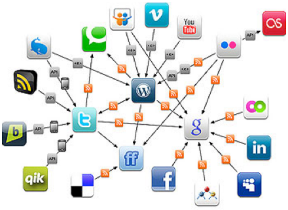
Graph Signal Processing

Gonzalo Mateos
Dept. of ECE and Goergen Institute for Data Science
University of Rochester
gmateosb@ece.rochester.edu
http://www.hajim.rochester.edu/ece/sites/gmateos/

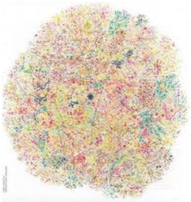
April 2, 2023

Network Science analytics

Online social media



Internet



Clean energy and grid analytics

- Network as graph $G(V, E)$: encode pairwise relationships
- Desiderata: Process, analyze and learn from network data [Kolaczyk'09]
  - $\Rightarrow$ Use G to study graph signals, data associated with nodes in V
- Ex: Opinion profile, buffer congestion levels, neural activity, epidemic

Roadmap

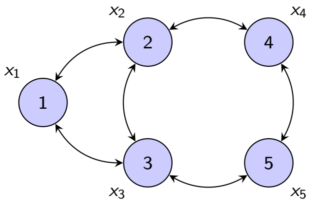
Graph signal processing: Motivation and fundamentals

Learning graphs from observations of smooth signals

Case study: Discriminative graph learning for emotion recognition

Graph signal processing (GSP)

- Graph G with adjacency matrix $\mathbf{A}$
  $\Rightarrow A_{ij}$ = proximity between i and j
- Signal $\mathbf{x} \in \mathbb{R}^{N_v}$ on top of the graph
  $\Rightarrow x_i$ = signal value at node i

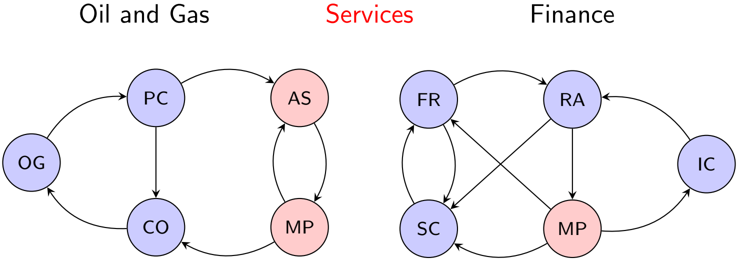


- Graph Signal Processing → Exploit structure encoded in $\mathbf{A}$ to process $\mathbf{x}$
- **Q:** Graph signals common and interesting as networks are?
- **Q:** Why do we expect the graph structure to be useful in processing $\mathbf{x}$?

A. Ortega et al, "Graph signal processing: Overview, challenges, and applications," *Proc. IEEE*, 2018

Network of economic sectors of the United States

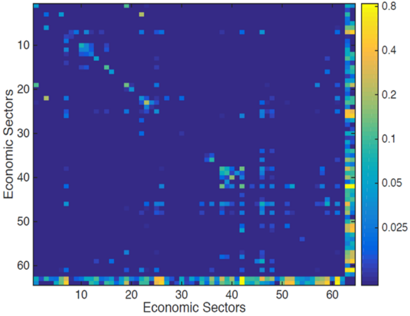
- Bureau of Economic Analysis of the U.S. Department of Commerce
  - A_{ij} = Output of sector i that becomes input to sector j (62 sectors)



- Oil extraction (OG), Petroleum and coal products (PC), Construction (CO)
- Administrative services (AS), Professional services (MP)
- Credit intermediation (FR), Securities (SC), Real state (RA), Insurance (IC)
- Only interactions stronger than a threshold are shown

Network of economic sectors of the United States

- Bureau of Economic Analysis of the U.S. Department of Commerce
  - A_{ij} = Output of sector i that becomes input to sector j (62 sectors)

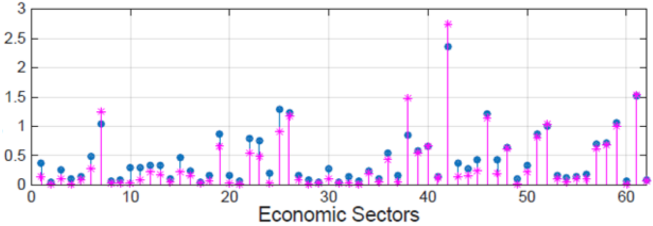


- A few sectors have widespread strong influence (services, finance, energy)
- Some sectors have strong indirect influences (oil)
- The heavy last row is final consumption

