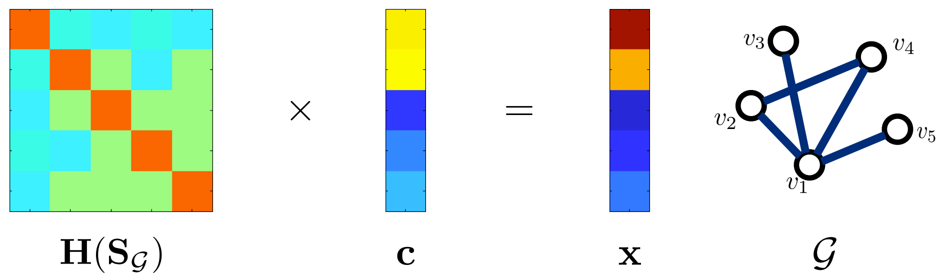


- Segarra et al. (2016)
  - $\mathbf{D}(\mathcal{G}) = \mathbf{H}(\mathbf{S}_\mathcal{G}) = \sum_{l=0}^{L-1} h_l \mathbf{S}_\mathcal{G}{}^l$

    (diffusion defined by a graph shift operator $\mathbf{S}_\mathcal{G}$ that can be arbitrary, but practically **W** or **L**)
  - **c** is a white signal

Model 2: Diffusion process

- Segarra et al. (2016)



  - $\mathbf{D}(\mathcal{G}) = \mathbf{H}(\mathbf{S}_{\mathcal{G}}) = \sum_{l=0}^{L-1} h_l {\mathbf{S}_{\mathcal{G}}}^l$

    (diffusion defined by a graph shift operator $\mathbf{S}_{\mathcal{G}}$ that can be arbitrary, but practically **W** or **L**)

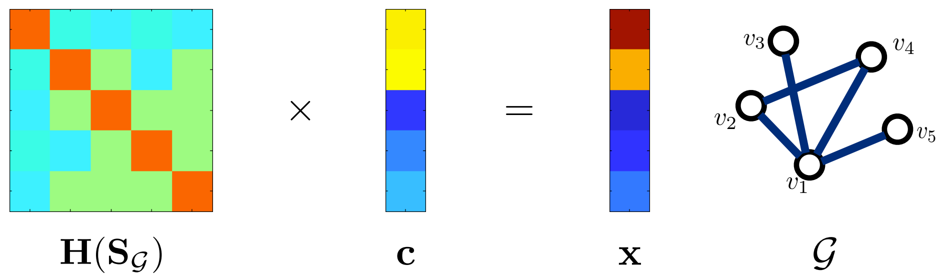
  - **c** is a white signal

  - Two-step approach:

    - Estimate eigenvector matrix: $\mathbf{\Sigma} = \mathbf{H}\mathbf{H}^T$

Model 2: Diffusion process

- Segarra et al. (2016)



  - $\mathbf{D}(\mathcal{G}) = \mathbf{H}(\mathbf{S}_{\mathcal{G}}) = \sum_{l=0}^{L-1} h_l {\mathbf{S}_{\mathcal{G}}}^l$

    (diffusion defined by a graph shift operator $\mathbf{S}_{\mathcal{G}}$ that can be arbitrary, but practically **W** or **L**)

  - **c** is a white signal

  - Two-step approach:

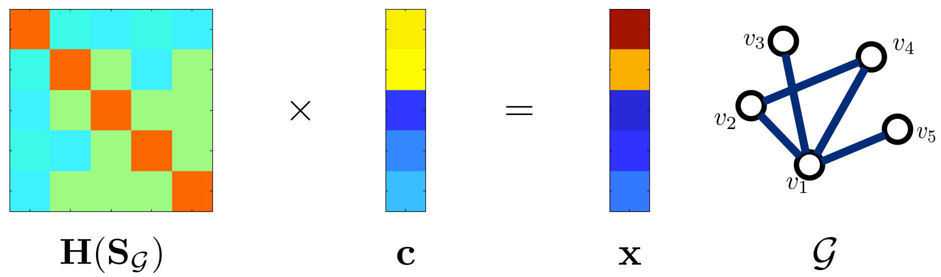
    - Estimate eigenvector matrix: $\mathbf{\Sigma} = \mathbf{H}\mathbf{H}^T$

    - Select eigenvalues that satisfy constraints of $\mathbf{S}_{\mathcal{G}}$:

$$\min_{\mathbf{S}_{\mathcal{G}}, \lambda} ||\mathbf{S}_{\mathcal{G}}||_0 \quad \text{s.t.} \quad \mathbf{S}_{\mathcal{G}} = \sum_{n=1}^{N} \lambda_n \mathbf{v}_n \mathbf{v}_n^T$$

"spectral templates" (eigenvectors)

Model 2: Diffusion process



- Segarra et al. (2016)
  - $\mathbf{D}(\mathcal{G}) = \mathbf{H}(\mathbf{S}_\mathcal{G}) = \sum_{l=0}^{L-1} h_l {\mathbf{S}_\mathcal{G}}^l$

    (diffusion defined by a graph shift operator $\mathbf{S}_\mathcal{G}$ that can be arbitrary, but practically **W** or **L**)
  - **c** is a white signal
  - Two-step approach:
    - Estimate eigenvector matrix: $\mathbf{\Sigma} = \mathbf{H}\mathbf{H}^T$
    - Select eigenvalues that satisfy constraints of $\mathbf{S}_\mathcal{G}$:

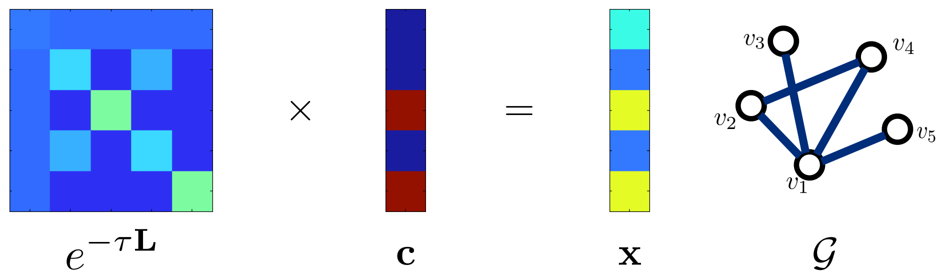
$$\min_{\mathbf{S}_\mathcal{G}, \lambda} ||\mathbf{S}_\mathcal{G}||_0 \quad \text{s.t.} \quad \mathbf{S}_\mathcal{G} = \sum_{n=1}^{N} \lambda_n \mathbf{v}_n \mathbf{v}_n^T$$

"spectral templates" (eigenvectors)

Similar in spirit to Pasdeloup, same assumption on stationarity but different inference framework due to different D

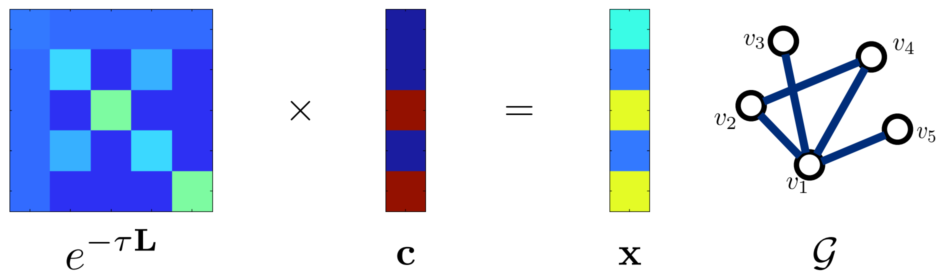
Can handle noisy or incomplete information on spectral templates

