IMPUTATION OF COUPLED TENSORS AND GRAPHS

V. N. Ioannidis, A. S. Zamzam*, G. B. Giannakis*, and N. D. Sidiropoulos†*

* ECE Dept. and Digital Tech. Center, Univ. of Minnesota, Mpls, MN 55455, USA
† ECE Dept., Univ. of Virginia, Charlottesville, VA 22903, USA
E-mails: {ioann006,ahmedz , georgios}@umn.edu, nikos@virginia.edu

ABSTRACT

Joint analysis of data from different sources can potentially improve one's ability to reveal latent structure in heterogeneous datasets. For instance, social network activities and user demographic information can be leveraged to improve recommendations. However, the incompleteness and heterogeneity of the data challenge joint factorization of multiple datasets. Aspiring to address these challenges, the coupled graph tensor factorization model accounts for side information available in the form of correlation matrices or graphs. Here, a novel ADMM-based approach is put forth to impute missing entries and unveil hidden structure in the data. The iterative solver enjoys closed-form updates that result in reduced computational complexity. Numerical tests with synthetic and real data corroborate the merits of the proposed method relative to competing alternatives.

***Index Terms*—** Non-negative factorization, parallel factor (PARAFAC)/ canonical polyadic decomposition (CPD) model, imputation, recommender systems

1. INTRODUCTION

Multi-relational data emerge in diverse applications such as social networks, recommender systems, biomedical imaging, computer vision and communication networks, where pertinent datasets are typically modeled as high-order tensors [1]. In several real settings however, only a subset of the data can be observed due to application-specific restrictions. For example, the ratings of new users in recommender systems are missing; in social applications individuals may be reluctant to share personal information due to privacy concerns; brain data may contain misses due to inadequate spatial resolution. In this context, a task of paramount importance is the imputation of the missing entries given the available data.

Oftentimes, side information is available that captures relations among items in a mode of the tensor. For instance, item correlation matrices may be available or estimated, and these correlations reflect an underlying graph structure [2]. Analyzing data from multiple sources jointly endows the imputation task with extra prediction capabilities. In recommender systems one may benefit from available user-user interactions over a social network to impute the missing ratings and facilitate profitable recommendations to new costumers. The present paper develops a novel algorithm for joint factorization of tensors and graphs with missing entries.

Related work. The so-termed coupled matrix tensor factorization (CMTF) involves matrices and tensors that are assumed to share factors [2–4]. Typically, CMTF approaches assume a low-rank model for the tensor that captures the "regularity" present among data in order to recover the missing entries [5]. Misses in both the side information and the tensor were handled in [2,3], but the case of graph matrices was not considered – and it requires judicious modeling. Similarity matrices have been used as regularizers for tensor factorization problems [5]. Assuming that the underlying low-rank factors follow a simplified distribution allows for incorporation of the correlation information in a Bayesian framework [5]. Albeit interesting this approach assumes that the similarity matrices are fully observable, which is not the case in several applications. In a social network for example, not all users will provide their social network connections. On the other hand, [6] accounts for graph data and allows for misses in both the matrices and the tensor. Unfortunately, [6] pursues the factors utilizing a first-order method which slows down the convergence rate and becomes computationally expensive when as the size of tensors and side information grows large.

In this paper, a novel algorithm based on the alternating direction method of multipliers (ADMM) for coupled graph and tensor factorization (CGTF) is proposed in order to impute missing entries in both the matrices and the tensors. The proposed model can handle the so called *cold start* challenge where an entire slab is missing from the tensor. The merits of our approach are as follows. First, during each ADMM iteration, we avoid solving constrained optimization problems, resulting in considerably lower computational complexity per iteration compared to constrained least-squares based algorithms. Second, our ADMM algorithm is suitable for accelerated solvers in [3] which exploit data sparsity. Finally, our approach can easily incorporate many other types of constraints on the latent factors, such as sparsity.

The work in this paper was supported by NSF grants 171141, 1500713, and 1442686.