- This is an interesting network $\Rightarrow$ Signals on this graph are as well

Disaggregated GDP of the United States

- Signal $\mathbf{x}$ = output per sector = disaggregated GDP
  - $\Rightarrow$ Network structure used to, e.g., reduce GDP estimation noise

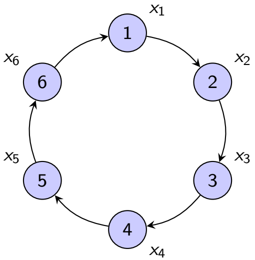


- Signal is as interesting as the network itself. Arguably more
  - Same is true for brain connectivity and fMRI brain signals, ...
  - Gene regulatory networks and gene expression levels, ...
  - Online social networks and information cascades, ...

Importance of signal structure in time

- Signal and Information Processing is about exploiting signal structure

- Discrete time described by cyclic graph
  - $\Rightarrow$ Time n follows time $n-1$
  - $\Rightarrow$ Signal value x_n similar to x_{n-1}
- Formalized with the notion of frequency

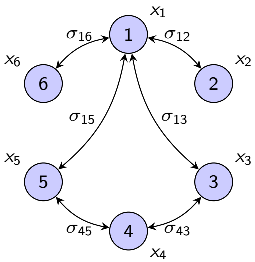


- Cyclic structure $\Rightarrow$ Fourier transform $\Rightarrow \tilde{\mathbf{x}} = \mathbf{F}^H \mathbf{x}$ $\left(F_{kn} = \dfrac{e^{j2\pi kn/N_v}}{\sqrt{N_v}} \right)$
- **Fourier transform** $\Rightarrow$ Projection on eigenvector space of cycle

Covariances and principal components

- Random signal with mean $\mathbb{E}[\mathbf{x}] = 0$ and covariance $\mathbf{C}_x = \mathbb{E}[\mathbf{x}\mathbf{x}^H]$
  $\Rightarrow$ Eigenvector decomposition $\mathbf{C}_x = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^H$

- Covariance matrix $\mathbf{A} = \mathbf{C}_x$ is a graph
  $\Rightarrow$ Not a very good graph, but still
- Precision matrix $\mathbf{C}_x^{-1}$ a common graph too
  $\Rightarrow$ Conditional dependencies of Gaussian $\mathbf{x}$



- Covariance matrix structure $\Rightarrow$ Principal components (PCA) $\Rightarrow \tilde{\mathbf{x}} = \mathbf{V}^H\mathbf{x}$
- **PCA transform** $\Rightarrow$ Projection on eigenvector space of (inverse) covariance
- **Q:** Can we extend these principles to general graphs and signals?

Graph Fourier Transform

- Adjacency $\mathbf{A}$, Laplacian $\mathbf{L}$, or, generically graph shift $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$

  $\Rightarrow$ $S_{ij} = 0$ for $i \neq j$ and $(i,j) \notin E$ (captures local structure in G)
- The Graph Fourier Transform (GFT) of $\mathbf{x}$ is defined as

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

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

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

  $\Rightarrow$ Eigenvectors $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_{N_v}]$ are the frequency basis (atoms)
- Additional structure

  $\Rightarrow$ If $\mathbf{S}$ is normal, then $\mathbf{V}^{-1} = \mathbf{V}^H$ and $\tilde{x}_k = \mathbf{v}_k^H\mathbf{x} = < \mathbf{v}_k, \mathbf{x} >$

  $\Rightarrow$ Parseval holds, $\|\mathbf{x}\|^2 = \|\tilde{\mathbf{x}}\|^2$
- **GFT** $\Rightarrow$ Projection on eigenvector space of graph shift operator $\mathbf{S}$

