PU Matrix Completion with Graph Information

Nagarajan Natarajan
Department of Computer Science
University of Texas at Austin

Nikhil Rao
Department of Computer Science
University of Texas at Austin

Inderjit Dhillon
Department of Computer Science
University of Texas at Austin

***Abstract*—Motivated by applications in recommendation systems and bioinformatics, we consider the problem of completing a low rank, partially observed binary matrix with graph information. We show that the corresponding problem can be set up in a positive and unlabeled data learning (referred to as *PU learning* in literature) framework. We make connections to convex optimization and show that existing greedy methods can be used to solve the problem. Experiments on simulated data as well as gene-disease associations data from bioinformatics show that using graphs, and adapting matrix completion in the PU learning setting, yield advantages over the standard binary matrix completion.**

I. Introduction and Problem Setup

Matrix completion plays a key role in collaborative filtering applications, where by modeling the target user-item preference matrix as low rank, one encodes the prior belief that the user and the item variables share a low dimensional subspace. Often, additional information other than the partially observed target matrix is available, either in the form of user and item attributes [1] or graphs that encode pairwise relationships between users and items [11]. In the recent past, many methods that incorporate such additional information have been shown to perform much better than the traditional, low rank matrix completion [1], [19], [11] .

Recovering a low rank matrix with graph constraints arises in several applications. In the context of recommender systems, we may have access to social network between users, or product co-purchasing graphs for items, or a graph that relates music and movies by genre, release date and other attributes. Similarly, in the context of gene-disease associations, graphs correspond to gene regulatory networks and relationship between diseases. The goal is not only to infer the missing entries of the matrix, but to do so in a manner that incorporates additional constraints that the graphs may give rise to. Matrix completion with graph information has been widely studied in [19], [11], [18], [10], [15], and references therein, and has been shown to outperform standard baselines that look to recover a low rank matrix without incorporating additional information.

In this paper, we consider a modification of the problem of recovering a graph constrained low rank matrix. Firstly, the underlying matrix that we want to recover is *binary*. Secondly, we assume that only a subset of the positive entries of the target matrix is observed; this setting is referred to as *positive and unlabeled* (PU) learning in literature. PU learning has been studied in the context of binary classification [6] and recently explored in collaborative filtering setting [8].

The problem of recovering a low rank binary matrix from PU observations along with graph structured information arises in several applications, two of which we outline below:

1) **Gene-Disease Prediction:** Discovering novel gene-disease associations is a problem of great importance in computational biology (see recent survey [7]). Public gene-disease data repositories report whether or not a gene *is associated* with a disease. We have network information such as functional gene-interaction networks and disease similarity networks that can be used for improving predictions [2], [16], [12]. Furthermore the repositories only report whether a gene is associated with a disease [16].
2) **Music Recommendation:** In online music recommender systems such as Pandora or targeted advertisements, users can only "like" or "dislike" a song. The songs themselves might be related via a graph, based on genre, artists, etc. Moreover, the users might be part of a social network: information that can very easily be obtained. Also, user *clicking* on recommended articles or songs is often used as a proxy for positive feedback, whereas there is no explicit negative feedback.

In this paper, we first visit the problem of recovering a binary preference matrix with graph information. We cast the problem as a convex optimization problem. In particular, we show that the problem can in fact be cast as an atomic norm constrained optimization, allowing us to use methods that have been developed specifically for this purpose [13]. Then, we suggest a modified optimization problem for the PU learning setting when there are no negative observations. We experimentally show that incorporating auxiliary graph information allows for better 1-bit matrix completion performance, in regular as well as PU settings, compared to the standard low-rank matrix completion.

