Discrete Signal Processing on Graphs (DSP_G): Big Data Processing

Deepak Goyadi, Jashan Singhal, Utkarsh Sharma and Mustafa Lokhandwala

Department of Electrical Engineering, IIT Bombay

Abstract

DSP_G extends signal processing concepts and methodologies from the classical signal processing theory to data indexed by general graphs.
We review fundamental concepts of DSP_G, including graph signals, graph filters, graph Fourier transform and compare them with their counterparts from the classical signal processing theory.

Graph Signal

A graph signal is a map from the set $\mathcal{V}$ of nodes into the set of complex numbers $\mathcal{C}$, written as

$$\mathbf{s} = (s_0 \quad s_1 \quad \dots \quad s_{N-1})^T$$

each element s_n being indexed by node v_n.[2]

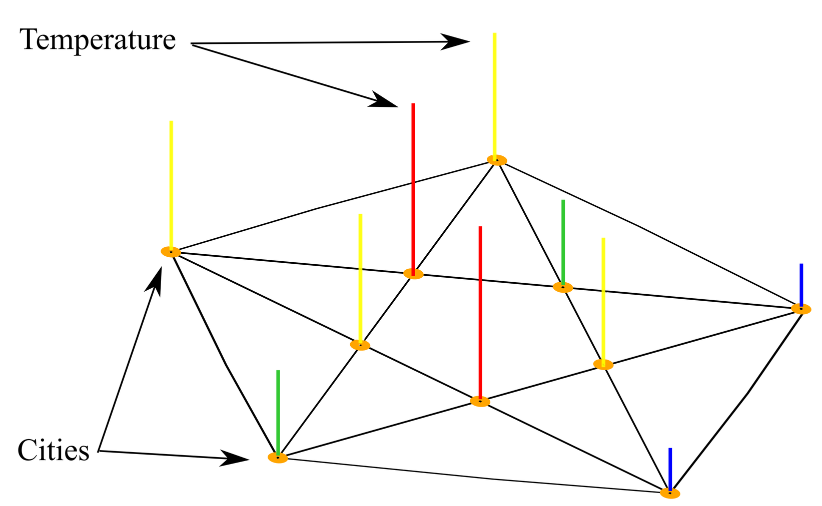

Figure 1: An example of a graph with 10 nodes representing geographical locations and the daily average temperature values at different nodes forming the graph signal.

Graph Shift

Graph shift is a generalization of the time shift or delay that delays the input signal $\mathbf{s}$ by one sample $\tilde{s}_n = s_{n-1}\ (mod\ N)$. For a general graph $G = (\mathcal{V}, \mathbf{A})$, the graph shift is realized by replacing the sample s_n at node v_n with the weighted linear combination of the signal samples at its neighbors:

$$\tilde{s}_n = \sum_{m=0}^{N-1} \mathbf{A}_{n,m} s_m$$

where $\mathcal{V} = \{v_0, ..., v_n\}$ is the set of nodes and $\mathbf{A}$ is the weighted adjacency matrix of the graph.[2]

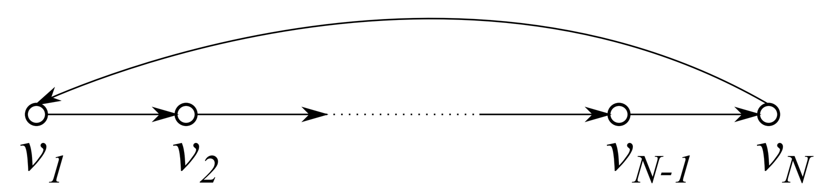

Figure 2: This graph represents a finite, periodic discrete time series. All edges are directed and have the same weight 1, reflecting the causality of a time series and the edge from v_N to v_0 reflects its periodicity.

The adjacency matrix of this graph for $N = 3$ is a 3×3 cyclic permutation matrix,

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

If the value of the graph signal at an instant n is $s_n = [3 \quad -4 \quad 5]$ at nodes $[v_1 \quad v_2 \quad v_3]$, then according to the structure of the graph, $s_{n+1} = [5 \quad 3 \quad -4]$, which can also be written as $s_{n+1} = \mathbf{A}s_n$. This alludes to the notion of the shift to general graph signals s where the relational dependencies among the data are represented by an arbitrary graph $G = (\mathcal{V}, \mathbf{A})$.