Frequency modes of the Laplacian

- Total variation of signal $\mathbf{x}$ with respect to $\mathbf{L}$

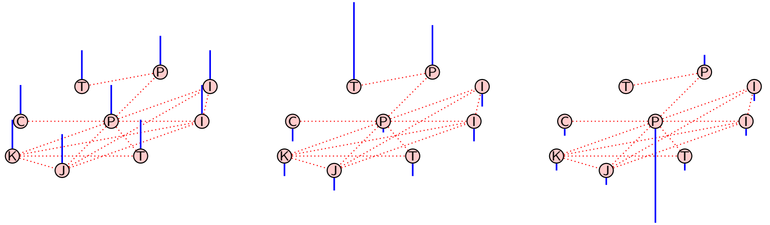
$$\text{TV}(\mathbf{x}) = \mathbf{x}^\top \mathbf{L}\mathbf{x} = \sum_{i,j=1,j>i}^{N_v} A_{ij}(x_i - x_j)^2$$

$\Rightarrow$ Smoothness measure on the graph G (Dirichlet energy)

- For Laplacian eigenvectors $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_{N_v}]$ $\Rightarrow$ $\text{TV}(\mathbf{v}_k) = \lambda_k$

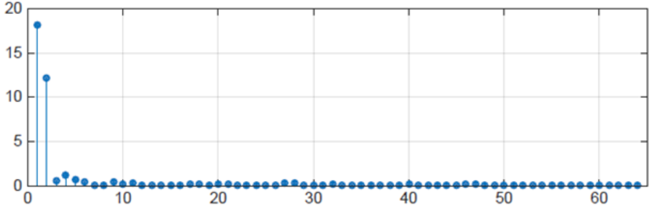
$\Rightarrow$ Can view $0 = \lambda_1 < \cdots \leq \lambda_{N_v}$ as frequencies

- Ex: gene network, $N_v = 10$, $k = 1$, $k = 2$, $k = 9$



Is this a reasonable transform?

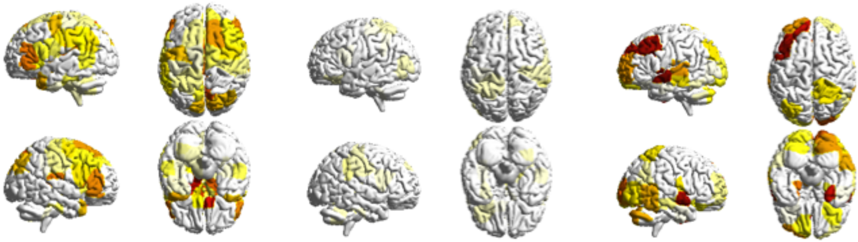
- Particularized to cyclic graphs $\Rightarrow$ GFT $\equiv$ Fourier transform
- Also for covariance graphs $\Rightarrow$ GFT $\equiv$ PCA transform
- But really, this is an empirical question. GFT of disaggregated GDP



- Spectral domain representation characterized by a few coefficients
  - $\Rightarrow$ Notion of bandlimitedness: $\mathbf{x} = \sum_{k=1}^{K} \tilde{x}_k \mathbf{v}_k$
  - $\Rightarrow$ Sampling, compression, filtering, pattern recognition

Graph frequency analysis of brain signals

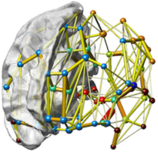
- GFT of brain signals during a visual-motor learning task [Huang et al'16]
  - ⇒ Decomposed into low, medium and high frequency components



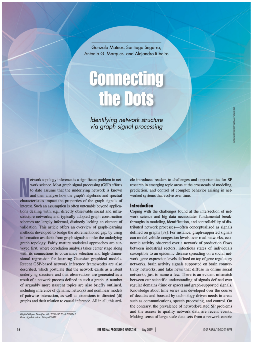
- Brain: Complex system where regularity coexists with disorder [Sporns'11]
  - ⇒ Signal energy mostly in the low and high frequencies
  - ⇒ In brain regions akin to the visual and sensorimotor cortices

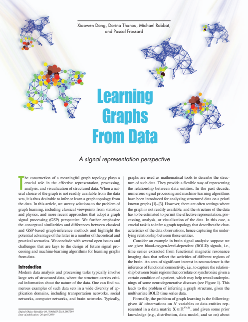
What is this class about?