A. Problem Setup

Let $\boldsymbol{M}^\star$ be the underlying real-valued matrix of size $m \times n$ and of rank k (indicating user-item preferences or gene-disease associations) . In many applications, we do not observe samples directly from $\boldsymbol{M}^*$, but only from a 1-bit quantization of its entries. Let $\boldsymbol{Y} \in \{0,1\}^{m \times n}$ such that $\boldsymbol{Y}_{ij} = 1$ if $\boldsymbol{M}^\star \geq q$ or $\boldsymbol{Y}_{ij} = 0$ otherwise; we can think of $\boldsymbol{M}^\star_{ij}$ as a preference score and $\boldsymbol{Y}_{ij}$ then corresponds to thresholding the score at q. (Alternatively, one can model $\boldsymbol{M}^\star$ as a probability matrix and define $\boldsymbol{Y}_{ij} = f(\boldsymbol{M}^\star_{ij})$ for suitably defined f [5]). Let Ω denote the set of observed indices of $\boldsymbol{Y}$. The 1-bit matrix completion problem can then be stated as follows:

$$\hat{\boldsymbol{M}} = \arg\min_{\boldsymbol{M}} \frac{1}{2}\|\mathcal{P}_\Omega(\boldsymbol{Y} - \boldsymbol{M})\|_F^2 \quad \textbf{s.t.} \quad \|\boldsymbol{M}\|_* \leq \tau \qquad (1)$$

where $\|.\|_*$ is the nuclear or the trace norm defined as $\|Z\|_* = \sqrt{Tr(Z^T Z)}$, and is a tight convex relaxation of

This research was supported by NSF grants CCF-1117055 and CCF-1320746, and a grant from Adobe.

the non-convex rank constraint. $\mathcal{P}_\Omega(\cdot)$ is the projection of the matrix onto the set Ω. When the observations Ω consist only of positive associations in the PU setting, i.e. $\boldsymbol{Y}_{ij} = 1, \forall (i,j) \in \Omega$, formulation (1) may yield degenerate results. For example, if $\boldsymbol{Y}$ corresponds to PU observations in (1), $\hat{\boldsymbol{M}} = \mathbf{1}\mathbf{1}^T$ is optimal but, of course, is not interesting. Hence we consider the following *biased* low rank matrix completion formulation [8]:

$$\hat{\boldsymbol{M}} = \arg\min_{\boldsymbol{M}} \alpha\|\mathcal{P}_\Omega(\boldsymbol{Y}-\boldsymbol{M})\|_F^2 + (1-\alpha)\|\mathcal{P}_{\bar{\Omega}}(\boldsymbol{Y}-\boldsymbol{M})\|_F^2 \quad \textbf{s.t.} \quad \|\boldsymbol{M}\|_* \leq \tau \tag{2}$$

where $\bar{\Omega} := \{(i,j) : (i,j) \notin \Omega\}$ is the set of unobserved entries, and we let $\boldsymbol{Y}_{ij} = 0$ for $(i,j) \in \bar{\Omega}$. The parameter $\alpha \in (0,1)$ is the *bias* that appropriately weighs observed and unobserved entries. In experiments, α and τ are estimated from data. We obtain an estimate of the underlying binary matrix $\boldsymbol{Y}$ by thresholding $\hat{\boldsymbol{M}}$.

Our focus is on the matrix completion setting where there is access to graphs $\boldsymbol{L}^u \in \mathrm{R}^{m\times m}$ and $\boldsymbol{L}^v \in \mathrm{R}^{n\times n}$ that encode pairwise relationships between the variables corresponding to the rows and the columns of $\boldsymbol{M}^\star$ respectively. While the aforementioned trace norm constraint forces the solution to have low rank, it completely ignores the additional information provided by the graphs. We would thus like to consider a different constraint than the ones in (1) and (2) that not only yields a low rank $\hat{\boldsymbol{M}}$ but also accounts for the graphs.

The rest of the paper is organized as follows: in the next section, we show that the 1-bit matrix completion problem on graphs can be cast as a convex program. We make connections to atomic norms and show that recovering graph constrained low rank matrices is equivalent to solving a weighted nuclear norm constrained program. In Section III we show that the problem is amenable to greedy algorithms that are orders of magnitude more efficient than corresponding proximal point methods. We report results on simulated as well as real-world datasets in Section IV, and conclude the paper in Section V.

