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

Acknowledgements

- Collaborators
  - **Sunil Narang** (Microsoft), **Godwin Shen** (Northrop-Grumman), **Eduardo Martínez Enríquez** (Univ. Carlos III, Madrid)
  - **Akshay Gadde, Jessie Chao, Aamir Anis, Yongzhe Wang, Eduardo Pávez, Hilmi Egilmez, Joanne Kao** (USC)
  - Marco Levorato (UCI), Urbashi Mitra (USC), Salman Avestimehr (USC), Aly El Gamal (USC/Purdue), Eyal En Gad (USC), Niccolo Michelusi (USC/Purdue), Gene Cheung (NII), Pierre Vandergheynst (EPFL), Pascal Frossard (EPFL), David Shuman (Macalaster College), Yuichi Tanaka (TUAT), David Taubman (UNSW).
- Funding
  - NASA (AIST-05-0081), NSF (CCF-1018977, CCF-1410009, CCF-1527874)
  - MERL, Samsung, LGE, Google
  - Sabbatical: Japan Society for Promotion of Science (JSPS), UNSW, NII, TUAT

Goals

- Some background and basic concepts
- A bit of history and what's going on now (at the workshop!)
- FAQs and challenges

- Disclaimers

Goals

- Some background and basic concepts
- A bit of history and what's going on now (at the workshop!)
- FAQs and challenges

- Disclaimers
  - Partial overview!

Goals

- Some background and basic concepts
- A bit of history and what's going on now (at the workshop!)
- FAQs and challenges

- Disclaimers
  - Partial overview!
  - More questions than answers

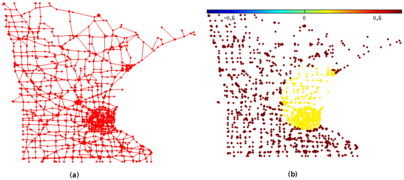
Outline

This is the 2nd GSP workshop

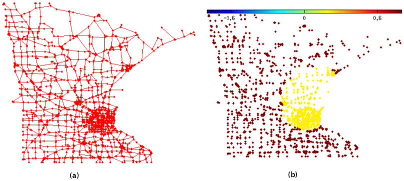
- First workshop was held with IEEE GlobalSIP, Dec 2013.
- Much more activity in this field!
- More than double the number of presentations, 1 day vs 3 days.

Graph signal processing: why now?

1796 Philadelphia roadmap, Library of Congress



Graph signal processing: why now?



1796 Philadelphia roadmap, Library of Congress

Standard questions: What is the shortest path? What is safest path?

Graph signal processing: why now?

- Going from physical graphs to information graphs:
  - From:
    - Roads and rail
    - Telephone Networks
  - To:
    - Web
    - Online social network

Graph signal processing: why now?

- Going from physical graphs to information graphs:
  - From:
    - Roads and rail
    - Telephone Networks
  - To:
    - Web
    - Online social network
- Information links were always there but they were not an obvious graph
  - encyclopedia vs wikipedia

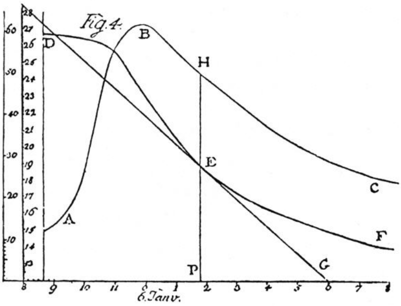
Graph signal processing: why now?

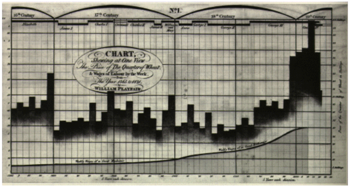
- Going from physical graphs to information graphs:
  - From:
    - Roads and rail
    - Telephone Networks
  - To:
    - Web
    - Online social network
- Information links were always there but they were not an obvious graph
  - encyclopedia vs wikipedia
- Sensing technology allows us to measure “on a graph”

Graph signal processing: why now?

- Going from physical graphs to information graphs:
  - From:
    - Roads and rail
    - Telephone Networks
  - To:
    - Web
    - Online social network
- Information links were always there but they were not an obvious graph
  - encyclopedia vs wikipedia
- Sensing technology allows us to measure "on a graph"
- Where is GSP being used?
  - Physical networks
  - Information networks
  - Regular signals

Graph signal processing: why now?

