MATRIX COMPLETION AS GRAPH BANDLIMITED RECONSTRUCTION

Weiyu Huang[†]*, Antonio G. Marques*[‡] *, and Alejandro Ribeiro*[†]

ABSTRACT

This paper develops new designs for recommender systems inspired by recent advances in graph signal processing. Recommender systems aim to predict unknown ratings by exploiting the information revealed in a subset of user-item observed ratings. Leveraging the notions of graph frequency and graph filters, we demonstrate that linear latent factor models, such as low-rank matrix completion, can be viewed as *bandlimited* interpolation algorithms that operate in a frequency domain given by the spectrum of a joint user and item network. This new interpretation paves the way to new methods for enhanced rating prediction. We propose a low complexity method by exploiting the eigenvector of correlation matrices constructed from known ratings. In the MovieLens 100k dataset, our designs reduce the root mean squared error compared to the ones in benchmark matrix completion by 0.6% and benchmark nearest neighbor methods by 4.2%.

Index Terms— Collaborative filtering, recommender systems, graph signal processing, graph filters, matrix completion

1. INTRODUCTION

The widespread deployment of the Internet technologies has generated a massive enrollment of online customers in web services, propelling the need for recommender systems (RS) to assist customers in making decisions. In a succinct way, RS are algorithms that collect information about how users of a particular service rate items. The collected information is then exploited, along with additional sources of exogenous information, to provide customers with recommendations for the unrated items [2, 3].

Research on RS includes the so-called content filtering approach, which starts by defining a set of features that characterize users and items and then uses those to perform predictions [2, 3]. It also includes the collaborative filtering (CoFi) approach, which relies mostly on past user behavior and carries out predictions without defining features. Although CoFi comes with certain disadvantages, it typically requires less assumptions and yields a superior performance in real datasets [3]. As a result, it has emerged as the central approach for RS. A common technique to design CoFi algorithm is latent linear factor models (LLFM), which tries to model the rating user-item function by identifying a number of latent factors. Although nonlinear latent factor models exist, the linear models based on matrix factorization [4] combine tractability with good practical performance [3]. In particular, by arranging the available ratings in a matrix form with one of the dimensions corresponding to users and the other one to items, LLFM typically carries out low-rank singular value decomposition (SVD) that jointly maps users and items to latent factor linear space of reduced dimensionality [3]. The rating user-item function is then modeled simply as inner products in the reduced subspace defined by the SVD.

The goal in this paper is to reinterpret CoFi algorithms using tools from graph signal processing (SP). In simple words, graph SP addresses the problem of analyzing and extracting information from data defined on irregular domains that can be represented by a graph. The tacit assumption is that the network structure defines a notion of proximity or dependence among the nodes of the graph [5, 6]. The theory and application of graph SP is growing rapidly [7–19]. This paper designs new and more general recommendation schemes, but equally relevant unveils important connections between CoFi and graph SP. Graph SP has been used as an extra term to improve performance of CoFi [17, 18] without changing the scheme, while our work tries to reinterpret and provides new algorithms for CoFi. More precisely, we show that matrix factorization methods can be reinterpreted as interpolation algorithms that, given a subset of signal observations (ratings), recover the full signal under the assumption that the ratings are bandlimited in a particular spectral domain. This interpretation opens the door to the design of more advanced algorithms leading to a better accuracy. In a nutshell, the contributions of this paper are: (a) To demonstrate how the standard CoFi approach based on LLFM can be interpreted as particular types of graph SP algorithms that model the rating signal as bandlimited. (b) To exploit this interpretation to design more general algorithms for LLFM. (c) To show that the proposed methods indeed produce notable improvement regarding rating prediction accuracy in the publicly available MovieLens 100k dataset [20][1].

2. FUNDAMENTALS OF COFI AND GRAPH SP

Consider an RS setup with U users indexed by u, and I items indexed by i. The rating that user u has given to item i is represented as $X_{u,i}$. For mathematical convenience, such ratings can be collected either into the rating matrix $\mathbf{X} \in \mathbb{R}^{U\times I}$, or into the rating vector $\mathbf{x} = \text{vec}(\mathbf{X}) \in \mathbb{R}^{UI}$. Additionally, vectors $\mathbf{x}_u = [X_{u,1}, ..., X_{u,I}]^\top \in \mathbb{R}^I$ and $\mathbf{x}^i = [X_{1,i}, ..., X_{U,i}]^\top \in \mathbb{R}^U$ represent the ratings by the u-th user and i-th item, respectively. To account for the fact that not all ratings are available, let $\mathcal{S}$ denote the set of indexes that identify user-item pairs whose rating

[†]Department of Electrical and Systems Engineering, University of Pennsylvania. [‡]Department of Signal Theory and Communications, King Juan Carlos University. Supported by Spanish MINECO TEC2013-41604-R and TEC2016-75361-R, and USA ARO W911NF1710438. Previous contribution is in [1].