2. MODEL AND PROBLEM FORMULATION

Consider a tensor $\underline{\mathbf{X}}$ of order N and size $I_1 \times I_2 \times \cdots \times I_N$. An entry of $\underline{\mathbf{X}}$ is denoted by $\underline{\mathbf{X}}(i_1, i_2, \cdots, i_N)$, where each index i_k refers to the kth-mode of the tensor. The focus of this paper is tensors with positive entries that appear in diverse applications such as recommendation systems, finance, or biology. The mode-n matricization of $\underline{\mathbf{X}}$ is denoted by the matrix $\mathbf{X}_n$, which arranges the mode-n one-dimensional fibers as columns of the resulting matrix; see [1] for details. Without loss of generality, we will focus on the case of 3-way tensors $\underline{\mathbf{X}} \in \mathbb{R}_+^{I_1 \times I_2 \times I_3}$. In several real settings, tensors are low rank and hence are expressed via the well-known parallel factor (PARAFAC) decomposition [1, 5]

$$\underline{\mathbf{X}}(i_1, i_2, i_3) = \sum_{r=1}^{R} \mathbf{A}_1(i_1, r)\mathbf{A}_2(i_2, r)\mathbf{A}_3(i_3, r) + \underline{\mathbf{E}}(i_1, i_2, i_3)$$

where R is the rank of the tensor, $\{\mathbf{A}_n \in \mathbb{R}_+^{I_n \times R}\}_{n=1}^3$ represent the low-rank factor matrices corresponding to the three modes of the tensor and $\underline{\mathbf{E}} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ captures model mismatch. The PARAFAC model is written in tensor-matrix form as follows

$$\underline{\mathbf{X}} = \big[\!\big[\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3\big]\!\big] + \underline{\mathbf{E}} \tag{1}$$

where $\big[\!\big[\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3\big]\!\big]$ is the outer product of these matrices resulting in a tensor. Oftentimes, only a subset of entries of $\underline{\mathbf{X}}$ is observable. Thus, we write $\underline{\mathbf{X}} = \underline{\mathbf{X}}^A + \underline{\mathbf{X}}^M$, where $\underline{\mathbf{X}}^A$ contains the observed tensor entries and otherwise is zero, and $\underline{\mathbf{X}}^M$ holds the missing values and zeros elsewhere.

Additionally, consider a set of graph (similarity) matrices $\{\mathbf{G}_n \in \mathbb{R}_+^{I_n \times I_n}\}_{n=1}^3$. The (i, j)-th entry of $\mathbf{G}_n$ reflects the similarity between the i-th and j-th data items of the n-th tensor mode, and thus $\mathbf{G}_n$ can be viewed as an adjacency matrix capturing the connectivity of an undirected graph. This prior information for the tensor items is well-motivated since network data are available in numerous disciplines including sociology, biology, neuroscience and engineering. In these applications, subsets of items form communities in the sense that they exhibit dense intra-connections and sparse inter-connections, which is captured by $\mathbf{G}_n$. For example, this property is common in social networks, where friends tend to form dense clusters [7]. Thus, we adopt a symmetric nonnegative matrix factorization (SNMF) model [8], that enables finding identifiable factors and recovering clusters of nodes. Specifically, we advocate the following diagonally-scaled SNMF model

$$\mathbf{G}_n = \mathbf{A}_n \operatorname{diag}(\mathbf{d}_n)\mathbf{A}_n^\top + \mathbf{V}_n,\ n = 1, 2, 3 \tag{2}$$

where $\{\mathbf{V}_n \in \mathbb{R}^{I_n \times I_n}\}_n$ capture modeling error and $\{\mathbf{d}_n \in \mathbb{R}_+^{R \times 1}\}_n$ weight the factor matrices. Notice that the factors $\{\mathbf{A}_n\}_n$ are shared between the tensor and the graph of the corresponding item justifying the name of the proposed model as coupled graph tensor factorization. Whereas classical CMTF approaches model the side information as $\mathbf{A}_n\mathbf{B}_n^\top$, the novel CGTF captures the graph structure by employing $\mathbf{A}_n \operatorname{diag}(\mathbf{d}_n)\mathbf{A}_n^\top$. Adding the diagonal loading matrices endows the model with the ability to adjust the relative weighting differences between the tensor and the side information matrices. Unfortunately, the network topologies may contain missing entries (links), which can be attributed to privacy concerns in social networks, or sampling constraints in massive networks. Hence, the graph matrices are modeled as $\mathbf{G}_n = \mathbf{G}_n^A + \mathbf{G}_n^M$, where $\mathbf{G}_n^A$ contains the observed links and $\mathbf{G}_n^M$ holds the missing values.

