SIGNAL PROCESSING TECHNIQUES FOR INTERPOLATION IN GRAPH STRUCTURED DATA

Sunil K Narang, Akshay Gadde and Antonio Ortega

Ming Hsieh Department of Electrical Engineering
University of Southern California
kumarsun@usc.edu, agadde@usc.edu,ortega@sipi.usc.edu

ABSTRACT

In this paper, we propose a novel algorithm to interpolate data defined on graphs, using signal processing concepts. The interpolation of missing values from known samples appears in various applications, such as matrix/vector completion, sampling of high-dimensional data, semi-supervised learning etc. In this paper, we formulate the data interpolation problem as a signal reconstruction problem on a graph, where a *graph signal* is defined as the information attached to each node (scalar or vector values mapped to the set of vertices/edges of the graph). We use recent results for sampling in graphs to find classes of bandlimited (BL) graph signals that can be reconstructed from their partially observed samples. The interpolated signal is obtained by projecting the input signal into the appropriate BL graph signal space. Additionally, we impose a 'bilateral' weighting scheme on the links between known samples, which further improves accuracy. We use our proposed method for collaborative filtering in recommendation systems. Preliminary results show a very favorable trade-off between accuracy and complexity, compared to state of the art algorithms.

***Index Terms*—** Graph signal processing, sampling in graphs, spectral graph theory, recommendation systems

1. INTRODUCTION

Graphs are a natural tool to represent data in many domains such as protein interaction networks [1], recommendation systems [2] and social networks [3]. A graph, represented as $G = (\mathcal{V}, E)$, consists of a set of nodes $\mathcal{V}$ and a set of links E connecting these nodes. The data on the graphs is often represented as a scalar or vector valued function attached to the vertices of the graph (see [4] for a comprehensive review). An important area of research is the interpolation problem in graph structured data. This arises in many guises, such as in semi-supervised learning of categorical data (see [5]), ranking problems [6], and missing value prediction such as matrix-completion problems [7]. A common theme in all these applications is that the goal is to predict the property of some nodes (class, ranking or function), by interpolating the property values from a known set of nodes. The accuracy of all linear and non-linear interpolation methods on graphs rely on the implicit assumption that nodes close to each other (in terms of the similarity captured by link-weights in the graph) would usually have similar signal values. For example, in an item-item graph in a recommendation system, a typical user would rate two similar items with similar ratings. In the same way, when predicting the functions of unannotated proteins based on a protein network, one relies on some notions of 'closeness' or 'distance' among the nodes. In other words, the graph functions of interest are slowly-varying or "smooth" on the graph. Smoothness is well studied in the classical signal processing domain, where it is measured in terms of the frequency contents of a signal. For example, there are well known sampling results for recovering smooth bandlimited signals from only a few of their samples via interpolation. In this paper, we focus on the development of similar interpolation techniques for graph signals.

The main objective in designing an interpolation method is of course achieving high accuracy. A major challenge in achieving this goal is managing to do so with reasonable computational complexity. Graph based interpolation approaches can be broadly divided into two categories: a) local methods and b) global methods. In local interpolation methods, such as kNN methods, the predicted value at an unknown node is computed as a weighted combination of k-nearest known samples [8]. Although the method is computationally simple and efficient, it does not capture global information, as well as the dependencies that exist between known samples. On the other hand, global methods, such as [9, 10], predict the value of all unknown nodes at once, by selecting as solution a function that matches the values at known nodes while satisfying certain global "smoothness" conditions. Global methods are computationally more expensive but provide more accurate results. However, it is not clear how to optimize the choice of objective function. Although our approach leads to a similar least-squares based interpolation as in [9], the signal processing perspective allows us to choose an optimal objective criterion, which will be shown to lead to better interpolation. Another problem is the choice of graph to interpolate data upon. The example in Figure 1 demonstrates the choice of either operating on a bare-bones star graph or a complete graph with all connections between movies. Interpolation can be defined on both graphs, so it is unclear which of the two (or any other graph) should be chosen.