[1] **Notation:** Generically, the entries of a matrix $\mathbf{X}$ and a vector $\mathbf{x}$ will be denoted as X_{ij} and x_i; however, when contributing to avoid confusion, the alternative notation $[\mathbf{X}]_{ij}$ and $[\mathbf{x}]_i$ will be used. In general, $\tilde{\mathbf{x}}$ denotes the frequency coefficients of $\mathbf{x}$, while $\hat{\mathbf{x}}$ is the estimate of $\mathbf{x}$. The notation $^\top$ stands for transpose; † stands for pseudo-inverse; $\lambda_{\max}(\mathbf{X})$ is the largest eigenvalue of the matrix $\mathbf{X}$; $\|\mathbf{X}\|_*$ is the nuclear norm of $\mathbf{X}$; $\|\mathbf{x}\|_p$ is the p-norm of $\mathbf{x}$; $\odot$ and $\otimes$ are, respectively the Khatri-Rao and Kronecker product of two matrices; and $|\mathcal{X}|$ is the cardinality of the set $\mathcal{X}$.

is known. Similarly, $\mathcal{S}_u$ denotes a set containing the indexes of the items that user u has rated. We can then use $\bar{\mathbf{x}}_{\mathcal{S}} \in \mathbb{R}^{|\mathcal{S}|}$ to denote a vector containing the *known* ratings. The problem of interest is as follows: Given the ratings $\bar{\mathbf{x}}_{\mathcal{S}}$ for the item-user pairs in $\mathcal{S}$, estimate the full rating vector $\mathbf{x}$ (matrix $\mathbf{X}$).

2.1. CoFi via LLFM

Latent factor models try to approximate the rating user-item matrix by identifying a few latent features (factors) and then characterize both users and items in terms of those factors. In the linear case, this amounts to project the original user ratings $\mathbf{x}_u$ and item ratings $\mathbf{x}^i$ into a feature vector space of dimension F. The ratings are then obtained as inner products in such a space. Since in CoFi the features are not known a priori but learned from the data, the SVD decomposition plays a key role in identifying the latent factors as well as the underlying mapping. To be more specific, let us rely on the SVD factorization to write the rating matrix as $\mathbf{X} = \mathbf{W}\text{diag}(\boldsymbol{\sigma})\mathbf{Z}^\top$. Next, use the singular vectors to define $\tilde{\mathbf{x}}_u = \mathbf{Z}^\top \mathbf{x}_u$ as the feature profile of user u and $\tilde{\mathbf{x}}^i = \mathbf{W}^\top \mathbf{x}^i$ as the feature profile of item i. With these notational conventions, any rating $X_{u,i}$ can be obtained as $X_{u,i} = \sum_{f=1}^{F} \sigma_f [\tilde{\mathbf{x}}_u]_f [\tilde{\mathbf{x}}^i]_f^\top$, with the f-th singular value σ_f representing the weight of factor f in explaining the rating. While the value of F is related to the the rank of the rating matrix, in real scenarios one expects F to be small, so that $\mathbf{X}$ can be approximated as a low-rank matrix.

Low-rank can be achieved by computing the SVD and keeping only the largest singular values. A major hurdle in this approach is that computation of the SVD requires full knowledge of $\mathbf{X}$, but only the rating values in $\mathcal{S}$ are known. This implies that one must solve instead the problem of minimizing rank$(\mathbf{X})$ subject to $X_{ij} = \bar{X}_{ij}, \forall\ (u,i) \in \mathcal{S}$. Since the rank function renders the problem non-convex, a widely used approach is to relax the rank using the nuclear-norm [4] and set $\hat{\mathbf{X}}$ as the solution to

$$\begin{aligned} \min_{\mathbf{X}} \quad & \|\mathbf{X}\|_* \\ \text{s.t.} \quad & |X_{u,i} - \bar{X}_{u,i}|^2 \leq \epsilon, \quad \forall\ (u,i) \in \mathcal{S}, \end{aligned} \tag{1}$$

where a tolerance ϵ to account for potential observation noise in the known ratings has been included.