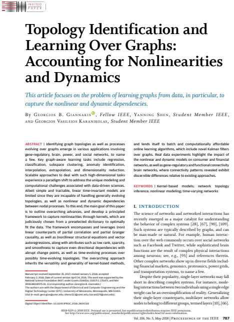
- Learning graphs from nodal observations
- Key in neuroscience
  - $\Rightarrow$ Functional network from fMRI signals



- Most GSP works: how known graph $\mathbf{S}$ affects signals and filters
- Here, reverse path: how to use GSP to infer the graph topology?
  - Graphical models [Egilmez et al'16], [Rabbat'17], [Kumar et al'19], ...
  - Smooth signals [Dong et al'15], [Kalofolias'16], [Sardellitti et al'17], ...
  - Graph filtering models [Shafipour et al'17], [Thanou et al'17], ...
  - Stationary signals [Pasdeloup et al'15], [Segarra et al'16], ...
  - Directed graphs [Mei-Moura'15], [Shen et al'16], ...







- Recent tutorials on learning graphs from data (with a GSP flavor)
  - IEEE Signal Processing Magazine and Proceedings of the IEEE

Learning graphs from smooth signals

Graph signal processing: Motivation and fundamentals

Learning graphs from observations of smooth signals

Case study: Discriminative graph learning for emotion recognition

Problem formulation

Rationale

- Seek graphs on which data admit certain regularities
  - Nearest-neighbor prediction (a.k.a. graph smoothing)
  - Semi-supervised learning
  - Efficient information-processing transforms
- Many real-world graph signals are smooth
  - Graphs based on similarities among vertex attributes
  - Network formation driven by homophily, proximity in latent space

Problem statement

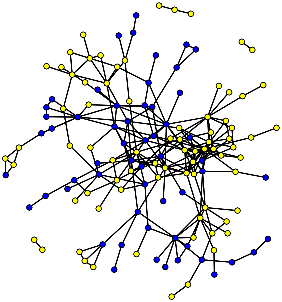
> Given observations $\mathcal{X} := \{\mathbf{x}_p\}_{p=1}^{P}$, identify a graph G such that signals in $\mathcal{X}$ are smooth on G.

- Criterion: Dirichlet energy on the graph G with Laplacian $\mathbf{L}$

$$\text{TV}(\mathbf{x}) = \mathbf{x}^\top \mathbf{L} \mathbf{x}$$

Example: Predicting protein function

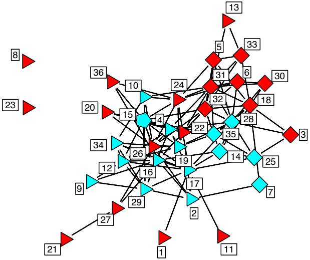
- Baker's yeast data, formally known as *Saccharomyces cerevisiae*
  - Graph: 134 vertices (proteins) and 241 edges (protein interactions)



- Signal: functional annotation intracellular signaling cascade (ICSC)
  - Signal transduction, how cells react to the environment
  - $x_i = 1$ if protein i annotated ICSC (yellow), $x_i = 0$ otherwise (blue)

Example: Predicting law practice

- Working relationships among lawyers [Lazega'01]
  - Graph: 36 partners, edges indicate partners worked together



- Signal: various node-level attributes $\mathbf{x} = \{x_i\}_{i \in V}$ including

  $\Rightarrow$ Type of practice, i.e., litigation (red) and corporate (cyan)

- Suspect lawyers collaborate more with peers in same legal practice

  $\Rightarrow$ Knowledge of collaboration useful in predicting type of practice

Laplacian-based factor analysis model

- Consider an unknown graph G with Laplacian $\mathbf{L} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^\top$
  - $\Rightarrow$ Adopt GFT basis $\mathbf{V}$ as signal representation matrix
- Factor analysis model for the observed graph signal

$$\mathbf{x} = \mathbf{V}\boldsymbol{\chi} + \boldsymbol{\epsilon}$$

  - $\Rightarrow$ Latent variables $\boldsymbol{\chi} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda}^\dagger)$ ($\approx$ GFT coefficients)
  - $\Rightarrow$ Isotropic error term $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$
- Smoothness: prior encourages low-pass bandlimited $\mathbf{x}$
  - $\Rightarrow$ Small eigenvalues of $\mathbf{L}$ (low freq.) $\rightarrow$ High-power factor loadings