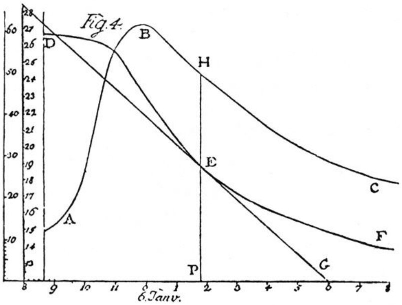


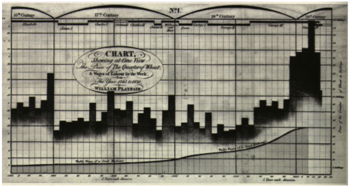


Lambert, 1765, Playfair ca. 1820

See (E. Tufte, The Visual Display of Quantitative Information, '83)

Graph signal processing: why now?



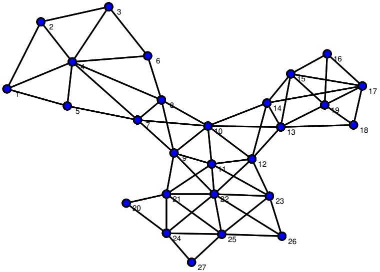


Lambert, 1765, Playfair ca. 1820

See (E. Tufte, The Visual Display of Quantitative Information, '83)
Do we know how to think about and visualize graph signals?

Outline

Multiple algebraic representations



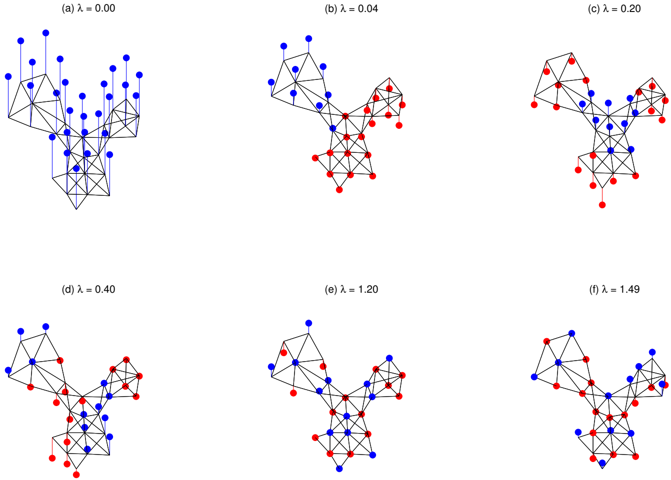
- Graph $G = (\mathcal{V}, E, w)$.
- Adjacency $\mathbf{A}$, a_{ij}, a_{ji} = weights of links between i and j (could be different if graph is directed.)
- Degree $\mathbf{D} = diag\{d_i\}$, in case of undirected graph.
- Various algebraic representations
  - normalized adjacency $\frac{1}{|\lambda_{max}|}\mathbf{A}$
  - Laplacian matrix $\mathbf{L} = \mathbf{D} - \mathbf{A}$.
  - Symmetric normalized Laplacian $\mathcal{L} = \mathbf{D}^{-1/2}\mathbf{L}\mathbf{D}^{-1/2}$
- Graph Signal $\mathbf{f} = \{f(1), f(2), ..., f(N)\}$

- Discussion:
  1. Undirected graphs easier to work with
  2. Some applications require directed graphs
  3. Graphs with self loops are useful

Graph spectrum, GFT

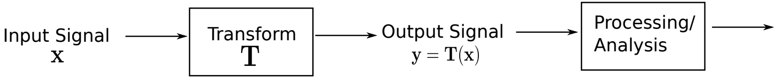
- Different results/insights for different choices of operator
- Laplacian $\mathbf{L} = \mathbf{D} - \mathbf{A} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}'$
- Eigenvectors of $\mathbf{L}$: $\mathbf{U} = \{\mathbf{u}_k\}_{k=1:N}$
- Eigenvalues of $\mathbf{L}$: $diag\{\boldsymbol{\Lambda}\} = \lambda_1 \leq \lambda_2 \leq ... \leq \lambda_N$
- **Eigen-pair system $\{(\lambda_k, \mathbf{u}_k)\}$ provides Fourier-like interpretation — Graph Fourier Transform (GFT)**