Model 2: Diffusion process

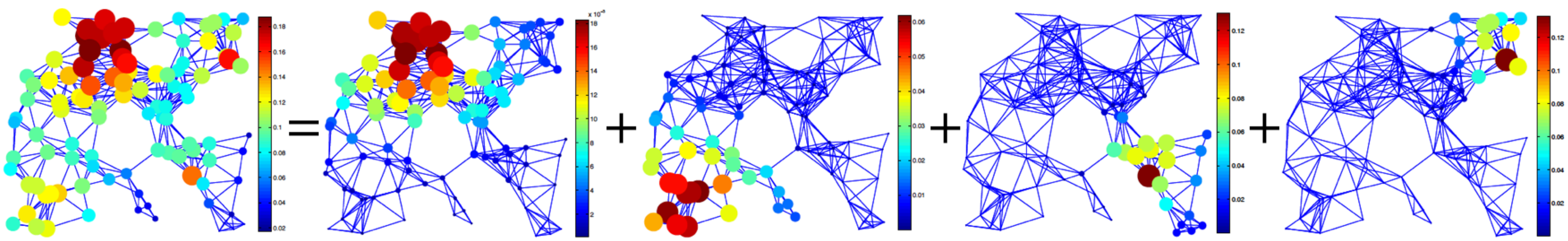


- Thanou et al. (2016)
  - $\mathbf{D}(\mathcal{G}) = e^{-\tau\mathbf{L}}$ (localization in vertex domain)
  - Sparsity assumption on $\mathbf{c}$

Model 2: Diffusion process

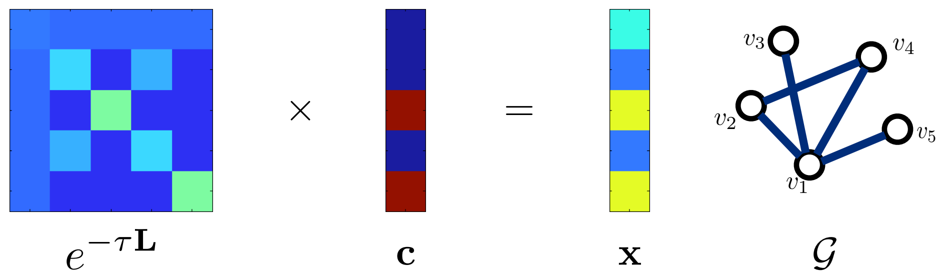


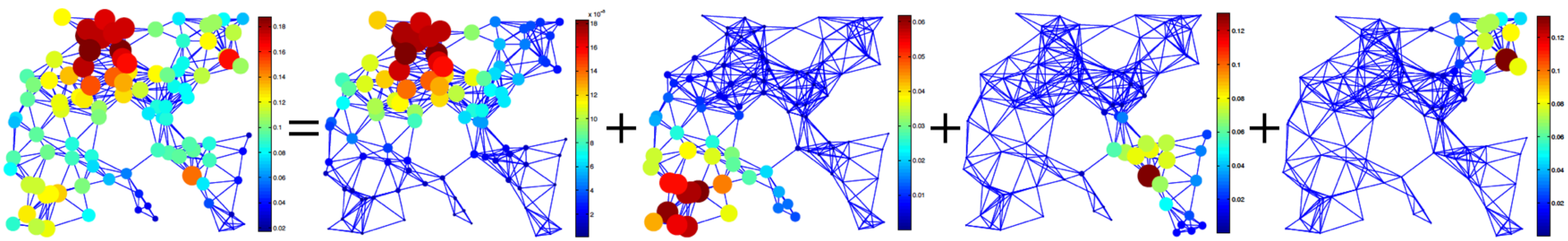
- Thanou et al. (2016)
  - $\mathbf{D}(\mathcal{G}) = e^{-\tau\mathbf{L}}$ (localization in vertex domain)
  - Sparsity assumption on $\mathbf{c}$
  - Each signal is a combination of several heat diffusion processes at time τ



Model 2: Diffusion process

- Thanou et al. (2016)
  - $\mathbf{D}(\mathcal{G}) = e^{-\tau \mathbf{L}}$ (localization in vertex domain)
  - Sparsity assumption on $\mathbf{c}$
  - Each signal is a combination of several heat diffusion processes at time τ

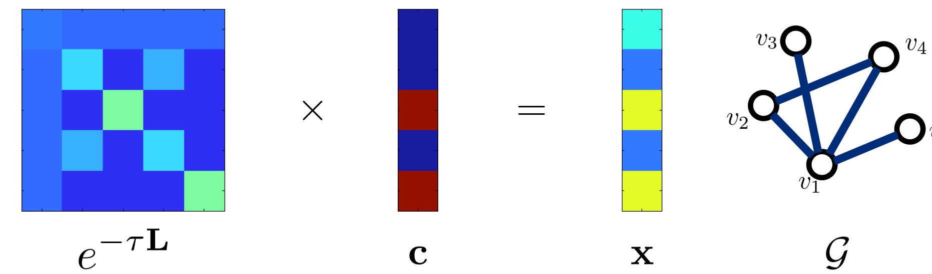




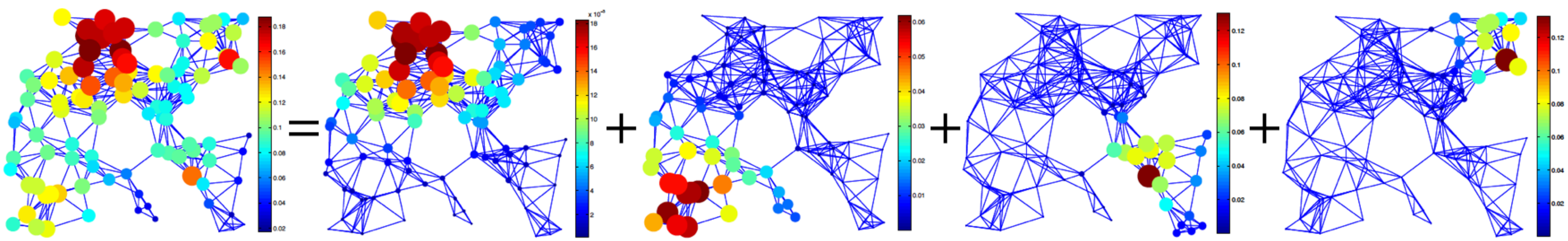
$$\min_{\mathbf{L},\mathbf{C},\tau} ||\mathbf{X} - \mathbf{D}(\mathbf{L})\mathbf{C}||_F^2 + \alpha \sum_{m=1}^{M} ||\mathbf{c}_m||_1 + \beta ||\mathbf{L}||_F^2 \quad \text{s.t.} \quad \mathbf{D} = [e^{-\tau_1 \mathbf{L}}, ..., e^{-\tau_S \mathbf{L}}]$$

Model 2: Diffusion process

- Thanou et al. (2016)



  - $\mathbf{D}(\mathcal{G}) = e^{-\tau \mathbf{L}}$ (localization in vertex domain)
  - Sparsity assumption on $\mathbf{c}$
  - Each signal is a combination of several heat diffusion processes at time τ

