Graph-based Reconstruction of Time-varying Spatial Signals

Kai Qiu*‡, Xiaohan Wang*, Tiejian Li†, and Yuantao Gu*
*Department of Electronic Engineering, Tsinghua University, Beijing 100084, CHINA
†State Key Laboratory of Hydroscience and Engineering, Tsinghua University, Beijing 100084, CHINA
‡Aviation University Air Force, Changchun 130022, CHINA

***Abstract*—Signal processing on graphs is an emerging field studying signals in irregular domains, and has been applied to many applications such as sensor networks and recommendation systems. In this paper, a novel method for the recovery of time-varying spatial signal based on graph is proposed. A graph is established according to the spatial position of the signal. Unlike the previous works, the smoothness of the temporal differential signal on the graph rather than the smoothness of the signal itself is used to help reconstruction. Two experiments of real-world are conducted. The first experiment of sea surface temperature data shows that the proposed algorithm achieves less reconstruction error than other algorithms, and the second experiment of sensor network data demonstrates the rationality and superiority of the proposed algorithms from intuition.**

***Index Terms*—signal processing on graphs; graph signals; signal reconstruction**

I. Introduction

High-dimensional time series analysis plays an important role in many applications, such as video processing, sensor networks, and climatology [1, 2]. In this paper, time-varying spatial signal is considered, with the position of each component known. Time-varying spatial signal is usually collected over a large time frame at some fixed positions, and the data may include missing values due to sensor malfunctions, human errors or conservation of resource. the missing values is called unsampled points, and the purpose of this paper is to reconstruct them based on the sampled points.

The classical methods to handle time-varying spatial signals with missing values are scattered data interpolations including Shepards interpolation, Kernel regression, Moving least-squares and so on [3]. Natural neighbor interpolation [4] is a typical technique that uses a Voronoi diagram of the sampled points to formalize the notion of neighbor. Two points share a common boundary in the Voronoi diagram are defined as neighbors. The unsampled points are reconstructed by the weighted average of their neighbors, where the weights are volume-based.

In recent years, low-rank matrix completion methods have been successfully applied in problems such as remote sensing [5] and recommender systems [6]. It is proved [7, 8] that the exact recovery of low-rank matrix in the case of random uniformly sampling without noise is achieved as long as the number of sampled entries is large enough. A robust algorithm named Fixed Point Continuation with Approximate SVD (FPCA) is proposed in [9]. Usually, the time-varying spatial signal may be approximately low-rank if it changes not too fast, thus the missing values could be reconstructed by low-rank matrix completion.

Signal processing on graphs is an emerging field mainly solving problems in irregular domain [10], such as neuronal networks, social networks, and transportation problem. Sampling and reconstruction of graph signal has been widely studied recently [11–13]. An algorithm named Iterative Least Squares Reconstruction is proposed to reconstruct band-limited graph signals from partially observed samples by projecting the input signal into the appropriate band-limited graph signal space [14, 15]. If the graph signals tend to be smooth but not exactly band-limited, a graph regularization framework is used. Most of the time the positions of a spatial signal may be distributed irregularly, so it is very effective to resort to signal processing on graphs. The corresponding graph could be constructed according to the distance between the points.

In general, for most real-world datasets such as temperatures of cities across the country, they are not smooth enough on graph, thus the results by graph regularization framework [15] may not be satisfactory. However, we find that the differential signals along the time direction often demonstrate good smoothness on graph than the original signals. Thus, we propose a new algorithm for the reconstruction of time-varying spatial signal based on graph, which utilizes the smoothness of temporal differential signals.

The rest of this article is organized as follows. Section 2 describes the problem and introduces several existing reconstruction methods for the problem. Section 3 proposes the new algorithm. In Section 4, the proposed algorithm is tested and compared with the existing methods. The conclusion is drawn in Section 5.

II. Problem description

A time-varying spatial signal can be denoted as a matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_T]$. Each row of $\mathbf{X}$ is corresponding to a position, which can be viewed as a time series, and each column of $\mathbf{X}$ denotes the values of all the positions at a certain moment. At each time i, only partial points are sampled, which

This work was partially supported by National Natural Science Foundation of China (NSFC 61371137, 61571263, 61531166005, 51459003) and Tsinghua University Initiative Scientific Research Program (Grant 2014Z01005). The corresponding author of this paper is Yuantao Gu (gyt@tsinghua.edu.cn).

is denoted as $\mathbf{y}_i = \mathbf{J}_i \mathbf{x}_i$. $\mathbf{J}_i$ is consisted of only 1 and 0, and each row contains only one 1.