II. Convex Formulation: Weighted Atomic Norm

In this section, we show that the problem of matrix completion on graphs can be cast as a convex program with an appropriate regularizer. Consider the factorization: $\boldsymbol{M} = \boldsymbol{U}\boldsymbol{V}^T$, where $\boldsymbol{U}$ (resp. $\boldsymbol{V}$) are the latent factors corresponding to the rows (resp. columns) of matrix $\boldsymbol{M}$. Given a graph $\mathcal{G}_u = \{\mathcal{V}_u, \mathcal{E}_u\}$ that encodes relationships between the rows, it is natural to assume that the stronger the weight $w_{(a,b)}$ between two nodes (a,b) in the graph, the more similar the corresponding latent factors $\boldsymbol{u}_a, \boldsymbol{u}_b$ are. As a concrete example, if two users are friends in a social network, then their features (and hence their preference score for the same item) are likely to be similar. One way to enforce such a constraint is to minimize:

$$\sum_{(a,b)\in\mathcal{E}_u} w_{(a,b)}\|\boldsymbol{u}_a - \boldsymbol{u}_b\|^2 \quad = Tr(\boldsymbol{U}^T\boldsymbol{L}^u\boldsymbol{U}) \tag{3}$$

where $\boldsymbol{L}^u$ is the graph Laplacian; a similar objective can be formed for the column factors $\boldsymbol{V}$. Noting that the low rank constraint itself can be written as

$$\|\boldsymbol{M}\|_* = \inf_{\boldsymbol{U},\boldsymbol{V}} \frac{1}{2}\left(Tr(\boldsymbol{U}^T\boldsymbol{U}) + Tr(\boldsymbol{V}^T\boldsymbol{V})\right)$$

and defining $\boldsymbol{L}_u = \boldsymbol{L}^u + \boldsymbol{I}$ (and similarly for $\boldsymbol{L}_v$), we can formulate the problem of recovering a graph constrained low rank matrix as

$$\hat{\boldsymbol{U}}, \hat{\boldsymbol{V}} = \arg\min_{\boldsymbol{U},\boldsymbol{V}} \frac{1}{2}\|\mathcal{P}_\Omega(\boldsymbol{Y} - \boldsymbol{U}\boldsymbol{V}^T)\|_F^2 \quad \textbf{s.t.} \quad \inf_{\boldsymbol{U},\boldsymbol{V}} \left(Tr(\boldsymbol{U}^T\boldsymbol{L}_u\boldsymbol{U}) + Tr(\boldsymbol{V}^T\boldsymbol{L}_v\boldsymbol{V})\right) \leq \tau \tag{4}$$

The above formulation can be solved using non convex alternating minimization methods. However, we now show that the re-formulation (4) is closely connected to atomic norms, and hence can be cast as a convex optimization problem.

The nuclear norm is the gauge function induced by the atomic set: $\mathscr{A}_* = \{\boldsymbol{w}_i\boldsymbol{h}_i^T \;:\; \|\boldsymbol{w}_i\| = \|\boldsymbol{h}_i\| = 1\}$ (see [4]). Note that all rank-1 matrices in $\mathscr{A}_*$ have unit Frobenius norm. Now, assume that the symmetric, positive definite matrix $\boldsymbol{L}_u$ has the eigendecomposition $\boldsymbol{L}_u = \boldsymbol{Z}_u\boldsymbol{\Sigma}_u\boldsymbol{Z}_u^T$. Similarly, let $\boldsymbol{L}_v = \boldsymbol{Z}_v\boldsymbol{\Sigma}_v\boldsymbol{Z}_v^T$. Let $\boldsymbol{A} = \boldsymbol{Z}_u\boldsymbol{\Sigma}_u^{-\frac{1}{2}}$, and $\boldsymbol{B} = \boldsymbol{Z}_v\boldsymbol{\Sigma}_v^{-\frac{1}{2}}$.

Consider the following "weighted" atomic set:

$$\mathscr{A} := \{\mathrm{a}_i = \boldsymbol{w}_i\boldsymbol{h}_i^T : \boldsymbol{w}_i = \boldsymbol{A}\boldsymbol{p}_i, \boldsymbol{h}_i = \boldsymbol{B}\boldsymbol{q}_i, \|\boldsymbol{p}_i\| = \|\boldsymbol{q}_i\| = 1\}. \tag{5}$$

Clearly, each atom a in $\mathscr{A}$ has non-unit Frobenius norm. We will see that this atomic set allows for biasing of the solutions towards certain atoms. We now define a corresponding atomic norm:

$$\|\boldsymbol{M}\|_{\mathscr{A}} = \inf \sum_{\mathrm{a}_i\in\mathscr{A}} |c_i| \quad \text{s.t. } \boldsymbol{M} = \sum_{\mathrm{a}_i\in\mathscr{A}} c_i\mathrm{a}_i. \tag{6}$$

It is not hard to verify that $\|\boldsymbol{M}\|_{\mathscr{A}}$ is a norm and $\{\boldsymbol{M} : \|\boldsymbol{M}\|_{\mathscr{A}} \leq \tau\}$ is closed and convex.