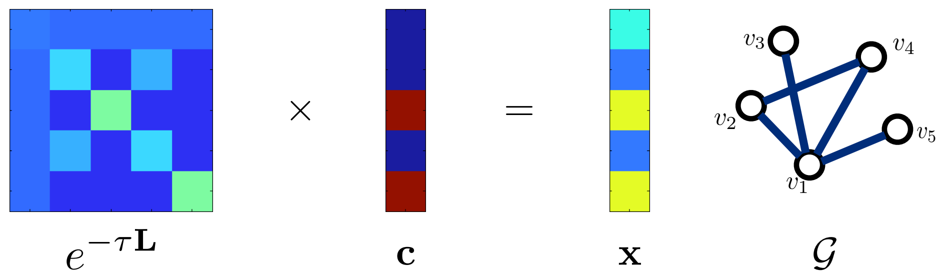


$$\min_{\mathbf{L},\mathbf{C},\tau} ||\mathbf{X} - \mathbf{D}(\mathbf{L})\mathbf{C}||_F^2 + \alpha \sum_{m=1}^{M} ||\mathbf{c}_m||_1 + \beta ||\mathbf{L}||_F^2 \quad \text{s.t.} \quad \mathbf{D} = [e^{-\tau_1 \mathbf{L}}, ..., e^{-\tau_S \mathbf{L}}]$$

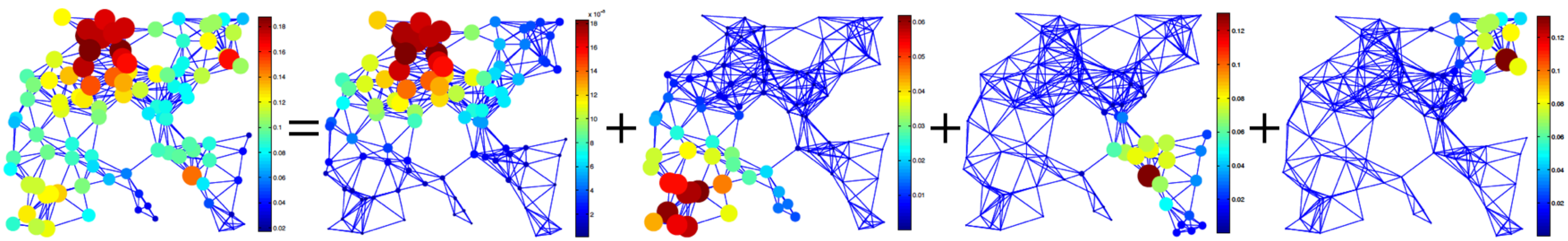
data fidelity — sparsity on $\mathbf{c}$ — regularization

Model 2: Diffusion process

- Thanou et al. (2016)



  - $\mathbf{D}(\mathcal{G}) = e^{-\tau \mathbf{L}}$ (localization in vertex domain)
  - Sparsity assumption on $\mathbf{c}$
  - Each signal is a combination of several heat diffusion processes at time τ



$$\min_{\mathbf{L},\mathbf{C},\tau} ||\mathbf{X} - \mathbf{D}(\mathbf{L})\mathbf{C}||_F^2 + \alpha \sum_{m=1}^{M} ||\mathbf{c}_m||_1 + \beta ||\mathbf{L}||_F^2 \quad \text{s.t.} \quad \mathbf{D} = [e^{-\tau_1 \mathbf{L}}, ..., e^{-\tau_S \mathbf{L}}]$$

data fidelity　　sparsity on $\mathbf{c}$　　regularization

Still diffusion-based model, but more "signal-centric"

No assumption on eigenvectors/stationarity, but on signal structure and sparsity

Can be extended to general polynomial case (Maretic et al. 2017)

Model 3: Time-varying observations

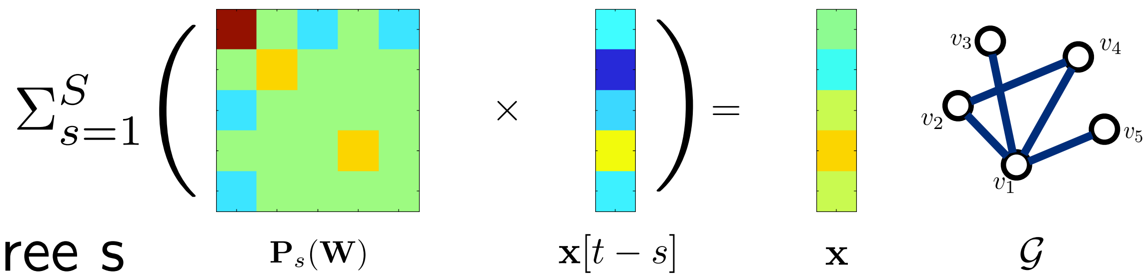
- Signals are time-varying observations that are causal outcome of current or past values (mixed degree of smoothness depending on previous states)

- Example: Evolution of individual behavior due to influence of different friends at different timestamps

- Characterized by an autoregressive model or a structural equation model (SEM)

Model 3: Time-varying observations

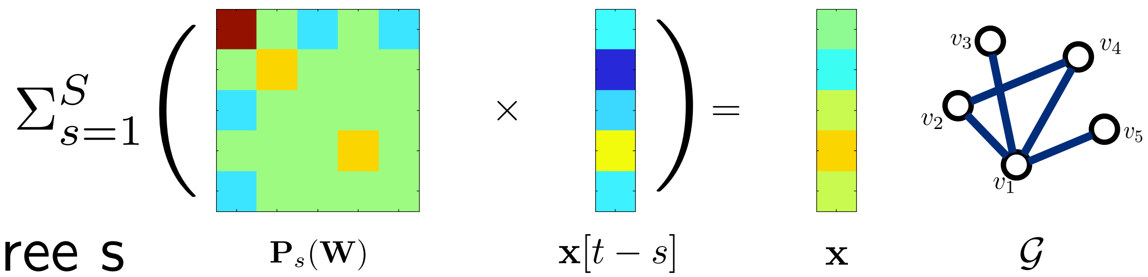
- Mei and Moura (2015)



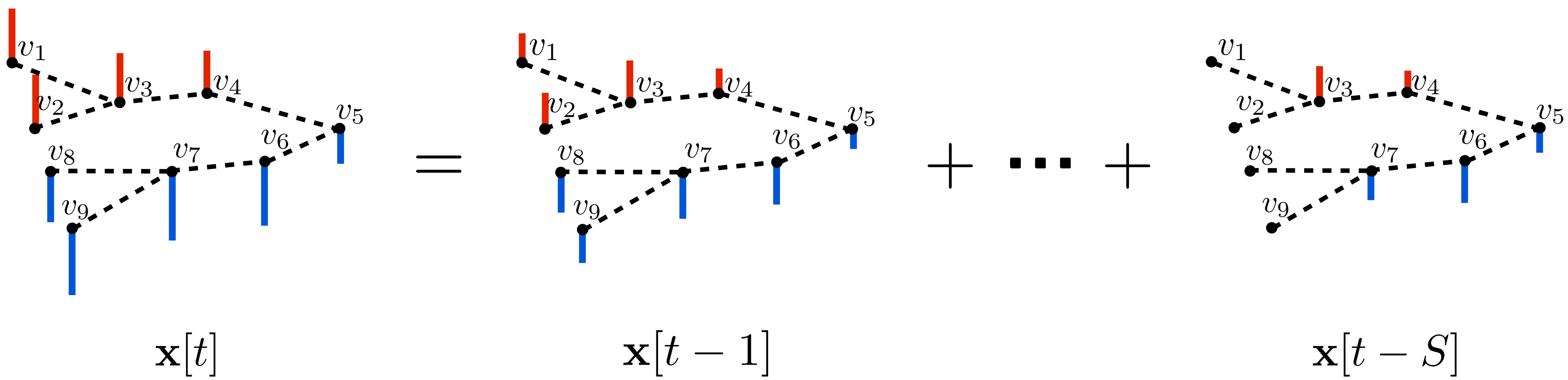
  - $\mathbf{D}_s(\mathcal{G}) = \mathbf{P}_s(\mathbf{W})$: polynomial of **W** of degree s
  - Define $\mathbf{c}_s$ as $\mathbf{x}[t-s]$