Problem statement. The broad goal of this paper is the joint recovery of $\underline{\mathbf{X}}^M$ and $\{\mathbf{G}_n^M\}_{n=1}^3$ from the available values $\underline{\mathbf{X}}^A$ and $\{\mathbf{G}_n^A\}_{n=1}^3$ by employing the proposed CGTF model.

3. COUPLED GRAPH TENSOR FACTORIZATION

Given (1) and (2), this section develops a novel approach to infer $\underline{\mathbf{X}}^M$ and $\{\mathbf{G}_n^M\}_{n=1}^3$. In order to find the latent factors that jointly approximate the tensor and the graph matrices, the following optimization problem is put forth

$$\begin{aligned} \underset{\underline{\mathbf{X}}^M, \{\mathbf{A}_n, \mathbf{d}_n, \mathbf{G}_n^M\}_{n=1}^3}{\text{minimize}} \quad & \|\underline{\mathbf{X}} - \big[\!\big[\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3\big]\!\big]\|_F^2 \\ & + \mu \sum_{n=1}^{3} \|\mathbf{G}_n - \mathbf{A}_n \operatorname{diag}(\mathbf{d}_n)\mathbf{A}_n^\top\|_F^2 \\ \text{s. t.} \quad & \mathbf{A}_n \geq \mathbf{0},\ \mathbf{d}_n \geq \mathbf{0}, \\ & \underline{\mathbf{X}} = \underline{\mathbf{X}}^A + \underline{\mathbf{X}}^M,\ \mathbf{G}_n = \mathbf{G}_n^A + \mathbf{G}_n^M, \\ & \mathcal{P}_{\underline{\Omega}}(\underline{\mathbf{X}}^M) = \underline{\mathbf{0}},\ \mathcal{P}_{\Omega_n}(\mathbf{G}_n^M) = \mathbf{0},\ n = 1, 2, 3 \end{aligned} \tag{3}$$

where $\mu > 0$ tunes the relative importance of the fit between the tensor and the graph matrices. The first term accounts for the LS fitting error of the PARAFAC model (1), and the second sum of LS costs accounts for the SNMF model (2). The positivity constraints stem from prior knowledge related to the factor and diagonal matrices. The equality conditions constrain $\underline{\mathbf{X}}$ and $\{\mathbf{G}_n\}_{n=1}^3$ to be equal to $\underline{\mathbf{X}}^A$ and $\{\mathbf{G}_n^A\}_{n=1}^3$ at the observed entries and to the optimization variables $\underline{\mathbf{X}}^M$ and $\{\mathbf{G}_n^M\}_{n=1}^3$ otherwise. The operators $\mathcal{P}_{\underline{\Omega}}$ and $\mathcal{P}_{\Omega_n}$ force the optimization variables to be zero at the observed entries.

The optimization problem in (3) is non-convex due to the trilinear terms $\big[\!\big[\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3\big]\!\big]$ and $\mathbf{A}_n \operatorname{diag}(\mathbf{d}_n)\mathbf{A}_n^\top$. The next section develops an efficient solver for (3) based on the ADMM [1].

3.1. ADMM for CGTF

First notice that the optimization problem (3) is even non-convex for each $\mathbf{A}_n$ separately due to the product of factor matrices in the SMNF model. This poses an additional challenge to an ADMM algorithm that iteratively pursues per block minimizers of the augmented Lagrangian. Hence, we introduce $\{\bar{\mathbf{A}}_n\}_n$ auxiliary variables and rewrite the SMNF cost

as $\|\mathbf{G}_n - \mathbf{A}_n \operatorname{diag}(\mathbf{d}_n)\bar{\mathbf{A}}_n^\top\|_F^2$. Furthermore, to handle the positivity constraints we introduce