The following result shows that the constraint in (4) can be expressed as an atomic norm:

Theorem II.1. *Given Laplacian matrices $\boldsymbol{L}_u, \boldsymbol{L}_v$, and $\boldsymbol{A} = \boldsymbol{Z}_u\boldsymbol{\Sigma}_u^{-\frac{1}{2}}$, $\boldsymbol{B} = \boldsymbol{Z}_v\boldsymbol{\Sigma}_v^{-\frac{1}{2}}$, and corresponding weighted atomic set $\mathscr{A}$,*

$$\|\boldsymbol{M}\|_{\mathscr{A}} = \inf_{\boldsymbol{U},\boldsymbol{V}} \quad \frac{1}{2}\{\|\boldsymbol{A}^{-1}\boldsymbol{U}\|_F^2 + \|\boldsymbol{B}^{-1}\boldsymbol{V}\|_F^2\} \text{ s.t. } \boldsymbol{M} = \boldsymbol{U}\boldsymbol{V}^T.$$

Proof: For all $\boldsymbol{M} = \sum_i c_i\mathrm{a}_i$ with $\|\boldsymbol{A}\|_{\mathscr{A}} = \sum_i |c_i|$, where $\mathrm{a}_i = \boldsymbol{A}\boldsymbol{p}_i\boldsymbol{q}_i^T\boldsymbol{B}^T$, we can construct the i-th column of $\boldsymbol{U}$ and $\boldsymbol{V}$ as

$$\boldsymbol{u}_i = \sqrt{|c_i|}\boldsymbol{A}\boldsymbol{p}_i \quad \text{and} \quad \boldsymbol{v}_i = \sqrt{|c_i|}\boldsymbol{B}\boldsymbol{q}_i.$$

Clearly, we have $\boldsymbol{M} = \boldsymbol{U}\boldsymbol{V}^T$ and

$$\|\boldsymbol{A}^{-1}\boldsymbol{U}\|_F^2 = \|\boldsymbol{B}^{-1}\boldsymbol{V}\|_F^2 = \sum_i |c_i|$$

Thus, if follows that LHS $\geq$ RHS. Oh the other hand, for a matrix $\boldsymbol{M} = \boldsymbol{U}\boldsymbol{V}^T$, we can construct

$$\boldsymbol{p}_i = \frac{\boldsymbol{A}^{-1}\boldsymbol{u}_i}{\|\boldsymbol{A}^{-1}\boldsymbol{u}_i\|} \quad \text{and} \quad \boldsymbol{q}_i = \frac{\boldsymbol{B}^{-1}\boldsymbol{v}_i}{\|\boldsymbol{B}^{-1}\boldsymbol{v}_i\|},$$

and $c_i = \|\boldsymbol{A}^{-1}\boldsymbol{u}_i\|\|\boldsymbol{B}^{-1}\boldsymbol{v}_i\|$. Clearly, we have $\boldsymbol{u}_i\boldsymbol{v}_i^T = c_i\boldsymbol{A}\boldsymbol{p}_i\boldsymbol{q}_i\boldsymbol{B}^T$ and $\boldsymbol{M} = \sum_i c_i\boldsymbol{A}\boldsymbol{p}_i\boldsymbol{q}_i^T\boldsymbol{B}^T$. We also have

$$|c_i| = \|\boldsymbol{A}^{-1}\boldsymbol{u}_i\|\|\boldsymbol{B}^{-1}\boldsymbol{v}_i\| \leq \frac{1}{2}\left(\|\boldsymbol{A}^{-1}\boldsymbol{u}_i\|^2 + \|\boldsymbol{B}^{-1}\boldsymbol{v}_i\|^2\right)$$

by AM-GM inequality. Thus, we have LHS $\leq$ RHS. ∎

Theorem II.1 allows us to write (4) as the following convex program:

$$\hat{\boldsymbol{M}} = \arg\min_{\boldsymbol{M}} \frac{1}{2}\|\mathcal{P}_\Omega(\boldsymbol{Y}-\boldsymbol{M})\|_F^2 \quad \textbf{s.t.} \quad \|\boldsymbol{M}\|_{\mathscr{A}} \leq \tau \tag{7}$$