Model 3: Time-varying observations

- Mei and Moura (2015)

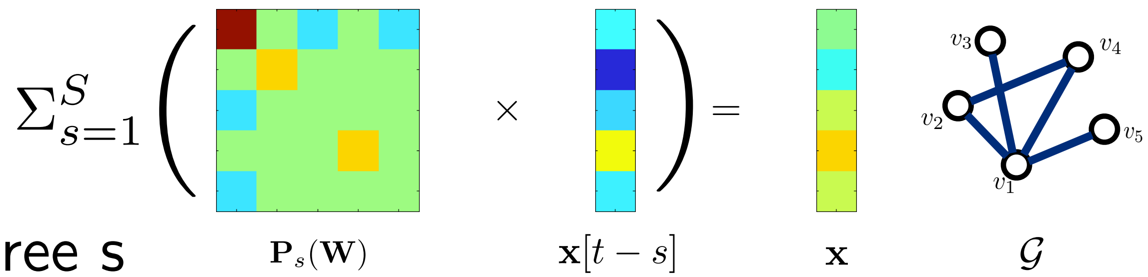


  - $\mathbf{D}_s(\mathcal{G}) = \mathbf{P}_s(\mathbf{W})$: polynomial of **W** of degree s
  - Define $\mathbf{c}_s$ as $\mathbf{x}[t-s]$

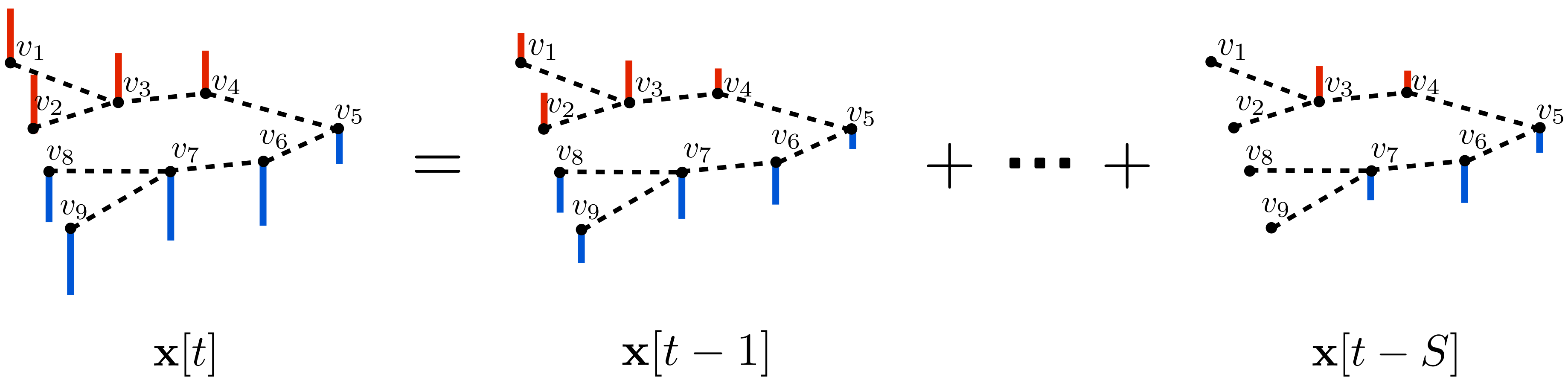


Model 3: Time-varying observations

- Mei and Moura (2015)



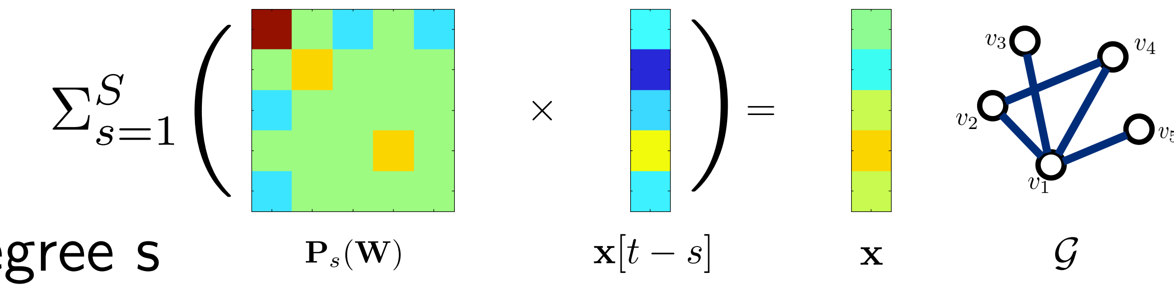
  - $\mathbf{D}_s(\mathcal{G}) = \mathbf{P}_s(\mathbf{W})$: polynomial of **W** of degree s
  - Define $\mathbf{c}_s$ as $\mathbf{x}[t-s]$



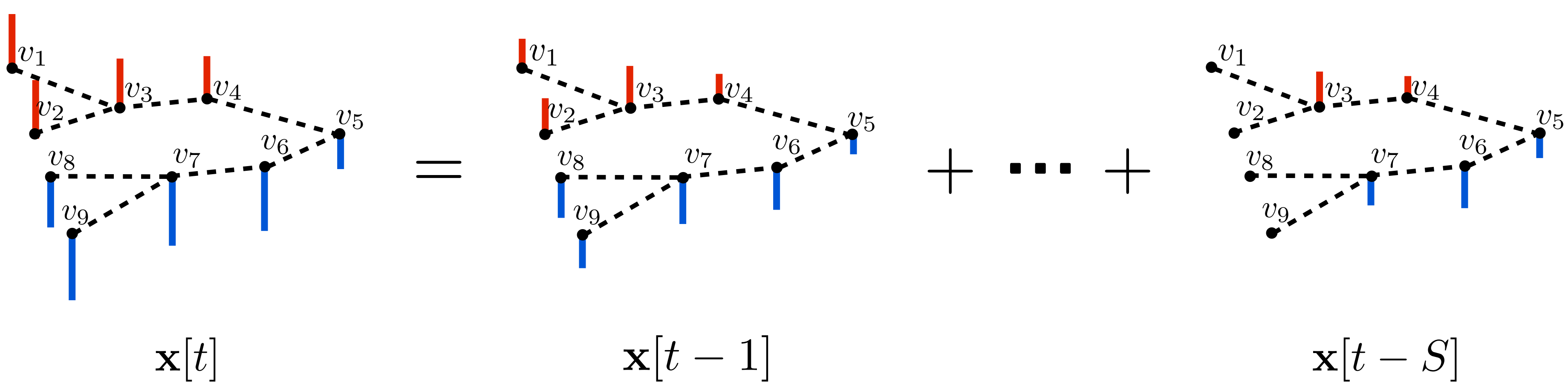
$$\min_{\mathbf{W},\mathbf{a}} \frac{1}{2} \sum_{k=S+1}^{K} ||\mathbf{x}[k] - \sum_{s=1}^{S} \mathbf{P}_s(\mathbf{W})\mathbf{x}[k-s]||_2^2 + \lambda_1 ||\mathrm{vec}(\mathbf{W})||_1 + \lambda_2 ||\mathbf{a}||_1$$

Model 3: Time-varying observations

- Mei and Moura (2015)



  - $\mathbf{D}_s(\mathcal{G}) = \mathbf{P}_s(\mathbf{W})$: polynomial of **W** of degree s
  - Define $\mathbf{c}_s$ as $\mathbf{x}[t-s]$