Graph Filter

In classical DSP, filters are systems that take a signal as input and produce another signal as output. Similar to traditional DSP, we can represent filtering on a graph using matrix-vector multiplication
A linear, shift-invariant graph filter $\mathbf{H}$ is a polynomial in graph shift $\mathbf{A}$ with possibly complex coefficients $h_l \in \mathcal{C}$ such that[2]:

$$\mathbf{H} = h_0\mathbf{I} + h_1\mathbf{A} + \dots + h_L\mathbf{A}^L$$

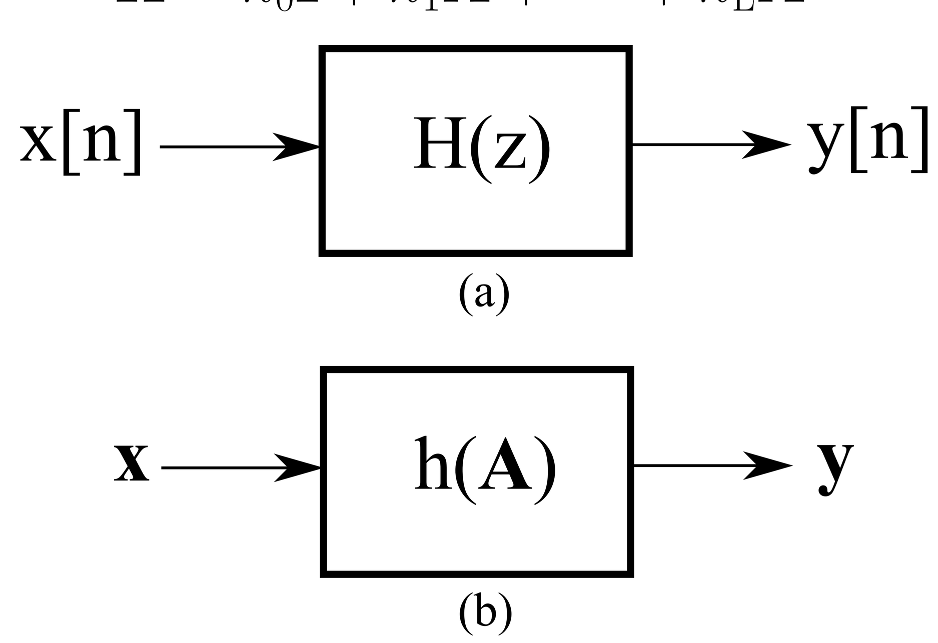

Figure 3: Similar to the function of a traditional DSP filter, a graph filter takes in a graph signal as an input and produces another graph signal at the output.

Big Data

- DSP_G is particularly motivated by the need to extend traditional signal processing methods to datasets with complex and irregular structure.[3]
- Graphs provide a versatile data abstraction for multiple types of data, including sensor network measurements, text documents, image and video databases, social networks, and others.[3]