2.2. Graph SP

Consider a directed graph $\mathcal{G}$ with a set of N nodes or vertices $\mathcal{N}$ and a set of links $\mathcal{E}$, such that if node n is connected to m, then $(n,m) \in \mathcal{E}$. For any graph we define the adjacency matrix $\mathbf{A}$ as a sparse $N \times N$ matrix with non-zero elements $A_{m,n}$ if and only if $(n,m) \in \mathcal{E}$. The value of $A_{m,n}$ captures the strength of the connection from n to m.

The focus of graph SP is on graph signals defined on the set of nodes $\mathcal{N}$. Formally, each of these signals can be represented as a vector $\mathbf{z} \in \mathbb{R}^N$ where the n-th element represents the value of the signal at node n. To facilitate the connections with NNM, in this work we chose as shift the adjacency matrix $\mathbf{A}$; however, our results can be easily generalized for other choices such as Laplacians [5]. We assume $\mathbf{S}$ is diagonalizable, so that $\mathbf{S} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{-1}$ with $\boldsymbol{\Lambda} = \text{diag}(\boldsymbol{\lambda}) \in \mathbb{C}^{N \times N}$ being diagonal and $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_N]$. When $\mathbf{S}$ is symmetric we have that $\mathbf{V}$ is real and unitary, which implies $\mathbf{V}^{-1} = \mathbf{V}^\top$.

Graph filters are a particular class of *linear* graph-signal operators able to be represented as matrix polynomials of $\mathbf{S}$ [6]

$$\mathbf{H} := \sum_{l=0}^{L-1} h_l \mathbf{S}^l. \tag{2}$$

For a given input $\mathbf{z}$, the output of the filter is simply $\mathbf{y} = \mathbf{H}\mathbf{z}$. The filter coefficients are collected into $\mathbf{h} := [h_0, \ldots, h_{L-1}]^\top$, with $L-1$ denoting the filter degree. The eigendecomposition of $\mathbf{S}$ is used to define the frequency representation of graph signals and filters. For a signal $\mathbf{z} \in \mathbb{R}^N$ and a graph shift operator $\mathbf{S} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^{-1} \in \mathbb{R}$. The vectors

$$\tilde{\mathbf{z}} = \mathbf{V}^{-1}\mathbf{z} \quad \text{and} \quad \mathbf{z} = \mathbf{V}\tilde{\mathbf{z}} \tag{3}$$

form a Graph Fourier Transform (GFT) pair [5, 6].

The GFT encodes a notion of variability for graph signals akin to one that the Fourier transform encodes for temporal signals [6, 21]. The smaller the distance between λ_p and $|\lambda_{\max}|$ in the complex spectrum, the lower the frequency it represents.

3. COFI FROM A GRAPH SP PERSPECTIVE

In this section, we show that if the ratings $\mathbf{x}$ are viewed as a graph signal defined on particular user-to-user and item-to-item networks, then LLFM predict signals $\hat{\mathbf{x}}$ that are *bandlimited* in the frequency domain defined by those particular networks. That is, signals that can be expressed as a combination of few eigenvectors of the graph shift operator. Let us consider either the rating matrix $\mathbf{X}$ or its centered version $\mathbf{X} - \boldsymbol{\mu}_{\mathcal{U}}\boldsymbol{\mu}_{\mathcal{I}}/\mu$, where $\boldsymbol{\mu}_{\mathcal{U}} = \frac{1}{I}\mathbf{X}\mathbf{1}$, $\boldsymbol{\mu}_{\mathcal{U}} = \frac{1}{U}\mathbf{X}^\top\mathbf{1}$ and $\mu = \frac{1}{UI}\sum_{ui} X_{ui}$. The underlying idea behind LLFM is to model the ratings as a matrix $\mathbf{X} = \mathbf{W}\text{diag}(\boldsymbol{\sigma})\mathbf{Z}^\top$ with rank $F = \|\boldsymbol{\sigma}\|_0$ small. The goal in here is to interpret LLFM from a graph SP perspective, showing that predictions based on LLFM give rise to bandlimited signals.

To that end, let us assume that $\mathbf{X}$ has been centered and consider the graph shift operators given by the user-to-user and item-to-item covariance matrices

$$\mathbf{D} = \frac{1}{I}\mathbf{X}\mathbf{X}^\top \ \text{ and } \ \mathbf{E} = \frac{1}{U}\mathbf{X}^\top\mathbf{X}. \tag{4}$$

Since the shifts $\mathbf{S} = \mathbf{D}$ and $\mathbf{S} = \mathbf{E}$ are symmetric, they admit a normal eigen-decomposition as

$$\mathbf{D} = \mathbf{V}\boldsymbol{\Lambda}_{\mathbf{D}}\mathbf{V}^\top \ \text{ and } \ \mathbf{E} = \mathbf{U}\boldsymbol{\Lambda}_{\mathbf{E}}\mathbf{U}^\top. \tag{5}$$