Similarly, letting $\boldsymbol{M} = \boldsymbol{U}\boldsymbol{V}^T$, the biased matrix completion objective (2), under the constraint in (4), we obtain a convex program for the corresponding PU setting as well.

III. Dual Norm and Greedy Algorithm

The results in the previous section allow us to obtain the dual weighted atomic norm for a matrix $\boldsymbol{M}$:

$$\|\boldsymbol{M}\|_{\mathscr{A}}^* = \|\boldsymbol{A}^T\boldsymbol{M}\boldsymbol{B}\,\| = \|\boldsymbol{\Sigma}_u^{-\frac{1}{2}}\boldsymbol{Z}_u^T\boldsymbol{M}\boldsymbol{Z}_v\boldsymbol{\Sigma}_v^{-\frac{1}{2}}\| \tag{8}$$

which is a weighted spectral norm. The dual norm can be derived as follows:

$$\begin{aligned}
&\|\boldsymbol{M}\|_{\mathscr{A}}^* = \sup_{\mathrm{a}\in\mathscr{A}} \langle \mathrm{a}, \boldsymbol{M}\rangle \\
=&\sup_{\boldsymbol{u},\boldsymbol{v}} \langle \boldsymbol{A}\boldsymbol{u}\boldsymbol{v}^T\boldsymbol{B}^T, \boldsymbol{M}\rangle, \quad \text{s.t. } \|\boldsymbol{u}\| = \|\boldsymbol{v}\| = 1 \\
=&\sup_{\boldsymbol{u},\boldsymbol{v}} \mathrm{Tr}\left(\boldsymbol{B}\boldsymbol{v}\boldsymbol{u}^T\boldsymbol{A}^T\boldsymbol{M}\right), \quad \text{s.t. } \|\boldsymbol{u}\| = \|\boldsymbol{v}\| = 1 \\
=&\sup_{\boldsymbol{u},\boldsymbol{v}} \boldsymbol{u}^T\boldsymbol{A}^T\boldsymbol{M}\boldsymbol{B}\boldsymbol{v}, \quad \text{s.t. } \|\boldsymbol{u}\| = \|\boldsymbol{v}\| = 1 \\
=&\|\boldsymbol{A}^T\boldsymbol{M}\boldsymbol{B}\| \ .
\end{aligned}$$

Note that we can then write

$$\|\boldsymbol{M}\|_{\mathscr{A}} = \|\boldsymbol{A}^{-1}\boldsymbol{M}\boldsymbol{B}^{-T}\|_* = \|\boldsymbol{\Sigma}_u^{\frac{1}{2}}\boldsymbol{Z}_u^{-1}\boldsymbol{M}\boldsymbol{Z}_v^{-T}\boldsymbol{\Sigma}_v^{\frac{1}{2}}\|_* \tag{9}$$

(8) shows that the dual atomic norm is merely the spectral norm of a weighted matrix. The spectral norm can be easily computed, by computing the leading singular values and vectors of the weighted matrix. Note that this operation is an order of magnitude cheaper than computing the entire SVD; hence, using greedy methods that only involve dual norm computations [17], [13], [14], [9] is significantly more efficient than solving a proximal point based first order method (variants of [3]). In this paper, we will use the method developed in [13] for solving (7):

$$\hat{\boldsymbol{X}} = \arg\min_{\boldsymbol{X}} \|\boldsymbol{y} - \boldsymbol{\Phi}\ \mathrm{vec}(\boldsymbol{X})\|_2^2 \quad \textbf{s.t.} \quad \|\boldsymbol{X}\|_{\mathscr{A}} \leq \tau \tag{10}$$

where $\boldsymbol{y} = \mathrm{vec}(\mathcal{P}_\Omega(\boldsymbol{Y}))$ and $\boldsymbol{\Phi} \in [0,1]^{|\Omega|\times mn}$ is the matrix that maps $\mathrm{vec}(\boldsymbol{X}) \in \mathrm{R}^{mn}$ to the observations $\boldsymbol{y}$.