X. Dong et al, "Learning Laplacian matrix in smooth graph signal representations," *IEEE Trans. Signal Process.*, 2016

Inference as denoising via graph kernel regression

- Maximum a posteriori (MAP) estimator of the latent variables χ

$$\hat{\chi}_{\text{MAP}} = \arg\min_{\chi} \left\{ \|\mathbf{x} - \mathbf{V}\chi\|^2 + \alpha\chi^\top \mathbf{\Lambda}\chi \right\}$$

  $\Rightarrow$ Parameterized by the unknown $\mathbf{V}$ and $\mathbf{\Lambda}$

- Define predictor $\mathbf{y} := \mathbf{V}\chi$, regularizer expressible as

$$\chi^\top \mathbf{\Lambda}\chi = \mathbf{y}^\top \mathbf{V}\mathbf{\Lambda}\mathbf{V}^\top \mathbf{y} = \mathbf{y}^\top \mathbf{L}\mathbf{y} = \text{TV}(\mathbf{y})$$

  $\Rightarrow$ Laplacian-based TV denoiser of $\mathbf{x}$, smoothness prior on $\mathbf{y}$

  $\Rightarrow$ Kernel-ridge regression with unknown $\mathbf{K} := \mathbf{L}^\dagger$ (graph filter)

- **Idea:** jointly search for $\mathbf{L}$ and denoised representation $\mathbf{y} = \mathbf{V}\chi$

$$\min_{\mathbf{L},\mathbf{y}} \left\{ \|\mathbf{x} - \mathbf{y}\|^2 + \alpha\mathbf{y}^\top \mathbf{L}\mathbf{y} \right\}$$

Formulation and algorithm

- Given signals $\mathcal{X} := \{\mathbf{x}_p\}_{p=1}^{P}$ in $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_P] \in \mathbb{R}^{N_v \times P}$, solve

$$\min_{\mathbf{L},\mathbf{Y}} \left\{ \|\mathbf{X} - \mathbf{Y}\|_F^2 + \alpha \text{trace}\left(\mathbf{Y}^\top \mathbf{L}\mathbf{Y}\right) + \frac{\beta}{2}\|\mathbf{L}\|_F^2 \right\}$$
$$\text{s. to} \quad \text{trace}(\mathbf{L}) = N_v,\ \mathbf{L}\mathbf{1} = \mathbf{0},\ L_{ij} = L_{ji} \leq 0,\ i \neq j$$

  $\Rightarrow$ Objective function: Fidelity + smoothness + edge sparsity

  $\Rightarrow$ Not jointly convex in $\mathbf{L}$ and $\mathbf{Y}$, but bi-convex

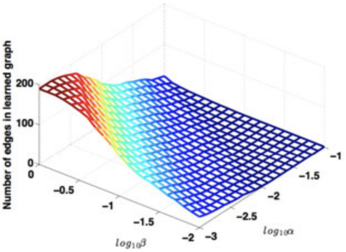
- Algorithmic approach: alternating minimization (AM), $O(N_v^3)$ cost
  - (S1) Fixed $\mathbf{Y}$: solve for $\mathbf{L}$ via interior-point method, ADMM (more soon)
  - (S2) Fixed $\mathbf{L}$: low-pass, graph filter-based smoother of the signals in $\mathbf{X}$

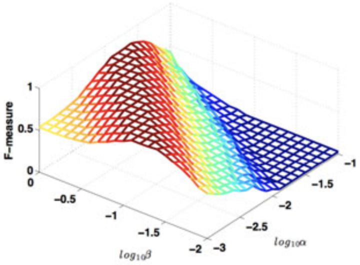
$$\mathbf{Y} = (\mathbf{I} + \alpha\mathbf{L})^{-1}\mathbf{X}$$

Impact of regularizers on sparsity and accuracy

- Generate multiple signals on a synthetic Erdős-Rényi graph

  $\Rightarrow$ Recover the graph for different values of α and β



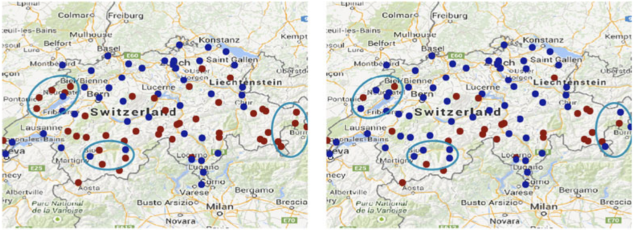


- More edges promoted by increasing β and decreasing α
- In the low noise regime, the ratio β/α determines behavior