For the shift $\mathbf{S} = \mathbf{D}$, the frequency representation of the ratings for a given item $\mathbf{x}^i$ is therefore given by $\tilde{\mathbf{x}}^i = \mathbf{V}^\top\mathbf{x}^i$ and the matrix collecting the frequency representations for all item signals is $\mathbf{V}^\top\mathbf{X}$. Similarly, for $\mathbf{S} = \mathbf{E}$ the frequency representation of $\mathbf{x}_u$ is given by $\tilde{\mathbf{x}}_u = \mathbf{U}^\top\mathbf{x}_u$, with matrix $\mathbf{U}^\top\mathbf{X}^\top$ collecting the frequency signal for all users. Notice that the two GFTs can be applied jointly, giving rise to the joint (two-dimensional) frequency representation of the user-item rating matrix as

$$\tilde{\mathbf{X}} = \mathbf{V}^\top\mathbf{X}\mathbf{U}. \tag{6}$$

The previous matrix can be vectorized to yield

$$\tilde{\mathbf{x}} = \text{vec}\left(\mathbf{V}^\top\mathbf{X}\mathbf{U}\right) = \mathbf{U}^\top \otimes \mathbf{V}^\top\mathbf{x}. \tag{7}$$

The equation shows that the if the ratings are expressed in the vector form $\mathbf{x}$, the unitary matrix $\mathbf{U}^\top \otimes \mathbf{V}^\top$ represents the associated GFT. This also implies that one can view $\mathbf{x} \in \mathbb{R}^{UI}$ as a signal defined on a shift $\mathbf{S} \in \mathbb{R}^{UI \times UI}$ that has $\mathbf{U} \otimes \mathbf{V}$ as eigenvectors. Two natural choices for such a shift are $\mathbf{S} = \mathbf{E} \otimes \mathbf{D}$, which is the Kronecker graph product graph of the two shifts in (4), and $\mathbf{S} = \mathbf{E} \otimes \mathbf{I} + \mathbf{I} \otimes \mathbf{D}$, the Cartesian graph product of the shifts in (4). We refer [22] for further details on product graphs.

Since LLFM try to approximate the ratings matrix using the SVD factorization $\mathbf{X} = \mathbf{W}\text{diag}(\boldsymbol{\sigma})\mathbf{Z}^\top$ with $F = \|\boldsymbol{\sigma}\|_0$ small, from the expressions in (4)-(6) it follows readily that:

(a) The eigenvectors of $\mathbf{D}$ and $\mathbf{E}$ are the singular vectors of $\mathbf{X}$.

(b) The corresponding GFTs are $\mathbf{V}^\top = \mathbf{W}^\top$ and $\mathbf{U}^\top = \mathbf{Z}^\top$.

(c) The matrix collecting the frequency coefficients in (6) is $\tilde{\mathbf{X}} = \text{diag}(\boldsymbol{\sigma})$.

Therefore, since $\tilde{\mathbf{X}}$ is diagonal with rank F, *vector* $\tilde{\mathbf{x}} \in \mathbb{R}^{UI}$ in (7) *will have at most* F non-zero entries, with $F \ll UI$. In words, LLFM is essentially *modeling the ratings* $\mathbf{x}$ *as* a signal that is *bandlimited* in a frequency domain associated with the product graph of the *covariance shifts* given in (4).

From a practical point of view, it is important to know that since only the subset of ratings in $\mathcal{S}$ is known, unless a priori information exists, the covariance matrices in (4) cannot be computed beforehand. This implies that the LLFM formulation in (1) is searching jointly for a sparse $\tilde{\mathbf{x}}$ as well as for the eigenvectors of the user-to-user and item-to-item covariance matrices.

4. ENHANCING COFI VIA GRAPH SP

Using definitions and tools from graph SP, the previous section demonstrated that the rating predictions generated by LLFM can be understood as signals that are sparse in a graph frequency domain. In this section, we illustrate how these interpretations can be leveraged to design novel graph-SP-based CoFi methods with enhanced prediction performance.

