Current Approaches to Graph Signal Processing



Pavel Loskot

pavelloskot@intl.zju.edu.cn





SIGNAL 2021: The Sixth International Conference on Advances in Signal, Image and Video Processing

May 30, 2021 to June 03, 2021 - Valencia, Spain

Авоит Ме



Pavel Loskot joined the ZJU-UIUC Institute as Associate Professor in January 2021. He received his PhD degree in Wireless Communications from the University of Alberta in Canada, and the MSc and BSc degrees in Radioelectronics and Biomedical Electronics, respectively, from the Czech Technical University of Prague. He is the Senior Member of the IEEE, Fellow of the HEA in the UK, and the Recognized Research Supervisor of the UKCGE.

In past 25 years, he was involved in numerous industrial and academic collaborative projects in the Czech Republic, Finland, Canada, the UK, Turkey, and China. These projects concerned mainly wireless and optical telecommunication networks, but also genetic circuits, air transport services, and renewable energy systems. This experience allowed him to truly understand the interdisciplinary workings, and crossing the disciplines boundaries.

His current research focuses on statistical signal processing and importing methods from Telecommunication Engineering and Computer Science to model and analyze systems more efficiently and with greater information power.

OBJECTIVES

- 1. Survey mainstream approaches to graph signal processing
- this tutorial deliberately does not add any new ideas or contributions to existing knowledge
- 2. Provide a starting point for researchers wishing to explore this area
- a few key ideas are identified that are exploited in many graph signal processing problems

OUTLINE

- 1. Why graph signal processing
- 2. Key ideas in graph signal processing
- 3. Survey of problems and tasks in graph signal processing
- 4. Recommended readings on graph signal processing

Part 1: Why graph signal processing?

GROWING INTEREST IN COMPLEX SYSTEMS

Networks seem to be suddenly appearing everywhere ...

Graphs

- mathematical models of networks
 → capture relations between pairs of objects
- many applications, well developed theory \rightarrow Graph Theory, Network Science
 - \rightarrow less so in Signal Processing

Complex systems

- appear to be designed as networks

 → distributing and pooling resources
 must be a fundamental principle of Nature
- tools to work with graphs
 - \rightarrow understanding complex systems





UNDERSTANDING GRAPHS

Graphs as data structure

• discrete mathematical structure

Graphs as computing models

- Gaussian networks
- Bayesian networks
- Markov random fields
- hidden Markov models
- finite state machines
- data flows (Tensorflow, Spark)

Graphs as system model

- physical networks
- social networks
- flow networks (telecommunications, transport, utilities)





Analysis strategy

- known tuples (structure, function) to train ML classifier (e.g. Deep Learning)
- common in biochemistry to predict protein function
- can replace Deep Packet Inspection with cheaper and faster ML classifiers

Synthesis strategy

• little explored Engineering territory





available measurements constrain possible applications



application determines required measurements

DATA PROCESSING

Signals and Systems

• concerned about creating and analyzing mathematical models

Signal processing

- models ←→ measurements (equivalence)
- numbers arranged into regular structures (scalars, vectors, matrices)
 → most algebraic operations defined for these regular data structures
- outcome is an algorithm

Machine Learning

- arbitrary data structures, mostly heuristic approaches
- the aim is discovery of patterns, relationships and knowledge

Computer programming

• different types of data structures and associated algorithms

STRUCTURED SIGNALS

Sets of scalar numbers

- V_1, V_2, V_3, \dots
- $V_1(t), V_2(t), V_3(t), \dots$ scalar index $t \in \mathcal{R}$
- $V_1(\boldsymbol{x}), V_2(\boldsymbol{x}), V_3(\boldsymbol{x}), \dots$ vector index $\boldsymbol{x} \in \mathcal{R}^K$

Discrete structures

• matrix vs. tensor: the latter has rank and additional constraints

Main idea

- map scalars V_1, V_2, V_3, \ldots to elements of discrete structures
- such mapping should reflect mutual relationships among V_1, V_2, V_3, \ldots

GRAPH SIGNALS

Scalars at nodes

- example scenarios: social networks, gene interaction networks
 - \rightarrow nodes has properties/attributes/features
 - \rightarrow edges indicate pairwise relationships
- measurements at each network node $v \in V$, |V| = N
- graph signal: $\mathbf{v} = [v_1, v_2, \dots, v_N]$ (special case, $N \to \infty$)

Terminology

Graph	Network	System		
vertex	node	component		
edge	link	interaction		

GRAPH SIGNALS (CONT.)

Scalars at edges

- example scenario: road and Internet traffic flows
 - \rightarrow edges has properties/attributes/features
 - \rightarrow edges are typically directed
- measurements at network edges $e \in E$
- graph signal: $e = [e_1, e_2, ..., e_{|E|}]$

Node	FUNCTION	FLOW BALANCE			
sink	absorbed flows	inflows > outflows			
source	generates flows	inflows < outflows			
router	mix and split flows	inflows = outflows			

GRAPH SIGNALS (CONT.)

Scalars at simplexes

- *k*-simplex: a closed-path object with *k* edges and k+1 nodes
 - \rightarrow 0-simplexes are nodes
 - \rightarrow 1-simplexes are edges (pairwise relations or flows)
 - \rightarrow 2-simplexes are open/closed triangles (triplewise relations)

Simplical complexes of order k

graph partitioning (a.k.a. graph cuts)
 → often aim to find the minimum cut

MATHEMATICAL REPRESENTATION OF GRAPHS

Adjacency matrix

$$[\mathbf{A}]_{ij} = \begin{cases} 1 & e_{ij} = 1 \text{ (edge from node } v_i \text{ to } v_j) \\ 0 & e_{ij} = 0 \text{ (no edge between } v_i \text{ and } v_j) \end{cases}$$

Weight matrix

$$[\boldsymbol{W}]_{ij} = \begin{cases} w \in \mathcal{R} & i \neq j \text{ (implies fully connected graph)} \\ 0 & i = j \end{cases}$$

Degree matrix

$$[\mathbf{D}]_{ij} = \begin{cases} \deg(v_i) & i = j \\ 0 & i \neq j \end{cases}$$

Laplacian matrix

$$\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}$$
$$\boldsymbol{L} = \operatorname{diag}(\boldsymbol{W} \cdot \boldsymbol{1}) - \boldsymbol{W}, \quad \boldsymbol{1} = [1, \dots, 1]^T$$

MATHEMATICAL REPRESENTATION OF GRAPHS (CONT.)

Incidence matrix

MATHEMATICAL REPRESENTATION OF GRAPHS (CONT.)

Undirected graphs

- **A**, **W**, **L** and **B**₁ all uniquely define the graph
- *A*, *W* and *L* are symmetric
- $L \in \mathbb{R}^{N \times N}$ is positively semi-definite i.e. its SVD:

 $\boldsymbol{L} = \boldsymbol{U} \boldsymbol{\Lambda} \boldsymbol{U}^T$

$$\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_N), \lambda_i \ge 0$$

$$\boldsymbol{U} \boldsymbol{U}^T = \boldsymbol{I}_N \text{ (identity matrix)}$$

- eigenvector \boldsymbol{u}_i corresponding to eigenvalue λ_i : $\boldsymbol{L}\boldsymbol{u}_i = \lambda_i \boldsymbol{u}_i$
- eigenvalues are obtained by solving the roots of $det(L \lambda I) = 0$
- if $\lambda_1 < \lambda_2 < \cdots < \lambda_N$ and the graph consists of *K* connected components, then $\lambda_1 = \lambda_2 = \cdots = \lambda_K = 0$, and always, $\lambda_1 = 0$ and $\boldsymbol{u}_1 = [1, \dots, 1]^T / \sqrt{N}$
- the number of walks of length K between nodes v_i and v_j is equal to $[\mathbf{A}^K]_{ij}$
- the number of walks of length at most *K* between v_i and v_j is $[\sum_{k=1}^{K} \mathbf{A}^k]_{ij}$ \rightarrow can be used to enumerate node neighbors at distance at most *K* steps

Part 2:

Key ideas in graph signal processing

CIRCULAR GRAPH

 $Au_k = \lambda_k u_k \iff [u_i]_{n-1} = \lambda_k [u_k]_n \text{ (delay operator!)}$

$$\Rightarrow \quad [\boldsymbol{u}_k]_n = \frac{1}{\sqrt{N}} e^{j2\pi nk/N}, \quad \lambda_k = e^{-j2\pi k/N} \text{ (this is DFT)}$$

Note that

$$\boldsymbol{A}^{n} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{-1}\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{-1}\cdots\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{-1} = \boldsymbol{U}\boldsymbol{\Lambda}^{n}\boldsymbol{U}^{-1}$$

so a (N-1)-degree polynomial $f(\mathbf{A}) = h_0 \mathbf{A}^0 + h_1 \mathbf{A} + \dots + h_{N-1} \mathbf{A}^{N-1} = \mathbf{U} f(\mathbf{A}) \mathbf{U}^{-1}$