In this paper, we take inspiration from signal processing techniques to formulate the partially known graph function as a downsampled-upsampled (DU) signal. In the regular signal domain, the original signal is recovered from its DU signal by applying a low-pass filter. Similarly, in the graph domain, we design low-pass filters to recover the original graph signal from its DU signal. The proposed approach is based on recent results for sampling in bandlimited graph signals by Pesenson [11] and our own prior work in [12]. In this approach, given a graph and the set of nodes for which the data is known we compute an optimal cut-off frequency, such that the reconstruction is exact if the original graph signal is bandlimited to this frequency, and a stable reconstruction is obtained for non-bandlimited signals. The errors can be further minimized by increasing smoothness of the function using a "bilateral- weighting"

This work was supported in part by NSF under grant CCF-1018977.

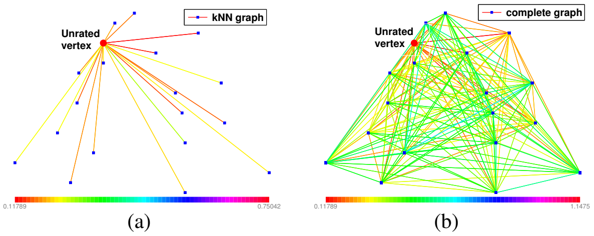

Fig. 1: An instance of predicting ratings of an unknown movie node (in red) using ratings of a known set of movie nodes (in blue), in MovieLens $100k$ dataset: (a) star graph commonly used in kNN [8] prediction methods, which ignores all the links between known movie nodes. (b) alternative graph that contains the star graphs and all the links between movies in the known set of movies.

step on the underlying graphs. To the best of our knowledge, this is a novel perspective for addressing graph based interpolation problems, and our preliminary results indicate a promising advantage over existing methods. The rest of the paper is organized as follows: in Section 2, we introduce the basic theory needed to understand the proposed work, in Section 3, we describe our proposed interpolation method. In Section 4, we demonstrate our proposed method on a real movie dataset, and finally in Section 5, we conclude and describe future work.

2. BASIC THEORY

2.1. Notations

An undirected simple graph $G = (\mathcal{V}, E)$ is a collection of nodes $\mathcal{V} = \{1, 2, ...N\}$ connected together by set of links $E = \{(i, j, w_{ij})\}$, $i, j \in \mathcal{V}$ where (i, j, w_{ij}) denotes the link of weight w_{ij} between node i and j. The adjacency matrix $\mathbf{W}$ of the graph is an $N \times N$ matrix such that $W(i, j) = w_{ij}$, the degree d_i a node i is the sum of link-weights connected to node i, the degree matrix $\mathbf{D} = diag\{d_1, d_2, ..., d_N\}$ is a diagonal matrix, and the combinatorial Laplacian matrix is $\mathbf{L} = \mathbf{D} - \mathbf{W}$. We use the normalized form of adjacency matrix $\boldsymbol{\mathcal{W}} = \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}$ and Laplacian matrix $\boldsymbol{\mathcal{L}} = \mathbf{D}^{-1/2}\mathbf{L}\mathbf{D}^{-1/2} = \mathbf{I} - \boldsymbol{\mathcal{W}}$. $\boldsymbol{\mathcal{L}}$ is a positive semi-definite matrix, and has an orthogonal set of eigenvectors denoted $\mathbf{U} = \{\mathbf{u}_1, \mathbf{u}_2, ...\mathbf{u}_n\}$ corresponding to eigenvalues $\sigma(G) = \{\lambda_1, \lambda_2, ..., \lambda_N\}$, respectively. We denote $< \mathbf{a},\ \mathbf{b} >= \mathbf{a}^t\mathbf{b}$ as the inner-product of vectors $\mathbf{a}$ and $\mathbf{b}$, where $\mathbf{a}^t$ is the transpose of $\mathbf{a}$, and $\mathcal{S}^c = \mathcal{V} - \mathcal{S}$, the complement set of $\mathcal{S}$ on graph G. Define $(\mathbf{A})_{\mathcal{S}_1,\mathcal{S}_2}$ to be the submatrix of any matrix $\mathbf{A}$ corresponding to rows corresponding to $\mathcal{S}_1$ and columns corresponding to $\mathcal{S}_2$. Simply denote $(\mathbf{A})_{\mathcal{S},\mathcal{S}} = (\mathbf{A})_{\mathcal{S}}$ for brevity. We denote $\mathbf{0}$ and $\mathbf{1}$ as the all zero and all one functions of size N. In order to use signal processing tools, we define a graph function as a scalar valued discrete signal $f : \mathcal{V} \to \mathbb{R}$, such that $f(i)$ is the value of the sample at node i. The downsampling-upsampling (DU) operation on the graph signal is defined as discarding the values of signal not on a selected subset of nodes (say $\mathcal{S}$) and replacing them with 0. By an appropriate permutation, let us choose an indexing such that $\mathcal{S} = \{1, 2, ...|\mathcal{S}|\}$, and $\mathcal{S}^c = \{|\mathcal{S}| + 1, |\mathcal{S}| + 2, ..., N\}$. Then the DU signal is denoted as $\mathbf{f}_\mathcal{S} = [\mathbf{f}(\mathcal{S})^t \ \ \mathbf{0}(\mathcal{S}^c)^t]^t$.