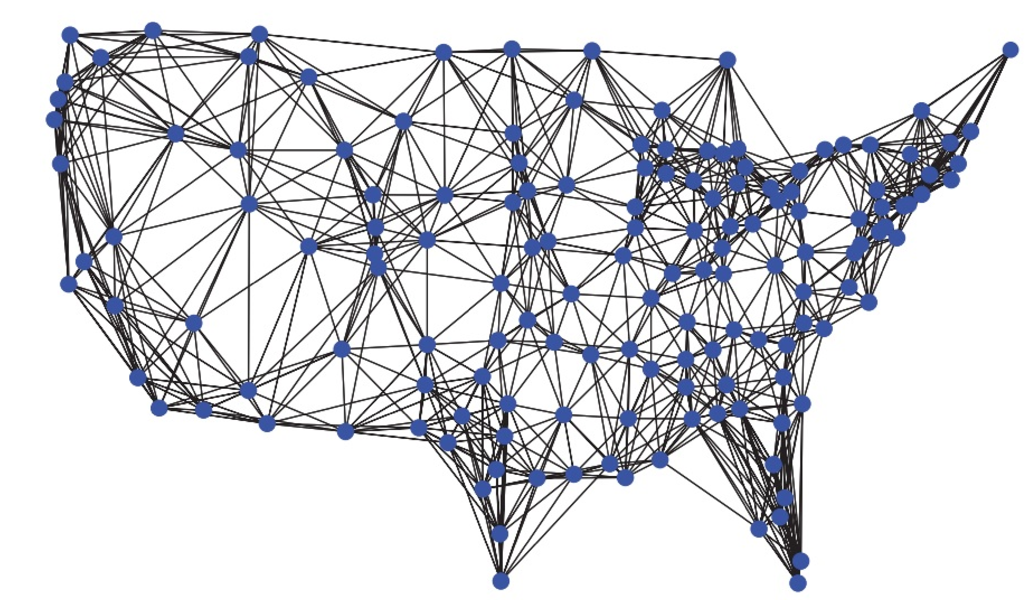

Figure 4: US weather stations are the nodes and the temperature value at each node constitutes the graph signal. The adjacency matrix is formed using the distance between the nodes.[3]

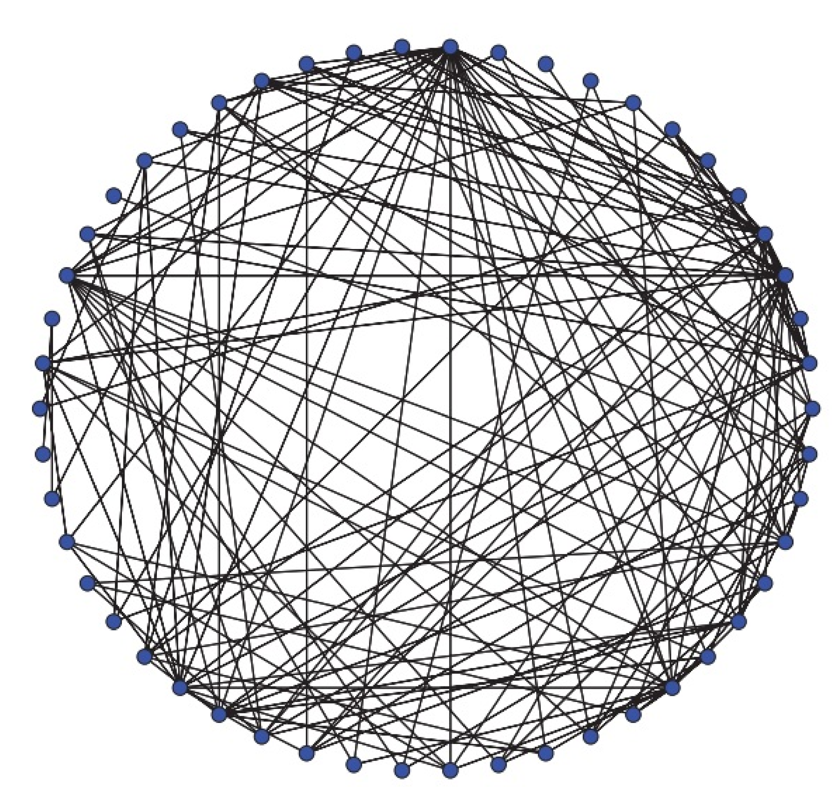

Figure 5: The nodes stand for webpages while their interconnections represent hyperlinks between pages. The pattern of interlinking of the nodes (webpages) leads to their classification.[3]

Graph Fourier Transform

- Fourier Transform requires that a given vector be decomposed into orthogonal components. Eigenvectors of the adjacency matrix form such an orthogonal set.[2]
- The eigenvectors evolve independent of one another when time evolution is carried out through the adjacency matrix. Therefore, this formulation apes the behaviour of Fourier transform.[2]