$$g(\mathbf{M}) = \begin{cases} 0, & \text{if } \mathbf{M} \geq \mathbf{0} \\ \infty, & \text{otherwise} \end{cases} \tag{4}$$

and the auxiliary variables $\{\tilde{\mathbf{A}}_n, \tilde{\mathbf{d}}_n\}_n$. Next, we rewrite (3) as the following equivalent optimization problem

$$\begin{aligned} \underset{\substack{\underline{\mathbf{X}}^M, \{\mathbf{A}_n, \bar{\mathbf{A}}_n, \tilde{\mathbf{A}}_n, \\ \mathbf{d}_n, \tilde{\mathbf{d}}_n, \mathbf{G}_n^M\}_{n=1}^3}}{\text{minimize}} & \|\underline{\mathbf{X}} - \big[[\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3]\big]\|_F^2 + \sum_{f=1}^3 g(\tilde{\mathbf{A}}_n) \\ + \mu \sum_{n=1}^3 & \|\mathbf{G}_n - \mathbf{A}_n \operatorname{diag}(\mathbf{d}_n)\bar{\mathbf{A}}_n^\top\|_F^2 + \sum_{f=1}^3 g(\tilde{\mathbf{d}}_n) \\ \text{s. t.} \quad & \mathbf{A}_n = \bar{\mathbf{A}}_n,\ \mathbf{A}_n = \tilde{\mathbf{A}}_n,\ \mathbf{d}_n = \tilde{\mathbf{d}}_n, \\ & \underline{\mathbf{X}} = \underline{\mathbf{X}}^A + \underline{\mathbf{X}}^M,\ \mathbf{G}_n = \mathbf{G}_n^A + \mathbf{G}_n^M, \\ & \mathcal{P}_{\underline{\Omega}}(\underline{\mathbf{X}}^M) = \underline{\mathbf{0}},\ \mathcal{P}_{\Omega_n}(\mathbf{G}_n^M) = \mathbf{0},\ n = 1, 2, 3. \end{aligned} \tag{5}$$

Even though (5) is still non-convex in all the variables, it is convex with respect to each block variable separately. Towards deriving an ADMM solver, we introduce the dual variables $\{\mathbf{Y}_{\bar{\mathbf{A}}_n} \in \mathbb{R}^{I_n \times R}, \mathbf{Y}_{\tilde{\mathbf{A}}_n} \in \mathbb{R}^{I_n \times R}, \mathbf{y}_{\tilde{\mathbf{d}}_n} \in \mathbb{R}^{I_n \times R}\}_n$ and the penalty terms $\{\rho_{\bar{\mathbf{A}}_n} > 0, \rho_{\tilde{\mathbf{A}}_n} > 0, \rho_{\tilde{\mathbf{d}}_n} > 0\}_n$.

The augmented Lagrangian is given by (6), at the bottom of the page, where $f(\cdot)$ is the objective function in (5). For ease of notation, no ADMM superscripts will be used in the following equations. Without loss of generality, only the ADMM updates for $n = 1$ will be presented. The update for $\mathbf{A}_1$ can be obtained by taking the derivative of L_ρ with respect to $\mathbf{A}_1$ and equating it to zero to obtain

$$\begin{aligned} &\hat{\mathbf{A}}_1(\mathbf{M}_1^\top\mathbf{M}_1 + \mu\mathbf{D}_1\bar{\mathbf{A}}_1^\top\bar{\mathbf{A}}_1\mathbf{D}_1 + (\rho_{\tilde{\mathbf{A}}_1} + \rho_{\bar{\mathbf{A}}_1})\mathbf{I}) \\ &:= \mathbf{X}_1\mathbf{M}_1 + \mu\mathbf{G}_1\bar{\mathbf{A}}_1\mathbf{D}_1 + \rho_{\bar{\mathbf{A}}_1}\bar{\mathbf{A}}_1 + \rho_{\tilde{\mathbf{A}}_1}\tilde{\mathbf{A}}_1 - \mathbf{Y}_{\tilde{\mathbf{A}}_1} - \mathbf{Y}_{\bar{\mathbf{A}}_1} \end{aligned} \tag{7a}$$