Similar to classical Fourier transform, the eigenvectors and eigenvalues of the Laplacian matrix $\boldsymbol{\mathcal{L}}$ provide a spectral interpretation of the graph signals. The eigenvalues $\{\lambda_1, \lambda_2, ..., \lambda_N\}$ can be treated as graph frequencies, and are always situated in the interval $[0, 2]$ on the real line. The eigenvectors of the Laplacian matrix demonstrate increasing oscillatory behavior as the magnitude of the graph frequency increases (see [13] for details). The *graph Fourier transform* (GFT) of a signal $\mathbf{f}$ is defined as its projection onto the eigenvectors of the graph, i.e., $\tilde{f}(\lambda_i) =< \mathbf{f},\ \mathbf{u}_i >$, or in matrix form $\tilde{\mathbf{f}} = \mathbf{U}^t\mathbf{f}$.

2.2. Sampling Theory for Graph Signals

We start by revisiting the theory of downsampling graph signals in the recent work by Pesenson [11], as well as its links to our prior work in [12]. A signal is said to be bandlimited to the graph frequency band $[0, \omega)$ on a graph G, if its GFT has support only at frequencies $[0, \omega)$. The space of ω-bandlimited signals is called the *Paley-Wiener space* and is denoted by $PW_\omega(G)$. It is easy to prove that if $\mathbf{f} \in PW_\omega(G)$, then:

$$\|\boldsymbol{\mathcal{L}}\mathbf{f}\| \leq \omega\|\mathbf{f}\| \qquad (1)$$

where $\|.\|$ denotes the l_2 norm. The theory presented by Pesenson [11], describes a method to reconstruct any signal in $PW_\omega(G)$ space from a special subset of nodes called the *uniqueness set*. This is defined as follows:

Definition 1 (Uniqueness Set) *A subset of nodes $\mathcal{S} \subset \mathcal{V}$ is a uniqueness set for a space $PW_\omega(G), \omega > 0$, if for any two signals from $PW_\omega(G)$, the fact that they coincide on $\mathcal{S}$ implies that they coincide on $\mathcal{V}$.*

The above definition implies that it is *sufficient* to know the value of a ω-bandlimited graph signal only on the uniqueness set $\mathcal{S}$. The uniqueness set is in turn found by finding its complement set $\mathcal{S}^c$, which is a Λ-set with $\Lambda = 1/\omega$. A Λ*-set* is defined as follows:

Definition 2 (Λ-set) *A subset of nodes $\mathcal{Q} \subset \mathcal{V}$ is a Λ-set if any $\phi \in L_2(\mathcal{Q})$ admits a Poincaré inequality with a constant $\Lambda > 0$, i.e.,*

$$\|\phi\| \leq \Lambda\|\boldsymbol{\mathcal{L}}\phi\|, \quad \phi \in L_2(\mathcal{Q}) \qquad (2)$$

where $L_2(\mathcal{Q})$ is the space of all graph signals that are zero everywhere except at the subset of nodes $\mathcal{Q} \subset \mathcal{V}$.