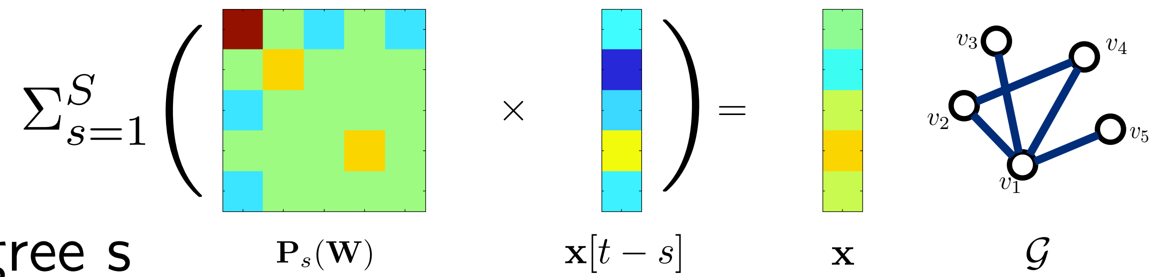


$$\min_{\mathbf{W},\mathbf{a}} \frac{1}{2} \sum_{k=S+1}^{K} ||\mathbf{x}[k] - \sum_{s=1}^{S} \mathbf{P}_s(\mathbf{W})\mathbf{x}[k-s]||_2^2 + \lambda_1 ||\mathrm{vec}(\mathbf{W})||_1 + \lambda_2 ||\mathbf{a}||_1$$

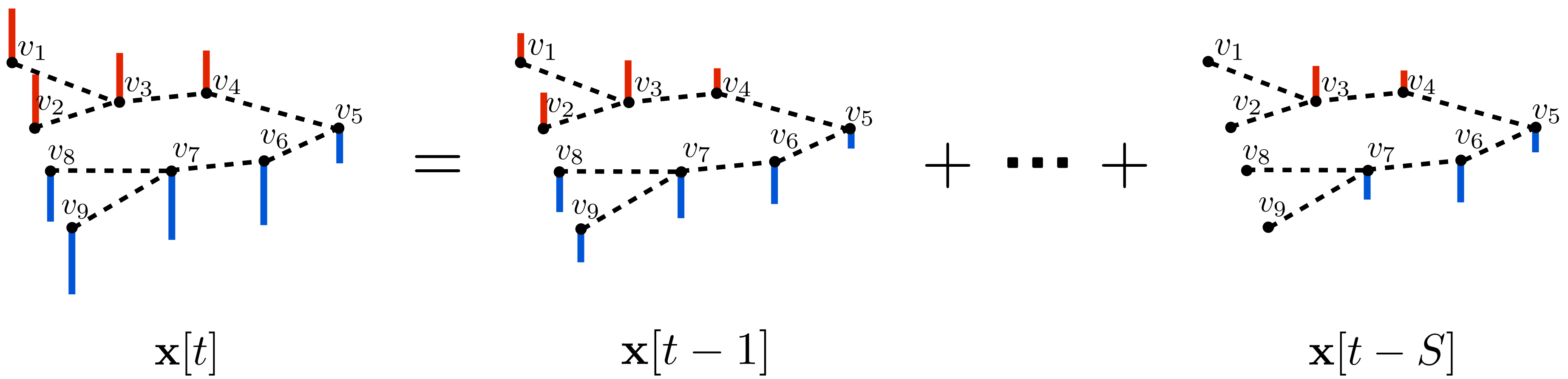
data fidelity — sparsity on **W** — sparsity on **a**

Model 3: Time-varying observations

- Mei and Moura (2015)



  - $\mathbf{D}_s(\mathcal{G}) = \mathbf{P}_s(\mathbf{W})$: polynomial of **W** of degree s
  - Define $\mathbf{c}_s$ as $\mathbf{x}[t-s]$



$$\min_{\mathbf{W},\mathbf{a}} \frac{1}{2} \sum_{k=S+1}^{K} ||\mathbf{x}[k] - \sum_{s=1}^{S} \mathbf{P}_s(\mathbf{W})\mathbf{x}[k-s]||_2^2 + \lambda_1 ||\text{vec}(\mathbf{W})||_1 + \lambda_2 ||\mathbf{a}||_1$$

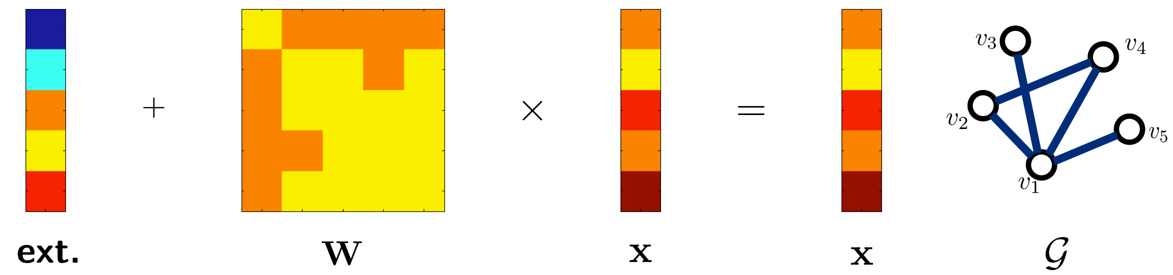
data fidelity   sparsity on **W**   sparsity on **a**

Polynomial design similar in spirit to Pasdeloup and Segarra

Good for inferring causal relations between signals

Kernelized version (nonlinear): Shen et al. (2016)

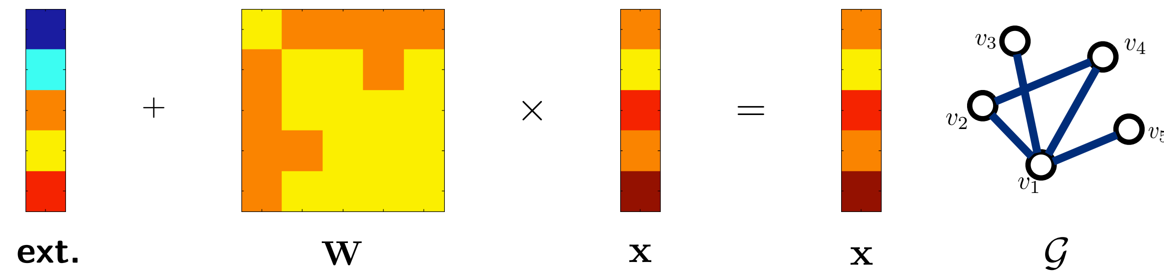
Model 3: Time-varying observations



- Baingana and Giannakis (2016)
  - $\mathbf{D}(\mathcal{G}) = \mathbf{W}^{\mathbf{s}(t)}$: Graph at time **t**

    (topologies switch at each time between S discrete states)
  - Define **c** as **x**

Model 3: Time-varying observations



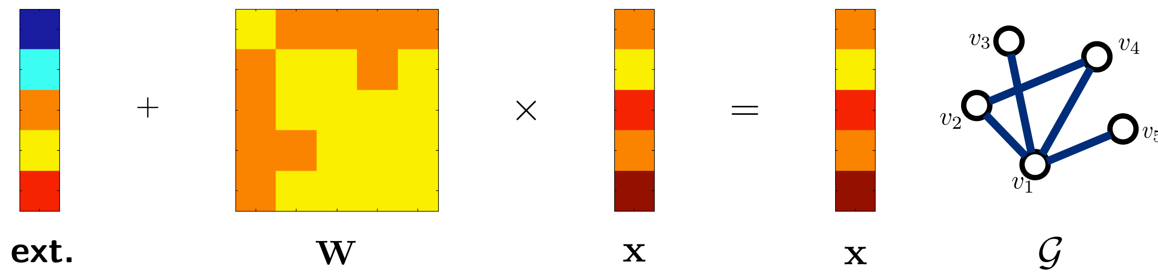
- Baingana and Giannakis (2016)
  - $\mathbf{D}(\mathcal{G}) = \mathbf{W}^{\mathbf{s}(t)}$: Graph at time **t**