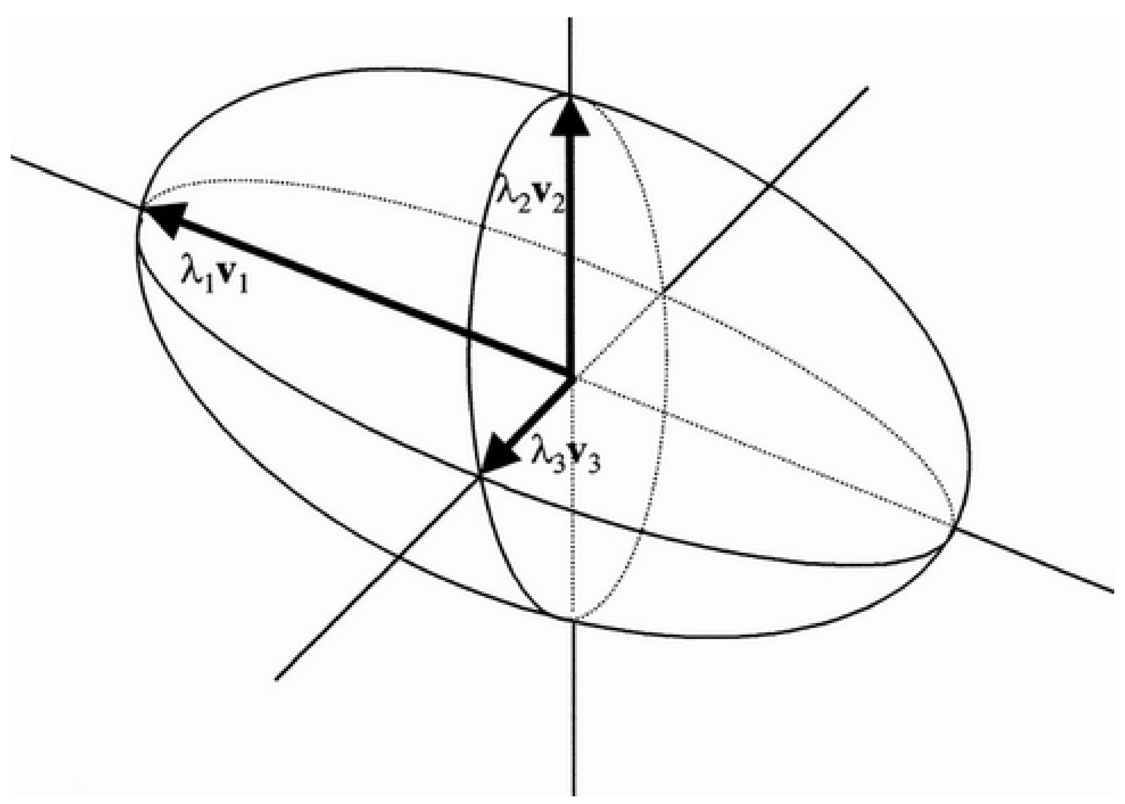

Figure 6: Eigenvectors of the adjacency matrix are orthogonal vectors that evolve independently and are therefore used as the basis for Fourier decomposition.

- Given the generalized eigenvector matrix $\mathbf{V}$ (of the adjacency matrix $\mathbf{A}$) and the value of the graph signal $\mathbf{s}$ at any instant, the Fourier transform of $\mathbf{s}$ can be easily found as,

$$\mathcal{F}(\mathbf{s}) = \mathbf{V}^{-1}\mathbf{s}$$

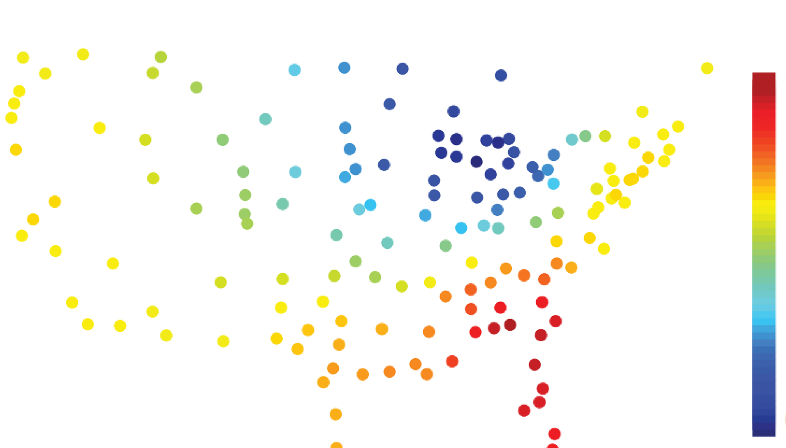

Figure 7: The Fourier basis vector that captures most energy of temperature measurements reflects the relative distribution of temperature across the mainland United States.[2]