EIGEN-DECOMPOSITION OF A GENERAL GRAPH

Example

 $\boldsymbol{A} = \boldsymbol{U} \boldsymbol{\Lambda} \boldsymbol{U}^T$

FILTERING OF GRAPH SIGNALS

Linear combining

$$\mathbf{x}(t) = \mathbf{A}\mathbf{x}(t-1), \quad vx(t-1) = [x_1(t-1), x_2(t-1), \dots, x_N(t-1)]^T$$

i.e., if N_i denotes direct neighbors of node i,

$$x_i(t) = \sum_{x_j \in \mathcal{N}_i} x_j(t-1)$$

Delay of graph signal

$$\mathbf{x}(t) = \mathbf{A}\mathbf{x}(t-1) = \mathbf{A}^{t}\mathbf{x}(0), \quad t = 1, 2, \dots$$

Classical linear-time invariant filter

$$y(t) = h(t) * x(t) = \sum_{k=0}^{M-1} h(k) x(t-k)$$

Linear-time invariant filter for graph signals

$$\mathbf{y}(t) = h(0)\mathbf{x}(t) + h(1)\mathbf{x}(t-1) + \dots + h(M-1)\mathbf{x}(t-M+1) = \underbrace{\sum_{k=0}^{M-1} h(k)\mathbf{A}^{t-k} \mathbf{x}(0)}_{H(\mathbf{A})} = H(\mathbf{A})\mathbf{x}(0)$$

- but, A may not be invertible, so 'shift' via A cannot be undone

FOURIER TRANSFORM OF GRAPH SIGNALS

Recall

$$\boldsymbol{A} = \boldsymbol{U} \boldsymbol{\Lambda} \boldsymbol{U}^{-1}, \quad \boldsymbol{x} = [x_1, x_2, \dots, x_N]^T$$

- eigenvector \boldsymbol{u}_k corresponds to eigenvalue λ_k - for circular graphs,

Define

$$\hat{\boldsymbol{x}} = \boldsymbol{U}^{-1}\boldsymbol{x}, \quad \text{i.e.}, \quad \boldsymbol{x} = \boldsymbol{U}\hat{\boldsymbol{x}}, \quad \hat{\boldsymbol{x}} = [\hat{x}_1, \hat{x}_2, \dots, \hat{x}_N]^T$$

 $\Rightarrow \quad \hat{x}_k = \sum_{n=1}^N x_n [\boldsymbol{u}_k]_n \quad \text{and} \quad x_n = \sum_{k=1}^N \hat{x}_k [\boldsymbol{u}_k]_n$

Polynomial filter

$$\mathbf{y} = H(\mathbf{A})\mathbf{x} = H(\mathbf{U}\mathbf{\Lambda}\mathbf{U}^{-1})\mathbf{x} = \mathbf{U}H(\mathbf{\Lambda})\mathbf{U}^{-1}\mathbf{x}$$
$$\Rightarrow \quad \underbrace{\mathbf{U}^{-1}\mathbf{y}}_{\hat{\mathbf{y}}} = H(\mathbf{\Lambda})\underbrace{\mathbf{U}^{-1}\mathbf{x}}_{\hat{\mathbf{x}}} \quad \Rightarrow \quad \hat{y}(k) = (h(0)\lambda_k^0 + \dots + h(M-1)\lambda_k^{M-1}) \ \hat{x}(k)$$

FILTERING OF GRAPH SIGNALS (CONT.)

Task

• design a graph filter $\mathbf{h} = [h_0, h_1, \dots, h_{M-1}]^T$ having the desired frequency response G_k at frequency λ_k , for $k = 0, 1, \dots, N-1$

Frequency domain filtering

- 1. obtain GDFT of graph signal: $\hat{x} = U^{-1}x$
- 2. frequency domain filtering: $\hat{y} = G(\Lambda)\hat{x}$
- 3. recover the output graph signal: $y = U\hat{y}$

Vertex domain filtering

1. if $N \ge M$, solve the linear set of equations for filter impulse response **h**

$$H(\lambda_k) = h_0 + h_1 \lambda_k^1 + \dots + h_{M-1} \lambda_k^{M-1} = G_k, \ \forall k = 0, 1, \dots, N-1$$

 $\Rightarrow V_{\lambda}h = G_{\lambda} \Rightarrow h = V_{\lambda}^{-1}G_{\lambda} \text{ (invert Vandermonde matrix)}$

LAPLACIAN SPECTRAL DOMAIN

Let

$$y = Lx \iff y(t) = 2x(t) - x(t-1) - x(t+1)$$

 $L = U\Lambda U^{-1}$ and $\hat{x} = U^{-1}x \iff x = U\hat{x}$

Graph signal smoothness

$$Lu = \lambda u \implies u^T Lu = \lambda u^T u = \lambda$$
$$u^T Lu = \frac{1}{2} \sum_{i,j=0}^{N} [A]_{i,j} (u(i) - u(j))^2 = \lambda$$

i.e. eigenvectors with small eigenvalues represent low-pass components of graph signal \boldsymbol{x}

Ideal low-pass filter

$$H(\lambda) = \begin{cases} 1 & \forall \lambda < \lambda_c \\ 0 & \text{otherwise} \end{cases}$$
 (frequency response

General filter

$$\mathbf{y} = \sum_{m=0}^{M-1} h_m \mathbf{L}^m \mathbf{x}$$

FOURIER ANALYSIS OF GRAPH SIGNALS

Attempted definitions of

- convolution in vertex and spectral domains
- Z-transform for graph signals
- Parseval's theorem for graph signals
- shift in frequency domain
- ... often not intuitive extensions, but more 'can be considered as'

Other weaknesses

- signal shift using A or L may not be invertible
- since graph signal is non-periodic (what would a periodic extension be?), Discrete-time Fourier Transform and not DFT should be assumed
- sampling should make the spectrum periodic
- ... indicative that other different approaches should be attempted

Smoothing Filters of Graph Signals

De-noising application

- input: noisy graph signal *x*
- output: smoothed output graph signal *y*
- solution:

$$\min J = \frac{1}{2} ||\mathbf{y} - \mathbf{x}||_2^2 + \alpha \mathbf{y}^T \mathbf{L} \mathbf{y}$$

$$\Rightarrow \quad \frac{\partial}{\partial \mathbf{y}^T} J \stackrel{!}{=} 0 \quad \Rightarrow \quad \mathbf{y} = (\mathbf{1} + 2\alpha \mathbf{L})^{-1} \mathbf{x} \quad \leftrightarrow \quad \hat{\mathbf{y}} = \underbrace{(\mathbf{1} + 2\alpha \mathbf{\Lambda})^{-1}}_{H(\mathbf{\Lambda})} \hat{\mathbf{x}}$$

• alternative formulation:

$$\min J = \frac{1}{2} ||\mathbf{y} - \mathbf{x}||_2^2 + \alpha \mathbf{y}^T \mathbf{L} \mathbf{y} + \beta \underbrace{\mathbf{y}^T \mathbf{L}^2 \mathbf{y}}_{||\mathbf{L}\mathbf{y}||_2^2}$$

$$\Rightarrow \quad \frac{\partial}{\partial \mathbf{y}^T} J \stackrel{!}{=} 0 \quad \Rightarrow \quad \mathbf{y} = \left(\mathbf{1} + 2\alpha \mathbf{L} + 2\beta \mathbf{L}^2\right)^{-1} \mathbf{x}$$

• it is also possible to consider sparsity instead of smoothness

GRAPH SAMPLING AND COMPRESSION

Task

• represent or reconstruct graph signal from small number of its samples

K-sparse graph signal

$$\hat{\boldsymbol{x}} = [\hat{x}(0), \hat{x}(1), \dots, \hat{x}(K-1), \underbrace{0, \dots, 0}_{N-K}]^T \implies x(n) = \sum_{k=0}^{K-1} \hat{x}(k) \boldsymbol{u}_k(n)$$

• assume *M* samples available at nodes $n_1, n_2, \ldots, n_{n_M}, K \le M < N$

$$\underbrace{ \begin{bmatrix} y(n_1) \\ y(n_2) \\ \vdots \\ y(n_M) \end{bmatrix} }_{y(n_M)} = \underbrace{ \begin{bmatrix} u_0(n_1) & u_1(n_1) & \cdots & u_{N-1}(n_1) \\ u_0(n_2) & u_1(n_2) & \cdots & u_{N-1}(n_2) \\ \vdots & & & \vdots \\ u_0(n_M) & u_1(n_M) & \cdots & u_{N-1}(n_M) \end{bmatrix} \underbrace{ \begin{bmatrix} \hat{x}(0) \\ \hat{x}(1) \\ \vdots \\ \hat{x}(K-1) \end{bmatrix} }_{\hat{x}(K)}$$
y (measurements)
$$M \text{ (measurement matrix)} \quad \hat{x}_{(K)}$$

• recovery from *M* > *K* samples

$$\hat{\boldsymbol{x}}_{(K)} = (\boldsymbol{M}^T \boldsymbol{M})^{-1} \boldsymbol{M}^T \boldsymbol{y} \implies \boldsymbol{x} = \boldsymbol{U} [\hat{\boldsymbol{x}}_{(K)}^T, 0 \cdots 0]^T$$

• this procedure can be used if positions of (N-K) zeros in \hat{x} are known

GRAPH SAMPLING AND COMPRESSION (CONT.)

