2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)

C^3APTION: Constrainted Coupled CP And PARAFAC2 Tensor Decomposition

Ekta Gujral
University of California Riverside
Email: egujr001@ucr.edu

Georgios Theocharous
Adobe Research
Email: theochar@adobe.com

Evangelos E. Papalexakis
University of California Riverside
Email: epapalex@cs.ucr.edu

***Abstract*—Given data from a variety of sources that share a number of dimensions, how can we effectively decompose them jointly into interpretable latent factors? The coupled tensor decomposition framework captures this idea by jointly supporting the decomposition of several CP tensors. However, coupling tends to suffer when one dimension of data is irregular, i.e., one of the dimensions of the tensor is uneven, such as in the case of PARAFAC2. In this work, we provide a scalable method for decomposing coupled CP and PARAFAC2 tensor datasets through non-negativity-constrained least squares optimization on a variety of objective functions. We offer the following contributions: (1) Our algorithm can perform coupled factorization with an active-set, block principal pivoting and least square optimization method including the Frobenius norm induced non-negative factorization. (2) C^3APTION scales to billions of non-zero elements in both the data and model. Comprehensive experiments on large data confirmed that C^3APTION is up to $5\times$ faster and $70-80\%$ accurate than several baselines. We present results showing the scalability of this novel implementation on a billion elements as well as demonstrate the high level of interpretability in the latent factors produced, implying that coupling is indeed a promising framework for large-scale, unsupervised pattern exploration and cluster discovery.**

I. INTRODUCTION

With the opportunity to handle large volumes and velocity of data as a result of recent technical developments, such as mobile connectivity [25], digital tools [24], biomedical technology [7] and modern medical testing techniques [11], we face multi-source and multi-view data [14], [15] sets. Suppose, for example, that we are given a health care record data, such as Centers for Medicare and Medicaid (CMS) [11], and we have information about patient who visited hospital, or who got what kind of diagnosis in which visit, and when. This data may be formulated as a three mode PARAFAC2 tensor. Suppose now that we also have some static features information pertaining to the patient, e.g. multi-aspect relation based on demographic information. This data may be formulated as a three mode CP tensor. This problem can be formulated as an example of a coupled factorization, where the two tensors of a 3-mode (visits, diagnosis, patients) PARAFAC2 and a 3-mode (patients, patients, aspect) CP tensor share a common dimension.

In many practical cases, we have multi-aspect information represented as tensors. Despite its attractiveness, individual tensor factorization suffers from robustness issues. Applying

IEEE/ACM ASONAM 2020, December 7-10, 2020
978-1-7281-1056-1/20/$31.00 © 2020 IEEE

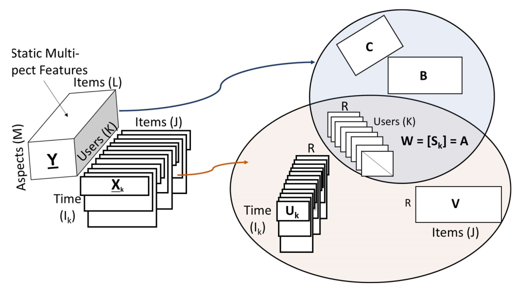

Figure. 1: Illustration of CAPTION decomposition. Each slice of $\underline{\mathbf{X}}_k$ represents the different clinical visits for patient k. CP tensor $\underline{\mathbf{Y}}$ includes the similarity CP tensor based on demographic information of patients. C^3APTION decomposes $\underline{\mathbf{X}}_k$ into three parts: $\mathbf{X}_k$, $\mathbf{W} = \mathbf{S}_k$, and $\mathbf{V}$. CP tensor $\underline{\mathbf{Y}}$ is decomposed into $\mathbf{W} = \mathbf{S}_k$, $\mathbf{B}$ and $\mathbf{C}$. Note that latent factor $\mathbf{W}$ is shared between both tensors.

coupled tensors and matrices to heterogeneous datasets from multiple sources has been a topic of interest in many areas. Coupled tensor decomposition gives an equivalent representation of multi-way data by a set of factors and parts of the factors are shared for coupled data. In literature, fusion and coupled methods [3], [4], [5], [10], [13] reported so far ignore the underlying irregular nature of the data in at least one of the modalities among the data like health care data. Acar et al. proposed an all-at-once coupled gradients based optimization approach, called CMTF-OPT [3]. The advanced version of CMTF-OPT, ACMTF-OPT [4], places additional constraints on the model to force good performance when distinguishing between shared and unshared data latent factors. Many researchers have subsequently made improvements [27], [2], [19] to CMTF for large-scale data. In [10], paper proposed fusion or soft coupling of both EEG and fMRI PARAFAC2 data and provides insights on presence of shifts in the ERPs per subject. Similarly, [13], proposed robust coupling of two CP tensors via measuring distance between factors. Recently, Afshar et al. [5] proposed method based on block principle pivot, namely TASTE, for coupling between PARAFAC2 tensor and matrix and this method provides valuable insights for phenotyping of electronic health records. But, these prior work has either focused on a specific type of coupled factorization (two CP tensors or two PARAFAC2 tensors or tensor-matrix) or a specific objective function, thus having a limited range of potential applications where two different format of data is required.

To handle the these limitations and inspired by the work by Afshar et al. [5], we proposed a scalable method namely C^3APTION that couple CP and PARAFAC2 tensor which incorporates non-negative constraints with multiple update settings of latent factors as shown in figure 1. We demonstrate, with synthetic and real data, the advantage of the proposed method over baseline methods in terms of accuracy and computation time. A preliminary version of this work appeared in [12] as a short paper. In this paper, we extend those preliminary results by (i) providing a detailed description of all proposed methods to handle limitations of previous work, (ii) provide thorough experimentation on synthetic and real data, (iii) provide detailed case studies in real data using our proposed method, and (iv) we conduct a scalability analysis, demonstrating that our proposed method can scale up to billions of non-zero elements in data and 5× faster and 70-80% accurate than any state-of-art method. Our contributions are summarized as follows:

- **Novel and Scalable Algorithm**: We propose C^3APTION, a method of coupling the CP and PARAFAC2 tensor with various optimization update rules with non-negative constraint. Our proposed method is efficient, scalable and provides stable decompositions than baselines.
- **Fast and Accurate Algorithm**: Our proposed fitting algorithm is up to 5× faster than the state of the art baseline. At the same time, C^3APTION preserves model accuracy better than baselines while maintaining interpretability.
- **Experimental Evaluation**: we show experimental results on both synthetic and real datasets.