The downsampling results given by Pesenson [11, Theorem 1.2 and 4.1] are formally:

Theorem 1 *A set $\mathcal{S} \subset \mathcal{V}$ is a uniqueness set for any signal in space $PW_\omega(G)$, if its complement $\mathcal{S}^c$ is a Λ-set with $0 < \omega < 1/\Lambda$, which implies that there exists a frame $\{\Theta_s\}_{s\in\mathcal{S}}$ in the subspace $PW_\omega(G)$ such that the following reconstruction formula holds true for all signals $\mathbf{f} \in PW_\omega(G)$:*

$$\mathbf{f}(v) = \sum_{s\in\mathcal{S}} \mathbf{f}(s)\Theta_s(v), v \in \mathcal{V} \qquad (3)$$

Theorem 1 provides a sufficient condition for reconstructing any signal in $PW_\omega(G)$ from its DU signal. The reconstruction method is described in Section 3. A corollary of Theorem 1 provides the uniqueness set for the special case when $\omega = 1$ [11, Lemma 3.5].

Corollary 1 *Any vertex cover* [1] *$\mathcal{S} \subset \mathcal{V}$ of the graph G is a uniqueness set for all signals $\mathbf{f} \in PW_\omega(G)$ with $\omega = 1$.*

[1] A vertex cover for a graph G is a subset $\mathcal{S} \subset \mathcal{V}$ such that each edge of the graph is incident on at least one vertex in $\mathcal{S}$.

This result is also consistent with the result in our prior work on bipartite graphs in [12, 14]. *Thus, an interpolation on a graph can be posed as the problem of first defining the set of nodes with known sample values as a uniqueness set $\mathcal{S}$, then identifying the maximum ω such that $\mathcal{S}$ is a uniqueness set for signals in $PW_\omega(G)$, and then reconstructing the signal values on the complement set $\mathcal{S}^c$ by using (3).*

While Pesenson's work proves the existence of such ω, it does not provide a method to compute it, and only considers how to interpolate unknown values (in $\mathcal{S}^c$) in the case when the signal is ω-bandlimited. In the next section, we provide a result to compute maximum ω given known set $\mathcal{S}$, as well as interpolation in case of non-bandlmited signals.

3. PROPOSED INTERPOLATION

To make a practical use of Pesenson's result in Theorem 1, we first present a result to compute the maximum ω, such that any signal in $PW_\omega(G)$ can be reconstructed, given a subset of known samples in $\mathcal{S}$ on any arbitrary graph G.

Proposition 2 *Given a graph G with normalized Laplacian matrix $\boldsymbol{\mathcal{L}}$, known set $\mathcal{S}$ and unknown set $\mathcal{S}^c$, let $(\boldsymbol{\mathcal{L}}^2)_{\mathcal{S}^c}$ be the submatrix of $\boldsymbol{\mathcal{L}}^2$ containing only the rows and columns corresponding to unknown set $\mathcal{S}^c$. Then the known set $\mathcal{S}$ is a uniqueness set for all signals $\mathbf{f} \in PW_{\omega_\mathcal{S}}(G)$ with $\omega^*_\mathcal{S} = \sigma_{min}$, where σ^2_{min} is the smallest singular value of $(\boldsymbol{\mathcal{L}}^2)_{\mathcal{S}^c}$.*

Proof Referring to Definition 2, let ϕ be a signal in $L_2(\mathcal{S}^c)$, i.e., $\phi = [\mathbf{0}(\mathcal{S})^t \;\; \phi(\mathcal{S}^c)^t]^t$. We have

$$\frac{||\boldsymbol{\mathcal{L}}\phi||^2}{||\phi||^2} = \frac{< \phi(\mathcal{S}^c)\;(\boldsymbol{\mathcal{L}}^2)_{\mathcal{S}^c}\phi(\mathcal{S}^c) >}{||\phi(\mathcal{S}^c)||^2} = R(\boldsymbol{\mathcal{L}}^2_{\mathcal{S}^c}) \tag{4}$$

where $R(.)$ is the Rayleigh quotient of a matrix. It can be shown that $R((\boldsymbol{\mathcal{L}}^2)_{\mathcal{S}^c})$ is always greater than the minimum eigenvalue σ^2_{min} of the matrix $(\boldsymbol{\mathcal{L}}^2)_{\mathcal{S}^c}$. Thus,

$$\frac{||\boldsymbol{\mathcal{L}}\phi||^2}{||\phi||^2} \geq \sigma^2_{min} \Rightarrow ||\phi|| \leq \frac{1}{\sigma_{min}}||\boldsymbol{\mathcal{L}}\phi|| \tag{5}$$

Thus, $\mathcal{S}^c$ is a Λ-set with $\Lambda = \frac{1}{\sigma_{min}}$. By Theorem 1, $\mathcal{S}$ is a uniqueness set for all signals $\mathbf{f} \in PW_\omega(G)$ with $\omega^*_\mathcal{S} = \sigma_{min}$.