K-sparse graph signal

- neither positions nor the number of zeros in \hat{x} is known
- reconstruction:

$$\min \|\boldsymbol{X}\|_0 \quad \boldsymbol{s}.\boldsymbol{t}. \quad \boldsymbol{y} = \boldsymbol{M}\boldsymbol{X}$$

- possible solution:
 - 1. estimate zeros and their positions as (N K) largest values in $X = M^T y$
 - 2. knowing K zero positions, use previous procedure to estimate x

Example

- signal measured at vertices 2,3,4,5,7: y = [0.224, 1.206, 1.067, 1.285, 1.116]
- it is estimated that there are (N K) = 2 non-zero components in $\mathbf{X} = \mathbf{M}^T \mathbf{y}$

GRAPH SAMPLING AND COMPRESSION (CONT.)

Aggregate sampling

• measurements using combining matrix C

$$y = Cx = \underbrace{CU}_{M} \hat{x}$$

- reconstruction: see previous methods
- can assume sampling at one vertex only

$$y_0 = [x]_n, \quad y_1 = [Ax]_n, \quad \dots, \quad y_{N-1} = [A^{N-1}x]_n$$

then

$$\boldsymbol{C} = \begin{bmatrix} 0 \cdots 1 \cdots 0 \\ \operatorname{row}_n(\boldsymbol{A}) \\ \vdots \\ \operatorname{row}_n(\boldsymbol{A}^{N-1}) \end{bmatrix}$$

GRAPH SAMPLING AND COMPRESSION (CONT.)

Task

- sub-sample large graph signals to reduce storage requirements and data processing complexity
- downsampled graph to retain certain similarity to original graph
 - \rightarrow connectivity distribution
 - \rightarrow spectral properties

Methods

- random selection of nodes (equi-probable, proportional to the node degrees or ranks) is satisfactory in very large graphs
- random selection of edges tends to produce disconnected graphs
- random walk and maximum spanning tree based
- many other methods in literature

FLOWS ON GRAPHS

Diffusion

$$\frac{\partial}{\partial t} \boldsymbol{x}(t) = -\boldsymbol{L}\boldsymbol{x}(t)$$

• backward difference approximation

$$\mathbf{x}(t+1) - \mathbf{x}(t) = -\alpha \mathbf{L}\mathbf{x}(t+1) \quad \Rightarrow \quad \mathbf{x}(t+1) = (\mathbf{I} + \alpha \mathbf{L})^{-1}\mathbf{x}(t), \ t = 0, 1, \dots$$

• forward difference approximation

$$\boldsymbol{x}(t+1) - \boldsymbol{x}(t) = -\alpha \boldsymbol{L} \boldsymbol{x}(t) \quad \Rightarrow \quad \boldsymbol{x}(t+1) = (\boldsymbol{I} - \alpha \boldsymbol{L}) \boldsymbol{x}(t), \ t = 0, 1, \dots$$

Alternative interpretation

• minimize the smoothness of *x*

$$\min_{\mathbf{x}} \mathbf{x}^T \mathbf{L} \mathbf{x} \implies \frac{\partial}{\partial \mathbf{x}} \mathbf{x}^T \mathbf{L} \mathbf{x} = 2\mathbf{x}^T \mathbf{L}$$

 $\Rightarrow \mathbf{x}(t+1) = \mathbf{x}(t) - \alpha \mathbf{L}\mathbf{x}(t) = (\mathbf{I} - \alpha \mathbf{L})\mathbf{x}(t) \quad \text{(steepest descent)} \\\Rightarrow \hat{\mathbf{x}}(t+1) = (\mathbf{I} - \alpha \mathbf{\Lambda})\hat{\mathbf{x}}(t) \quad \text{(spectral domain)}$

Building Graph Models

Graph model

- what do node represent?
- when to add edges between nodes?

Node modeling

- obvious cases: social networks, agent networks, routers
- less obvious cases: time series data, complex systems
- challenges:
 - \rightarrow merging/clustering nodes
 - \rightarrow time-varying scenarios

Edge modeling

- obvious cases: social networks, flow networks
- challenges in other cases:
 - \rightarrow weight metric selection
 - \rightarrow no edge can also mean missing data
 - \rightarrow time-varying scenarios

Building Graph Models (cont.)

Euclidean distance metric

- used when mutual geographical location r of nodes matters, i.e., the distance $r_{nm} = ||\mathbf{r}_n \mathbf{r}_m||_2$
- edge weight metrics with parameters $\alpha \ge 0$, $r_0 > 0$ and $c \ge 0$

$$W_{nm} = \begin{cases} e^{-\alpha (r_{nm})^{c}} & r_{nm} \le r_{0} \\ 0 & r_{nm} > r_{0} \end{cases}$$

• these weights can produce exponentially weighted moving average of neighboring nodes for linear graph model

$$\boldsymbol{x}(t+1) = (\boldsymbol{I} + \boldsymbol{W}) \, \boldsymbol{x}(t)$$

Other edge metrics

- any similarity metric can be considered
- total variance, correlation (Pearson, Spearman)
- binary (relationship exists/does not exist)
- learning/extrapolating relationships and weights

LEARNING GRAPHS FROM DATA

Task

- given measurements $x_1, x_2, ..., x_N$ at *N* nodes, construct a graph representing the data
- exploit smoothness of the graph signal rather than internal correlations

Define

- data matrix $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_N]$
- graph signal smoothness

$$\operatorname{tr}\left\{\boldsymbol{X}^{T}\boldsymbol{L}\boldsymbol{X}\right\} = \frac{1}{2}\sum_{n,m} [\boldsymbol{W}]_{n,m} \|\boldsymbol{x}_{n} - \boldsymbol{x}_{m}\|^{2} = \frac{1}{2}\operatorname{tr}\left\{\boldsymbol{W}\boldsymbol{\Delta}\boldsymbol{X}\right\}, \quad [\boldsymbol{\Delta}\boldsymbol{X}]_{n,m} = \|\boldsymbol{x}_{n} - \boldsymbol{x}_{m}\|^{2}$$

Optimization problem

$$\min_{\boldsymbol{W}} \operatorname{tr}\{\boldsymbol{W} \Delta \boldsymbol{X}\} + \alpha h(\boldsymbol{W}) \quad s.t. \quad \operatorname{diag}(\boldsymbol{W}) = \boldsymbol{0}, \ \boldsymbol{W} = \boldsymbol{W}^{T}$$

- regularization $h(\boldsymbol{W})$: $\|\boldsymbol{W}\|_1$, $\|\boldsymbol{W}\|_2^2$, $\log \det \boldsymbol{W}$
- additional constraint diag(W1) = I (weights normalization), or as objective

$$\min_{\boldsymbol{W}} \operatorname{tr}\{\boldsymbol{W} \Delta \boldsymbol{X}\} + \alpha h(\boldsymbol{W}) - \beta \mathbf{1}^T \log \boldsymbol{W} \mathbf{1} \quad \boldsymbol{s}.\boldsymbol{t}. \quad \operatorname{diag}(\boldsymbol{W}) = \mathbf{0}, \ \boldsymbol{W} = \boldsymbol{W}^T$$

Noisy data

- learn the graph and de-noise the measurements
- optimization problem

 $\min_{\boldsymbol{W},\boldsymbol{Y}} \|\boldsymbol{Y} - \boldsymbol{X}\|_{2}^{2} + \alpha \operatorname{tr} \{\boldsymbol{W} \Delta \boldsymbol{Y}\} + \beta h(\boldsymbol{W}) \quad s.t. \operatorname{diag}(\boldsymbol{W}) = \boldsymbol{0}, \ \boldsymbol{W} = \boldsymbol{W}^{T}, \ [\Delta \boldsymbol{Y}]_{n,m} = \|\boldsymbol{y}_{n} - \boldsymbol{y}_{m}\|^{2}$

Gaussian assumption

- graph signal as Gaussian-Markov random field
- \bullet optimization problem (gdet is generalized determinant) is defined as maximum-likelihood estimation of L

 $\max_{\boldsymbol{L}} \log \operatorname{gdet} \boldsymbol{L} - \operatorname{tr} \{ \boldsymbol{CL} \} - \alpha h(\boldsymbol{L}) \quad \text{s.t.} \quad \boldsymbol{L1} = \boldsymbol{0}, \ \boldsymbol{L} = \boldsymbol{L}^{T}$

where *C* is estimated covariance matrix $C = \frac{1}{n-1} \sum_{i=1}^{n} \operatorname{col}_{i} (X - \bar{X}) \operatorname{col}_{i} (X - \bar{X})^{T}$

$$\operatorname{tr}\{\boldsymbol{C}\boldsymbol{L}\} \propto \operatorname{tr}\{\boldsymbol{X}^T\boldsymbol{L}\boldsymbol{X}\}$$