Section 3 revealed that, when interpreted from a graph SP perspective, LLFM builds on two fundamental assumptions: i) the rating signal $\mathbf{x}$ is bandlimited, and ii) $\mathbf{x}$ is defined on top of a graph which can be obtained by combining the user-to-user and item-to-item covariance matrices. As a result, one can reinterpret the prediction carried out by an LLFM as a sampling and reconstruction problem (cf. Problem 1): given $\bar{\mathbf{x}}_\mathcal{S}$, our goal is to recover the full signal $\mathbf{x}$; and the key assumption is that $\mathbf{x}$ is bandlimited and can be written as a linear combination of a few columns of $\mathbf{U} \otimes \mathbf{V}$. To be rigorous, recall that $\mathbf{x}$ stands for the centered ratings, and (7) stated that $\tilde{\mathbf{x}} = \mathbf{U}^\top \otimes \mathbf{V}^\top \mathbf{x}$ and we have that $\tilde{\mathbf{x}} = \text{vec}(\tilde{\mathbf{X}}) = \text{vec}(\text{diag}(\boldsymbol{\sigma}))$. Since only F of the singular values in $\boldsymbol{\sigma}$ are non-zero, the support of $\tilde{\mathbf{x}}$ will have cardinality F and correspond to the F first diagonal elements of $\tilde{\mathbf{X}}$. This readily implies that only F frequencies will be active, those corresponding the Kronecker product of the f-th column of $\mathbf{U}$ with the f-th column of $\mathbf{V}$. Using $\odot$ to denote the Khatri-Rao product and $\mathbf{U}_F$ to represent the F first columns of $\mathbf{U}$, we then have

$$\mathbf{x} = (\mathbf{U}_F \odot \mathbf{V}_F)\tilde{\mathbf{x}}_F, \;\; \text{with} \;\; \tilde{\mathbf{x}}_F \in \mathbb{R}^F. \tag{8}$$

If the eigenvectors are known, the procedure is clear [7, 23]: Given $|\mathcal{S}| \geq F$ samples of $\mathbf{x}$, invert (8) to estimate $\tilde{\mathbf{x}}_F$, and then use the estimate $\hat{\tilde{\mathbf{x}}}_F$ to recover the full rating signal as $\hat{\mathbf{x}} = (\mathbf{U}_F \odot \mathbf{V}_F)\hat{\tilde{\mathbf{x}}}_F$. Moreover, if $(\mathbf{U}_F \odot \mathbf{V}_F)_\mathcal{S}$ denotes a submatrix of $\mathbf{U}_F \odot \mathbf{V}_F$ formed by the rows corresponding to the observed ratings, the least squares estimate of the frequency components is $\hat{\tilde{\mathbf{x}}}_F = (\mathbf{U}_F \odot \mathbf{V}_F)^\dagger_\mathcal{S} \bar{\mathbf{x}}_\mathcal{S}$ and the overall prediction becomes

$$\hat{\mathbf{x}} = \mathbf{U}_F \odot \mathbf{V}_F (\mathbf{U}_F \odot \mathbf{V}_F)^\dagger_\mathcal{S} \bar{\mathbf{x}}_\mathcal{S}. \tag{9}$$

This interpretation is useful not only to come up with new reconstruction schemes, but also to apply results from sampling of graph signals to RS. For example, different papers have shown that when the number of observations is small, the set of sampled nodes plays a critical role on recovery performance [7, 23]. Consequently, schemes to select the nodes to be sampled have been developed [7]. This will amount to identifying user-item pairs that, if known, would contribute to increase the prediction performance. In this context, one can envision active sampling schemes where some users are exposed to particular items so that the overall prediction of the RS improves.

The rest of the section is devoted to leverage the interpretation to present different alternatives for LLFM. The main problem to implement is that the full covariances $\mathbf{D}$ and $\mathbf{E}$ in (4), which give rise to $\mathbf{V}$ and $\mathbf{U}$, are not known. If the available ratings are uniformly distributed across users and items, a simple strategy is to replace the covariances in (4) with the approximations $\boldsymbol{\Sigma}^\mathcal{U}$ and $\boldsymbol{\Sigma}^\mathcal{I}$ using only the observed ratings defined as

$$\begin{aligned} \Sigma^\mathcal{U}_{uv} &:= \frac{1}{|\mathcal{S}_{uv}|} \sum_{i \in \mathcal{S}_{uv}} (\bar{X}_{ui} - \mu)(\bar{X}_{vi} - \mu), \\ \Sigma^\mathcal{I}_{ij} &:= \frac{1}{|\mathcal{S}_{uv}|} \sum_{u \in \mathcal{S}^{ij}} (\bar{X}_{uj} - \mu)(\bar{X}_{ui} - \mu). \end{aligned} \tag{10}$$

Let $\hat{\mathbf{V}}$ and $\hat{\mathbf{U}}$ represent the approximated eigenvectors, the recovery problem to solve in this case is

$$\begin{aligned} &\text{find} \quad \tilde{\mathbf{x}}_F \\ &\text{s.t.} \quad \left| \bar{x}_n - \left[\hat{\mathbf{U}}_F \odot \hat{\mathbf{V}}_F \tilde{\mathbf{x}}_F \right]_n \right|^2 \leq \epsilon, \quad \forall\, n \in \mathcal{S}, \end{aligned} \tag{11}$$

where the main difference is that the solution in the sampled set is not forced to coincide with the original observations.