Let $\mathbf{x}$ be the vectorization of matrix $\mathbf{X}$, $\mathbf{J} = \text{diag}\{\mathbf{J}_1, \mathbf{J}_2, \cdots, \mathbf{J}_T\}$, then the total samples can be denoted as $\mathbf{y} = \mathbf{J}\mathbf{x}$. Obviously, it is underdetermined to recover $\mathbf{m}$ given $\mathbf{y}$. The problem of this work is that if the time-varying spatial signal changes slowly with time, how to recover the original signal by only partial samples. We introduce several existing reconstruction methods for the problem at first.

A. Scattered data interpolation

Considering a point set $\{\mathbf{p}_k | k = 1, 2, \cdots, N\}$, the value of point $\mathbf{p}_k$ is denoted as $f(\mathbf{p}_k)$. The approximation of point $\mathbf{p}$ by natural neighbor interpolation [3] is written as

$$\hat{f}(\mathbf{p}) = \sum_{k \in S} \omega_k(\mathbf{p}) f(\mathbf{p}_k), \tag{1}$$

where S is the set of indices associated with the neighbors of point $\mathbf{p}$ and $\omega_k(\mathbf{p})$ are weight functions.

B. Recovery by FPCA

The precondition of using matrix completion is that matrix $\mathbf{X}$ is low-rank. Under the premise of signal changing slowly with time, this condition is easy to be satisfied. Algorithm FPCA [9] formulates the reconstruction problem as

$$\min_{\mathbf{X}} \mu ||\mathbf{X}||_* + \frac{1}{2} ||\mathbf{J}\mathbf{x} - \mathbf{y}||_2^2, \tag{2}$$

where μ is the regularization parameter. Nuclear norm $||\mathbf{X}||_*$ is the sum of the singular values of $\mathbf{X}$, which is the approximation of the rank of $\mathbf{X}$.

C. Recovery by graph regularization

An undirected graph is denoted as $\mathcal{G}(\mathcal{V}, \mathcal{E}, \mathbf{W})$, where $\mathcal{V}$ is the vertex set, $\mathcal{E}$ is the edge set, and $\mathbf{W}$ is the weighted adjacency matrix of the graph. If one real number is associated with each vertex, these numbers of all the vertices are collectively referred as a graph signal. The graph Laplacian is defined as

$$\mathbf{L} = \mathbf{D} - \mathbf{W}, \tag{3}$$

where the degree matrix $\mathbf{D}$ is a diagonal matrix, and the ith diagonal element is the degree of vertex i.

Considering the time-varying spatial signal at time i, because the position of each point is known, we can establish a graph whose weights are related to the geometric distance between every two points. If signal $\mathbf{x}_i$ is smooth on the graph, $\mathbf{x}_i^{\mathrm{H}} \mathbf{L} \mathbf{x}_i$ will be small. Then a graph regularization framework [15] can be used for the recovery of signal $\mathbf{x}_i$. The problem is formulated as

$$\min_{\mathbf{x}_i} ||\mathbf{J}_i \mathbf{x}_i - \mathbf{y}_i||_2^2 + \mu \mathbf{x}_i^{\mathrm{H}} \mathbf{L} \mathbf{x}_i, \quad \forall i \tag{4}$$

where μ is the regularization parameter.

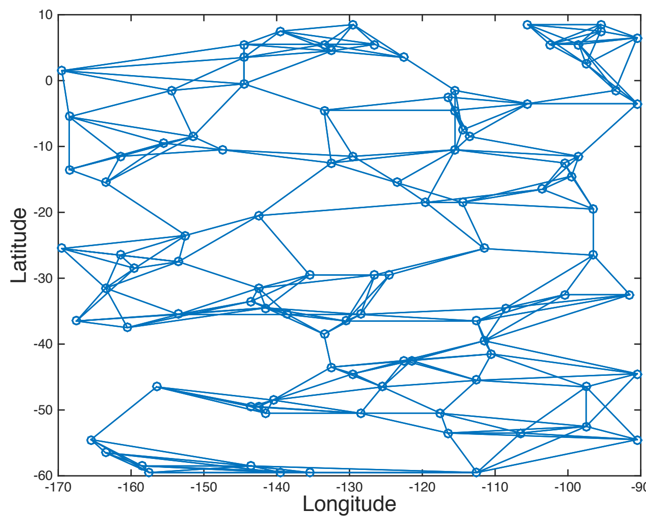

Fig. 1. Graph established by 5-nearest neighbors.