    (topologies switch at each time between S discrete states)
  - Define **c** as **x**

$$\mathbf{x}[t] = \mathbf{W}^{\mathbf{s}(t)}\mathbf{x}[t] + \mathbf{B}^{\mathbf{s}(t)}\mathbf{y}[t]$$

internal (neighbors) external

Model 3: Time-varying observations



- Baingana and Giannakis (2016)
  - $\mathbf{D}(\mathcal{G}) = \mathbf{W}^{\mathbf{s}(t)}$: Graph at time **t**

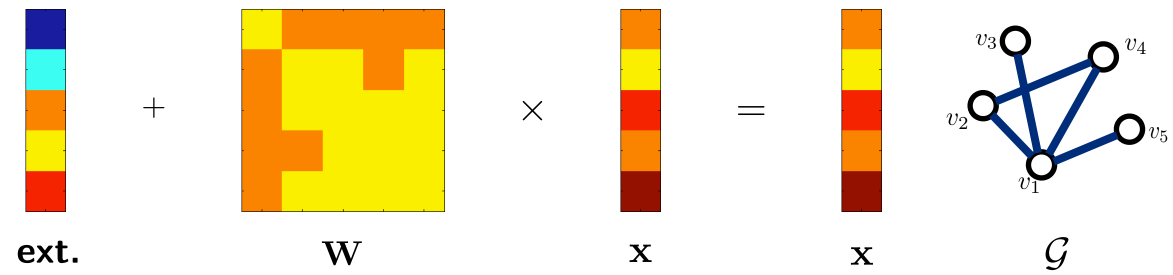
    (topologies switch at each time between S discrete states)
  - Define **c** as **x**

$$\mathbf{x}[t] = \underbrace{\mathbf{W}^{\mathbf{s}(t)}\mathbf{x}[t]}_{\text{internal (neighbors)}} + \underbrace{\mathbf{B}^{\mathbf{s}(t)}\mathbf{y}[t]}_{\text{external}}$$

  - Solve for all states of **W**:

$$\min_{\{\mathbf{W}^{\mathbf{s}(t)},\mathbf{B}^{\mathbf{s}(t)}\}} \frac{1}{2}\sum_{t=1}^{T} \underbrace{||\mathbf{x}[t] - \mathbf{W}^{\mathbf{s}(t)}\mathbf{x}[t] - \mathbf{B}^{\mathbf{s}(t)}\mathbf{y}[t]||_F^2}_{\text{data fidelity}} + \sum_{s=1}^{S} \lambda_s \underbrace{||\mathbf{W}^{\mathbf{s}(t)}||_1}_{\text{sparsity on } \mathbf{W}}$$

Model 3: Time-varying observations



- Baingana and Giannakis (2016)
  - $\mathbf{D}(\mathcal{G}) = \mathbf{W}^{\mathbf{s}(t)}$: Graph at time **t**

    (topologies switch at each time between S discrete states)
  - Define **c** as **x**

$$\mathbf{x}[t] = \mathbf{W}^{\mathbf{s}(t)}\mathbf{x}[t] + \mathbf{B}^{\mathbf{s}(t)}\mathbf{y}[t]$$

internal (neighbors) external

  - Solve for all states of **W**:

$$\min_{\{\mathbf{W}^{\mathbf{s}(t)}, \mathbf{B}^{\mathbf{s}(t)}\}} \frac{1}{2}\sum_{t=1}^{T} ||\mathbf{x}[t] - \mathbf{W}^{\mathbf{s}(t)}\mathbf{x}[t] - \mathbf{B}^{\mathbf{s}(t)}\mathbf{y}[t]||_F^2 + \sum_{s=1}^{S} \lambda_s ||\mathbf{W}^{\mathbf{s}(t)}||_1$$

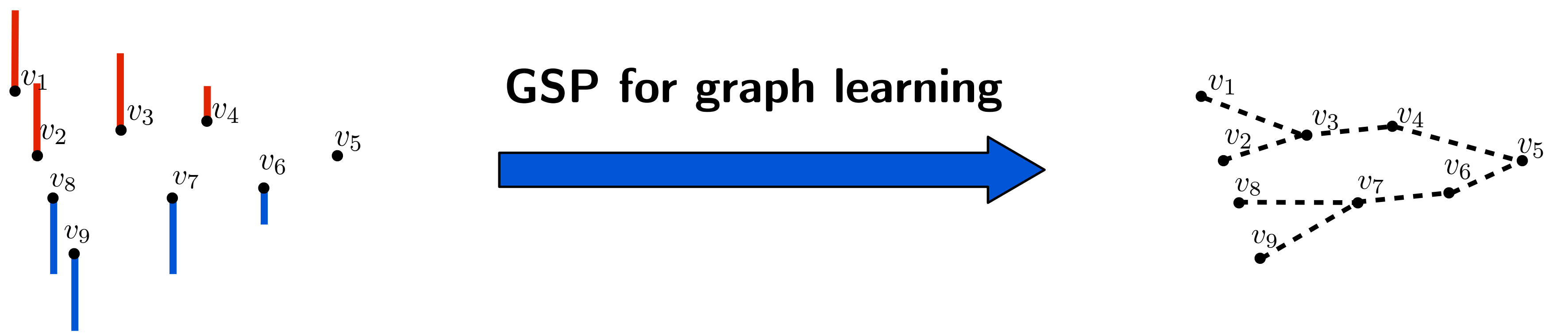
data fidelity sparsity on **W**

Good for inferring causal relations between signals as well as dynamic topologies

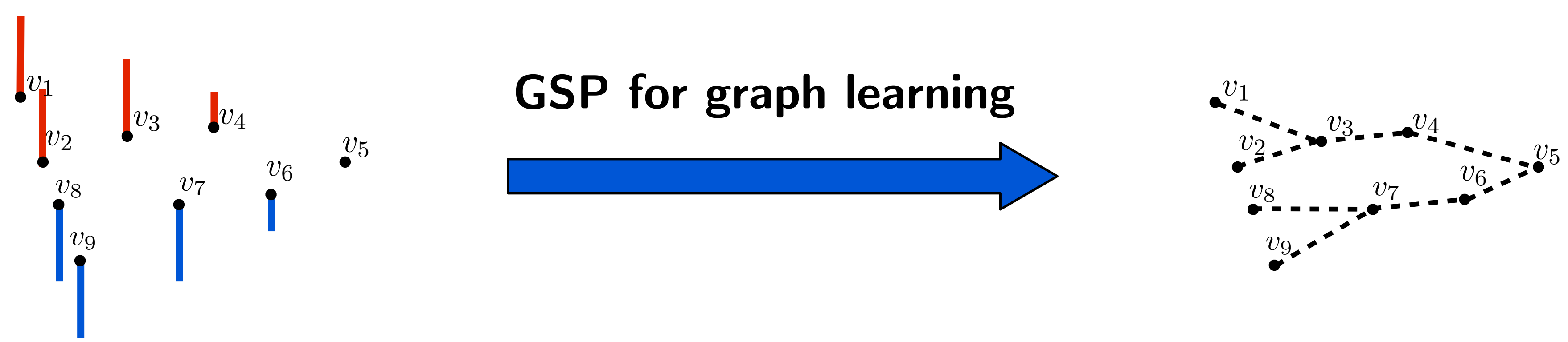
Comparison of different methods

Methods	Signal model	Assumption	Learning output	Edge direction	Inference
Dong (2015)	Global smoothness	Gaussian	Laplacian	Undirected	Signal-centric
Kalofolias (2016)	Global smoothness	Gaussian	Adjacency	Undirected	Signal-centric
Egilmez (2016)	Global smoothness	Gaussian	Generalized Laplacian	Undirected	Signal-centric
Chepuri (2016)	Global smoothness	Gaussian	Adjacency	Undirected	Graph-centric
Pasdeloup (2015)	Diffusion by Adj.	Stationary	Normalized Adj./ Laplacian	Undirected	Graph-centric
Segarra (2016)	Diffusion by Graph shift operator	Stationary	Graph shift operator	Undirected	Graph-centric
Thanou (2016)	Heat diffusion	Sparsity	Laplacian	Undirected	Signal-centric
Mei (2015)	Time-varying	Dependent on previous states	Adjacency	Directed	Signal-centric
Baingana (2016)	Time-varying	Dependent on current int/ext info	Time-varying Adjacency	Directed	Signal-centric