Note

 these methods pose no constraints on graph structure except the number of graph nodes

Learning specific graph structure

- grid, tree, bipartite, multi-component, regular, densely connected etc.
- challenge is how to specify the graph structure as optimization constraint

Strategy

• spectral constraints on eigenvalues or eigenvectors of L or W

$$\Lambda(L) \in S_{\lambda}$$
 or $\Lambda(W) \in S_{\lambda}$

• algorithms for such non-convex problems were developed in literature

Example: *k*-component graph

• learn k clusters, i.e., $\lambda_1 = \cdots = \lambda_k = 0$, and additional constraint

$$\Rightarrow \operatorname{rank}(\boldsymbol{L}) = N - k \quad \Leftrightarrow \quad \sum_{i=1}^{k} \lambda_i(\boldsymbol{L}) = 0$$

• unforeseen problem: *k*-component graph learning generates isolated nodes \rightarrow add constraint(s) to avoid zero-degree nodes, e.g. diag(W1) = I

Learning correlation graph

• for n = 1, 2, ..., N, sequentially determine the edge weights W_{nm}

$$\min_{\operatorname{row}_{n}(\boldsymbol{W})} \left\| \boldsymbol{x}_{n} - \sum_{\substack{m=1\\m\neq n}}^{N} W_{nm} \boldsymbol{x}_{m} \right\|_{2}^{2} + \alpha \|\operatorname{row}_{n}(\boldsymbol{W})\|_{1} \quad \text{(LASSO minimization)}$$

LASSO minimization

$$\min_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \alpha \|\boldsymbol{x}\|_1$$

is solved iteratively as

$$\boldsymbol{x}_{k+1} = \operatorname{soft} \left(2\beta \boldsymbol{A}^T (\boldsymbol{y} - \boldsymbol{A} \boldsymbol{x}_k) + \boldsymbol{x}_k, \beta \alpha \right), \ k = 1, 2, \dots$$

where the learning step β

 $0 < \beta < 1/(2\lambda_{max}), \lambda_{max}$ is max eigenvalue of $\mathbf{A}^T \mathbf{A}$

and

$$\operatorname{soft}(y, u) = \begin{cases} y+u, & y < -u \\ 0, & |y| \le u \\ y-u, & y > u \end{cases}$$

Non-Gaussian data

• tweak tr{*CL*} term in objective function, e.g. majorization-minimization iterations [Palomar et al. 2017]

tr
$$\{\boldsymbol{C}^{k}\boldsymbol{L}\}$$
 where $\boldsymbol{C}^{k} = \frac{1}{n}\sum_{i=1}^{n}\frac{\operatorname{col}_{i}(\boldsymbol{X})\operatorname{col}_{i}(\boldsymbol{X})^{T}}{\operatorname{col}_{i}(\boldsymbol{X})^{T}\boldsymbol{L}^{k}\operatorname{col}_{i}(\boldsymbol{X})}, \ k = 1, 2, \dots$

Time-varying data

- divide data to T chunks, estimate graph L_t for every data chunk t = 1, 2, ..., T
- add regularization term $\sum_{t=2}^{T} d(L_t, L_{t-1})$ to objective function to maintain graph consistency in time, e.g.

$$d(L_t, L_{t-1}) = ||L_t - L_{t-1}||_2^2$$

• *L_t* can be estimated sequentially to reduce complexity

SPECTRAL CLUSTERING

Task

- partition the graph in *K* clusters of approximately equal size
- edges between clusters should have small weights

Solution

• any clustering algorithm e.g. K-means

Define graph cut

- graph weights W_{ij} between nodes $v_i \in V$ and $v_j \in V$

 $V_1 \cup V_1^c = V, V_1 \cap V_1^c = \emptyset \implies \operatorname{cut}(V_1, V_1^c) = \sum_{i \in V_1, j \in V_1^c} W_{ij}$ (other operators: avg, min)

- for K disjoint node subsets

$$\bigcup_{k=1}^{K} V_k = V, \ V_k \cap V_l = \emptyset \ \forall k, l \quad \Rightarrow \quad \operatorname{cut}(V_1, \dots, V_K) = \sum_{k=1}^{K} \operatorname{cut}(V_k, V_k^c)$$

- normalization to encourage equal-size clusters

$$\Rightarrow \operatorname{cut}(V_1, \dots, V_K) = \sum_{k=1}^K \frac{\operatorname{cut}(V_k, V_k^c)}{|V_k|}$$

SPECTRAL CLUSTERING (CONT.)

Optimum clustering via graph signal (case K = 2)

given V_1 , let $a = \sqrt{\frac{|V_1^c|}{|V_1|}}$, and define graph signal $f(v_i) = f_i = \begin{cases} a & v_i \in V_1 \\ 1/a & v_i \in V_1^c \end{cases}$

the smoothness, $\boldsymbol{f}^T \boldsymbol{L} \boldsymbol{f} = \frac{1}{2} \sum_{i,j=1}^N W_{ij} (f_i - f_j)^2 = \frac{N}{2} \sum_{k=1}^2 \frac{\operatorname{cut}(V_k, V_k^c)}{|V_k|}$

$$\Rightarrow \min_{V_1 \subset V} \boldsymbol{f}^T \boldsymbol{L} \boldsymbol{f} \quad \boldsymbol{s}. \boldsymbol{t}. \quad \boldsymbol{f}^T \boldsymbol{1} = 0, \ \|\boldsymbol{f}\|_2^2 = N \quad (\mathsf{NP-hard problem})$$

Applications of graph clustering

- graph partitioning
 - \rightarrow graph signals often smooth within clusters
 - \rightarrow nodal domain of f is a sub-graph where f in all nodes have the same sign
- hierarchical graph representation
- graph reduction
 - \rightarrow replace small clusters with single node
- graph visualization
- graph coloring

GRAPH DOWNSAMPLING

Task

- remove less important nodes and/or edges
 - \rightarrow how to select those?
 - \rightarrow e.g. account for edge weights and node degrees
- aim: reduce the graph size by half

Solution using the largest eigenvector

• let u_{max} be the eigenvector of L corresponding to eigenvalue λ_{max}

$$V_{\text{keep}} = \begin{cases} v_i \in V : \\ v_i \in V : \\ N/2\text{-largest } |\boldsymbol{u}_{\max}(i)| \end{cases}$$

• note that V_{keep} and V_{keep}^c represents the graph cut

RANDOM SIGNALS ON GRAPHS

Stationary graph signal

• can be expressed as

$$\boldsymbol{x} = \underbrace{\boldsymbol{x}_0}_{\text{deterministic}} + \underbrace{\sum_{l=0}^{N-1} h_l \boldsymbol{L}^l}_{\boldsymbol{H}} \boldsymbol{w}$$

where w are zero-mean uncorrelated random samples

- corresponding covariance $C_x = E[xx^T] = HH^T$
- if $L = U\Lambda L^T$, then $s = \text{diag}(U^T C_x U)$ is power spectral density vector

Non-stationary graph signal

- w does not need to be white i.e. $C_w = E[ww^T] \neq I$
- consequently, $\boldsymbol{U}^T \boldsymbol{C}_x \boldsymbol{U}$ is not a diagonal matrix
- one strategy: if C_w is known, find H, so that $\hat{C}_x = \frac{1}{M} \sum_{m=1}^M x_m x_m^T$ and $C_x = E[xx^T] = HC_w H^T$ are close in some sense assuming symmetry $H = H^T$

GRAPH OBSERVABILITY

Linear graph

- $a_{ij}(t)$ is edge weight between nodes *i* and *j*
- x_i is weight of node *i*, so the internal state $\mathbf{x}(t) = [x_1(t), \dots, x_n(t)]^T$
- if time-invariant, state-space description

$$\dot{\mathbf{x}}(t) = \mathbf{A} \mathbf{x}(t) + \mathbf{B} \mathbf{u}(t)$$
$$\mathbf{y}(t) = \mathbf{C} \mathbf{x}(t)$$

... this is a linear MIMO system!

Observability

- a state (values of internal variables) can be obtained from finite observations \rightarrow find state trajectory $\mathbf{x}(t)$ from any initial state $\mathbf{x}(0)$ to the current state
- strongly connected components (there is path between any two nodes)
 → have to observe at least one node from each SCC
- for linear model, can use maximum matching to find minimum # sensors
 → often much larger than # SCC (due to model symmetries)
- surprisingly, for <u>non-linear model</u>, sensors predicted by SCCs are necessary as well as sufficient (since model symmetries are rare)

GRAPH CONTROLABILITY

Kalman's condition

• controlability matrix D must be full rank i.e. rankD = n where n = |x|

$$\boldsymbol{D} = [\boldsymbol{B}, \boldsymbol{A}\boldsymbol{B}, \boldsymbol{A}^2\boldsymbol{B}, \dots, \boldsymbol{A}^{n-1}\boldsymbol{B}]$$

Driver nodes

• for linear networks, maximum matching again helps

- surprisingly, driver nodes tend to avoid hubs
 - \rightarrow average degree of driver nodes is smaller than average degree of graph
 - \rightarrow # driver nodes mainly determined by degree distribution

"Sparse and heterogeneous networks are harder to control than dense and homogeneous networks."