Another alternative is to enlarge the set of active frequencies both inside and outside the diagonal of $\tilde{\mathbf{X}}$. Suppose that the frequency support $\mathcal{F}$ is small and known, the formulation is then

$$\begin{aligned} &\text{find} \quad \tilde{\mathbf{x}} \\ &\text{s.t.} \quad \left| \bar{x}_n - \left[\hat{\mathbf{U}} \otimes \hat{\mathbf{V}} \tilde{\mathbf{x}} \right]_n \right|^2 \leq \epsilon, \quad \forall\, n \in \mathcal{S}, \\ &\qquad \tilde{x}_f = 0, \quad f \notin \mathcal{F}, \end{aligned} \tag{12}$$

and the predicted ratings are simply $\hat{\mathbf{x}} = (\hat{\mathbf{U}} \otimes \hat{\mathbf{V}})\tilde{\mathbf{x}}^*$, with $\tilde{\mathbf{x}}^*$ being the solution to (12). If $\mathcal{F}$ is not known, a regularizer $\|\tilde{\mathbf{x}}\|_0$ penalizing the number of nonzero coefficients can be added to the optimization. Since the ℓ_0 norm is non-convex, the convex surrogate $\|\tilde{\mathbf{x}}\|_1$ is used instead to yield

$$\begin{aligned} &\min_{\tilde{\mathbf{x}}} \quad \|\tilde{\mathbf{x}}\|_1 \\ &\text{s.t.} \quad \left| \bar{x}_n - \left[\hat{\mathbf{U}} \otimes \hat{\mathbf{V}} \tilde{\mathbf{x}} \right]_n \right|^2 \leq \epsilon, \quad \forall\, n \in \mathcal{S}. \end{aligned} \tag{13}$$

Table 1. Summarizing table of performance and the improvement of the proposed methods compared to benchmarks.

Methods	RMSE	Improvement of the proposed from benchmarks
Proposed	0.8674	
User-based NNM	0.9116	4.85%
Item-based NNM	0.9053	4.19%
Conventional matrix completion	0.8723	0.56%
Matrix completion w/ smoothing	0.8688	0.16%

The above optimization is a classical sparse recovery problem whose performance depends on the number of observed ratings $|\mathcal{S}|$, the tolerance ϵ, as well as on the properties (including coherence) of the so-called sensing matrix $\hat{\mathbf{U}} \otimes \hat{\mathbf{V}}$ [24].

The last algorithm in this section is inspired by the possibility to use different graphs for each user and item. This can be incorporated to the current sampling setup. Suppose that the focus is on predicting $x_n = X_{ui}$. Then, denote $\mathbf{B}_i$ and $\mathbf{C}_u$ as the user-to-user similarity and item-to-item similarity network for item i and user u respectively, and let $\mathbf{V}_i$ and $\mathbf{U}_u$ be the eigenvectors of $\mathbf{B}_i$ and $\mathbf{C}_u$, respectively. The problem (12) then becomes

$$\begin{aligned} \text{find} \quad & \tilde{\mathbf{x}} \\ \text{s.t.} \quad & \left|\bar{x}_{n'} - [\mathbf{U}_u \otimes \mathbf{V}_i \tilde{\mathbf{x}}]_{n'}\right|^2 \leq \epsilon, \quad \forall\, n' \in \mathcal{S}, \\ & \tilde{x}_f = 0, \quad f \notin \mathcal{F}, \end{aligned} \tag{14}$$

The main difficulty with this approach is that (14) needs to be solved for every n not in $\mathcal{S}$. On the positive side, matrices $\mathbf{B}_i$ and $\mathbf{C}_u$ tend to be very sparse and only a few of their eigenvectors are required (the ones associated with the largest eigenvalues), so those eigenvectors can be found efficiently.

Summarizing, by reinterpreting LLFM as the recovery of a bandlimited graph signal from a subset of samples, a number of novel prediction algorithms have been proposed. All the considered algorithms proceed in two steps. First, the user-to-user and item-to-item networks are built and their eigenvectors are found. Second, using those eigenvectors as input, the prediction is formulated as a sparse recovery problem. The different algorithms correspond to different ways to build the similarity shifts as well as different formulations of the sparse recovery. Clearly, alternative definitions for the shifts and modifications in the sparse optimization are also worth considering, but left as future work.