Example: Temperature graph in Switzerland

- $N_v = 89$ stations measuring monthly temperature averages (1981-2010)

  $\Rightarrow$ Learn a graph G on which the temperatures vary smoothly
- Geographical distance not a good idea $\Rightarrow$ different altitudes



- Recover altitude partition from spectral clustering on G

  $\Rightarrow$ Red (high stations) and blue (low stations) clusters
- K-means applied directly to the temperatures (right) fails

Signal smoothness meets edge sparsity

- Recall $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_P] \in \mathbb{R}^{N_v \times P}$, let $\bar{\mathbf{x}}_i^\top \in \mathbb{R}^{1 \times P}$ denote its i-th row
  - $\Rightarrow$ Euclidean distance matrix $\mathbf{Z} \in \mathbb{R}_+^{N_v \times N_v}$, where $Z_{ij} := \|\bar{\mathbf{x}}_i - \bar{\mathbf{x}}_j\|^2$
- Neat trick: link between smoothness and sparsity

$$\sum_{p=1}^{P} \text{TV}(\mathbf{x}_p) = \text{trace}(\mathbf{X}^\top \mathbf{L} \mathbf{X}) = \frac{1}{2} \|\mathbf{A} \circ \mathbf{Z}\|_1$$

  - $\Rightarrow$ Sparse E when data come from a smooth manifold
  - $\Rightarrow$ Favor candidate edges (i, j) associated with small Z_{ij}
- Shows that edge sparsity on top of smoothness is redundant
- Parameterize graph learning problems in terms of $\mathbf{A}$ (instead of $\mathbf{L}$)
  - $\Rightarrow$ Advantageous since constraints on $\mathbf{A}$ are decoupled

V. Kalofolias, "How to learn a graph from smooth signals," *AISTATS*, 2016

Scalable topology identification framework

- General purpose model for learning graphs [Kalofolias'16]

$$\min_{\mathbf{A}} \left\{ \|\mathbf{A} \circ \mathbf{Z}\|_1 - \alpha \mathbf{1}^\top \log(\mathbf{A}\mathbf{1}) + \frac{\beta}{2} \|\mathbf{A}\|_F^2 \right\}$$
$$\text{s. to} \quad \text{diag}(\mathbf{A}) = \mathbf{0},\ A_{ij} = A_{ji} \geq 0,\ i \neq j$$

$\Rightarrow$ Logarithmic barrier forces positive degrees

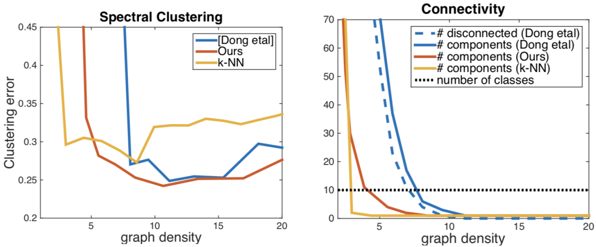
$\Rightarrow$ Penalize large edge-weights to control sparsity

- Primal-dual solver amenable to parallelization, $O(N_v^2)$ cost
- Laplacian-based factor analysis encore. Tackle **(S1)** as

$$\min_{\mathbf{A}} \left\{ \|\mathbf{A} \circ \mathbf{Z}\|_1 - \log(\mathbb{I}\left\{\|\mathbf{A}\|_1 = N_v\right\}) + \frac{\beta}{2} \left(\|\mathbf{A}\mathbf{1}\|^2 + \|\mathbf{A}\|_F^2\right) \right\}$$
$$\text{s. to} \quad \text{diag}(\mathbf{A}) = \mathbf{0},\ A_{ij} = A_{ji} \geq 0,\ i \neq j$$

Example: Learning the graph of USPS digits

- 1001 images of the 10 digits, but highly imbalanced ($2.6i^2$)
  - $\Rightarrow$ 10 classes via graph recovery plus spectral clustering
- Compare two methods based on smoothness and k-NN graph



- Performance more robust to graph density
  - $\Rightarrow$ Likely attributable to non-singleton nodes