III. The Proposed Algorithm

For most real-world datasets such as spatial signals, they are not smooth enough on graph, thus the results by graph regularization framework [15] may not be satisfactory. However, we find that the differential signals along the time direction often demonstrate good smoothness on graph than the original signals, and the reconstruction of the time-varying spatial signal can be formulated as

$$\min_{\mathbf{x}} \frac{1}{2} (\mathbf{T}\mathbf{x})^{\mathrm{H}} \mathcal{L} \mathbf{T}\mathbf{x}, \quad \text{subject to} \quad \mathbf{y} = \mathbf{J}\mathbf{x}, \tag{5}$$

where $\mathcal{L} = \text{diag}\{\mathbf{L}, \mathbf{L}, \cdots, \mathbf{L}\}$ is a block diagonal matrix, and the number of $\mathbf{L}$ is T. $\mathbf{T}$ is the differential operator of signal $\mathbf{x}$ along the time direction. A gradient projection algorithm with backtracking line search can be used to solve this problem.

Reformulated to an unconstrained optimization problem, problem (5) is written as

$$\min_{\mathbf{x}} \frac{1}{2} ||\mathbf{J}\mathbf{x} - \mathbf{y}||_2^2 + \frac{\lambda}{2} (\mathbf{T}\mathbf{x})^{\mathrm{H}} \mathcal{L} \mathbf{T}\mathbf{x}, \tag{6}$$

where λ is the regularization parameter.

The detailed algorithm of problem (6) is listed in Table I. A nonlinear conjugate gradient descent algorithm with backtracking line search is used, where $f(\mathbf{x})$ is the cost function as defined in (6). The stopping criterion could be a maximum number of iterations, or $||\mathbf{g}_k||_2$ less than a threshold. The gradient of $f(\mathbf{x})$ is

$$\nabla f(\mathbf{x}) = \mathbf{J}^{\mathrm{H}} (\mathbf{J}\mathbf{x} - \mathbf{y}) + \lambda \mathbf{T}^{\mathrm{H}} \mathcal{L} \mathbf{T}\mathbf{x}. \tag{7}$$

Because problem (6) is convex, the algorithm will converge to the global optimal solution after sufficient iterations.

IV. Experiments and Discussions

The proposed algorithm is applied on two datasets: the sea surface temperature from Earth System Research Laboratory [16], and the sensor network data of Intel Berkeley Research Lab [17]. The comparison algorithms include scattered data

TABLE I
THE PROCEDURE OF PROPOSED ALGORITHM

Input:
- $\mathbf{y}$ - sampled data
- $\mathbf{J}$ - sampling matrix
- $\mathbf{L}$ - Laplacian matrix
- λ - regularization parameter
- α, β - line search parameters

Output:
- $\mathbf{x}$ - reconstructed signal

Initialization:

$\mathbf{x}_0 = 0; \quad \mathbf{g}_0 = \nabla f(\mathbf{x}_0); \quad \Delta\mathbf{x}_0 = -\mathbf{g}_0$

Iteration:

1: Backtracking line-search.
  $t = 1$;
  while $f(\mathbf{x}_k + t\Delta\mathbf{x}_k) > f(\mathbf{x}_k) + \alpha t \mathbf{g}_k^{\mathrm{H}} \Delta\mathbf{x}_k$
    $t = \beta t$;
  $\mathbf{x}_{k+1} = \mathbf{x}_k + t\Delta\mathbf{x}_k$;

2: Update search direction.
  $\mathbf{g}_{k+1} = \nabla f(\mathbf{x}_{k+1})$;
  $\gamma = \frac{||\mathbf{g}_{k+1}||_2^2}{||\mathbf{g}_k||_2^2}$;
  $\Delta\mathbf{x}_{k+1} = -\mathbf{g}_{k+1} + \gamma\Delta\mathbf{x}_k$;
  $k = k + 1$;

3: Repeat 1 and 2 until stopping criterion is met.