Eigenvectors of graph Laplacian



- Basic idea: increased variation **on the graph**, e.g., $\mathbf{f}^t\mathbf{L}\mathbf{f}$, as frequency increases

Graph Transforms and Filters

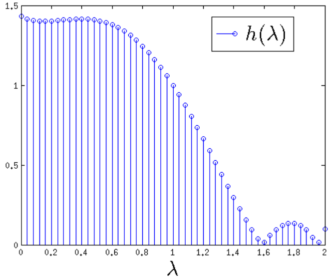


- Properties
  - Invertible
  - Critically sampled/overcomplete
  - Orthogonal/near orthogonal/frames
- **What makes these "graph transforms"?**
  - **Frequency interpretation**
    - Operation is diagonalized by $\mathbf{U}$
  - **Vertex localization**: polynomial of the operator ($\mathbf{A}$, $\mathbf{L}$, etc)
    - Polynomial degree is small (vs a polynomial of degree $N - 1$).

Frequency interpretation

- Spectral Wavelet transforms (Hammond et al, CHA'09):

  Design spectral kernels: $h(\lambda) : \sigma(G) \to \mathbb{R}$.



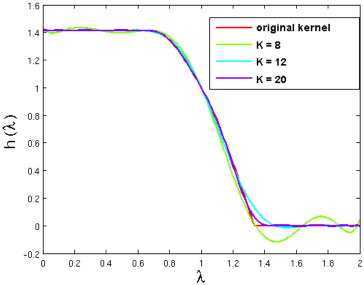
$$\mathbf{T}_h = h(\mathcal{L}) = \mathbf{U} h(\mathbf{\Lambda}) \mathbf{U}^t$$

$$h(\mathbf{\Lambda}) = diag\{h(\lambda_i)\}$$

- Analogy: FFT implementation of filters

Vertex Localization: SGWT

- Polynomial kernel approximation:



$$h(\lambda) \approx \sum_{k=0}^{K} a_k \lambda^k$$

$$\mathbf{T}_h \approx \sum_{k=0}^{K} a_k \mathcal{L}^k$$

- Note that **A** and **L** are both 1-hop operations *K*-hop localized: no spectral decomposition required.

Summary

- Different types of graphs (directed, undirected, with/without self loops)
- Multiple algebraic representations of graphs ($\mathbf{A}, \mathbf{L}, \ldots$)
- Their eigenvectors/eigenvalues induce a notion of variation
- The corresponding operators can be viewed as "shifts" or "elementary operators"
- Polynomials of these operators represent "local" processing on the graph
- Graph filtering: polynomial/diagonal in vertex/frequency

Outline

Graphs Signal Processing: a bit of history

Key point: many threads lead to graph signal processing

- spectral graph theory
- image processing
- semi-supervised learning
- block transforms
- vertex domain transforms
- frequency domain transforms

Spectral graph theory

- Results emerging in '50s and '60s linking algebraic graph structure to graph properties, initial work in Math, later interest in CS
- Classic works
  - (Cvetkovic, Doobs and Sachs, '80)
  - (Chung, '96)
  - (Spielman, '01)
- Primary concerns are linking spectrum and graph properties, no signals

- Link to GSP is weaker than it should be: because we are interested in working on arbitrary graphs (more on this later)

Linking graph spectrum to graph structure

DCT and KLT

- (Ahmed, Natarajan, Rao, T. on Computers, '74) Optimality of DCT for high correlation random vectors (close to 1)
- (Strang, SIAM'99) Graph interpretation (eigenvectors of line graphs with weight one), connection to DST
- (Püschel & Moura, SIAM'03) Generalization,
- (Püschel & Moura, TSP'08) General Algebraic signal processing perspective:
  - DCT as basis for a signal space of finite signals under different boundary conditions (Sandryhaila & Moura, '14) (see afternoon talk!)
- (Shen et al, PCS'10) regular graphs with irregular weights (use GFT of the graph)
- (Zhang, Florencio & Chou, SPL'13), General case of graphs obtained from precision matrices corresponding to Gauss Markov Random Fields.