where $\mathbf{M}_1 := \mathbf{A}_3 \odot \mathbf{A}_2$, and $\mathbf{D}_1 := \operatorname{diag}(\mathbf{d}_1)$. The update for $\mathbf{d}_1$ can be obtained in a similar manner as

$$\begin{aligned} &((\bar{\mathbf{A}}_1 \odot \mathbf{A}_1)^\top(\mu\bar{\mathbf{A}}_1 \odot \mathbf{A}_1) + \rho_{\tilde{\mathbf{d}}_1}\mathbf{I})\hat{\mathbf{d}}_1 \\ &:= \mu(\bar{\mathbf{A}}_1 \odot \mathbf{A}_1)^\top\mathbf{g}_1 + \rho_{\tilde{\mathbf{d}}_1}\tilde{\mathbf{d}}_1 - \mathbf{y}_{\tilde{\mathbf{d}}_1}. \end{aligned} \tag{7b}$$

Accordingly, the update for $\bar{\mathbf{A}}_1$ is given by

$$\begin{aligned} &\hat{\bar{\mathbf{A}}}_1(\mu\mathbf{D}_1\mathbf{A}_1^\top\mathbf{A}_1\mathbf{D}_1 + \rho_{\bar{\mathbf{A}}_1}\mathbf{I}) \\ &:= \mu\mathbf{G}_1^\top\mathbf{A}_1\mathbf{D}_1 + \rho_{\bar{\mathbf{A}}_1}\mathbf{A}_1 + \mathbf{Y}_{\bar{\mathbf{A}}_1}. \end{aligned} \tag{7c}$$

The auxiliary variables $\tilde{\mathbf{A}}_1, \tilde{\mathbf{d}}_1$ are updated by projecting to the nonnegative orthant; that is,

$$\begin{aligned} \hat{\tilde{\mathbf{A}}}_1 &:= \left(\mathbf{A}_1 + \frac{1}{\rho_{\tilde{\mathbf{A}}_1}}\mathbf{Y}_{\tilde{\mathbf{A}}_1}\right)_+ \\ \hat{\tilde{\mathbf{d}}}_1 &:= \left(\mathbf{d}_1 + \frac{1}{\rho_{\tilde{\mathbf{d}}_1}}\mathbf{y}_{\tilde{\mathbf{d}}_1}\right)_+. \end{aligned} \tag{7d}$$

Using the estimated factors $\{\hat{\mathbf{A}}_n\}_n$ the updates for the missing tensor elements are found as

$$\hat{\underline{\mathbf{X}}}^M := \mathcal{P}_{\underline{\Omega}}(\big[[\hat{\mathbf{A}}_1, \hat{\mathbf{A}}_2, \hat{\mathbf{A}}_3]\big]). \tag{7e}$$

Similarly, the missing links in $\mathbf{G}_1$ can be obtained as

$$\hat{\mathbf{G}}_1^M := \mathcal{P}_{\Omega_1}(\hat{\mathbf{A}}_1 \operatorname{diag}(\hat{\mathbf{d}}_1)\hat{\bar{\mathbf{A}}}_1^\top). \tag{7f}$$

Finally, the updates for the Lagrange multipliers are given by

$$\begin{aligned} \mathbf{Y}_{\bar{\mathbf{A}}_1} &= \mathbf{Y}_{\bar{\mathbf{A}}_1} + \rho_{\bar{\mathbf{A}}_1}(\mathbf{A}_1 - \bar{\mathbf{A}}_1) \\ \mathbf{Y}_{\tilde{\mathbf{A}}_1} &= \mathbf{Y}_{\tilde{\mathbf{A}}_1} + \rho_{\tilde{\mathbf{A}}_1}(\mathbf{A}_1 - \tilde{\mathbf{A}}_1) \\ \mathbf{y}_{\tilde{\mathbf{d}}_1} &= \mathbf{y}_{\tilde{\mathbf{d}}_1} + \rho_{\tilde{\mathbf{d}}_1}(\mathbf{d}_1 - \tilde{\mathbf{d}}_1). \end{aligned} \tag{7g}$$

The steps of our CGTF algorithm are listed in Algorithm 1. Since (5) is a non-convex problem, a judicious initialization of $\{\mathbf{A}_n\}_n$ is required. Towards that end, we adopt an efficient algorithm for SNMF, see e.g. [9], to initialize the factor matrices using only the available elements in the corresponding graphs $\{\mathbf{G}_n^A\}$, while $\{\mathbf{d}_n\}$ are initialized by setting all entries equal to one. Since SNMF is unique under certain conditions, the initialization is likely to be a good one [9]. The ADMM algorithm stops when the primal residuals and the dual feasibility residuals are sufficiently small.