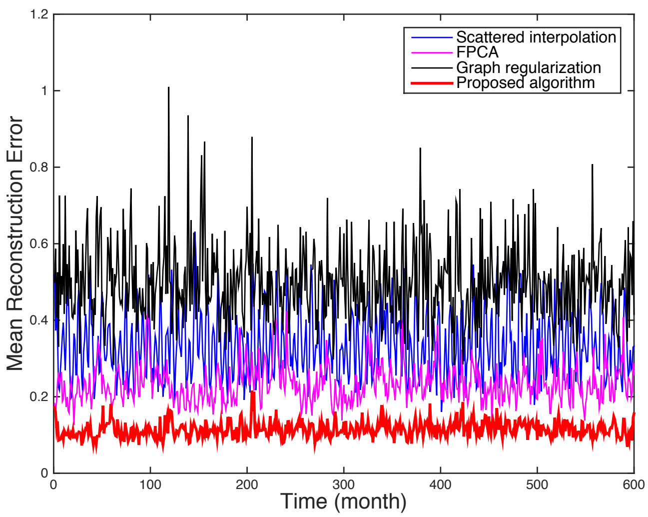

Fig. 2. The mean reconstruction error of each time.

interpolation [3], a low rank matrix completion method named FPCA [9], and the graph regularization framework [15].

Mean Absolute Error (MAE) is used for result evaluation:

$$\mathrm{MAE} = \frac{||\hat{\mathbf{x}} - \mathbf{x}||_1}{N_x} \tag{8}$$

where $\mathbf{x}$ is the original signal, $\hat{\mathbf{x}}$ is the reconstructed signal, and N_x is the length of the signals.

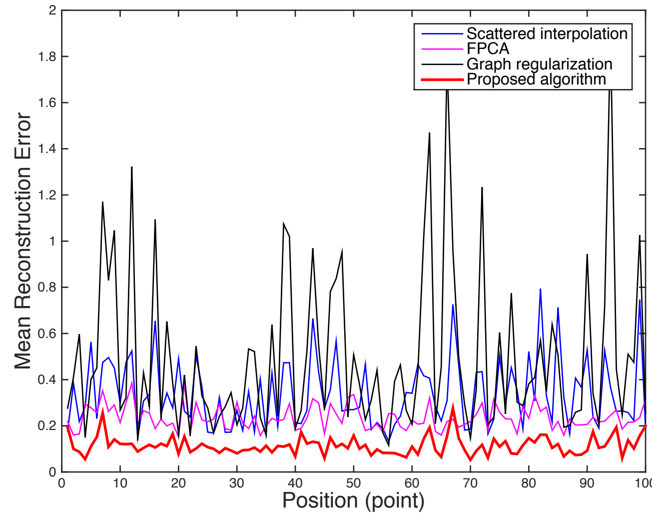

Fig. 3. The mean reconstruction error of each point.

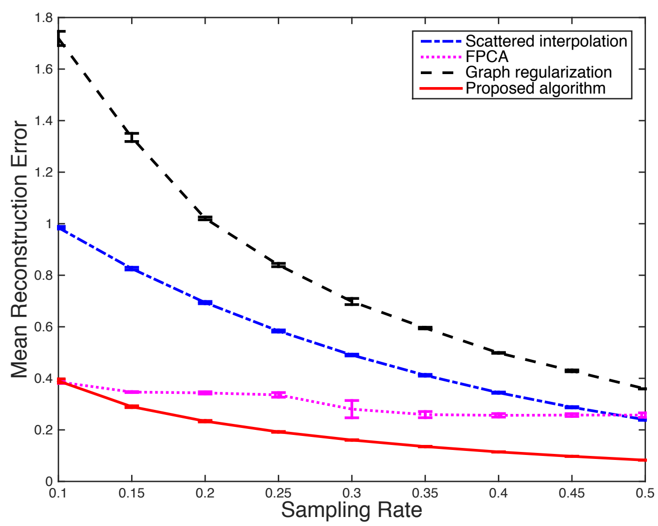

Fig. 4. The mean reconstruction error under different sampling rate.

A. The sea surface temperature dataset

The sea surface temperature data is collected monthly from 1870 to 2014, and the spatial resolution is 1 degree latitude by 1degree longitude global grid. We randomly select 100 points on the Pacific for simulation, with the time length of 600 months. The selected data range from -0.01 degree to 30.72 degree, and the mean temperature is 19.15 degree. Besides, we also have the latitude and longitude position of the selected 100 points.

Intuitively, these data are related to each other more or less. For example, usually the temperature difference of two close points is relatively small. Thus it is reasonable to construct a graph to describe the relation of these points. In the experiment, the graph is established by 5-nearest neighbors, and the weights are inversely proportional to the square of geodesic distance, as shown in Figure 1.