Eigenvectors of graphs with regular connectivity and unequal weights/self-loops

Image processing

- (Wu & Leahy, PAMI, '93), (Shi & Malik, PAMI, '00): graph cuts for image segmentation, smaller edge weights across image boundaries
- (Tomasi & Manduchi, '98) Bilateral filtering, filter weights function of pixel and photometric distances
- (Elmoataz, Lezoray, Bougleux, TIP'08), (Osher et al, SIAM'07) Graph Laplacians for regularization
- (Milanfar, SPM'13) Various signal dependent image filters from a graph perspective

Weighted graphs with edges a function of pixel distance and intensity differences

Semi-supervised learning

- Learning from labeled and unlabeled training data
  - Estimate labels for unlabeled data
  - Decide what data to label
  - Consider kNN data graph
- (Belkin, Niyogi, '03 NIPS), (Zhou et al, NIPS'04), (Smola & Kondor, COLT'03), (Zhu et al, ML'03) Regularization on graphs, semi-supervised learning, label propagation
  - Generally use $\mathbf{L}$ to favor smooth signals on the graph
- (Anis et al, KDD'14) Graph signal sampling interpretation
- Why should a "label" signal be smooth?

Graphs connecting datapoints in feature space (e.g., kNN), labels should be slow varying

Graph Transforms: vertex domain approaches

- (Schroeder & Sweldens, '95), (DeRose & Salesin, '95) Transforms for attributes defined on meshes, often use lifting based techniques
- Network graphs, (Crovella& Kolaczyk, INFOCOM'03), graphs with arbitrary connectivity, analysis tool, overcomplete
- Sensor networks, (Baraniuk et al, IPSN'06), (Wang & Ramchandran, '06), (Ciancio, et al, IPSN'06) (Shen & Ortega, IPSN'08, TSP'10)
  - Emphasis on distributed operation, approaches sometimes use structure (e.g., trees, tesselations)
- Graph lifting (Narang & Ortega, APSIPA'09), (Janson et al, Royal Statistical Society'09)
  - Even/odd assignment in regular signals correspond to bipartite approximation of a graph

Vertex domain approaches require graph partitions

Graph Transforms: frequency domain approaches

- Diffusion wavelets (Coiffman & Maggioni'06):
  - Use successive applications of a diffusion to create subspaces of lower graph frequency content (and less localization in vertex domain)
  - Eigendecomposition of powers of $\mathbf{L}$
  - No exact localization in vertex domain
- Spectral Graph Wavelets (Hammond et al, CHA '09)
  - Spectral domain design (kernel having desirable properties and its scalings)
  - Polynomial approximation for localization, no need to explicit frequency decomposition
  - Nice vertex/frequency interpretation, overcomplete
- Filterbanks (Narang & Ortega, TSP'12, TSP'13)
  - Critically sampled, orthogonal/bi-orthogonal solutions
  - Exact solutions, only bi-partite graphs

Different trade-offs possible depending on whether critical sampling is required

Sampling

- Irregular sampling in regular domains, e.g., properties that guarantee reconstructions (Gröchenig, 92), (Aldroubi & Gröchenig, '01)
  - Focus is on reconstruction based on regular domain properties (frequency)
- Optimality conditions for combinatorial graphs (Pesenson'08)
- Various approaches for sample set selection and reconstruction (Anis et al, '14), (Shomorony & Avestimehr, 14), (Chen et al, '14)

Selection of nodes that are most informative

Conference topics

- Distributed processing
- Graph learning
- Filtering
- Fundamentals
- Sampling
- Statistical Graph Signal Processing
- Applications

Graph Filter Design

- Classic problem in DSP
- Goal: Design filters with different properties in terms of localization, orthogonality, etc.
- Different types of filters:
  - Moving average, graph-temporal
  - Graph diffusion
  - graph-temporal
  - lifting approaches
  - representations using dictionaries

Graph Learning

- Goal: learn a graph from data
- Multiple cases: covariance,propagating graph signals, etc.
- Examples: estimate a sparse inverse covariance (precision) matrix, estimate a Laplacian that makes a observed data smooth on average
- Question: what is the advantage of using a graph vs PCA?
  - Interpretation, approximate KLT with polynomial graph operations