Perspective



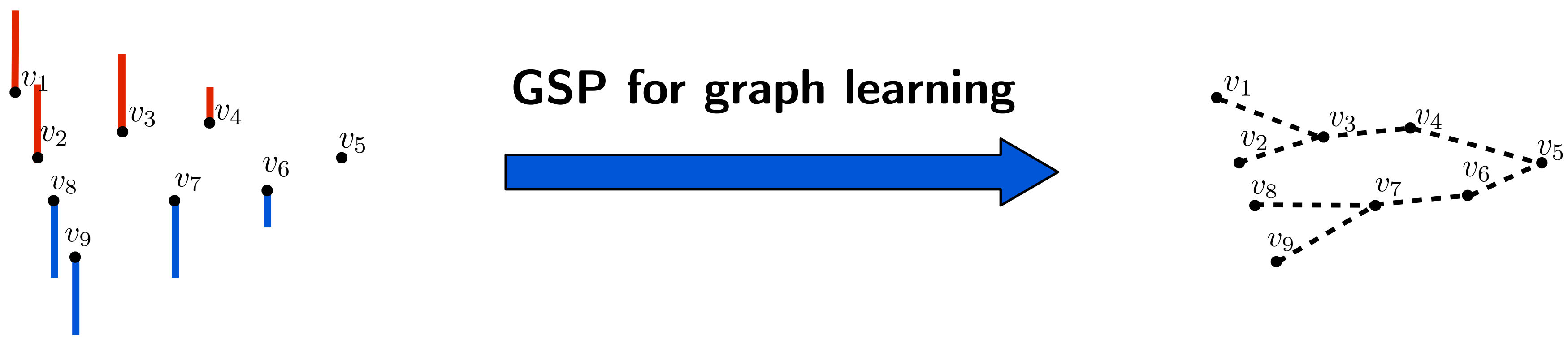
Perspective



Learning input

- **missing observations**
- **partial observations, e.g., by sampling**

Perspective



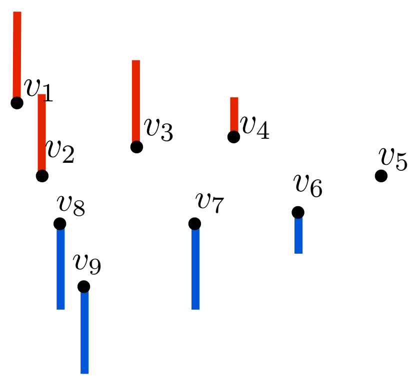
Learning input

- **missing observations**
- **partial observations, e.g., by sampling**

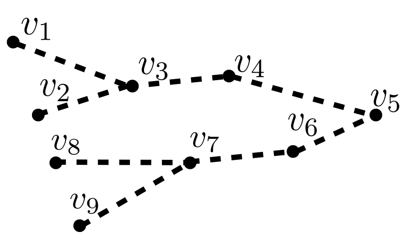
Learning output

- **directed graphs (Shen 2017)**
- **time-varying graphs (Kalofolias 2017)**
- **multi-layer graphs**
- **subgraphs or “ego-networks”**
- **intermediate graph representation**

Perspective



GSP for graph learning



Learning input

- **missing observations**
- **partial observations, e.g., by sampling**

Signal/graph model

- **beyond smoothness: localization in vertex-frequency domain, bandlimited (Sardellitti 2017)**

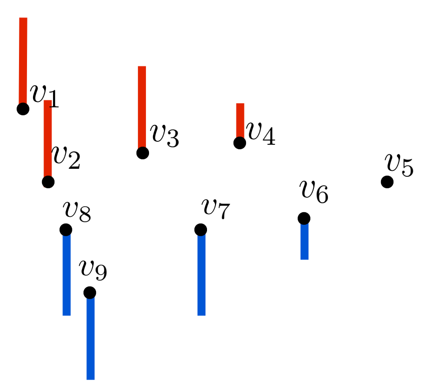
Learning output

- **directed graphs (Shen 2017)**
- **time-varying graphs (Kalofolias 2017)**
- **multi-layer graphs**
- **subgraphs or "ego-networks"**
- **intermediate graph representation**

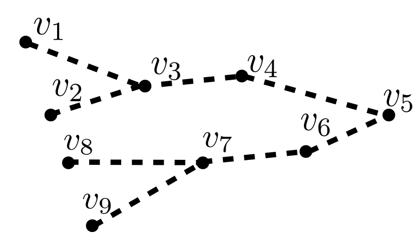
Perspective

Theoretical consideration

- performance guarantee (Rabbat 2017)
- computational efficiency



GSP for graph learning



Learning input

- missing observations
- partial observations, e.g., by sampling

Signal/graph model

- beyond smoothness: localization in vertex-frequency domain, bandlimited (Sardellitti 2017)

Learning output

- directed graphs (Shen 2017)
- time-varying graphs (Kalofolias 2017)
- multi-layer graphs
- subgraphs or "ego-networks"
- intermediate graph representation

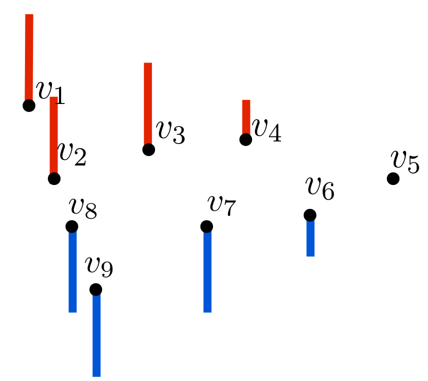
Perspective

Learning objective

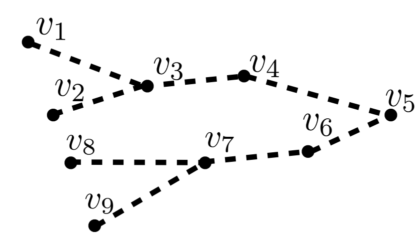
- for what SP applications? e.g., classification (Yankelevsky 2016), coding and compression (Rotondo 2015, Fracastoro 2016)
- for traditional graph-based learning, e.g., clustering, dim. reduction, ranking

Theoretical consideration

- performance guarantee (Rabbat 2017)
- computational efficiency



GSP for graph learning



Learning input

- missing observations
- partial observations, e.g., by sampling

Signal/graph model

- beyond smoothness: localization in vertex-frequency domain, bandlimited (Sardellitti 2017)

Learning output

- directed graphs (Shen 2017)
- time-varying graphs (Kalofolias 2017)
- multi-layer graphs
- subgraphs or "ego-networks"
- intermediate graph representation