To promote reproducibility, we make our MATLAB implementation publicly available at link[1].

II. Preliminaries and Background

A tensor [22] is a higher order generalization of a matrix. An N-mode[2] tensor $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_1 \cdots \times I_N}$ is the outer product of N vectors, as given in equation 1,

$$\underline{\mathbf{X}} = \mathbf{a}_1 \circ \mathbf{a}_2 \cdots \circ \mathbf{a}_N \tag{1}$$

essentially indexed by N variables i.e. $(\mathbf{a}_1, \mathbf{a}_2 \ldots, \mathbf{a}_N)$. Table I contains the symbols used throughout the paper.

TABLE I: Symbols and definitions

Symbol	Definition
$\underline{\mathbf{X}}, \mathbf{X}, \mathbf{x}, x$	Tensor, Matrix, vector, scalar
$\mathbf{X}(:, \mathbf{i})$/ $\mathbf{X}(\mathbf{i}, :)$	Spans the entire i^{th} column/row of $\mathbf{X}$ (same for tensors)
$diag(\mathbf{X})$	Diagonal of matrix $\mathbf{X}$
$\mathbf{X}_k$	k^{th} frontal slice of tensor $\underline{\mathbf{X}}$
$\{\mathbf{X}_k\}$	the set of $\underline{\mathbf{X}}_k$
$*, \otimes, \odot$	Hadamard, Kronecker and Khatri-Rao product

A. Tensor and its Decomposition

Here we briefly provide insights on: (1) CP Decomposition and (2) PARAFAC2 Decomposition.

[1]http://www.cs.ucr.edu/~egujr001/ucr/madlab/src/caption_code.zip

[2]Notice that the literature (and thereby this paper) uses the above terms as well as "order" interchangeably.

*1) **CP decomposition**:* The CP decomposition [6], [9], [17] of a N-mode tensor $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_1 \cdots \times I_N}$ with Rank R is defined as the sum of outer product rank-1 components:

$$\underline{\mathbf{X}} = \sum_{r=1}^{R} \mathbf{a}_1^{(r)} \circ \mathbf{a}_2^{(r)} \cdots \circ a_N^{(r)} \tag{2}$$

Given a three-mode tensor $\underline{\mathbf{X}} \in \mathbb{R}^{I \times J \times K}$, using the CP decomposition, the tensor is decomposed into a sum of rank-one tensors, i.e., a sum of outer products of three vectors and in order to compute the decomposition, we need to solve the following problem which minimizes the sum-of-squares of the residuals:

$$\min_{\mathbf{A},\mathbf{B},\mathbf{C}} \|\underline{\mathbf{X}} - \sum_r \mathbf{A}(:,r) \circ \mathbf{B}(:,r) \circ \mathbf{C}(:,r)\|_F^2 \tag{3}$$

where $\mathbf{A} \in \mathbb{R}^{I \times R}, \mathbf{B} \in \mathbb{R}^{J \times R}, \mathbf{C} \in \mathbb{R}^{K \times R}$.

*2) **PARAFAC2 Decomposition**:* PARAFAC2 model [18] differs from CP/PARAFAC [6], [9], [17] where a low-rank trilinear model is not required. The CP decomposition applies the same factors across all the different modes, whereas PARAFAC2 allows for non-linearities such that variation across the values and/or the size of one mode as shown in Fig 1. PARAFAC2 can be written w.r.t. the frontal slices of the tensor $\underline{\mathbf{X}}$ as:

$$\underline{\mathbf{X}}_k = \mathbf{U}_k \mathbf{S}_k \mathbf{V}^T \tag{4}$$

where $k = 1, \ldots, K$, $\mathbf{U}_k \in \mathbb{R}^{I_k \times R}$, $\mathbf{S}_k = diag(W(k,:)) \in \mathbb{R}^{R \times R}$ is diagonal and $\mathbf{V} \in \mathbb{R}^{J \times R}$. Given the above modeling, the standard algorithm to solve PARAFAC2 for data $\underline{\mathbf{X}}$ tackles the following optimization problem:

$$\min_{\{\mathbf{U_k}\},\{\mathbf{S_k}\},\mathbf{V}} \sum_{k=1}^{K} \|\underline{\mathbf{X}}_k - \mathbf{U}_k \mathbf{S}_k \mathbf{V}^T\|_F^2 \tag{5}$$

subject to $\mathbf{U}_k = \mathbf{Q}_k \mathbf{H}$, $\mathbf{Q}_k^T \mathbf{Q}_k = \mathbf{I}$, and $\mathbf{S}_k$ is diagonal. The $\mathbf{U}_k$ decomposed into two matrices, $\mathbf{Q}_k$ that has orthonormal columns and $\mathbf{H}$ which is invariant regardless of k.

The Equ (5) in form of orthogonal form can be re-written as :

$$\begin{aligned} \mathcal{L} = \underset{\mathbf{Q}}{\operatorname{argmin}} \frac{1}{2} ||\underline{\mathbf{X}}_k - \mathbf{Q}_k \mathbf{H} \mathbf{S}_k \mathbf{V}^T||_2^F \quad \forall k \\ \text{subject to} \quad \mathbf{Q}_k \mathbf{Q}_k^T = \mathbf{I}_r \end{aligned} \tag{6}$$