Sampling

- Goal: decide which graph signal samples (values associated to vertices) should be observed so that we can reconstruct the others
- Assumptions about smoothness of signal
- (almost) any random sampling works when signals are exactly bandlimited and noise free
- Noise and non-bandlimited behavior make things complicated
  - Robust sampling methods (randomize, iterative, with/without knowledge of the GFT)
  - New criteria for signals to be sampled (piecewise smooth)
  - Distributed sampling

Statistical GSP

- Random signals on graphs
- Definition of Stationary Graph Signals
- Time varying signals over graphs
- PCA on graphs

Applications

- Image regularization
- Compression
- Computer vision applications (e.g., motion analysis)
- Origin-Destination traffic matrices
- Tracing of outbreaks
- Wireless network optimization
- Brain connectivity

Outline

Scale

- Can we really apply our tools to large scale datasets?
  - Facebook 1G+ nodes
- How to interpret results in a large scale graph
- How to interpret locality:
  - Graph diameter vs average?
  - We often do not consider what the real footprint is for a polynomial of degree K.
- Large scale implementation
  - Parallelization
  - GraphLab
    - Basic primitive: each node communicating with its neighbors
    - Algorithms to distribute nodes across processors to preserve locality
    - Example: node requests information from neighbors and computes an output ($\mathbf{Lx}$) or apply this recursively ($\mathbf{L}(\mathbf{Lx})$), i.e. every node stores $\mathbf{Lx}$ and then $\mathbf{L}^2\mathbf{x}$, etc.
  - Should our community contribute?

How to choose a graph

- In some applications graph is given (e.g., social networks)
- In some it is a function of some known information (e.g., distance in a sensor network)
  - How to select weights?
  - e.g., bilateral filter, etc
  - are there optimality results?
- Designing graphs from data
  - Sparse inverse covariance: why is a graph representation of a dataset better?
  - Advantages vs other methods

Shifts and localization

- Most current filtering schemes use an operator (Laplacian or Adjacency)
- Should it be considered a "shift" (note that sometimes signals vanish after being shifted).
- The effect of a shift depends on the eigenvalue associated with it: graphs with same eigenvectors, but different eigenvalues? $\mathbf{L} = \sum_i \lambda_i \mathbf{u}_i \mathbf{u}_i^t$, different graph connectivity, but same frequency interpretations (Gavili & Zhang, Arxiv'15)
- Classes of equivalent graphs?
- Bounds on frequency-vertex domain localization
- Specialize these bounds to specific graph types
- Complicated because of properties of shift
- Several contributions in this workshop

Need to consider special characteristics of graphs

- Example: how to deal with high multiplicity eigenvalues (high dimensional subspace with the same graph frequency) (Zeng et al, ICASSP'16)
- How to assess the impact of "removing" edges?
- What is the best way to approximate a graph?
- Graph reductions/simplifications
- More generally there could be interesting results that apply only to certain classes of graphs?
  - bipartite (Narang & Ortega, '11), circulant (Ekambaran et al, '13), M-block cyclic (Teke & Vaidyanathan, '16)

Datasets and community

- Should we have a set of standard datasets?
- Matlab code: GSP Toolbox EPFL (Perraudin & Paratte)
- Anything else we should do?

What is the killer app?

- Understand in what cases a graph-based approach is better than directly working with signals in $\mathbb{R}^N$.
- Many cases
  - Graph is given (web, social network)
  - Irregular measurements (sensor networks)
  - Graph approaches are an alternative (e.g., images)
  - Data driven methods (e.g., machine learning).
  - Large scale systems (e.g., finite state machines)
- **GSP methods are closely linked to existing approaches**
- New perspectives on existing topics
- perhaps an emerging new way to understand problems

Outline

Conclusions

- GSP has deep roots in the Signal Processing community
- A lot of progress, interesting results
- Many open questions!
- Outcomes
  - Work with massive graph-datasets: potential benefits of localized "frequency" analysis
  - Novel insights about traditional applications (image/video processing)
  - Promising results in machine learning, image processing, among other areas
- To get started (Shuman et al, SPM'13), (Sandryhaila & Moura, SPM'14)
- Enjoy the workshop!