The advantage of introducing the auxiliary variables is twofold. First, by employing $\bar{\mathbf{A}}_n$, we bypass solving the non-convex SNMF that would require a costly iterative algorithm per factor update. Second, by introducing $\{\tilde{\mathbf{A}}_n, \tilde{\mathbf{d}}_n\}$, we avoid solving a constrained optimization problem, resulting in a more computationally affordable update compared to constrained least-squares based algorithms. In a nutshell, our novel reformulation allows for closed-form updates per step of the ADMM solver. Even though $\{\tilde{\mathbf{A}}_n, \tilde{\mathbf{d}}_n\}_n$ are by construction non-negative, $\{\mathbf{A}_n, \bar{\mathbf{A}}_n, \mathbf{d}_n\}_n$ are not necessarily non-negative, but they become so upon convergence. Extensive simulation on synthetic and real data validate that Algorithm 1 converges, while theoretical guarantees for convergence will be provided in a subsequent journal version.

$$\begin{aligned} L_\rho\big(\underline{\mathbf{X}}^M, \{\mathbf{A}_n, \bar{\mathbf{A}}_n, \tilde{\mathbf{A}}_n, \mathbf{d}_n, \tilde{\mathbf{d}}_n, \mathbf{G}_n^M, \mathbf{Y}_{\bar{\mathbf{A}}_n}, \mathbf{Y}_{\tilde{\mathbf{A}}_n}, \mathbf{y}_{\tilde{\mathbf{d}}_n}\}_{n=1}^3\big) &:= f\big(\underline{\mathbf{X}}^M, \{\mathbf{A}_n, \bar{\mathbf{A}}_n, \tilde{\mathbf{A}}_n, \mathbf{d}_n, \tilde{\mathbf{d}}_n, \mathbf{G}_n^M\}_{n=1}^3\big) \\ + \sum_{f=1}^3 \big(\operatorname{Tr}(\mathbf{Y}_{\bar{\mathbf{A}}_n}\mathbf{A}_n^\top) + \frac{\rho_{\bar{\mathbf{A}}_n}}{2}\|\mathbf{A}_n - \bar{\mathbf{A}}_n\|_F^2 + \operatorname{Tr}(\mathbf{Y}_{\tilde{\mathbf{A}}_n}\mathbf{A}_n^\top) + \frac{\rho_{\tilde{\mathbf{A}}_n}}{2}\|\mathbf{A}_n - \tilde{\mathbf{A}}_n\|_F^2 + \operatorname{Tr}(\mathbf{y}_{\tilde{\mathbf{d}}_n}\mathbf{d}_n^\top) + \frac{\rho_{\tilde{\mathbf{d}}_n}}{2}\|\mathbf{d}_n - \tilde{\mathbf{d}}_n\|_F^2\big). \end{aligned} \tag{6}$$

Algorithm 1 ADMM for CGTF

Input: $\underline{\mathbf{X}}^A$ and $\{\mathbf{G}_n^A\}_{n=1}^3$

Intialization: SNMF for $\{\mathbf{A}_n\}_n$ using [9].
while iterates not converge **do**
  Update $\mathbf{A}_n$ using (7a).
  Update $\mathbf{d}_n$ using (7b).
  Update $\bar{\mathbf{A}}_n$ using (7c).
  Update $\{\tilde{\mathbf{A}}_n, \tilde{\mathbf{d}}_n\}$ using (7d).
  Update $\hat{\underline{\mathbf{X}}}^M$ using (7e).
  Update $\hat{\mathbf{G}}_n^M$ using (7f).
  Update Lagrangian multipliers using (7g).
end while

Output: $\hat{\underline{\mathbf{X}}}^M, \{\hat{\mathbf{A}}_n, \hat{\mathbf{d}}_n, \hat{\mathbf{G}}_n^M\}_n$

4. NUMERICAL TESTS

This section evaluates the performance of the proposed CGTF on synthetic and real data. The compared approaches include the CANDECOMP/PARAFAC Weighted OPTimization (CP_WOPT) algorithm [4]; the nonnegative tensor factorization (NTF) implemented in [10]; and the CMTF [2]. The algorithms were initialized using the proposed SNMF scheme, which enhances the performance of all methods.