To solve Eq (6), most common method is Alternating Least Square (ALS) that updates $\mathbf{Q}_k$ by fixing other factor matrices i.e $\mathbf{H}, \mathbf{S}_k$, and $\mathbf{V}$. The orthogonal coupling matrix $\mathbf{Q}_k$ can be obtained by Singular Value decomposition (SVD) of $(\mathbf{H}\mathbf{W}\mathbf{V}^T \underline{\mathbf{X}}_k^T) = [\mathbf{P}_n, \Sigma_n, \mathbf{Z}_n^T]$. With $\mathbf{Q}_k^T = \mathbf{P}_n \mathbf{Z}_n^T$ fixed, the rest of factors can be obtained as:

$$\begin{aligned} \mathcal{L} = \underset{\mathbf{H},\mathbf{S}_k,\mathbf{V}}{\operatorname{argmin}} \frac{1}{2} ||\mathbf{Q}_k^T \underline{\mathbf{X}}_k - \mathbf{H} \mathbf{S}_k \mathbf{V}^T||_2^F s.t. \mathbf{Q}_k \mathbf{Q}_k^T = \mathbf{I}_r \\ \underset{\mathbf{H},\mathbf{S}_k,\mathbf{V}}{\operatorname{argmin}} \frac{1}{2} ||\underline{\mathbf{Y}} - \mathbf{H} \mathbf{S}_k \mathbf{V}^T||_2^F \end{aligned} \tag{7}$$

The Eq. (7) is equivalent of solving CP decomposition of $\underline{\mathbf{Y}}$ using ALS method.

III. Proposed method: C^3APTION

A generalized CP and PARAFAC2 approach is appealing from several perspectives including the ability to use different aspect or information of data, improved interpretability of decomposed factors, and more reliable and robust results. We propose C^3APTION, a scalable and coupled CP and PARAFAC2 model, to impose non-negativity constraints on the factors.

A. General Framework for C^3APTION

Given PARAFAC2 tensor $\underline{\mathbf{X}}$ and CP tensor $\underline{\mathbf{Y}}$ coupled in its 3^{rd} mode, this section proposes three different settings of the coupled tensor decomposition C^3APTION in order to factorize the multi-aspect graph or tensor into its constituent community-revealing factors. We focus on a third-order tensor $\underline{\mathbf{X}} \in \mathbb{R}^{I_k \times J}$ ($\forall k \in [1, K]$) and $\underline{\mathbf{Y}} \in \mathbb{R}^{K \times L \times M}$ for our problem and its loss function formulation is given by:

$$\begin{aligned}\mathcal{LS} =& _{\mathbf{Q}_k,\mathbf{U}_k,\mathbf{H},\mathbf{V},\mathbf{S}_k,\mathbf{B},\mathbf{C}} \frac{1}{2}||\underline{\mathbf{X}}_k - \mathbf{U}_k\mathbf{S}_k\mathbf{V}^T||_F^2 + \\ & \frac{\lambda}{2}||\underline{\mathbf{Y}} - \mathbf{W}(\mathbf{C}\odot\mathbf{B})^T]||_F^2 + \sum_{k=1}^{K}\left(\frac{\mu_k}{2}||\mathbf{U}_k - \mathbf{Q}_k\mathbf{H}||_F^2\right) \\ & \textbf{s. t. } \mathbf{Q}_k^T\mathbf{Q}_k = \mathbf{I}, \mathbf{U}_k \geq 0, \mathbf{S}_k \geq 0, \\ & \quad \mathbf{W}(k,0) = diag(\mathbf{S}_k), \mathbf{B} \geq 0, \mathbf{C} \geq 0 \quad \forall k \in [1, K]\end{aligned} \tag{8}$$

For PARAFAC2, we re-write the first part $\mathcal{LS}_1$ of minimization of $\mathcal{LS}$ function in terms of $\mathbf{Q}_k$ as $Trace(\underline{\mathbf{X}}_k^T\underline{\mathbf{X}}_k) + Trace(\mathbf{V}\mathbf{S}_k\mathbf{H}^T\mathbf{Q}_k^T\mathbf{Q}_k\mathbf{H}\mathbf{S}_k\mathbf{V}^T) - 2 * Trace(\underline{\mathbf{X}}_k^T\mathbf{Q}_k\mathbf{H}\mathbf{S}_k\mathbf{V}^T)$. The first and second terms are constant and by rearranging the rest, it is equivalent to

$$\begin{aligned}\mathcal{LS}_1 = & _{\mathbf{Q}_k,\mathbf{U}_k,\mathbf{H},\mathbf{V},\mathbf{S}_k} \frac{1}{2}||\mathbf{Q}_k^T\underline{\mathbf{X}}_k - \mathbf{H}\mathbf{S}_k\mathbf{V}^T||_F^2 \\ & \textbf{s. t. } \mathbf{Q}_k^T\mathbf{Q}_k = \mathbf{I}, \mathbf{U}_k \geq 0, \mathbf{S}_k \geq 0, \quad \forall k \in [1, K]\end{aligned} \tag{9}$$

Thus, the constrained coupled tensor decomposition objective $\mathcal{LS}$ is of the form:

$$\begin{aligned}\mathcal{LS} =& _{\mathbf{Q}_k,\mathbf{U}_k,\mathbf{H},\mathbf{V},\mathbf{S}_k,\mathbf{B},\mathbf{C}} \frac{1}{2}||\mathcal{X} - \mathbf{H}\mathbf{S}_k\mathbf{V}^T||_F^2 + \\ & \frac{\lambda}{2}||\underline{\mathbf{Y}} - \mathbf{W}(\mathbf{C}\odot\mathbf{B})^T]||_F^2 + \sum_{k=1}^{K}\left(\frac{\mu_k}{2}||\mathbf{U}_k - \mathbf{Q}_k\mathbf{H}||_F^2\right) \\ & \textbf{s. t. } \mathcal{X}(:,:,k) = \mathbf{Q}_k^T\underline{\mathbf{X}}_k, \mathbf{Q}_k^T\mathbf{Q}_k = \mathbf{I}, \mathbf{U}_k \geq 0, \mathbf{S}_k \geq 0, \\ & \quad \mathbf{W}(k,0) = diag(\mathbf{S}_k), \mathbf{B} \geq 0, \mathbf{C} \geq 0 \quad \forall k \in [1, K]\end{aligned} \tag{10}$$

B. Inference of Factors

We propose 3 types of algorithms to solve coupling namely C^3APTION-ASET (unconstrained version), C^3APTION-BPP (constrained), and C^3APTION-ALS (constrained). First two methods are natural extension of [5] tensor coupling. Equation 10 is non-convex, our method utilizes instances of the non-negativity constrained least squares (NNLS) framework to divide problem into sub-problems. We use optimization method block principal pivoting for C^3APTION-BPP [21], active set of Lawson and Hanson [20] for C^3APTION-ASET and least square method for C^3APTION-ALS to solve each sub-problem. Next, we summarize the solution for each latent factor.