Due to space limitations, we will only provide key insights into the atom selection step of the algorithm (for details, see [13]) which dominates the complexity of the overall method:

a) Greedy Atom Selection Step: The step that determines the computational complexity of the greedy method is the atom selection step that solves

$$\boldsymbol{a} = \arg\sup_{\mathrm{a}\in\mathscr{A}} \langle \mathrm{a}, \nabla f^t\rangle \tag{11}$$

where ∇f^t is the gradient of the objective function value at iteration t. Since it suffices to obtain the leading singular values and vectors of the gradient matrix, the complexity is $O(n^2)$ using power iterations, and $O\left(\frac{\mathrm{nnz}(\nabla f^t)}{\sqrt{\epsilon}}\right)$ using Lanczos iterations. Proximal methods, on the other hand, require $O\left(n^3\right)$ operations to compute the entire SVD of the matrix and are hence much less scalable.

A. Extension to PU Learning

The greedy scheme outlined above can be extended in a straightforward manner to handle PU samples in the training data, by modifying the loss function as given in formulation (2). Letting α be the bias factor as in (2), we see that:

$$\begin{aligned}
\mathcal{L}_{PU} &:= \alpha\|\mathcal{P}_\Omega(\boldsymbol{Y}-\boldsymbol{M})\|_F^2 + (1-\alpha)\|\mathcal{P}_{\bar{\Omega}}(\boldsymbol{Y}-\boldsymbol{M})\|_F^2 \\
&= \alpha\|\boldsymbol{y}_1 - \boldsymbol{\Phi}_1\boldsymbol{x}\|^2 + (1-\alpha)\|\boldsymbol{y}_0 - \boldsymbol{\Phi}_0\boldsymbol{x}\|^2 \\
&= \|\sqrt{\alpha}\boldsymbol{y}_1 - \sqrt{\alpha}\boldsymbol{\Phi}_1\boldsymbol{x}\|^2 \\
&\quad + \|\sqrt{(1-\alpha)}\boldsymbol{y}_0 - \sqrt{(1-\alpha)}\boldsymbol{\Phi}_0\boldsymbol{x}\|^2 \\
&= \|\boldsymbol{y}_c - \boldsymbol{\Phi}_c\boldsymbol{x}\|^2
\end{aligned} \tag{12}$$

where we define $\boldsymbol{y}_1$ to be the sub-vector of $\boldsymbol{y}$ that corresponds to the positive (observed) examples, and $\boldsymbol{\Phi}_1$ is the corresponding sub matrix of $\boldsymbol{\Phi}$; $\boldsymbol{y}_0, \boldsymbol{\Phi}_0$ are defined similarly. Finally, we let $\boldsymbol{y}_c := \begin{bmatrix}\sqrt{\alpha}\boldsymbol{y}_1^T & \sqrt{1-\alpha}\boldsymbol{y}_0^T\end{bmatrix}^T$, and $\boldsymbol{\Phi}_c$ to be the corresponding permutation of $\boldsymbol{\Phi}$.

From (12), we see that the PU learning formulation can be cast in the same way as (10), with a slight modification to the "sensing" matrix $\boldsymbol{\Phi}$ and the observations $\boldsymbol{y}$. We can use the same greedy procedure to solve the PU version of the graph constrained matrix completion problem:

$$\hat{\boldsymbol{X}} = \arg\min_{\boldsymbol{X}} \|\boldsymbol{y}_c - \boldsymbol{\Phi}_c\ \mathrm{vec}(\boldsymbol{X})\|_2^2 \quad \textbf{s.t.} \quad \|\boldsymbol{X}\|_{\mathscr{A}} \leq \tau \ . \tag{13}$$

IV. Experiments and Results

In this section, we test our proposed methods on both simulated and real-world data. We compare four methods:

1) **GMC :** Proposed matrix completion with graph information, (10)
2) **PU-GMC :** Proposed PU matrix completion with graph information, (13)
3) **MC :** The standard low-rank matrix completion, (1)
4) **PU-MC :** PU matrix completion, (2)

The training data remains the same for all the methods: a binary matrix with all the observed entries being positive. The test data consists of randomly selected entries (both 0's and 1's) from the binary ground-truth matrix $\boldsymbol{Y}$. We report the RMSE on the test data.