4.1. Synthetic Data

Synthetic tensor $\underline{\mathbf{X}} \in \mathbb{R}^{35\times35\times30}$ with $R = 4$ was generated according to the PARAFAC model (1), where the true factors $\{\mathbf{A}_n\}_{n=1}^3$ are drawn from a standard normal distribution. Matrices $\{\mathbf{G}_n\}_{n=1}^3$ were generated using the SMNF (2).

To evaluate the performance of the various factorization algorithms, the entries of $\underline{\mathbf{X}}$ were corrupted with random i.i.d Gaussian noise. Fig. 1 depicts the normalized mean squared error NMSE:$= \sum_{i_3=1}^{I_3} \|\hat{\underline{\mathbf{X}}}(:,:,i_3) - \underline{\mathbf{X}}(:,:,i_3)\|_F^2 / \sum_{i_3=1}^{I_3} \|\underline{\mathbf{X}}(:,:,i_3)\|_F^2$ against the signal to noise ratio (SNR). The novel CGTF exploits the structure of the graph matrices and achieves superior performance relative to the competing methods. The large performance difference can be explained since the NTF and CP_WOPT approaches do not exploit the side information, while the CMTF method does not exploit the special structure of graph matrices.

4.2. Real Data

A real recommendation dataset was employed that comprises a three-way tensor indicating the frequency of a user performing an activity at a location [11]. It contains information about 164 users, 168 locations and 5 activities. A binary tensor $\underline{\mathbf{X}}$ is constructed to represent the links between the users, the locations and the activities. In other words, $\underline{\mathbf{X}}(i_1, i_2, i_3)$ equals 1 if user i_1 visited location i_2 and performed activity i_3; otherwise, it is 0. Additionally, similarity matrices between

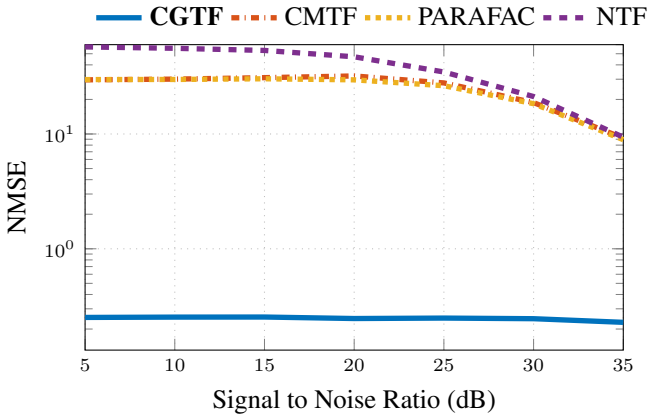

Fig. 1: Tensor imputation performance based on NMSE. $(\{\rho_{\bar{\mathbf{A}}_n} = 100, \rho_{\tilde{\mathbf{A}}_n} = 100, \rho_{\tilde{\mathbf{d}}_n} = 100\}_n, \mu = 1)$.

the users and the activities are provided. The similarity value between two locations is defined by inner product between the corresponding feature vectors. The dataset is missing social network information for 28 users, and feature vectors for 32 locations. For all approaches the tensor rank was set to $R = 4$.

Table 1 shows the NMSE for different percentages of missing data for the tensor. The CGTF model exploits judiciously the structure of the available graph information and hence our efficient ADMM solver outperforms competing alternatives, and facilitates improved recommendations.