*1) **Factor $\mathbf{Q}_k$ update**:* Consider first the update of factor $\mathbf{Q}_k$ obtained after fixing other factor matrices. For C^3APTION-ASET and C^3APTION-BPP, we update the $\mathbf{Q}_k$ by minimizing Equ. 8 using following method:

$$_{Q_k} \frac{\mu_k}{2}||\mathbf{U}_k - \mathbf{Q}_k\mathbf{H}||_F^2 \quad s.t. \quad \mathbf{Q}_k^T\mathbf{Q}_k = \mathbf{I}$$

$$_{Q_k} \mu_k(Tr(\mathbf{Q}_k\mathbf{H}\mathbf{H}^T\mathbf{Q}_k^T - 2\mathbf{Q}_k\mathbf{H}\mathbf{U}_k^T + \mathbf{U}_k\mathbf{U}_k^T)) = 0$$

Using $Tr(ABC) = Tr(CAB)$ property, we can re-write $Tr(\mathbf{Q}_k\mathbf{H}\mathbf{H}^T\mathbf{Q}_k^T) = Tr(\mathbf{Q}_k^T\mathbf{Q}_k\mathbf{H}\mathbf{H}^T)$. As $\mathbf{Q}_k^T\mathbf{Q}_k = \mathbf{I}$, we can reformulate above equation w.r.t. $\mathbf{Q}_k$ as follows :

$$_{Q_k} \mu_k\mathbf{Q}_k\mathbf{H}\mathbf{U}_k^T \quad s.t. \quad \mathbf{Q}_k^T\mathbf{Q}_k = \mathbf{I}$$

$$[\mathbf{Q}_k] = SVD[\mu_k\mathbf{H}\mathbf{U}_k^T] \quad s.t. \quad \mathbf{Q}_k^T\mathbf{Q}_k = \mathbf{I} \tag{11}$$

For C^3APTION-ALS, this factor is computed using the simple SVD as:

$$[\mathbf{Q}_k] = SVD[\mathbf{H} \times diag(\mathbf{W}(k,:)) \times (\underline{\mathbf{X}}_k \times \mathbf{V})^T] \tag{12}$$

Note that each $\mathbf{Q}_k$ can contain negative values.

*2) **Factor $\mathbf{H}$ update**:* We update $\mathbf{H}$ by fixing $\mathbf{V}$, $\mathbf{W}$ and $\mathbf{Q}_k$. We set derivative the loss $\mathcal{LS}$ w.r.t. $\mathbf{H}$ (Equ. 8, note that part 1 and part 2 are constant) to zero to find local minima as follows:

$$\frac{\delta\mathcal{LS}}{\delta\mathbf{H}} = \frac{\sum_{k=1}^{K}\frac{\mu_k}{2}Tr((\mathbf{U}_k - \mathbf{Q}_k\mathbf{H})(\mathbf{U}_k - \mathbf{Q}_k\mathbf{H})^T)}{\delta\mathbf{H}} = 0$$

$$\delta\left(\sum_{k=1}^{K}\mu_k Tr(\mathbf{Q}_k\mathbf{H}\mathbf{H}^T\mathbf{Q}_k^T - 2\mathbf{Q}_k\mathbf{H}\mathbf{U}_k^T + \mathbf{U}_k\mathbf{U}_k^T)\right)/\delta\mathbf{H} = 0$$

$$\sum_{k=1}^{K}\mu_k\mathbf{Q}_k^T\mathbf{Q}_k\mathbf{H} - \sum_{k=1}^{K}\mu_k\mathbf{Q}_k^T\mathbf{U}_k = 0$$

as $\mathbf{Q}_k^T\mathbf{Q}_k = \mathbf{I}$, the update rule for latent factor $\mathbf{H}$ is given below:

$$\begin{aligned}&\textbf{For } C^3\textbf{APTION-ASET:} \quad \mathbf{H} = \frac{\sum_{k=1}^{K}\mathbf{Q}_k^T\mathbf{U}_k}{\sum_{k=1}^{K}\mu_k} \\ &\textbf{For } C^3\textbf{APTION-BPP:} \quad \mathbf{H} = \frac{\sum_{k=1}^{K}\mathbf{Q}_k^T\mathbf{U}_k}{\sum_{k=1}^{K}\mu_k} \quad \textbf{s. t. } \mathbf{H} \geq 0\end{aligned} \tag{13}$$

For C^3**APTION-ALS**, we set $\mu_k = 0$ and derive update rule from Equ. 10 as follows:

$$\mathbf{H} = \frac{(\mathcal{X}\mathbf{V}) * \mathbf{W}^T}{(\mathbf{V}^T\mathbf{V} * \mathbf{W}^T\mathbf{W})} \quad \textbf{s. t. } \mathbf{H} \geq 0 \tag{14}$$

*3) **Factor $\mathbf{S}_k$ or $\mathbf{W}$ update**:* This mode of the PARAFAC2 tensor is coupled with CP tensor. The objective function 8 with respect to $\mathbf{W}$ can be rewritten as:

$$_{S_k} \frac{1}{2}||\underline{\mathbf{X}}_k - \mathbf{U}_k\mathbf{S}_k\mathbf{V}^T||_F^2 + \frac{\lambda}{2}||\underline{\mathbf{Y}} - \mathbf{W}(\mathbf{C}\odot\mathbf{B})^T]||_F^2 \tag{15}$$

For C^3**APTION-ASET** Equation 15 can be rewritten as:

$$_{\mathbf{S}_k} \frac{1}{2}\left|\left|\begin{bmatrix}(\mathbf{V}\odot\mathbf{U}_k) \\ \sqrt{\lambda}(\mathbf{C}\odot\mathbf{B})\end{bmatrix}\mathbf{W}(k,:)^T - \begin{bmatrix}\text{vec}(\underline{\mathbf{X}}_k) \\ \sqrt{\lambda}\text{vec}(\underline{\mathbf{Y}}(k,:,:))\end{bmatrix}\right|\right|_F^2 \tag{16}$$

For C^3**APTION-BPP**, Equation 16 can be computed such that $\mathbf{W}(k,:) \geq 0$. The Khatri-rao product operation is expensive that can be replaced by element-wise (hadamard) product and matrix to tensor product can be replaced by slice wise dot product with factor matrices [23].