Note that, $\omega^*_\mathcal{S}$ computed above is the maximum possible value that satisfies the sufficient conditions in Theorem 1. We term $\omega^*_\mathcal{S}$ as the *cut-off frequency* for reconstruction, since any graph-signal below this frequency can be perfectly reconstructed from its DU signal on $\mathcal{S}$.

3.1. Interpolation method

Given the cut-off frequency $\omega^*_\mathcal{S}$ computed from Proposition 2, the reconstruction of any graph signal is done by least-square projection of the corresponding DU signal onto the $PW_\omega(G)$ space. Let K^* be the number of eigenvalues of the Laplacian matrix $\boldsymbol{\mathcal{L}}$, less than $\omega^*_\mathcal{S}$. Define $\mathbf{U}_{K^*}$ as the matrix containing first K^* eigenvectors, and $(\mathbf{U}_{K^*})_\mathcal{S}$ as the submatrix of $\mathbf{U}_{K^*}$ containing rows corresponding to set $\mathcal{S}$. The first eigenvector of a connected graph is $\mathbf{u}_1 = \mathbf{D}^{1/2}\mathbf{1}$, which is not constant for graphs with irregular degrees. Therefore, for piecewise constant signals, a common practice is to find LS approximation of function $\mathbf{g} = \mathbf{D}^{1/2}\mathbf{f}$. Thus, we want to compute interpolated signal $\hat{\mathbf{g}}$ such that $(\hat{\mathbf{g}})_\mathcal{S} = (\mathbf{g})_\mathcal{S}$ and:

$$\hat{\mathbf{g}} = \sum_{k=1}^{K_\Lambda} x(k)\mathbf{u}_{\lambda_k} = \mathbf{U}_{K^*}\mathbf{x}, \tag{6}$$

where $x(k) = \tilde{g}(\lambda_k)$ is the k^{th} GFT coefficient of $\mathbf{g}$, and $\mathbf{x} = [x_1, x_2, ...x_{K_\Lambda}]$. Comparing only the known set $\mathcal{S}$ on both sides in (6), we obtain a linear system of equations: $\mathbf{g}(\mathcal{S}) = (\mathbf{U}_{K^*})_\mathcal{S}\mathbf{x}$. Theorem 1 ensures that $(\mathbf{U}_{K^*})_\mathcal{S}$ *is a stable frame operator*, and the solution can be found by computing the pseudo inverse of $(\mathbf{U}_{K^*})_\mathcal{S}$. Thus, the interpolated graph signal on the unknown set $\mathcal{S}^c$ is given by

$$\mathbf{g}^*(\mathcal{S}^c) = (\mathbf{U}_{K^*})_{\mathcal{S}^c}\left(((\mathbf{U}_{K^*})_\mathcal{S})^t(\mathbf{U}_{K^*})_\mathcal{S}\right)^{-1}((\mathbf{U}_{K^*})_\mathcal{S})^t\mathbf{f}(\mathcal{S}). \tag{7}$$

Finally, the interpolated signal is computed as: $\hat{\mathbf{f}} = \mathbf{D}^{-1/2}\hat{\mathbf{g}}$.

Note that, while other global methods such as [9, 10] also propose similar least-square reconstruction solutions, the choice of number of eigenvectors K in these methods is heuristic. In our proposed method K^* is chosen specifically to be the number of eigenvalues below the cut-off frequency $\omega^*_\mathcal{S}$ given in Proposition 2, which depends on the known set of nodes, and the topology of the graph. The optimality of choosing $\omega^*_\mathcal{S}$ as cut-off frequency can be justified as follows:

Let $\omega_\mathcal{S}$ is the chosen cut-off frequency. If signal $\mathbf{f} \in PW_{\omega_\mathcal{S}}(G)$, then the proposed reconstruction is perfect (loss-less), hence optimal. For $\mathbf{f} \notin PW_{\omega_\mathcal{S}}(G)$, our proposed method still provides a stable least-square solution. The solutions with $\omega_\mathcal{S} < \omega^*_\mathcal{S}$ are clearly suboptimal in this case since the solution space of $\omega_\mathcal{S}$ is contained in the solution space of $\omega^*_\mathcal{S}$. For $\omega_\mathcal{S} > \omega^*_\mathcal{S}$, the reconstruction may sometimes produce less error, but it is not guaranteed to be stable (i.e., matrix $(\mathbf{U}_{K^*})_\mathcal{S}$ might not be a frame). In Section 4, we show that choosing $\omega_\mathcal{S} > \omega^*_\mathcal{S}$ leads to poorer results.