Missing	NTF	CP_WOPT	CMTF	Algorithm 1
40%	2.0815	0.9517	0.975	**0.48**
50%	–	0.9574	1.001	**0.75**

Table 1: NMSE for different ratios of missing data. $(\{\rho_{\bar{\mathbf{A}}_n} = 100, \rho_{\tilde{\mathbf{A}}_n} = 100, \rho_{\tilde{\mathbf{d}}_n} = 100\}_n, \mu = 10^{-4})$.[1]

5. CONCLUSIONS AND FUTURE WORK

This paper investigates the imputation of missing entries in tensors and graphs based on a novel CGTF model. An efficient algorithm is developed to identify the factor matrices and recover the missing entries. The ADMM solver enjoys closed-form updates and is amenable to parallel and accelerated implementation. Also, the proposed method is the only one among the ones considered here that can overcome the so-called *cold-start* challenge, where the tensor has missing slabs or the similarity matrices are not complete. The novel algorithm makes accurate prediction of the missing values and can be employed in several real world settings, especially in recommendation systems. Our future research agenda includes refinement of the optimization algorithm to scale for tensors with large dimensions; extending the proposed model to handle anomalies in the data; and employing the factors recovered by the CGTF model for detecting communities in the graphs.

[1] The NTF model does not provide meaningful results for high percentages of missing value.

6. REFERENCES

[1] N. D. Sidiropoulos, L. D. Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos, "Tensor decomposition for signal processing and machine learning," *IEEE Trans. Sig. Process.*, vol. 65, no. 13, pp. 3551–3582, Jul. 2017.

[2] B. Ermiş, E. Acar, and A. T. Cemgil, "Link prediction in heterogeneous data via generalized coupled tensor factorization," *Springer, Data Mining and Knowledge Discovery*, vol. 29, no. 1, pp. 203–236, Jan. 2015.

[3] E. E. Papalexakis, C. Faloutsos, T. M. Mitchell, P. P. Talukdar, N. D. Sidiropoulos, and B. Murphy, "Turbo-SMT: Accelerating Coupled Sparse Matrix-Tensor Factorizations by 200x." in *Proc. of SIAM Intl. Conf. on Data Mining*, Philadelphia, PA, Apr. 2014, pp. 118–126.

[4] E. Acar, D. M. Dunlavy, T. G. Kolda, and M. Mørup, "Scalable Tensor Factorizations with Missing Data." in *Proc. of SIAM Intl. Conf. on Data Mining*, Columbus, OH, Apr. 2010, pp. 701–712.

[5] J. A. Bazerque, G. Mateos, and G. Giannakis, "Rank regularization and Bayesian inference for tensor completion and extrapolation," *IEEE Trans. Sig. Process.*, vol. 61, no. 22, pp. 5689–5703, Nov. 2013.

[6] A. S. Zamzam, V. N. Ioannidis, and N. D. Sidiropoulos, "Coupled graph tensor factorization," in *Proc. Asilomar Conf. Sig., Syst., Comput.*, Pacific Grove, CA, Nov. 2016, pp. 1755–1759.

[7] F. Sheikholeslami and G. B. Giannakis, "Overlapping community detection via constrained parafac: A divide and conquer approach," in *Proc. of IEEE Intl. Conf. on Data Mining*, New Orleans, LA, Nov. 2017, pp. 127–136.

[8] D. Kuang, H. Park, and C. H. Ding, "Symmetric Nonnegative Matrix Factorization for Graph Clustering." in *Proc. of SIAM Intl. Conf. on Data Mining*, vol. 12, Anaheim, CA, Apr. 2012, pp. 106–117.

[9] K. Huang, N. D. Sidiropoulos, and A. Swami, "Non-negative matrix factorization revisited: Uniqueness and algorithm for symmetric decomposition," *IEEE Trans. Sig. Process.*, vol. 62, no. 1, pp. 211–224, Oct. 2014.

[10] C. A. Andersson and R. Bro, "The N-way toolbox for MATLAB," *Chemometrics and Intelligent Laboratory Systems*, vol. 52, no. 1, pp. 1–4, Aug. 2000.

[11] V. W. Zheng, B. Cao, Y. Zheng, X. Xie, and Q. Yang, "Collaborative filtering meets mobile recommendation: A user-centered approach," in *Proc. of ACM Conf. on Artificial Intelligence*, Atlanta, GA, Jul. 2010.