For C^3**APTION-ALS**, we update $\mathbf{S}_k$ or $\mathbf{W}$ as:

$$\mathbf{S}_k = \frac{(\mathbf{U}_k^T\mathbf{U}_k * \mathbf{V}^T\mathbf{V}) + (\sqrt{\lambda}(\mathbf{C}^T\mathbf{C} * \mathbf{B}^T\mathbf{B})}{diag(\mathbf{U}_k^T\underline{\mathbf{X}}_k\mathbf{V}) + diag(\mathbf{B}^T\underline{\mathbf{Y}}(k,:,:)\mathbf{C})} \tag{17}$$

$$\mathbf{W}(k,:) = diag(\mathbf{S}_k) \quad \textbf{s. t. } \mathbf{W}(k,:) \geq 0, \forall k \in [1, K]$$

*4) **Factor*** $\mathbf{V}$ ***update**:* We solve Equation 8 with respect to $\mathbf{V}$ as given below:

$$_{\mathbf{V}}\frac{1}{2}||\underline{\mathbf{X}}_k - \mathbf{U}_k\mathbf{S}_k\mathbf{V}^T||_F^2 \tag{18}$$

For C^3**APTION-ASET**, Equ. (18) can be formulated as:

$$\mathbf{V}(:,k) = \frac{\mathbf{X}_k^T}{\mathbf{S}_k\mathbf{U}_k^T} \tag{19}$$

For C^3**APTION-BPP**, Equ. (19) can be easily updated such that by $\mathbf{V} \geq 0$.

For C^3**APTION-ALS**, we update $\mathbf{V}$ as given below:

$$\mathbf{V} = \frac{(\mathcal{X}^T\mathbf{H}) * \mathbf{W}^T}{(\mathbf{H}^T\mathbf{H} * \mathbf{W}^T\mathbf{W})} \quad \textbf{s. t.} \quad \mathbf{V} \geq 0 \tag{20}$$

*5) **Factor*** $\mathbf{B}$ ***or*** $\mathbf{C}$ ***update**:* Finally, factor matrices $\mathbf{B}$ and $\mathbf{C}$ represents the participation of CP tensor for user similarities. We solve Equation 8 w.r.t $\mathbf{B}$ as given below:

$$_{\mathbf{B}}\frac{1}{2}||\underline{\mathbf{Y}} - \mathbf{B}(\mathbf{C} \odot \mathbf{W})^T||_F^2 \quad \text{s. t.} \quad \mathbf{B} \geq 0 \tag{21}$$

which can be easily updated via all-set method [] for C^3APTION-ASET and via block principal pivoting [] for C^3APTION $- BPP$.

For C^3APTION-ALS, we update $\mathbf{B}$ as given below:

$$\mathbf{B} = \frac{\text{MTTKRP}(\underline{\mathbf{Y}}, \mathbf{C}, \mathbf{W})}{(\mathbf{C}^T\mathbf{C} * \mathbf{W}^T\mathbf{W})} \quad \textbf{s. t.} \quad \mathbf{B} \geq 0 \tag{22}$$

Similarly, we update $\mathbf{C}$ as:

$$\mathbf{C} = \frac{\text{MTTKRP}(\underline{\mathbf{Y}}, \mathbf{B}, \mathbf{W})}{(\mathbf{B}^T\mathbf{B} * \mathbf{W}^T\mathbf{W})} \quad \textbf{s. t.} \quad \mathbf{C} \geq 0 \tag{23}$$

*6) **Factor*** $\mathbf{U}_k$ ***update**:* For C^3APTION-ALS, this factor is computed using the simple multiplication $\mathbf{U}_k = \mathbf{Q}_k * \mathbf{H}$. For C^3APTION-ASET and C^3APTION-BPP (where $\mathbf{U}_k \geq 0$), the objective function with respect to $\mathbf{U}_k$ can be solved as:

$$_{\mathbf{U}_k}\frac{1}{2}||\begin{bmatrix}(\mathbf{V}\mathbf{S}_k)\\ \sqrt{\mu_k}\mathbf{I}\end{bmatrix}\mathbf{U}_k^T - \begin{bmatrix}\underline{\mathbf{X}}_k^T\\ \sqrt{\mu_k}\mathbf{H}^T\mathbf{Q}_k^T\end{bmatrix}||_F^2 \tag{24}$$

IV. EXPERIMENTS AND ANALYSIS

In this section we extensively evaluate the performance of C^3APTION on multiple synthetic and real datasets, and compare its performance with state-of-the-art approaches. We focus on answering the following:

Q1. Does C^3APTION preserve accuracy while being fast to compute and helps in Identifiability of latent factors?

Q2. How does C^3APTION scale for increasing number of users (K)?

Q4. How can we use C^3APTION for real-world utility?

Dataset	Statistics (K: Thousands M: Millions)			
	$[I_{max}, J, K]$	**[K, L, M]**	***R***	***#nnz***
SYN-I	$[500, 1K, 5K]$	$[5K, 5K, 500]$	40	$[0.5B, 1.5B]$
SYN-II	$[1K, 1K, 10K]$	$[10K, 10K, 1K]$	40	$[1.4B, 3.9B]$
SYN-III	$[1K, 5K, 50K]$	$[50K, 50K, 1K]$	10	$[6B, 9B]$
Collaboration	$[25, 10, 11K]$	$[11, 11K, 5]$	$5-50$	$[1M, 1.2M]$
Movielens	$[121, 4K, 6K]$	$[6K, 6K, 5]$	$5-50$	$[1M, 4.5M]$
Adobe	$[1K, 1K, 31K]$	$[31K, 31K, 5]$	$5-50$	$[1.7M, 6.3]$
CMS	$[250, 1K, 98K]$	$[98K, 98K, 5]$	$5-50$	$[9.6M, 9.7M]$

TABLE II: Details for the datasets.