When the sampling rate is 40%, the MAE of the comparison

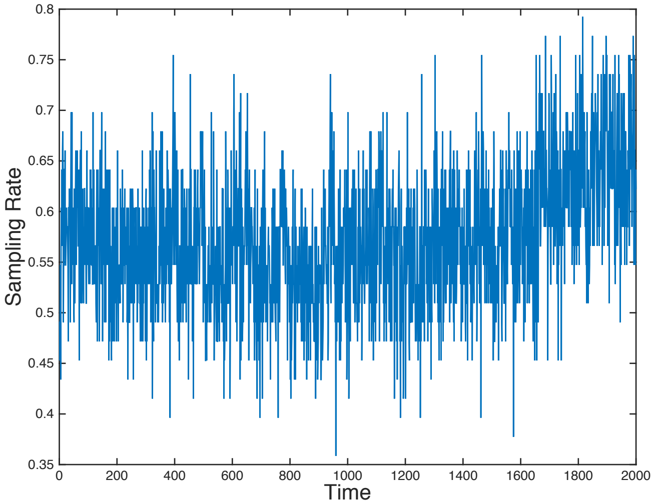

Fig. 5. Sampling rate of the sensor network data.

algorithms scattered data interpolation, FPCA, graph regularization, and the proposed algorithm are 0.3390, 0.23480, 0.4970 and 0.1146, respectively. Figure 2 is the MAE of each time, and Figure 3 is the MAE of each position. As can be seen, the performance the proposed algorithm is better than the other methods. There are two reasons for the improvement of the proposed algorithm. Firstly, we construct a graph to describe the relation of the points, which exploits the spatial correlation adequately. Secondly, we utilize the smoothness of the temporal differential signals on the graph, which is reasonable for real-world time-varying spatial signals.

Figure 4 gives the MAE of the overall signals under different sampling rate. The experiment has been run 100 times under each sampling rate. The error bars in the figure represent standard deviation of the MAE. We could find that the MAE of all algorithms drop as the sampling rate becomes higher. Meanwhile, the proposed algorithm is always superior to the others under sampling rate ranging from 0.1 to 0.5.

B. The sensor network dataset

The sensor network data is collected from 54 sensors distributed around the laboratory [17] every 30 seconds from February 28th, 2004, and the section between 01:06 and 17:56 on February 28th, 2004 is selected for simulation. The dimension of the selected data is 53×2000, excluding one fully damaged sensor. The graph is established by 4-nearest neighbors, with the weights inversely proportional to the square of geometric distance.

There is missing data due to various reasons such as sensor malfunctions and human errors. The missing values are called unsampled points. Figure 5 shows the real sampling rate at each time.

The reconstructed temperatures of the 14th sensor by the three comparison algorithms and the proposed algorithm are listed in Figure 6, where the circles denote the sampled points. The recovery of scattered interpolation coincides well with the sampled points, but fluctuates greatly at the unsampled points. On the contrary, the recovery of FPCA presents less fluctuation on the whole and large deviation with the sampled points. Because it does not utilize the temporal correlation of signal, the result of graph regularization is not smooth at all. In contrast, the result of the proposed algorithm demonstrates very good smoothness along time and coincides well with the sampled points.