GRAPH TOMOGRAPHY

Linear graph

- assume a graph with the weight $a_{ij} \in \mathcal{R}$ between nodes *i* and *j*
- let path $\mathcal{P} = \{(i_1i_2), (i_2i_3), \dots, (i_{d-1}i_d)\}$ from node i_1 to node i_d
 - \rightarrow accumulated weight along the path \mathcal{P} is $a_{i_1i_d} = \sum_{(ij)\in\mathcal{P}} a_{ij}$
 - \rightarrow in matrix notation

$$y = \boldsymbol{P} \cdot \boldsymbol{a} + \boldsymbol{w}$$

where rows of binary matrix $P \in \{0, 1\}^{m \times n}$ correspond to each measured path, a is graph adjacency vector, and m denotes the number of probes

Multiplicative weights

• if $a_{i_1i_d} = \prod_{(ij) \in \mathcal{P}} a_{ij}$, we can assume model

$$\log y_{i_1 i_d} = \sum_{(ij) \in \mathcal{P}} \log a_{ij} + w$$

GRAPH TOMOGRAPHY (CONT.)

Task 1

- given measurements y, find minimal path matrix P to recover k weights in $a \rightarrow$ this assumes that the graph structure is known
- the path matrix *P* is minimal if either the sum-length of all probing paths considered is minimum, or if the longest among these paths is minimized
- trivial (non-minimal) design: measure all weights a_{ij} one by one
 - \rightarrow determining all paths of a general graph is NP-hard
 - \rightarrow in practice, only a subset of nodes may be used for I/O or as gateways

Task 2

- knowing the graph structure, and given subset of paths P and the corresponding measurements y, recover k graph weights in a
 - $\rightarrow P$ may not be large enough to enable compressive sensing of a
 - \rightarrow some weights in *a* may be calculated if overdetermined system ($m \ge k$)
- different strategy:
 - \rightarrow determine weights change Δa if nominal weights a are known

$$y = P \cdot (a + \Delta a) + w = \underbrace{P \cdot a}_{\text{known}} + P \cdot \Delta a + w$$

assuming k'-sparse change vector Δa with $k' \ll k$

INFORMATION DIFFUSION¹

Diffusion in networks

- information, ideas, behaviors, epidemics, computer viruses, economic transactions, petitions etc.
- transfer occurs in a particular random time

Mathematical model

• from node *j* to *i* with a rate parameter $\alpha_{ji} \sim 1/\langle \Delta_{ij} \rangle$

 $f(\Delta_{ij}, \alpha_{ij})$ usually exponential or power-law but also multi-modal distributions

• information cascade: a set $\{t_i\}_i$ where

$$\Delta_{ji} = (t_j - t_i) \sim f(\Delta_{ij}, \alpha_{ij})$$

• node is activated once by the first parent

¹MG Rodriguez, Le Song, KDD tutorial, 2015.

Network inference problem

• given information cascades $\{t_i^1\}_i, \{t_i^2\}_i, \dots$ infer connection (adjacency) matrix and rate constants $\alpha_{ij} > 0$

 $\rightarrow \alpha_{ij} = 0$ if no connection from *i* to *j*

• assuming independence of individual diffusion events, we can obtain and then maximize the likelihood of observed cascades $c \in C$

$$\hat{\boldsymbol{A}} = \operatorname{argmax}_{\boldsymbol{A}} \sum_{c \in C} \log f(\{t_i^c\}_i, \boldsymbol{A})$$

can we recover completely recover A?
 → number of cascades |C| needs to be large (cf. compressive sensing)

$$|C| \approx O(d_{\max}^3 \log N)$$

where d_{max} is the maximum in-degree in a network of *N* nodes

Other inferences

• find probability $Pr(t_n \leq T | \mathbf{A})$

Content-sensitive diffusion

 $\Delta_{ji} \sim f(\Delta_{ji}, \alpha_{jim}), \quad m = 1, 2, 3, \dots$

• average rate
$$\bar{\alpha}_{ji} = \sum_{m} \alpha_{jim} \Pr(m)$$

Time-varying rates and connections

$$\Delta_{ji}(\tau) \sim f(\Delta_{ji}, \alpha_{ji}(\tau)), \, \alpha_{ji}(\tau) \geq 0$$

• tracking variations

$$\hat{A}(\tau) = \operatorname{argmax}_{A(\tau)} \sum_{c \in C} w_c(\tau) \log f(\{t_i^c\}_i, A(\tau))$$

where weights $w_c(\tau) = \exp(\tau/T)$

(i.e. forget history more in the past)

Practical challenges

- difficult to identify individual cascades in observed data
 → most cascades are single values (never propagates)
- solution: assume counting process (of a specific activity or event)

$$N_i(t) = \sum_{-\infty}^t \underline{\lambda_i(\tau)} \, \mathrm{d}\tau$$

where the process intensity $\lambda_i \propto \alpha_i$ has the unit [events/hour]

Exogenous activity

• drivers external to the network

Endogenous activity

• response to activities of other users within the network

Total intensity

$$\lambda(t) = \underbrace{\lambda^{0}(t)}_{\text{exogenous}} + \underbrace{\lambda^{*}(t)}_{\text{endogenous}}$$

Hawkes process:

$$\lambda^*(t) = \sum_{i:t_i < t} a_{uu_i} g(t - t_i)$$

- a_{uu_i} : influence of neighbor u_i on user u
- g(t): memory (lifetime of the event or activity)

Correlated events/activities

Maximizing the influence

• influence up to time *T* from nodes in *A*

$$\sigma(A;T) = \mathbb{E}[N(A;T)] = \sum_{n=1}^{N} \Pr(t_n \le T|A)$$

• maximization (NP-hard problem)

$$A^* = \operatorname{argmax}_{|A| \le K} \sigma(A; T)$$

 \rightarrow we can use a greedy algorithm to solve it or other approximate solutions, e.g.

Source localization

- observations often incomplete, can we still find the first node in a cascade?
- need to eliminate all unobserved events from the likelihood before maximizing it → computationally very difficult problem
 - \rightarrow importance sampling-like strategies seem to work well

Activity shaping

Influence	Fixed	One	time	same	Aim	is	maximum
maximization	incentive	information			adoption		
Activity	Variable	Multiple time, multiple			Many	/	different
shaping	incentive	info, recurrent		shaping tasks			

- incentivize few users to produce a given level of overall activity
 → exogenous activity → endogenous activity
- example objectives:
 - reach average activity $E[\lambda(t)]$ at time *t* where $\lambda(t) = \lambda^*(t) + \lambda^0(t)$
 - maximize utility $U(E[\lambda(t)])$
 - maximize activity of the least active user
 - maximize the total number of events in the network

Part 3: Survey of problems and tasks in graph signal processing

PROBLEMS IN NETWORK SCIENCE

Subgraphs

- community detection
- maximum clique identification
- strongly connected components

Path

- minimum spanning tree
- maximum/perfect matching
- shortest path between a pair of nodes
- longest path in a graph (diameter)
- traveling salesman problem

Many other

- drawing and visualization
 → graph coloring
- graph search
- linear ordering of nodes (for acyclic graphs)
 → more generally, ranking and sorting of nodes
- maximum flow/minimum cut

(mostly static graphs)

PROBLEMS IN GRAPH SIGNAL PROCESSING

Main concerns

- graph model given *a priory* or learned from data
- working with directed graphs
- how to assign variables to graph
- time-varying and/or random graph structure and variables
- linear and non-linear processing of graph signals
- scalability to very large graph signals

Graph processing tasks (graph structure only)

- graph algebra
 - \rightarrow graph fields
 - \rightarrow addition, multiplication, modulo operations
- graph conversions and transforms
 - \rightarrow different types of graph
 - \rightarrow to/from other discrete structures
- graph structure measures and description
 → well defined by Network Science (motifs, clustering etc.)
- graph embedding in high-dimensional vector space
 → often used in machine learning tasks

PROBLEMS IN GRAPH SIGNAL PROCESSING (CONT.)

Graph signal processing tasks (graph structure and variables)

- defining graph signals
 - \rightarrow mapping variables to graph (nodes, edges, simplexes)
 - \rightarrow learning graph from data (given type or general graph)
 - \rightarrow graph signal \longleftrightarrow manifold conversions
- orthonormal decomposition \rightarrow *a priory* basis selection or basis learned from data
- spectral (frequency domain) representation
 → Fourier analysis (discrete Fourier transforms)
 - \rightarrow wavelet transforms
- filtering
 - \rightarrow in time and in frequency domains
 - \rightarrow convolution, de-convolution, windowing
 - \rightarrow designing low-pass and high-pass filters
 - \rightarrow observability and controlability
- sampling, downsampling and compression
 - \rightarrow lossless or near-perfect reconstruction
 - \rightarrow graph signal tomography

PROBLEMS IN GRAPH SIGNAL PROCESSING (CONT.)