A. *Dataset*

We provide the datasets used for evaluation in Table II. Rank determination in the experiments is performed with the aid of the Core Consistency Diagnostic method [8], [26].

Synthetic Data: In order to fully explore the performance of C^3APTION, in our experiments we generate synthetic tensor with varying density. Those tensors are created from a known set of randomly generated factors, so that we have full control over the ground truth of the full decomposition. The specifications of synthetic datasets are given in Table II.

Real Data: In order to truly evaluate the effectiveness of C^3APTION, we test its performance against four real datasets that have been used in the literature. Those datasets are summarized in Table III and details are below.

Collaboration Data[1]: It is co-authorship network (where two authors are connected if they publish at least one paper together) of $11,176$ authors over years 1990-2015 for International Conference on Data Mining (ICDM), International Conference on Machine Learning (ICML), Knowledge Discovery and Data Mining (KDD) conference.

Movielens[16]: MovieLens-1M dataset is widely used in recent literature. For this dataset, we created tensor as year-by-movie-by-user i.e each year of ratings corresponds to a certain observation for each user's activity.

Adobe: Adobe dataset is sequential data and it consists of tutorial sequence of anonymous 7 million users. We selected users (31K) who watched at least unique 15 tutorials. We created PARAFAC2 tensor as sequence-by-tutorial-by-user [max 1k × 1k × 31k] and CP tensor as user-by-user-by-similarity. We have semi synthetic ground truth values for this dataset and we assigned each user to class based on the type of tutorial watched.

CMS [11]: This dataset is synthetically created by Centers for Medicare and Medicaid (CMS) by using 5% of real medicare data and includes 98K beneficiaries. We created PARAFAC2 tensor as visits-by-diagnosis-by-patient and CP tensor as user-by-user-by-similarity.

We create CP tensor using well known similarity methods such that cosine similarity, Jaccard similarty, LSH hasing [28], ABC hashing[28], K-mean and Edit distance.

B. *Baselines*

Here we briefly present the state-of-the-art baselines we used for comparison. Note that for each baseline we use the reported parameters that yielded the best performance in the respective publications.All comparisons were carried out over 10 iterations each, and each number reported is an average

Dataset	Metric	SCD [10]	RCTF [13]	TASTE [5]	C-BPP	C-ASET	C-ALS
SYN-I	RMSE	0.43(0.055)	0.38(0.068)	0.21(0.041)	0.20(0.068)	0.26(0.032)	**0.18 (0.026)**
	NMI	0.45(0.010)	0.49(0.074)	0.78(0.021)	0.79(0.019)	0.65(0.012)	**0.92 (0.034)**
	Time (mins)	490.01(14.07)	548.32(16.36)	357.23(34.59)	336.43(11.59)	348.56(56.43)	**301.87 (34.43)**
SYN-II	RMSE			0.30(0.023)	0.29(0.013)	0.36(0.092)	**0.25 (0.065)**
	NMI	[[OoM]]	[[OoM]]	0.65(0.044)	0.68(0.023)	0.61(0.049)	**0.75 (0.063)**
	Time (mins)			2109.11(89.75)	2021.48(56.35)	2090.41(134.67)	**1689.68 (101.23)**
SYN-III	RMSE			0.35(0.022)	0.38(0.035)	0.43(0.056)	**0.32 (0.081)**
	NMI	[[OoM]]	[[OoM]]	0.72(0.069)	0.70(0.013)	0.65(0.086)	**0.76 (0.081)**
	Time (mins)			2387.56(72.85)	2304.68(89.63)	2360.41(112.34)	**1959.68 (91.23)**

TABLE III: Performance of C^3APTION in terms of RMSE, NMI and CPU Time (mins) for synthetic data. Numbers where our proposed method outperforms other baselines are bolded. For each dataset, we report the standard deviation between two parentheses along with average score. Remarkably, C^3APTION-ALS better preserve accuracy which ultimately, improves task performance.

with a standard deviation attached to it. We compared the following algorithms for coupling CP and PARAFAC2 tensors.

- **TASTE** [5]: This method based on block principle pivot for coupling between PARAFAC2 tensor and matrix. We run algorithm for all slices of CP tensor $\underline{\mathbf{Y}}$.
- **Soft Coupled Decomposition** [10]: SCD method is soft coupling of two PARAFAC2 data.
- **Robust Coupled Tensor Factorization** [13]: RCTF is robust coupling of two CP tensors via measuring distance between factor.
- Our proposed methods:
  - C^3**APTION-ASET**: This is solving coupled tensor factorization via Active set methods with using the unconstrained least squares optimization.
  - C^3**APTION-BPP**: This is solving coupled tensor factorization via block principal pivoting method using the non negativity-constrained least squares optimization.
  - C^3**APTION-ALS**: This is solving coupled tensor factorization via alternating non-negative least squares optimization.

C. Evaluation Measures

We evaluate C^3APTION and the baselines using three quantitative criteria: Root Mean Square Error and CPU-Time (in minutes). Briefly,

- **Root Mean Square Error**: Performance is evaluated as the Root Mean Square Error (RMSE) which is a well known evaluation measure used in coupled tensor factorization literature. Mathematically,

$$RMSE = \sqrt{\frac{\sum_{k=1}^{K}\left(||\underline{\mathbf{X}}_k - \hat{\underline{\mathbf{X}}}_k||^2\right) + ||\underline{\mathbf{Y}} - \hat{\underline{\mathbf{Y}}}||^2}{\sum_{k=1}^{K}(I_k \times J) + (K \times L \times M)}} \tag{25}$$

- **CPU time (sec)**: indicates how much faster does the decomposition runs as compared to baselines. The average running time is measured in seconds, and is used to validate the time efficiency of an algorithm.
- **Normalized Mutual Information**: Normalized Mutual Information (NMI) is a good measure for determining the quality of clustering. Mathematically,

$$NMI(Y,C) = \frac{2 * I(Y;C)}{[H(Y) + H(C)]} \tag{26}$$

where I(Y, C) is mutual information between cluster Y and C, H(Y) and H(C) are entropy of cluster and classes.