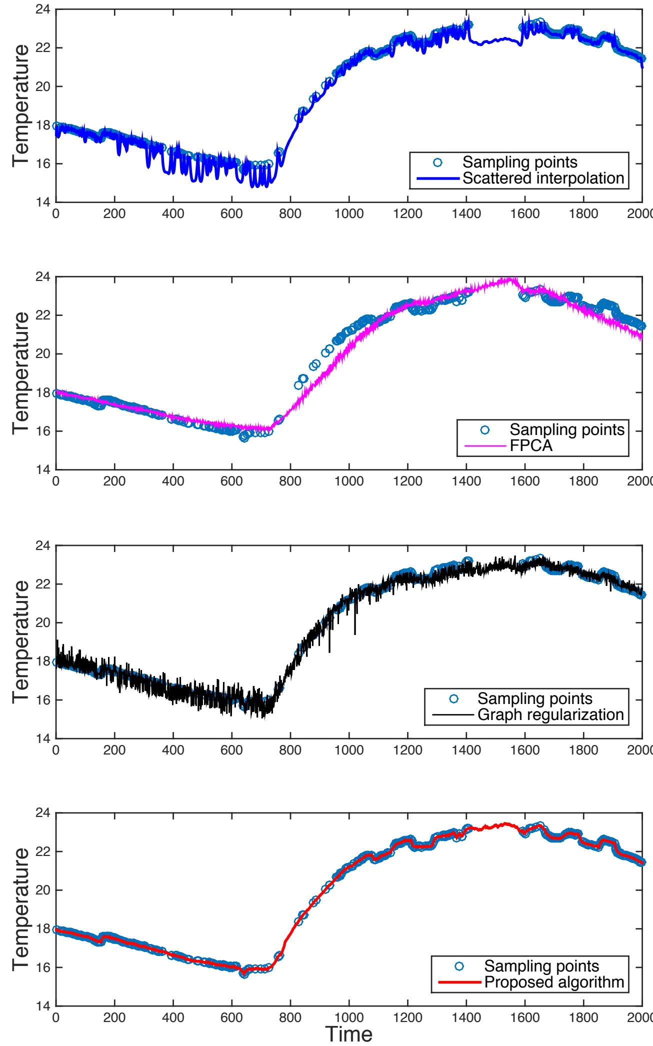

Fig. 6. The reconstructed temperature of one sensor.

V. Conclusion

In this article, we propose a novel method for the recovery of time-varying spatial signal based on graph. A graph is established according to the spatial position of the signal, and the smoothness of the temporal differential signals on the graph is used as the prior knowledge to help reconstruction. Experiments demonstrate that the proposed algorithm achieves less reconstruction error compared with other algorithms.

REFERENCES

[1] H. F. Yu, N. Rao, and I. S. Dhillon, "High-dimensional Time Series Prediction with Missing Values," *arXiv preprint arXiv:1509.08333*, 2015.

[2] O. Anava, E. Hazan, and A. Zeevi, "Online Time Series Prediction with Missing Data," *Proceedings of the 32nd International Conference on Machine Learning (ICML-15)*, pp. 2191-2199, 2015.

[3] K. Anjyo, J. P. Lewis, and F. Pighin, "Scattered data interpolation for computer graphics," *ACM SIGGRAPH 2014 Courses*, pp. 27, 2014.

[4] S. Robert, "A brief description of natural neighbor interpolation," *Interpreting Multivariate Data, chapter 2*, pp. 21-36, 1981.

[5] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276-280, 1986.

[6] V. Kalofolias, X. Bresson, M. Bronstein, and P. Vandergheynst, "Matrix completion on graphs," *arXiv preprint arXiv:1408.1717*, 2014.

[7] E. Candes and B. Recht, "Exact matrix completion via convex optimization," *FCM*, vol. 9, no. 6, pp. 717-772, 2009.

[8] L. Chen and Y. Gu, "Robust recovery of low-rank matrices via non-convex optimization," *International Conference on Digital Signal Processing (DSP)*, pp. 355-360, 2014.

[9] S. Ma, D. Goldfarb, and L. Chen, "Fixed point and Bregman iterative methods for matrix rank minimization," *Mathematical Programming*, vol. 128, no. 1, pp. 321-353, 2011.

[10] D. I. Shuman, S. K. Narang, P. Frossard, and A. Ortega, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 83-98, 2013.

[11] X. Wang, M. Wang, and Y. Gu, "A Distributed Tracking Algorithm for Reconstruction of Graph Signals," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 4, pp. 728-740, 2015.

[12] X. Wang, P. Liu, and Y. Gu Y, "Local-set-based graph signal reconstruction," *IEEE Transactions on Signal Processing*, vol. 63, no. 9, pp. 2432-2444, 2015.

[13] S. Chen, et al, "Discrete Signal Processing on Graphs: Sampling Theory," *IEEE Transactions on Signal Processing*, vol. 63, no. 24, pp. 6510-6523, 2015.

[14] S. K. Narang, A. Gadde, and A. Ortega, "Signal processing techniques for interpolation in graph structured data," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5445-5449, 2013.

[15] S. K. Narang, A. Gadde, E. Sanou, and A. Ortega, "Localized iterative methods for interpolation in graph structured data," *Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 491-494, 2013.

[16] http://www.esrl.noaa.gov/psd/data/gridded/data.noaa.oisst.v2.html.

[17] http://www.cs.cmu.edu/ guestrin/Research/Data/.