Graph signal processing tasks (cont.)

- invariant measures
 - \rightarrow graph signal shift in time and frequency domains
- clustering, classification, sorting, ranking samples of graph signals
- time-varying signals
 - \rightarrow flows and diffusion on graphs (information)
 - \rightarrow events spreading over graphs (epidemics)
 - \rightarrow event detection from small number of observations
- random graph signals
 - \rightarrow stationarity and ergodicity
 - \rightarrow de-noising and smoothing
 - \rightarrow parameter estimation from small number of observations

Driving applications of graph signal processing

- image processing
- brain functional networks
- semi-supervised and deep graph learning

... bottom line: taking into account signal (graph) structure can greatly improve the performance

Part 4:

Recommended readings on graph signal processing

Books

Petar M. Djurić and Cédric Richard Cooperative and Graph Signal Processing

Principles and Applications

SYNTHESIS LECTURES ON ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

Ronald J. Brachman, Francesca Rossi, and Peter Stone, Series Editors

BOOK CHAPTERS

Introduction to Graph Signal Processing

Ljubiša Stanković, Miloš Daković and Ervin Sejdić

Abstract Graph signal processing deals with signals whose domain, defined by a graph, is irregular. An overview of basic graph forms and definitions is presented first. Spectral analysis of graphs is discussed next. Some simple forms of processing signal on graphs, like filtering in the vertex and spectral domain, subsampling and interpolation, are given. Graph topologies are reviewed and analyzed as well. Theory is illustrated through examples, including few applications at the end of the chapter.

1 Introduction

Graph signal processing is an active research area in recent years resulting in many advanced solutions in various applications. In numerous practical cases the signal domain is not a set of equidistant instants in time or a set of points in space on a regular grid. The data sensing domain could be irregular and, in some cases, not related to the time or space. The data sensing domain is then related to other properties of the considered system/network. For example, in many social or web related networks, the sensing points and their connectivity are related to specific objects and their links. In some physical processes other properties than the space or time coordinates define the relation between points where the signal is sensed. Even for the data sensed in the well defined time and space domain, the introduction of new relations between the sensing points may produce new insights in the analysis and result in more advanced data processing techniques.

L. Stanković (E2) · M. Daković University of Montenegro, Podgorica, Montenegro e-mail: ljubisa@ac.me M. Daković e-mail: miles@ac.me E. Sejdić University of Pittsburg, Pittsburg, PA, USA e-mail: esejdic@iece.org

© Springer Nature Switzerland AG 2019 L. Stanković and E. Sejdić (eds.), Vertex-Frequency Analysis of Graph Signals, Signals and Communication Technology, https://doi.org/10.1007/978-3-030-03574-7_1

GRAPH SIGNAL PROCESSING

CHAPTER

José M.F. Moura® artment of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, United States'

8.1 INTRODUCTION

It is by now a cliché that data is everywhere, and we are and will be inundated by data. According to a 2014 study by IDC [1], in 2020 alone, all digital data created, replicated, and consumed will amount to 44 zetabytes. A zetabyte is 10²¹ bytes, so we will produce yearly data quantified by the number 44 followed by 21 zeros. To have a more physical feeling for this staggering amount of data, we compare it to a coarse estimate of the entire collection of books, reports, written texts, maps, images, audio recordings, and videos in the US Library of Congress (LOC) (charged with maintaining a copy of every book ever printed in the United States). This estimate may vary, but we consider it to be three petabytes [2]. Then, the 44 zetabytes of data produced worldwide in the year 2020 will be roughly equivalent to 15 million LOCs. IDC has recently updated this estimate to 163 zetabytes by 2025 [3,4]. But these data will be very different from the data we are traditionally concerned with in discipline such as statistics, signal or image processing, computer vision, or machine learning. Beyond traditional time series, speech, audio, radio, radio, time series, speech, audio, radio, ra the 11 billion internet-of-things (IOT) devices in 2016 that are estimated to triplicate to 30 billion in 2020, from the activity of billions of cell phone users of the many service providers; from public and private urban transportation systems; in health care from the digital records of patients, providers, visits, exams, tests, results, costs, insurance, hospital procedures; from interactions among social network agents, corporate financial data, hyperlinked blogs, or tweeters, metabolic networks, protein interaction networks, just to mention some examples. These Big Data are often characterized by three, five, or seven V: Variety, Volume, Velocity, Veracity, Variability, Value, and Visualization. In many contexts, these data are produced by scattered sources such as the thousands of webcams monitoring traffic in urban centers, i.e., the data are distributed. Besides numerical, the data can be Boolean, ordinal, or categorical. Finally, the data is unlikely to fit neatly in a table; in other words, the data are unstructured. The goal of this chapter is to introduce data analytic tools to process these data that are much like the cal tools used with time series, images, or videos, but now applied to the variety of avstructured Big Data of today.

"This work is partially supported by NSF grant CCF # 1513936. Corporative and Cough Signal Processing. https://doi.org/10.1016/0978-0-12-813677-5.00008-0 Copyright O 2018 Elsevier Rev. All rights reserved.

239

JOURNAL PAPERS

Topological Signal Processing over Simplicial Complexes

Sergio Barbarossa, Fellow, IEEE, Stefania Sardellitti, Member, IEEE

Mar

arXiv:1907.11577v2

Xiaowen Dong, Dorina Thanou, Laura Toni, Michael Bronstein, and Pascal Frossard

<text><text><section-header><text><text><text><text>

GRAPH SIGNAL PROCESSIN

A review and new perspectives

The effective preventions, proceeding, analysis, and visual, historical flags water intermediating everythyses and end-tis complex domains, such as surveixs and graphs, as more the bay quarkinos in modern machine learning. Orpoth signal processing (SDP), a shared branch of signal processing models and algorithms that are in building data surgeoid on proble-diates and the second states and the signal processing models data water and the signal processing states and the signal states and the signal processing states and the signal states and the signal processing states and the signal states and the signal states and the signal states and the states and the signal states and the signal states and the states and the signal states and the signal states and the states and the signal states and the signal states and the states and the signal states and the signal states and the states and the signal states and the signal states and the states and the signal states and the signal states and the states and the signal states and the signal states and the states and the signal states and the signal states are state the states and the signal states and the signal states and the states and the signal states and the signal states are states and the signal states and the signal states are states and the signal states and the signal states are states and the signal states and the signal states are states and the signal states and the signal states are states and the signal states and the signal states are states and the signal states and the signal states are states and the signal states are states and the signal states are states and the signal states and the signal states are states and the signal states are states and the signal states and the signal states are states and the signal states are states and the signal states and the signal states are states and

Introduction We live in a concerted society. Data collected from large-scale intenctive systems, such as biological, accial, and financial networks, become largely volitable. In particle, the part few decades have seen a significant anneatt of interest in the ma-chicola learning convensity for entworks (data processing and analysis, Networks have an intrinsis structure that conveys vary specific properties to data, e.g., intradependencies be-tween data entities in the form of pairwise traditionality. These encourses are motionicable control one without pairs of the pairs o

properties are traditionally captured by mathematical repre

oping fast. Let us consider, for example, a network of protein protein interactions and the expression level of individual genes at every point in time. Some typical tasks in network biology related to this type of data are 1) discovery of key genes (vin

protein grouping) affected by the infection and 2) prediction of how the host centerion reacts (in terms of some expression

117

ntations such as graphs. In this context, new trends and challenges have been devel-

Graph Signal Processing for Machine Learning

Paolo Di Lorenzo, Member, IEEE, Paolo Banelli, Member, IEEE, Elvin Isufi, Student Member, IEEE, Sergio Barbarossa, Fellow, IEEE, and Geert Leus, Fellow, IEEE

<text><text><section-header><text><text><text><text><text>

Wavelets on graphs via spectral graph theory

David K. Hammond^{a,*,1}, Pierre Vandergheynst^{b,2}, Rémi Gribonval^c ⁴ Neuralnformatics Contex, University of Diegon, Eugens, USA ^b Ecole Pulynchmigue Hiddrale de Lausanne, Lausanne, Switzerland ⁴ JMRA, Rennes, France

ABSTRACT

ARTICLE INFO

Article history: Received 14 Navember 2009 Revised 21 April 2010 Accepted 25 April 2010 Available online 28 April 2010 Communicated by Charles K. Chu Rywords: Graph theory Wavelets Yeavelets Spectral graph theory Overcomplete wavelet frames

© 2010 Elsevier Inc. All rights reserved.