A. Simulated Data

We first present results on simulated data. We set $m = n = 200$. For both the row and the column entities, we generate a random power law graph so that the i^{th} node has $\lceil i^{0.25}\rceil$ randomly selected neighbors. We compute the matrices $\boldsymbol{A}, \boldsymbol{B}$ using these graphs, and obtain the ground truth matrix as $\boldsymbol{M} = \boldsymbol{A}\boldsymbol{S}\boldsymbol{B}^T$ where $\boldsymbol{S}$ is a random $m \times n$ matrix of rank $k = 5$. We then threshold $\boldsymbol{M}_{ij}$ entries at 0 to obtain binary $\boldsymbol{Y}$, and subsample 25% of the entries of $\boldsymbol{Y}$ as the training set. We further subsample a fraction β of the positive entries in the training data to be revealed to the methods. The RMSE on the test set as a function of fraction $1-\beta$ of the positive entries hidden is shown in Fig. 1.

An interesting point to note from Fig. 1 is that the RMSE of PU-MC is higher than that of GMC, indicating that using graph information can yield much better performance even when we do not explicitly account for the fact that we observe PU data.

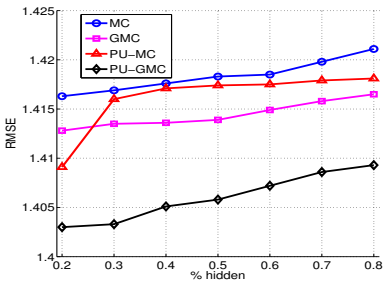

Fig. 1: Comparison of different methods on simulated data. We see that the proposed PU-GMC yields much lower RMSE on the test set compared to the other methods.

B. Gene Disease Prediction

We now consider the problem of predicting gene-disease associations in bioinformatics — here the data consists of positive associations between genes and diseases (indicating that a gene is causally linked to a disease), but no negative associations. We report results on the widely-used OMIM (Online Mendelian Inheritance in Man[1]) dataset (obtained from [8], [16]), consisting of associations between 12331 genes and 3215 diseases. The dataset also has a functional gene-gene interaction network and a similarity network of diseases, which we directly use for our graph based formulations. We use the symmetric graph laplacian $\boldsymbol{L} = \boldsymbol{I} - \boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{W}\boldsymbol{D}^{-\frac{1}{2}}$, where $\boldsymbol{W}$ denotes the (weighted) gene-gene interaction network (or the similarity network for diseases).

We only retain the genes and the diseases for which there is at least one association known, so that the standard matrix completion based methods can be applied. This results in a highly sparse associations matrix of 1907 genes and 2205 diseases, with 2785 non-zeros. We randomly set aside 33% of the non-zeros as the test set, and train our model on the remaining data. We varied τ, α over a grid: $\tau \in \{0.01, 0.1, 1, 10, 100\}$ and $\alpha \in [0.01, 1]$.

Table I shows the RMSE on the test set, for varying amounts of observed data. Note that as the $\%$ observed data reduces, GMC does even better than PU-MC. This again highlights the importance of incorporating auxiliary graph information: we get superior performance using the standard matrix completion with graph information, when compared to an optimized method for PU learning but ignoring graphs.

observed positives	MC	GMC	PU-MC	PU-GMC
50 %	1.4563	1.3210	1.2691	**1.2370**
25 %	1.3511	1.3005	1.3108	**1.2901**
10 %	1.3610	1.3205	1.3311	**1.3004**

TABLE I: Predicting gene-disease associations.