Graph learning at GSPW 2017

Thursday June 1st

8:30 – 10:10		Graph Structural Analysis and Topology Identification		
8:30	8:55	Oguzhan Teke and P.P. Vaidyanathan	Discrete Uncertainty Principles and Sparse Eigenvectors	Oguzhan Teke
8:55	9:20	Eduardo Pavez, Hilmi E. Eglimez, and Antonio Ortega	Learning Graphs with Structured Sparsity Properties: Theoretical Analysis and Algorithms	Eduardo Pavez
9:20	9:45	Hoi-To Wai, Anna Scaglione, Amir Leshem, Sissi Xiaoxiao Wu, Uzi Harush, and Barush Barzel	Network RADAR: Theory and Practice for Network Topology Inference from Perturbation Data	Hoi-To Wai
9:45	10:10	Paul Bogdan	Compact yet Accurate Mathematical Modeling: New Mathematical Tools for Graph Topology Inference	Paul Bogdan

Friday June 2nd

8:30 – 10:35		System Identification and Statistical Processing on Graphs		
8:30	8:55	Abhishek Deb, Nagaraj T. Janakiraman, and Krishna R. Narayanan	Exploring connections between Spectral Estimation for Graph Signals, Coding Theory and Compressed Sensing	Nagaraj T. Janakiraman
8:55	9:20	Antonio G. Marques and Santiago Segarra	Joint Inference of Multiple Networks from Stationary Graph Signals	Antonio Marques
9:20	9:45	Rasoul Shafipour, Santiago Segarra, Antonio G. Marques and Gonzalo Mateos	Network Topology Inference from Non-Stationary Graph Signals	Gonzalo Mateos
9:45	10:10	Fernando Gama and Alejandro Ribeiro	Optimal Graph Filter for Estimating the Mean of a WSS Graph Process	Alejandro Ribeiro
10:10	10:35	Arman Hasanzadeh, Xi Liu, Krishna Narayanan, Nick Duffield, Byron Chigoy and Shawn Turner	Congestion Detection and Traffic Prediction in Transportation Networks Using Graph Signal Processing	Arman Hasanzadeh

References

- B. Baingana and G. B. Giannakis. Tracking switching network topologies from propagating graph signals. In Graph Signal Processing Workshop, 2016.
- B. Baingana and G. B. Giannakis. Tracking switched dynamic network topologies from information cascades. IEEE Transactions on Signal Processing, 65(4):985-997, 2017.
- O. Banerjee, L. El Ghaoui, and A. d'Aspremont. Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data. The Journal of Machine Learning Research, 9:485-516, 2008.
- S. P. Chepuri, S. Liu, G. Leus, and A. O. Hero III. Learning sparse graphs under smoothness prior. arXiv:1609.03448, 2016.
- D. I Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. IEEE Signal Processing Magazine, 30(3):83-98, 2013.
- S. I. Daitch, J. A. Kelner, and D. A. Spielman. Fitting a graph to vector data. In Proceedings of the International Conference on Machine Learning, 201-208, 2009.
- A. P. Dempster. Covariance selection. Biometrics, 28(1):157-175, 1972.
- X. Dong, D. Thanou, P. Frossard, and P. Vandergheynst. Laplacian matrix learning for smooth graph signal representation. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 3736-3740, 2015.
- X. Dong, D. Thanou, P. Frossard, and P. Vandergheynst. Learning laplacian matrix in smooth graph signal representations. IEEE Transactions on Signal Processing, 64(23):6160-6173, 2016.

References

- H. E. Egilmez, E. Pavez, and A. Ortega. Graph learning from data under structural and laplacian constraints. arXiv:1611.05181, 2016.
- G. Fracastoro, D. Thanou, and P. Frossard. Graph transform learning for image compression. In Proceedings of the Picture Coding Symposium, 2016.
- J.Friedman, T.Hastie, and R.Tibshirani. Sparse inverse covariance estimation with the graphical lasso. Biostatistics, 9(3):432-441, 2008.
- D. K. Hammond, P. Vandergheynst, and R. Gribonval. Wavelets on graphs via spectral graph theory. Applied and Computational Harmonic Analysis, 30(2):129-150, 2011.
- C.-J. Hsieh, I. S. Dhillon, P. K. Ravikumar, and M. A. Sustik. Sparse inverse covariance matrix estimation using quadratic approximation. In Advances in Neural Information Processing Systems 24, 2330-2338, 2011.
- C. Hu, L. Cheng, J. Sepulcre, G. E. Fakhri, Y. M. Lu, and Q. Li. A graph theoretical regression model for brain connectivity learning of Alzheimer's disease. In Proceedings of the IEEE International Symposium on Biomedical Imaging, 616-619, 2013.
- V. Kalofolias. How to learn a graph from smooth signals. In Proceedings of the International Conference on Artificial Intelligence and Statistics, 920-929, 2016.
- V. Kalofolias, A. Loukas, D. Thanou, and P. Frossard. Learning time varying graphs. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 2826-2830, 2017.
- B. Lake and J. Tenenbaum. Discovering structure by learning sparse graph. In Proceedings of the Annual Cognitive Science Conference, 2010.

References

- H. P. Maretic, D. Thanou, and P. Frossard. Graph learning under sparsity priors. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 6523-6527, 2017.
- J. Mei and J. M. F. Moura. Signal processing on graphs: Estimating the structure of a graph. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 5495-5499, 2015.
- N. Meinshausen and P. Bühlmann. High-dimensional graphs and variable selection with the lasso. Annals of Statistics, 34(3):1436-1462, 2006.
- S. K. Narang and A. Ortega. Lifting based wavelet transforms on graphs. In Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, 441-444, 2009.
- B. Pasdeloup, V. Gripon, G. Mercier, D. Pastor, and M. G. Rabbat. Characterization and inference of weighted graph topologies from observations of diffused signals. arXiv:1605.02569, 2016.
- B. Pasdeloup, M. Rabbat, V. Gripon, D. Pastor, and G. Mercier. Graph reconstruction from the observation of diffused signals. In Proceedings of the Annual Allerton Conference, 1386-1390, 2015.
- P. Ravikumar, M. J. Wainwright, and J. Lafferty. High-dimensional Ising model selection using l1-regularized logistic regression. Annals of Statistics, 38(3):1287-1319, 2010.
- I. Rotondo, G. Cheung, A. Ortega, and H. E. Egilmez. Designing sparse graphs via structure tensor for block transform coding of images. In Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, 571-574, 2015.
- A. Sandryhaila and J. M. F. Moura. Discrete signal processing on graphs. IEEE Transactions on Signal Processing, 61(7):1644-1656, 2013.

References

- S. Segarra, A. G. Marques, G. Mateos, and A. Ribeiro. Network topology inference from spectral templates. arXiv:1608.03008, 2016.
- Y. Shen, B. Baingana, and G. B. Giannakis. Nonlinear structural vector autoregressive models for inferring effective brain network connectivity. arXiv:1610.06551, 2016.
- Y. Shen, B. Baingana, and G. B. Giannakis. Kernel-based structural equation models for topology identification of directed networks. IEEE Transactions on Signal Processing, 65(10):2503-2516, 2017.
- M. Slawski and M. Hein. Estimation of positive definite M-matrices and structure learning for attractive Gaussian Markov random fields. Linear Algebra and its Applications, 473:145-179, 2015.
- D. Thanou, D. I Shuman, and P. Frossard. Learning parametric dictionaries for signals on graphs. IEEE Transactions on Signal Processing, 62(15):3849-3862, 2014.
- D. Thanou, X. Dong, D. Kressner, and P. Frossard. Learning heat diffusion graphs. arXiv:1611.01456, 2016.
- Y. Yankelevsky and M. Elad. Dual graph regularized dictionary learning. IEEE Transactions on Signal and Information Processing over Networks, 2(4):611-624, 2016.
- X. Zhang, X. Dong, and P. Frossard. Learning of structured graph dictionaries. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 3373-3376, 2012.