5. NUMERICAL EXPERIMENTS

In this section, we illustrate how the proposed methods improve the rating accuracy in real data. For that purpose we use the MovieLens 100k dataset [20], which contains ratings from 943 users on 1,682 movies. The number of available ratings is 100,000, i.e., the 6.3% of the total number of user-item pairs. We randomly select 100 ratings as the testing set $\mathcal{X}_{\text{ts}}$, and use the rest as training set $\mathcal{X}_{\text{tr}}$. The networks and filter coefficients are only trained on the training set. As a performance metric we use the global root mean squared error (RMSE). User-based nearest-neighbor methods (NNM), item-based NNM, and conventional matrix completion are used as benchmark algorithms. Due to space limitations only two experiment results are presented. An exhaustive evaluation will be provided in the journal version of this paper.

We consider (13) because the support of frequency $\mathcal{F}$ is unknown. The RMSE achieved on the testing set is 0.8674, which is better than user-based NNM (RMSE 0.9116) by 4.85%, item-based NNM (RMSE 0.9053) by 4.19%, conventional matrix completion (RMSE 0.8723) by 0.56%, and slightly better than matrix completion with Laplacian quadratic term [17, 18] (RMSE 0.8688) by 0.16%; summarizing table illustrated in Table 1. This shows that the approximated eigenvectors $\hat{\mathbf{V}}$ and $\hat{\mathbf{U}}$ can be used to represent the actual eigenvectors well, and therefore the single-step matrix completion can be done separately by first evaluating the sample eigenvectors $\hat{\mathbf{V}}$ and $\hat{\mathbf{U}}$, then estimating the bandlimited frequency components $\tilde{\mathbf{x}}$. We emphasize this divide-and-conquer will greatly lower complexity, as sample eigenvectors $\hat{\mathbf{V}}$ and $\hat{\mathbf{U}}$ are fast to evaluate, and (13) can be written equivalently as

$$\min_{\tilde{\mathbf{x}}} \quad \sum_{n\in\mathcal{S}} \left|\bar{x}_n - \left[\hat{\mathbf{U}} \otimes \hat{\mathbf{V}}\tilde{\mathbf{x}}\right]_n\right|^2 + \lambda\|\tilde{\mathbf{x}}\|_1, \tag{15}$$

which can be efficiently solved by well-developed solvers, e.g. [25]. The next experiment relaxes the requirement $\tilde{\mathbf{X}} = \text{diag}(\boldsymbol{\sigma})$ by letting some off-diagonal elements to be non-zero. More specifically, if τ represents a small integer number, entries $\tilde{X}_{r,t}$ such that $|r-t| \leq \tau$ are allowed to be non-zero. This yields the following problem

$$\begin{aligned} \min_{\tilde{\mathbf{X}}} \quad & \left\|\text{vec}(\tilde{\mathbf{X}})\right\|_1 \\ \text{s.t.} \quad & \left|\bar{X}_{ui} - \left[\hat{\mathbf{U}}\tilde{\mathbf{X}}\hat{\mathbf{V}}\right]_{u,i}\right|^2 \leq \epsilon, \quad \forall\, (u,i) \in \mathcal{S}, \\ & \tilde{X}_{r,t} = 0, \quad \text{if } |r-t| > \tau. \end{aligned} \tag{16}$$

Setting $\tau = 0$ would reduce the problem (16) back to (13). We evaluated the performance of (16) for $\tau = 1$ and $\tau = 2$, and found that the off diagonal elements $\tilde{X}_{r,t}$ with $r \neq t$ generated by the solver were zero. In other words, the solution and performance was the same than that of (13). This implies that the model in (13) captures the key features of our data, with the estimated (sample) eigenvectors being close to the actual eigenvectors.

6. CONCLUSIONS

This paper exploited results from graph SP to propose new interpretations for RS methods. Our first contribution was to show that CoFi can be viewed as bandlimited interpolation algorithms that operate in a frequency domain of a joint user and item network. Leveraging this, we then proposed a new method for RS by first estimating eigenvectors, and then solving a sparse recovery problem. We also proposed a computationally efficient scheme to design the parameters that define our methods and assessed performance in the MovieLens-100k dataset. The results obtained showed that we reduced the RMSE by a rate of 4.19% compared to NNM and 0.56% compared to matrix completion. Future work would be to consider other types of graph filters and to investigate joint filter from both the user and item graph domain.

7. REFERENCES

[1] W. Huang, A. G. Marques, and A. Ribeiro, "Collaborative filtering and matrix completion via graph filtering," in *IEEE Eur. Signal Process. Conf.*, Aug. 2017, pp. 1744–1748.