D. Experimental Result

Q1a. Effectiveness and Run Time

We evaluate performance of the algorithm for community detection where each node in a graph is assigned to a single label. In our study, we perform hard clustering over latent factor matrices.We run each method for 10 different random initialization and provide the average and standard deviation of RMSE and CPU Time (min) as shown in Table III.

Synthetic Data: The baseline method SCD and RCTF unable to decompose SYN-II and SYN-III due to out of memory during intermediate computations. Our proposed methods C^3APTION-BPP and C^3APTION-ASET provide comparable accuracy and runtime when compared to TASTE method. Overall, C^3APTION-ALS achieves significant improvement on running time and average $3-8\%$ RMSE improvement. Therefore, our approach is the only one that achieves a fast and accurate solution.

For real dataset we do not have labels, so we provide only RMSE and CPU Time for these data as discussed below:

Collaboration Data: We observed that C^3APTION-ALS provides the high-quality communities as shown in Table IV. We see similar behaviour with C^3APTION-BPP also. We select top two communities based on size. Each community represents a group of scientists with the same research interests, such as Data Mining community (#3) and Information Retrieval and Web Mining community (#10) in Table V. Researchers like "Jiawei Han" and "Philip S. Yu", have published a large number of papers in collaboration with people from various research communities. These authors considered as tightly related to the same community in the network. We further analyze the outcome of baseline methods and observe that SCD and RCTF are not able to find few strongly connected communities and fails to merge the small groups even those share a strong connection. In terms of RMSE, C^3APTION-ALS, outperformed the baselines as shown in Table IV **Movielens Data**: We decompose movielens data using R=20 for all methods. Our proposed method C^3APTION-ALS outperformed w.r.t RMSE and computation time. Here, we explore interesting observations. First, we observe that there is a rapid growth of sci-fi movies beginning of 1970, a few months after the first Moon landing. Secondly, we observe

Dataset	Metric	SCD [10]	RCTF[13]	TASTE [5]	C-BPP	C-ASET	C-ALS
Collaboration	RMSE	0.39(0.075)	0.35(0.068)	0.17(0.041)	0.17(0.068)	0.23(0.032)	**0.14 (0.026)**
	Time (mins)	379.01(14.07)	427.88(16.36)	236.93(19.69)	210.43(11.59)	226.43(13.28)	**174.31 (11.41)**
Movielens	RMSE	0.24(0.021)	0.28(0.020)	0.19(0.082)	0.16(0.012)	0.21(0.093)	**0.14 (0.012)**
	Time (mins)	48.21(3.21)	45.20(2.34)	21.34(1.69)	20.89(4.83)	25.55(2.84)	**10.19 (1.36)**
Adobe	RMSE	0.29(0.033)	0.35(0.062)	0.28(0.015)	0.22(0.023)	0.26(0.042)	**0.20 (0.013)**
	NMI	0.42(0.05)	0.48(0.09)	0.53(0.02)	0.54(0.01)	0.49(0.06)	**0.58 (0.08)**
	Time (mins)	210.23(11.34)	198.20(16.96)	98.22(10.58)	93.23(23.52)	150.24(25.58)	**78.72 (21.74)**
CMS	RMSE	0.29(0.045)	0.34(0.098)	0.21(0.058)	**0.20 (0.052)**	0.24(0.021)	0.23(0.037)
	Time (mins)	466.23(13.44)	435.34(9.10)	150.24(10.92)	149.24(11.23)	202.24(19.13)	**112.33 (11.93)**

TABLE IV: Performance of C^3APTION in terms of RMSE and CPU Time (mins) for real data decomposed. Numbers where our proposed method outperforms other baselines are bolded. For each dataset, we report the standard deviation between two parentheses along with average score. *Note: we have semi-synthetic labels for the Adobe dataset only.

Community[#3]	Community[#10]
Jiawei Han	Rakesh Agrawal
Philip S. Yu	Ramakrishnan Srikant
Wei Fan	Panayiotis Tsaparas
Charu C. Aggarwal	H Lei Zhang
Jimeng Sun	Josh Attenberg
Jian Pei	Anitha Kannan
Bing Liu	Sreenivas Gollapudi
Bhavani M. Thuraisingham	Kamal Ali
Longbing Cao	Sunandan Chakraborty
Tanya Y. Berger Wolf	Rui Cai
Xindong Wu	Indu Pal Kaur

TABLE V: Top two communities (based on size) discovered by C^3APTION-ALS on *Collaboration* Dataset. Selected researchers are based on top 10 factor values of latent factor.

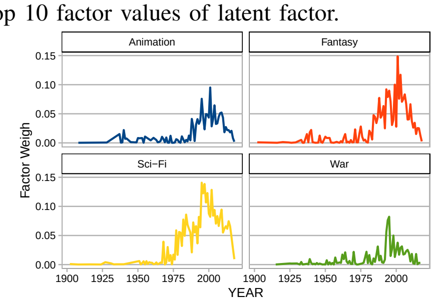

Figure. 2: Movielens Data exploratory analysis for top movie genre.

that the rise of popularity of animation movies as shown in fig 2, the reason could be the advancement of the computer animation technology which made the development of such movies much easier. Next, the most of the war/action movies were popular around the time of World War II, Vietnam War and war in Afghanistan and Iraq. It's interesting to observe that how the cinematography world reflected the viewership of the real world. Another interesting observation is that most of the salesman and programmers mostly loved adventurer movies and lawyers liked most of drama and fantasy movies.

Adobe Data: C^3APTION-ALS outperforms the baseline methods. Remarkably, it surpasses the baselines most when the data is sparse. In order to present the use of C^3APTION towards community detection, we focus our analysis on a subset of tutorials watched by each community in this dataset. Figure 3 presents the top 5 (based on size) community's most frequent tutorial(s) sequence watched. Conceptually, those users share similar interest in terms of domain knowledge, learning or interests. We observe from the factors of the CP tensor that these communities are connected strongly within the group

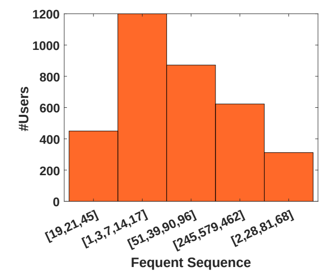

Figure. 3: Frequent sequence of tutorials watched for top communities based on size.

and have few connections outside the group. Nevertheless, C^3APTION-ALS achieves significantly good performance in terms of $NMI \approx 0.58$ as shown in Table IV.