1. Introduction

- unregeting network
 - unregeting network
 - Iman advances. International homogenetics (C. K. Hannenský, pieros svadeního poshlyki, d. (Y. Vadengheynet), renis galovaválhostal (G. Criborevá),
 - This work was performed wide (BDF was at CPR).
 - This work was performed wide (BDF was at CPR).
 - This work was performed wide (BDF was at CPR).
 - This work was performed wide (F. Criborevá), FPI Open propert (PF): SCT 225(91)-SIM4L; Sparse Models, Algorithms and Learning for Largo Star Mato.

1063-5203/5 - see from matter © 2010 Ebevier Inc. All rights reserved

A signal representation perspective

potential advantage of the latter in a number of theoretical and practical scenarios. We conclude with several open issues and challenges that are keys to the design of future signal pro-cessing and machine-learning algorithms for learning graphs from data.

Ireduction down data analysis and processing tasks typically involve gr sets of structured data, where the structure carries crit-information about the nature of the data. Does can find me-rrous examples of such data sets in a wide diversity of ap-cariton decomine, including transportation networks, social works, computer networks, and brain networks, Typically. Modern data ana large sets of struc cal information al

Introduction

The construction of a meaningful graph topology physa, manying, and visualizations of mercurrent data. Wens a may start, is in the effective representing, manying, and visualizations of mercurrent data. Wens a may start, is in start decimalities to infer or learning approximation of a mercurrent data. In this array, we survey subtimes to the problem graph tearning, including classical viseopoints from statistics and physics, and mercurrent approximations to the problem graph tearning, including classical viseopoints from statistics and physics, and mercurrent approximations and differences there encomparison infinitions and differences thereen changes protectial advantage of the latters in a number of theorem to prostrial advantage of the latters in a num

lying relationship between these estitutes. Consider an example in busin signal analysis: suppose we are given blood-oxygen-bevel-dependent (BOLD) signals, i.e., times series extracted from finewriticnal magnetic resonance imaging data that reflect the activities of different regions of the brain. An area of significant interest in neuroscience is the inference of functional concentivity, i.e., to capare the relationship between brain regions that correlate or syncheotice given a contain condition of a patient, which may help beyoud materpic-ning of some nearedogenerative diseases (see Figure 1). This leads to the problem of infering a graph structure, given the milbrariate BOLD time strice data. Formally, the problem of graph learning is the fellowing given M observations on N variables or data entities rep-resented in a data matrix X ge 20⁻²⁰, and given some prefor knowledge é.g., distribution, data model, and so on) about

REE SERVE PROCESSIG MADAZINE | May 2019 |

We reveal topology informers in significant problem in not-sure science, and so graph signal processing (USS) effects and the analyse how the graph signal processing (USS) effects and the analyse how the graph signals and spectral dramateristics impact the properties of the graph signals of interest, Sudan analyse probability of productions of these sensitives are appresented by the sensitive and the schemes are arguing informat, distinctly locing an element of schemes are largely informat, distinctly locing an element of schemes are largely informat, distinctly locing an element of schemes are largely informat, distinctly locing and schemes of graph locing schemes and the scheme scheme scheme and the scheme schemes and the scheme scheme scheme applit topology. Thy names catatrical schemes have also scheme schemes and schemes and schemes and schemes schemes and schemes result of a network process defined in such a graph. A number of arguably more nascent topics are also briefly outlined, luding inference of dynamic networks and nonlinear models

cle introduces readers to challenges and opportunities for SP research in emerging topic areas at the crossrouds of modeling, prediction, and control of complex behavior arising in net-

Introduction

Intraduction Copies with the challenges found at the intersection of ner-work science and big data measurants frandamental brank throughs in modeling instritutions, on accordinality of data throughs of the science of the science of the science of the through at the science of the science of the science of the data of the science of the science of the science of the data of the science of the science of the science of the brance of the science of the science of the science of the brance of the science of the science of the science of the brance industrial science, information or as science of the brance industrial science, information or a science of the sci type in the concepts and concept index cancel and plandar in the concepts of t a network process defined in such a graph. A mather between our scientific understanding of signals defined over between our scientific understanding of signals defined over graphiered of gramme netwices and noticeare models wise interaction, as well as extensions to directed (6) or purrows interaction, in were as determines to interact (or graphs and their relation to causal inference. All in all, this art supports and their relation to causal inference. All in all, this art based thus insure stirmmetry that prevalence of answer/certaid SP problems and the access to quality network data are necess events. Making sense of languescale data sets from a network-certaid

K3-4888.7000 2020ETT

REE SIGNAL PROCESSING MAGAZINE | November 2020 |

VERSITY. Downloaded on May 14,2021 at 09:51:30 UTC from IEEE Xplore. Restrictions apply.

We proper a new of method for constructing wavefst transforms of functions defined as the vertices of an altituty faits weight graph. Journal properties along the properties about the termination of the function density of the properties density in the properties density of the properties o

JOURNAL PAPERS (CONT.)

INVITED

Graph Signal Processing: Overview, Challenges, and Applications

This article presents methods to process data associated to graphs (graph signals) extending techniques (transforms, sampling, and others) that are used for conventional signals.

BY ANTONIO ORTEGA[®], Fellow IEEE, PASCAL FROSSARD, Fellow IEEE, JELENA KOVAČEVIĆ, Fellow IEEE, JOSE M. F. MOURA^O, Fellow IEEE, AND PIERRE VANDERGHEYNST

ABSTRUCT I Research in graph signal processing (GSP) aims to develop tools for processing data defined on impgate graph domains. In this paper, we first provide an overview of one ideas in GSP and their connection to conventional eligital signal processing. GGP and their connections to connectional digital signal processing, along with a lark filterical perspective to highlight how connects recently developed in GGP build on top of prior research in other mass. We thin samalite reconst alwances in developing basic GGP taols, landaring methods for standing. Ellering, or graph learning, text, we review process is several actionation areas using GGP individing processing and analysis of sensor retenois data, biological individing processing and analysis of sensor retenois data, biological landarian and applications to langer processing and machine learning.

KEYWORDS | Graph signal processing (GSP); network science and graphs; sampling: signal processing

I. INTRODUCTION AND MOTIVATION I. INTRODUCTION AND MOTIVATION Data is all around an admeniation of 1. Anonet every aper of human life is now being necessful at all the from the mating and necessful approaches the second present all data through heidh motiving devices and apps, minutia and human procession and approaches and approximation of the second procession and approx-ment that the data now reside on integralar and complex moments that due data one second second works.

I November 25, 2013; revised Wards 10, 2018; accepted Wards evention April 24, 2018; Contemponding author Antonio D to Entimentary of Southern California, Los Angeles, CA 100 with the towards of Statement, and the second and J. M. F. Manura and with Connects Mellow University Philadersh.

interceion among them, for campite, sourt on Texter coils models in noise with their territory constructions in models and the territory of the territory construction of the model and the territory of the territory of the territory of the start of the territory is a staff energy of a total start of the territory is a staff energy of the territory territory of the territory territory of the territory territory of the territory territory of the territory of the territory of the territory of the territory territory of the territo

Graphs offer the ability to model such data and complex

interactions among them. For example, users on Twitter can be modeled as nodes while their friend connections can be modeled

OC 2018 2820136 0018-0219 © 2018 EEE. Personal sur is premitted, but republication/sediat/bation requires EEE premission See http://www.seee.org/publications_standarb/publicationsit/glouindex.html for more information.

808 PROCEEDINGS OF THE IEEE [154, 306, No. 5, May 2018 Activitied lonned use immed to: SWINDER UNIVERSITY, Downloaded on May 14, 2021 at 09:44:28 UTC from IEEE Xplore. Restrictions apply

Graph topology inference based on sparsifying transform learning

Stefania Sardellitti, Member IEEE, Sergio Barbarossa, Fellow, IEEE, and Paolo Di Lonnoro, Member, IEEE

<text><text><section-header><text><text><text>

Accepted Manuscript

Random sampling of bandlimited signals on graphs Gilles Puy, Nicolas Tremblay, Rémi Gribonval, Pierre Vandergheynst

> \$1063-5203(16)30021-5 65 acba 2016 05 005 YACHA 1142

Applied and Computational Harmonic Analysis To appear in:

Received date: 23 November 2015 Revised date: 20 May 2016 Accepted date: 28 May 2016

FLOW SMOOTHING AND DENOISING: GRAPH SIGNAL PROCESSING IN THE EDGE-SPACE

Michael T. Schaub^{6,b} and Santiago Segarra⁶

Institute for Data, Systems, and Society, Massachusetts Institute of Technology, Cambridge, MA, USA ^bDepartment of Engineering Science, University of Oxford, Oxford, UK

<text><text><text><text>

— The Probability of the Pro

ata [14], or information flows in brain networks or other biological susse [15]-[10]. While spatially embedded networks provider many if the most innative examples, edge-flows are also the natural the London street network (Section V), hefore bield discussion on reveaus for future research bield discussion on reveaus for future research in the street street