[2] F. Ricci, L. Rokach, and B. Shapira, *Introduction To Recommender Systems Handbook*, Springer, 2011.

[3] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, vol. 42, no. 8, pp. 30–37, Aug. 2009.

[4] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Found. Comput. Math.*, vol. 9, no. 6, pp. 717–772, Jun. 2009.

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

[6] A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs," *IEEE Trans. Signal Process.*, vol. 61, no. 7, pp. 1644–1656, Apr. 2013.

[7] A. G. Marques, S. Segarra, G. Leus, and A. Ribeiro, "Sampling of graph signals with successive local aggregations," *IEEE Trans. Signal Process.*, vol. 64, no. 7, pp. 1832–1843, Apr. 2016.

[8] M. Rabbat, M. Coates, and S. Blouin, "Graph Laplacian distributed particle filtering," in *IEEE Eur. Signal Process. Conf.*, Aug. 2016, pp. 1493–1497.

[9] N. Tremblay and P. Borgnat, "Subgraph-based filterbanks for graph signals," *IEEE Trans. Signal Process.*, vol. 64, no. 15, pp. 3827–3840, Aug. 2016.

[10] B. Pasdeloup, R. Alami, V. Gripon, and M. Rabbat, "Toward an uncertainty principle for weighted graphs," in *IEEE Eur. Signal Process. Conf.*, Aug. 2015, pp. 1496–1500.

[11] W. Huang, L. Goldsberry, N. F. Wymbs, S. T. Grafton, D. S. Bassett, and A. Ribeiro, "Graph frequency analysis of brain signals," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 7, pp. 1189–1203, Oct. 2016.

[12] J. D. Medaglia et al., "Functional alignment with anatomical networks is associated with cognitive flexibility," *arXiv preprint arXiv:1611.08751*, Nov. 2016.

[13] R. Liu, H. Nejati, and N.-M. Cheung, "Simultaneous low-rank component and graph estimation for high-dimensional graph signals: application to brain imaging," *arXiv preprint arXiv:1609.08221*, Sep. 2016.

[14] J. Pang and G. Cheung, "Graph Laplacian regularization for inverse imaging: Analysis in the continuous domain," *arXiv preprint arXiv:1604.07948*, Apr. 2016.

[15] D. Thanou, P. A. Chou, and P. Frossard, "Graph-based compression of dynamic 3D point cloud sequences," *IEEE Trans. Image Process.*, vol. 25, no. 4, pp. 1765–1778, Apr. 2016.

[16] Y. Wang, A. Ortega, D. Tian, and A. Vetro, "A graph-based joint bilateral approach for depth enhancement," in *IEEE Int. Conf. on Acoustics, Speech and Signal Process.*, May 2014, pp. 885–889.

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

[18] P. Giménez-Febrer and A. Pagès-Zamora, "Matrix completion of noisy graph signals via proximal gradient minimization," in *IEEE Int. Conf. on Acoustics, Speech and Signal Process.*, Mar. 2017, pp. 4441–4445.

[19] J. Ma, W. Huang, S. Segarra, and A. Ribeiro, "Diffusion filtering for graph signals and its use in recommendation systems," in *IEEE Int. Conf. on Acoustics, Speech and Signal Process.*, Mar. 2016, pp. 4563 – 4567.

[20] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, "Grouplens: An open architecture for collaborative filtering of netnews," in *ACM Conf. on Comput. Support. Coop. Work*, Oct., 1994, pp. 175–186.

[21] A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs: Frequency analysis," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3042–3054, Jun. 2014.

[22] A. Sandryhaila and J. M. F. Moura, "Big data analysis with signal processing on graphs: Representation and processing of massive data sets with irregular structure," *IEEE Signal Process. Mag.*, vol. 31, no. 5, pp. 80–90, Sep. 2014.

[23] S. Chen, R. Varma, A. Sandryhaila, and J. Kovačević, "Discrete signal processing on graphs: Sampling theory," *IEEE Trans. Signal Process.*, vol. 63, no. 24, pp. 6510–6523, Dec. 2015.

[24] J. M. Duarte-Carvajalino and G. Sapiro, "Learning to sense sparse signals: Simultaneous sensing matrix and sparsifying dictionary optimization," *IEEE Trans. Image Process.*, vol. 18, no. 7, pp. 1395–1408, Jul. 2009.

[25] S. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, "A method for large-scale l1-regularised least squares problems with applications in signal processing and statistics," *IEEE J. Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 606–617, Dec. 2007.