3.2. Bilateral Link-weight Adjustment

Unlike regular signals, the smoothness of a graph signal depends both on the signal values and the underlying graph. This leads to the question of whether we can modify the graph to adapt to the given signal so that the signal is more band-limited on the simplified graph, thus leading to less interpolation error. The simplification makes sense in many cases such as in recommendation systems, where the underlying graph is the result of observing average correlation over a set of training users (multiple instances), and the signal corresponds to a single test user. We take inspiration from image processing where this kind of signal adaptive filtering is achieved by *bilateral filters* [15].

In our proposed method we use $\mathbf{g} = \mathbf{D}^{1/2}\mathbf{f}$ as the signal to be interpolated. From (1), we observe that $\mathbf{g}$ can be made more bandlimited by minimizing $||\boldsymbol{\mathcal{L}}\mathbf{g}||$. Define an error function:

$$\boldsymbol{\zeta} = \boldsymbol{\mathcal{L}}\mathbf{g} = (\mathbf{I} - \boldsymbol{\mathcal{W}})\mathbf{g} = \mathbf{D}^{1/2}(\mathbf{I} - \mathbf{D}^{-1}\mathbf{W})\mathbf{f}. \tag{8}$$

Clearly, minimizing $\boldsymbol{\zeta}$ at each node minimizes $||\boldsymbol{\mathcal{L}}\mathbf{g}||$. The value of $\boldsymbol{\zeta}$ at node i can be written as:

$$\zeta(i) = \sqrt{d_i}\left(f(i) - \frac{1}{d_i}\sum_j w_{ij}f(j)\right), \tag{9}$$

which is proportional to the difference of $f(i)$ with the weighted average (the weight being link-weights) of nodes directly connected to node i. Thus, by adapting the weights w_{ij} to be inversely proportional to the absolute difference $|(f(i) - f(j)|$ at every node, we can minimize the error $\boldsymbol{\zeta}$, and hence $||\boldsymbol{\mathcal{L}}\mathbf{g}||$. Following similar intuition as in bilateral filters, we modify the weights between nodes in $\mathcal{S}$ as:

$$\hat{w}_{ij} = w_{ij}.exp(-\frac{|f(i) - f(j)|^2}{\theta^2}) \ \forall \{i, j \in \mathcal{S}\}, \qquad (10)$$

where parameter θ is chosen to be the mean rating over all training samples. Note that in (10), we can only change the weights between two known nodes. Our proposed interpolation algorithm is given in Algorithm 1. The complexity of our method is $\mathcal{O}(K^*|\mathcal{V}|^2)$, primarily due to the least-square projection steps $(5 - 7)$ in Algorithm 1.

Algorithm 1 Proposed Graph Interpolation Method

```
Require: $G_0 = (\mathcal{V}, E)$: Initial Graph, $\mathbf{f}$: DU signal
1: Compute normalized Laplacian matrix $\boldsymbol{\mathcal{L}}$.
2: Compute $\omega^*_{\mathcal{S}}$ as the square-root of the smallest eigenvalue of $(\boldsymbol{\mathcal{L}}^2)_{\mathcal{S}^c}$.
3: Modify known nodes' link-weights as bilateral weights using (10).
4: Recompute normalized Laplacian matrix $\boldsymbol{\mathcal{L}}$.
5: Compute $K^*$ eigenvectors of $\boldsymbol{\mathcal{L}}$ corresponding to eigenvalues $\lambda < \omega^*_{\mathcal{S}}$.
6: Compute $\mathbf{g} = \mathbf{D}^{1/2}\mathbf{f}$.
7: Compute $\hat{\mathbf{g}}$: $(\hat{\mathbf{g}})_{\mathcal{S}} = (\mathbf{g})_{\mathcal{S}}$ and $(\hat{\mathbf{g}})_{\mathcal{S}^c}$ using (7).
8: Interpolated signal $\hat{\mathbf{f}} = \mathbf{D}^{-1/2}\hat{\mathbf{g}}$.
```