V. Conclusions

In this paper, we considered the problem of positive-unlabeled matrix completion where additional information about the entities in the form of graphs is known. We showed that the resulting constraint can be cast as an atomic norm, thus allowing us to use specialized solvers developed for this purpose. Experiments on both simulated and real data served to establish two things: a) using auxiliary graph information helps in achieving significantly superior performance compared to the standard methods, and b) biased matrix completion method with graph information is useful in practice and achieves low RMSEs in typical PU learning settings.

[1] http://omim.org

References

[1] Jacob Abernethy, Francis Bach, Theodoros Evgeniou, and Jean-Philippe Vert. Low-rank matrix factorization with attributes. *arXiv preprint cs/0611124*, 2006.

[2] Albert-László Barabási, Natali Gulbahce, and Joseph Loscalzo. Network medicine: a network-based approach to human disease. *Nature Reviews Genetics*, 12(1):56–68, 2011.

[3] Jian-Feng Cai, Emmanuel J Candès, and Zuowei Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization*, 20(4):1956–1982, 2010.

[4] Venkat Chandrasekaran, Benjamin Recht, Pablo A Parrilo, and Alan S Willsky. The convex geometry of linear inverse problems. *Foundations of Computational mathematics*, 12(6):805–849, 2012.

[5] Mark A Davenport, Yaniv Plan, Ewout van den Berg, and Mary Wootters. 1-bit matrix completion. *Information and Inference*, 3(3):189–223, 2014.

[6] Charles Elkan and Keith Noto. Learning classifiers from only positive and unlabeled data. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 213–220. ACM, 2008.

[7] Nivit Gill, Shailendra Singh, and Trilok C Aseri. Computational disease gene prioritization: an appraisal. *Journal of Computational Biology*, 21(6):456–465, 2014.

[8] Cho-Jui Hsieh, Nagarajan Natarajan, and Inderjit S Dhillon. PU learning for matrix completion. *arXiv preprint arXiv:1411.6081*, 2014.

[9] Martin Jaggi. Revisiting frank-wolfe: Projection-free sparse convex optimization. In *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, pages 427–435, 2013.

[10] Vassilis Kalofolias, Xavier Bresson, Michael Bronstein, and Pierre Vandergheynst. Matrix completion on graphs. (EPFL-CONF-203064), 2014.

[11] Hao Ma, Dengyong Zhou, Chao Liu, Michael R Lyu, and Irwin King. Recommender systems with social regularization. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pages 287–296. ACM, 2011.

[12] Nagarajan Natarajan and Inderjit S Dhillon. Inductive matrix completion for predicting gene–disease associations. *Bioinformatics*, 30(12):i60–i68, 2014.

[13] Nikhil Rao, Parikshit Shah, and Stephen Wright. Forward-backward greedy algorithms for atomic norm regularization. *arXiv preprint arXiv:1404.5692*, 2014.

[14] Nikhil Rao, Parikshit Shah, Stephen Wright, and Robert Nowak. A greedy forward-backward algorithm for atomic norm constrained minimization. In *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pages 5885–5889. IEEE, 2013.

[15] Nikhil Rao, Hsiang-Fu Yu, Pradeep Ravikumar, and Inderjit Dhillon. Collaborative filtering with graph information: Consistency and scalable methods. *NIPS (to appear)*, 2015.

[16] U Martin Singh-Blom, Nagarajan Natarajan, Ambuj Tewari, John O Woods, Inderjit S Dhillon, and Edward M Marcotte. Prediction and validation of gene-disease associations using methods inspired by social network analyses. 2013.

[17] Ambuj Tewari, Pradeep K Ravikumar, and Inderjit S Dhillon. Greedy algorithms for structurally constrained high dimensional problems. In *Advances in Neural Information Processing Systems*, pages 882–890, 2011.

[18] Zhou Zhao, Lijun Zhang, Xiaofei He, and Wilfred Ng. Expert finding for question answering via graph regularized matrix completion. *Knowledge and Data Engineering, IEEE Transactions on*, PP(99), 2014.

[19] Tinghui Zhou, Hanhuai Shan, Arindam Banerjee, and Guillermo Sapiro. Kernelized probabilistic matrix factorization: Exploiting graphs and side information. In *SDM*, volume 12, pages 403–414. SIAM, 2012.