978-1-7281-1295-4/18/\$31.00 @2018 IEEE

bird discussion on areanes for future meansch. Ind insoration in 1055 mersten discussion in 2003 neurosch and insoration in 1055 MT and anges. The future has no socia in design of this study, the marks provided levels and the nather views. This is made, the marks provided levels and the nather views. This is made, the marks provided levels and the nather views. This is made, the marks provided levels and the nather views. This is done, the nather views and the nather views

715

use limited to: SWANSEA UNIVERSITY. Downloaded on May 14,0021 at 09.45/59 UTC from REEE Xplore. Restrictions appl

Please cite this article in press as: G. Pay et al., Random sampling of bandlimited signals on graphs, Appl. Comput. Harmon. Anal. (2016), http://dx.doi.org/10.1016/j.acha.2016.05.005

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we an providing this early version of the manuscript, The manuscript will undergo copyediting, typeietting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain

On the Graph Fourier Transform for Directed Graphs

Stefania Sardellitti, Member, IEEE, Sergio Barbarossa, Fellow, IEEE, and Paolo Di Lorenzo, Member, IEEE

Advance—The analysis of signals defined over a graph in the section is any application, such as sold and errors in the section is any application, such as sold and errors in the section is any application, such as sold and errors in the section is any application, such as sold and errors in the section is any application, such as sold and errors in the section is any application, such as sold and errors in the section is any application, such as sold and errors in the section is any application, such as sold and the section is address of general cost of section is any application in the section is address of general cost of section is any application in the section is address of general cost of section is any application in the section is address of general cost of section is any application in the section is address of general cost of section is any application in the section is address of general cost of section is any application in the section is address of general cost of section is any application in the section is address of general cost of section is any application in the section is address of general cost of section is any application in the section is address of the section is any application is another provide the section is address of the section is any application is another provide the section is address of the section is any application in the section is address of method is section in the section of the section is address of method is provide in the restrict in the section is address of method is provide in the restrict in the section is address of method is provide in the restrict in the section is address of method is provide in the restrict in the section is address of method is provide in the restrict in the section is address of method is provide in the restrict in the section is address of method is provide in the restrict in the section is address of method is provide in the restrict in the section is address of method is address of the section is address of method is address of the sect

bids on the finding accession of the signed statistic st

Vertex-Frequency Analysis on Graphs

David I Shuman, Benjamin Ricaud, Pierre Vandergheynst^{1,2} ery (LTS2), Ecole Polytechnique Fédérale de Lausanne (EPFL), Lause

Abstract One of the hyst challenges in the zeros of signal processing on graphs is to design dictionaries and transform methods to identify and exploit structure in signals on weighter graphs. To do so, we need to account for the intrinsic generative structure of the underlying graph disk absolution. In this paper, we generative cost of the next inspirator signal processing tools, "wholewed Parier analysis is the graph setting. Our approach explore structure of the underlying graph disk and an individual graph hereality in the explore structure operative such as the challenge of the cost of the structure of the structure of the explore structure operative such as the challenge of the cost of the structure of the structure of the explore the structure of the structure of the structure of the structure of the spirator structure of the spirator structure of the spirator structure of the structure

E

Y.

GlobalSIP 2018

topologically-complex data domains. In nodro to reveal-becaust structural properties of such data on graphs and/or querely represent different classes of signals on graphs, we can construct distinuirs of atoms, and represent graph signals as linear combinations of the distionary stans. The dosigo of such distancies is one of the finadumtal problems of signal processing, and the literature in Effect with a vide range of distinuiries, including, e.g., Poszier, time-frequency, shortle, and bandle distinuiris (see, α_i) [1] for an excitate listatical correct

time frequency, curvels, shorts, and bandle dictionrics (see, s.g., 1) for an conclust historical surveys of dictionery disposed and signal transformation). See the start of the start of

Preprint submitted to Elsevie

Introduction to Graph Signal Processing Ljubiša Stanković, Miloš Daković, and Ervin Sejdić

Published in: L. Stanković and E. Sejdić (eds), Vertex-Frequency Analysis of Graph Signals Springer-Nature, 2019. ISBN 978-030-01573-0

Contents

- Introduction Gran 2.2 Properties of Graphs and Associated Matrices 2.3 Expersion be comparation of the Adjacency Matrix 2.4 Expersion be comparation of the Adjacency Matrix 2.5 Vertex Ordering, Cohering, and Segnetration 2.5 Vertex Ordering, Cohering, and Segnetration 2.3 Graph Found transformed Segnetration 3.1 Graph Found transformed Segnetration 3.4 Fibring in the Adjacency Matrix Special Domain 3.4.7 Spectra Domain 3.4.7

- 1.4.2 Spectral Danuas Pittering
 1.4.3 Spectral Danuas Pittering
 1.4.3 Spectral Danuas Pitter Design
 1.4.4 Spectral Danuas Pitter Design
 1.5 Ordering and Pittering in the Laplacias Spectral
 1.7 Systems on a Graph Defined Using the Laplacias
 3.1 Graph, stars Biguids
 3.1 Graph, stars Biguids
 3.1 Diagna, stars Biguid
 3.1 Diagna, stars
 4.2 Subampling of the
 3.2 System
 3.2 Diagna
- L. Stanković and M. Daković University of Montenegro, Podgorica, Montenegro E-mail: ljubisa@ac.me, milos@ac.me

- University of Pittsburg, Pittsburg, PA, USA E-mail: esricic@pitt.edu

Abstract 2013

Kepwerds: Signal processing on graphs; time-frequency analysis; generalized translation and modul spectral graph theory; localizatios; clustering

1. Introduction

In applications such as social networks, electricity networks, transportation networks, and sensor net-works, data naturally reside on the vertices of weighted graphs. Moreover, weighted graphs are a fields to that on the used to describe similarities between also points in statistical learning problems, func-tional connectivities between afferent regions of the brain, and the geometric structures of constites other topologically-compute data domains.

lices in a homogeneous way (i.e., the resulting dictionaries are invariant to perm beling). Unfortunately, weighted graphs are irregular structures that lack a shift-

Bend addresses david stansaforf i. a. (David 1 Himmo), bujania.r.icon40qr1.a. (Bonjunia Hiccord), picers subgraphysettyr i. a. (Press Vasilagderus) "Diaw out was expected by PTI-Ong and number 355331 (MLacX. "Pist of the was hyperted here was presented at the IEEE Statistical Signal Processing Workshop, Aspart 2011, Ann Arter, MJ.

58/63

Noname manuscript No. (will be inserted by the editor

PRESENTATIONS

Graph Signal Processing: An Introductory Overview

Antonio Ortega

Signal and Image Processing Institute Department of Electrical Engineering University of Southern California Los Angeles, California

May 25, 2016

Dictionary Design for Graph Signal Processing

David Shuman

May 25, 2016 Graph Signal Processing Workshop Philadelphia, PA

Special thanks and acknowledgement to my collaborators:

Andre Archer, Andrew Beveridge, Xiaowen Dong, Mohammad Javad Faraji, Stefan Faridani, Pascal Frossard, Nicki Holighaus, Jason McEwen, Daniel Kressner, Yan Jin, Sunil Narang, Antonio Ortega, Javier Pérez-Trufero, Nathanaël Perraudin, Benjamin Ricaud, Dorina Thanou, Pierre Vandergheynst, Elle Weeks, and Christoph Wesmeyr

Wavelets on Graphs, an Introduction

Pierre Vandergheynst and David Shuman

Ecole Polytechnique Fédérale de Lausanne (EPFL) Signal Processing Laboratory {pierre.vandergheynst,david.shuman}@epfl.ch

> Université de Provence Marseille, France

November 17, 2011

Geometric Deep Learning

For non-grid 3D images like point clouds and meshes, and inherently graphbased data

Version "Wed 13 September 2017" Petteri Teikari, PhD Boy/petteri reskari, com/

101 (0) (2) (2) 2 040

YOUTUBE PRESENTATIONS

YOUTUBE PRESENTATIONS (CONT.)

Welcome to the GSP workshop 2020! 2020 Graph Signal Processing Workshop

Sergio Barbarossa mation Engineering, Electronics and Telecommunications ize University of Rome

Topological Signal Processing: Uncovering petterns in the data relying on multiway reliefons

GSP2020 www.gipworkitop.org

Antonia G. Marquesi EXPOSE deneral Char

Conclusions

IN SUMMARY

Main observations

- there is clearly a need to work with graph signals
 - \rightarrow network like systems and applications
 - \rightarrow accounting for signal structure can significantly improve the performance
- (not surprisingly) the aim is to extend signal processing to graph signals
- most techniques for graph signal processing evolved around eigendecomposition of W or L

... however

- there are often several/many different approaches to accomplishing the same thing or task in graph signal processing

 → the field still appears far less mature than e.g. Network Science
 → a viable strategy can be transforms of graphs ←→ regular signals
- distributed signal processing seems to be appealing practical alternative to (centralized) graph signal processing

Thank you!

pavelloskot@intl.zju.edu.cn