4. EXPERIMENTS

We apply proposed interpolation method for collaborative filtering in recommendation systems. The input in this problem is a partially observed user-item rating matrix $\mathbf{R}$, such that $\mathbf{R}(u, m)$ is the rating given by user u to the movie m. Based on this information, the system predicts new user-movie ratings. For empirical evaluation, we choose the MovieLens $100k$ [16] dataset containing $100k$ user-movie-rating triplets from $N = 943$ users and $M = 1682$ movies. The ratings are integer values between 1 and 5. We use the 5-fold cross validation data available in [16], which consists of 5 disjoint random sets of $20k$ triplets. At each iteration, one of these sets is used for testing and remaining sets for training.

We first compute a movie graph G_0 from the training samples, by computing cosine similarity [17] between every pair of movies. For each test user u, we define $\mathcal{S}$ to be the set of movies with known ratings, and and $\mathcal{U}$ to be the set of test movies. We compute the subgraph $G_u = (\mathcal{S} \cup \mathcal{U}, E_u)$, corresponding to subset $\mathcal{S} \cup \mathcal{U}$ of nodes. We define DU signal $\mathbf{f}_u$ for u to be of size $|\mathcal{U} \cup \mathcal{S}|$, with $\mathbf{f}_u(\mathcal{U}) = 0$ and $\mathbf{f}_u(\mathcal{S})$ equal to known ratings. Subsequently, we compute interpolated signal $\hat{\mathbf{f}}_u$ by using proposed method given in Algorithm 1, with G_u and $\mathbf{f}_u$ as inputs. In our preliminary analysis, we compare the performance of our proposed method with two most popular algorithms for collaborative filtering: 1) kNN method [17] (with $k = 30$), and 2) probabilistic matrix factorization (PMF) [18] with 10 latent features. Our proposed method is similar to the least square (LS) interpolation of [9], except that we operate based on normalized Laplacian matrix, and choose K^* specified by $\omega^*_{\mathcal{S}}$. To show that this K is a good choice, we also implement the method in [9] with $K = K^* + 10$ (i.e., with 10 additional eigenvectors), respectively. We use RMSE between predicted values and actual values to measure performance. The predicted values less than 0 (more than 5) are set to 0 (5) before computing RMSE. Figure 2 plots the RMSE of prediction as a function of number of training samples available. Figure 2(a), shows cumulative RMSE of various implemented methods. Observe that kNN method perform the worst of all method, and PMF method performs the best. Our proposed methods both with or without bilateral weighting are very close to PMF method, with the interpolation with bilateral weights performing slightly better. We also observed that choosing $K = K^* + 10$ in LS method leads to poorer results. The Figure 2(b) shows another RMSE plot where users are grouped by the number of training samples, with x-axis showing those groups. We observe that both the LS method and kNN method perform significantly worse when the number of available training samples are small. The effect of applying bilateral weighting in our proposed methods is also most visible here.

The PMF method predicts the ratings of all movies for all users simultaneously by factorizing the whole $N \times M$ rating matrix. It is based on an iterative update rule and requires $\mathcal{O}(NMP)$ operations per iteration where P is the size of the latent space. Theoretically, any change to the rating matrix would require the PMF system to be retrained on all users. However, in our method, the process of computing the movie graph is decoupled from the process of predicting ratings for a given user. Accommodating a few new ratings into the systems is fast as it only affects a local portion of the movie graph. Once the movie graph is fixed, the proposed method allows us to predict the ratings of movies for each user separately in $\mathcal{O}(K^*M^2)$ operations. Thus, assuming $K^* \approx P$, the proposed method is faster $(i.e., M^2 < NM)$ than PMF, when ratings of items change frequently and the recommendations need to be calculated only when a user request them. Further , it may be possible to reduce complexity in our method by using simplified filtering operations.