Graph learning via edge subset selection

- **Idea:** parameterize the unknown topology via an edge indicator vector
- Complete graph on N_V nodes, having $M := \binom{N_v}{2}$ edges
  - $\Rightarrow$ Incidence matrix $\mathbf{B} := [\mathbf{b}_1, \ldots, \mathbf{b}_M] \in \mathbb{R}^{N_v \times M}$
- Laplacian of a candidate graph $G(V, E)$

$$\mathbf{L}(\boldsymbol{\omega}) = \sum_{m=1}^{M} \omega_m \mathbf{b}_m \mathbf{b}_m^\top$$

  - $\Rightarrow$ Binary edge indicator vector $\boldsymbol{\omega} := [\omega_1, \ldots, \omega_M]^\top \in \{0, 1\}^M$
  - $\Rightarrow$ Offers an explicit handle on the number of edges $\|\boldsymbol{\omega}\|_0 = |E|$

Problem: Given observations $\mathcal{X} := \{\mathbf{x}_p\}_{p=1}^P$, learn an unweighted graph $G(V, E)$ such that signals in $\mathcal{X}$ are smooth on G and $|E| = K$.

Cardinality-constrained Boolean optimization

- Natural formulation is to solve the non-convex problem

$$\min_{\boldsymbol{\omega}\in\{0,1\}^M} \text{trace}(\mathbf{X}^\top \mathbf{L}(\omega)\mathbf{X}), \quad \text{s. to } \|\boldsymbol{\omega}\|_0 = K$$

- Solution obtained through a simple rank-ordering procedure
  - Compute edge scores $c_m := \text{trace}(\mathbf{X}^\top(\mathbf{b}_m\mathbf{b}_m^\top)\mathbf{X})$
  - Set $\omega_m = 1$ for those K edges having the smallest scores
- More pragmatic AWGN setting where $\mathbf{x}_p = \mathbf{y}_p + \boldsymbol{\epsilon}_p$, $p = 1, \ldots, P$

$$\min_{\mathbf{Y},\boldsymbol{\omega}\in\{0,1\}^M} \left\{\|\mathbf{X}-\mathbf{Y}\|_F^2 + \alpha\text{trace}(\mathbf{Y}^\top \mathbf{L}(\omega)\mathbf{Y})\right\}, \quad \text{s. to } \|\boldsymbol{\omega}\|_0 = K$$

$\Rightarrow$ Tackle via AM or semidefinite relaxation (SDR)

S. Chepuri et al, "Learning sparse graphs under smoothness prior," *ICASSP*, 2017

Comparative summary

- Noteworthy features of the edge subset selection approach
  - ✓ Direct control on edge sparsity
  - ✓ Simple algorithm in the noise-free case
  - ✓ Devoid of Laplacian feasibility constraints
  - ✗ Does not guarantee connectivity of G
  - ✗ No room for optimizing edge weights

- Scalable framework in [Kalofolias'16] also quite flexible

$$\min_{\mathbf{A}} \{\|\mathbf{A} \circ \mathbf{Z}\|_1 + g(\mathbf{A})\}$$
$$\text{s. to} \quad \text{diag}(\mathbf{A}) = \mathbf{0},\ A_{ij} = A_{ji} \geq 0,\ i \neq j$$

$\Rightarrow$ Subsumes the factor-analysis model [Dong et al'16]

$\Rightarrow$ Recovers Gaussian kernel weights $A_{ij} := \exp\left(-\frac{\|\bar{\mathbf{x}}_i - \bar{\mathbf{x}}_j\|^2}{\sigma^2}\right)$ for

$$g(\mathbf{A}) = \sigma^2 \sum_{i,j} A_{ij}(\log(A_{ij}) - 1)$$

Case study

Graph signal processing: Motivation and fundamentals

Learning graphs from observations of smooth signals

Case study: Discriminative graph learning for emotion recognition

Case study: Classification of network data

- Labeled graph signals $\mathcal{X}_c := \{\mathbf{x}_p^{(c)}\}_{p=1}^{P_c}$ from C different classes
  - $\Rightarrow$ Signals in each class possess a very distinctive structure
- **As.:** Class c signals are smooth w.r.t. unknown $G_c(V, E_c)$
- Multiple linear subspace model
  - $\Rightarrow$ Signals spanned by few Laplacian modes (GFT components)
  - $\Rightarrow$ Like susbpace clustering [Vidal'11], but with supervision

Problem statement

Given training signals $\mathcal{X} = \bigcup_{c=1}^{C} \mathcal{X}_c$, learn discriminative graphs $\mathbf{A}_c$ under smoothness priors to classify test signals via GFT projections.

Discriminative graph learning

- Discriminative graph learning per class c

$$\min_{\mathbf{A}_c} \left\{ \|\mathbf{A}_c \circ \mathbf{Z}_c\|_1 - \alpha \mathbf{1}^\top \log(\mathbf{A}_c \mathbf{1}) + \frac{\beta}{2}\|\mathbf{A}_c\|_F^2 - \gamma \sum_{k \neq c}^{C} \|\mathbf{A}_c \circ \mathbf{Z}_k\|_1 \right\}$$

$$\text{s. to} \quad \text{diag}(\mathbf{A}_c) = \mathbf{0},\ [\mathbf{A}_c]_{ij} = [\mathbf{A}_c]_{ji} \geq 0,\ i \neq j$$

$\Rightarrow$ Capture the underlying graph topology (class c structure)

$\Rightarrow$ Discriminability to boost classification performance

- **Q:** Given graphs $\{\hat{\mathbf{A}}_c\}_{c=1}^{C}$, how do we classify a test signal $\mathbf{x}$?
- Pass $\mathbf{x}$ through a filter-bank with C low-pass filters (LPFs)

$$\tilde{\mathbf{x}}_{F,c} = \text{diag}(\tilde{\mathbf{h}})\hat{\mathbf{V}}_c^\top \mathbf{x} \quad \Rightarrow \quad \hat{c} = \underset{c}{\text{argmax}} \left\{ \|\tilde{\mathbf{x}}_{F,c}\|^2 \right\}$$

$\Rightarrow$ LPF frequency response $\tilde{\mathbf{h}}$, learned class-c GFT basis $\hat{\mathbf{V}}_c$

Emotion recognition from EEG records

- Discriminative graph learning for emotion recognition from EEG signals
- DEAP dataset $\Rightarrow$ 32 subjects watch music videos (40 trials each)
  - Asked to rate videos: valence, arousal, like/dislike, dominance
  - Focus on valence labels: low (1-5 rating) and high (6-10 rating)
  - Signals acquired from $N_v = 32$ EEG channels
- We perform a subject-specific valence classification task

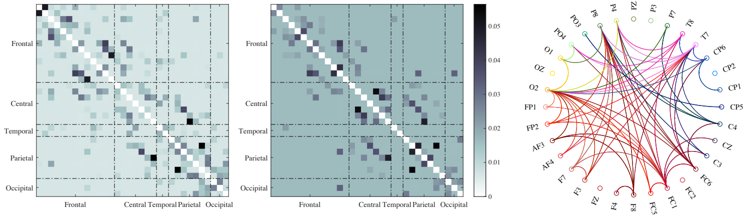
  $\Rightarrow$ Learn $C = 2$ graphs and project onto the 8 smoothest modes

  $\Rightarrow$ Report leave-one (trial)-out classification accuracy
- Mean classification accuracy over subjects is 92.73%

S. S. Saboksayr et al, "Online discriminative graph learning from multi-class smooth signals," *Signal Processing*, 2022

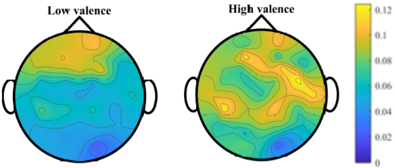
Valence classification

- **Q:** What information do we glean from the class-conditional graphs?



Low valence High valence Different connections (p=0.002)

- Connectivity increases with emotion intensity (frontal lobe links)



- Asymmetric frontal activity apparent from the 8 smoothest modes

Glossary

- Graph signal
- Graph signal processing
- Fourier transform
- Covariance matrix
- Principal component analysis
- Graph shift operator
- Graph Fourier transform
- Topology identification
- Smooth signal
- Dirichlet energy
- Factor analysis model
- Alternating minimization
- Euclidean distance matrix
- Edge subset selection
- Gaussian kernel graph
- Multiple subspace model
- Subspace clustering
- Discriminative graphs
- Low-pass graph filter
- Emotion recognition
- EEG signals
- Valence classification