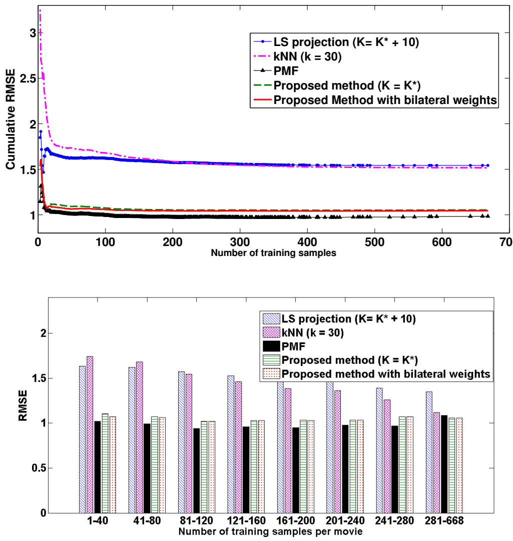

Fig. 2: RMSE of different prediction algorithms with the number of training samples on MovieLens dataset .

5. CONCLUSIONS AND FUTURE WORK

We proposed a novel method for interpolation of graph structured data using signal processing tools. We formulated the interpolation problem as reconstruction of a graph-signal from its DU sampled signal. The success of this method in preliminary experiments opens up many new opportunities and challenges. The future work includes improving interpolation accuracy by a) finding out both necessary and sufficient conditions for reconstruction (Theorem 1 is only a sufficient condition), and b) understanding and formulating bilateral weighting step as an optimization problem.

6. REFERENCES

[1] M. Weber and S. Kube, "Robust Perron cluster analysis for various applications in computational life science," in *CompLife*, 2005, pp. 57–66.

[2] Z. Huang, W. Chung, T. Ong, and H. Chen, "A graph-based recommender system for digital library," pp. 65–73, 2002.

[3] M. Girvan and M. E. Newman, "Community structure in social and biological networks.," *Proc Natl Acad Sci U S A*, June 2002.

[4] D. I Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "Signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular data domains," *to appear SPM 2012, also in arXiv'12*, , no. arXiv:1211.0053, Dec. 2012.

[5] R.I. Kondor and J. Lafferty, "Diffusion kernels on graphs and other discrete structures," in *Proc. ICML*, 2002, pp. 315–322.

[6] S. Hoche, P. Flach, and D. Hardcastle, "A fast method for property prediction in graph-structured data from positive and unlabelled examples," in *ECAI 2008*, 2008, pp. 162–166.

[7] J. Bennett, S. Lanning, and N. Netflix, "The netflix prize," in *In KDD Cup and Workshop in conjunction with KDD*, 2007.

[8] J. Chen, H. Fang, and Y. Saad, "Fast approximate KNN graph construction for high dimensional data via recursive lanczos bisection," *J. Mach. Learn. Res.*, vol. 10, pp. 1989–2012, dec 2009.

[9] M. Belkin and P. Niyogi, "Semi-supervised learning on riemannian manifolds," pp. 209–239, 2004.

[10] L. Grady and E. L. Schwartz, "Anisotropic interpolation on graphs: The combinatorial dirichlet problem," Tech. Rep., Boston University, 2003.

[11] I. Pesenson, "Sampling in Paley-Wiener spaces on combinatorial graphs," *Trans. Amer. Math. Soc*, vol. 360, no. 10, pp. 5603–5627, 2008.

[12] S.K. Narang and A. Ortega, "Downsampling graphs using spectral theory," in *ICASSP '11.*, May 2011.

[13] E. B. Davies, G. M. L. Gladwell, J. Leydold, and P. F. Stadler, "Discrete nodal domain theorems," *Linear Algebra and its Applications*, vol. 336, no. 1-3, pp. 51 – 60, 2001.

[14] S.K. Narang and Ortega A., "Perfect reconstruction two-channel wavelet filter-banks for graph structured data," *IEEE trans. on Sig. Proc. 2012*, vol. 60, no. 6, June 2012.

[15] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Sixth International Conference on Computer Vision*, Jan 1998, pp. 839–846.

[16] "MovieLens dataset," as of 2003, `http://www.grouplens.org/`.

[17] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," in *WWW '01*, 2001, pp. 285–295.

[18] Ruslan Salakhutdinov and Andriy Mnih, "Probabilistic matrix factorization," in *Advances in Neural Information Processing Systems*, 2008, vol. 20.