Q1b. Identifiability Analysis As we know that PARAFAC2 is known for the hardness in terms of optimization. In many instances, Alternate Least Square (ALS) based algorithms do not converge to good solutions, although the PARAFAC2 decomposition theoretically has identifiability. For identifiability analysis, PARAFAC2 tensor $\underline{\mathbf{X}} \in \mathbb{R}^{1000\times1000\times1000}$ and CP tensor $\underline{\mathbf{Y}} \in \mathbb{R}^{1000\times1000\times500}$ is constructed with fixed target rank $R = 10$. Our aim is to recover latent factors as similar as possible to original latent factors. To simplify, we discuss identifiability of latent factor matrix $\mathbf{W}$ only. We compute the dot product for all permutations of columns between original latent factors (W_{org}) and latent factors (W_{pred}) obtained after decomposition. If the computed dot product is higher than the threshold value (80%), the two factors match, and we consider them as recovered factor. If the dot product between a column in W_{org} and with all the columns in W_{pred} has a value less than the threshold, we consider it as an non-recovered factor. The Figure 4 shows that using coupled CP tensor data alongside PARAFAC2 data could help to alleviate the above discussed challenge and improve the identifiability of decomposition.

Q2. Scalability We also evaluate the scalability of our algorithm on synthetic dataset in terms of time needed for increasing load of input users (K). A PARAFAC2 tensors $\underline{\mathbf{X}} \in \mathbb{R}^{100\times100\times[1K-1M]}$ and CP tensor $\underline{\mathbf{Y}} \in \mathbb{R}^{[1K-1M]\times[1K-1M]\times5}$ are decomposed with fixed target rank $R = 40$. The time needed by C^3APTION increases very linearly with increase in non-zero elements. Our proposed method C^3APTION-ALS, successfully decomposed the large

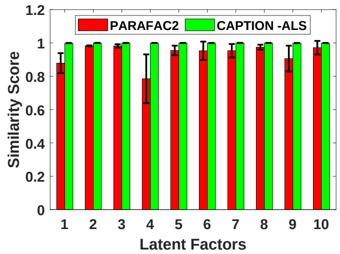

Figure. 4: Identifiability analysis with and without coupling of PARAFAC2. Higher the value, better the identifiability.

coupled tensors in reasonable time as shown in Figure 5(a - b) and is up to $5\times$ faster than baseline methods. Figure 5(c - h), we present the RMSE, and the computational time for the approaches under comparison for ML, Adobe and CMS data for increasing target rank from $5-50$. We remark that all methods achieve comparable RMSE values for three different data sets but proposed method C^3APTION-ALS is up to average $2.5\times$ faster than baselines for all data sets. We remark the favorable scalability properties of C^3APTION, rendering it practical to use for large tensors.

Q3. C^3APTION at Work

Centers for Medicare and Medicaid (CMS) data files were created to allow researchers to gain familiarity using Medicare claims data while protecting beneficiary privacy. The CMS data contains multiple files per year. The file contains synthesized data taken from a 5% random sample of Medicare beneficiaries in 2008 and their claims from 2008 to 2010. We created CP tensor $\underline{\mathbf{Y}}$ from files that contain demographic characteristics (sex, race, state etc) of the beneficiary. The PARAFAC2 tensor $\underline{\mathbf{Y}}$ is created from files that has clinical variables such as chronic conditions. We decompose CP and PARAFAC2 tensor jointly with rank $R = 40$.

Model Interpretation: We propose the following model interpretation towards the target challenge:

- **Diagnosis feature factor:** Each column of factor matrix $\mathbf{V}$ represents a cluster and each row indicates a medical feature. Therefore an entry $\mathbf{V}(i, j)$ represents the membership of medical feature i to the j.
- **Irregular dimension factor (visits):** The r^{th} column of $\mathbf{U}_k$ presents the evolution of cluster r for all clinical visits for patient k.
- **Coupled factor (patients):** The coupled latent factor $\mathbf{W} = diag(\mathbf{S}_k)$ and CP latent factor $\mathbf{B}$, indicates the R communities of the patient.
- **Similarity factor:** The factor $\mathbf{C}$ indicates the importance of similarities membership which is responsible for creating clusters.

Findings: In order to illustrate the use of C^3APTION towards clustering, we focus our analysis on a subset of patients those are classified as medically complex. We observe that cluster number 32 has most patients with respiratory disease. These are the patients with high utilization ($> 50\%$), multiple clinical visits (avg 67) and high severity (death rate 8-10%). Most of the patients share ICD-9 code 492 (Emphysema), 496 (Chronic airway obstruction) and 511 (Pleurisy). These codes are characterized by obstruction of airflow that interferes with normal breathing. It is observed that these conditions frequently co-exists in real world and are hard to treat. In Figure 6, we provide time-frame captured by C^3APTION-ALS for patient no. 11426 with chronic airway obstruction. In the patient's heath timeline, it shows that on the first few weeks visit, there is no sign of obstruction. The subsequent visits reflects a change in the patient's heath with a large number of diagnosis (day 84). Nevertheless, as shown in figure 5, C^3APTION-ALS achieves significantly good performance in terms of RMSE (avg 60% better) and computation time ($3\times$ faster) by leveraging the coupling between CP and PARAFAC2 tensor data. .

V. Conclusions and Future Work

This paper outlined our vision on exploring the coupling of CP and PARAFAC2 tensor decomposition using various optimization methods (BPP, ASET and ALS) to improve accuracy of factorization. We propose C^3APTION, a framework that is able to offer interpretable results, and we provide a experimental analysis on synthetic as well as real world dataset. Extensive experiments with large synthetic dataset have demonstrated that the proposed method is capable of handling larger dataset for coupling for which most of the baselines are not able to performs due to lack of memory.

Furthermore, this paper outlines a set of interesting future research directions:

- How can we couple one of auxiliary tensor with irregular mode of PARAFAC2 tensor to obtain better approximation?
- What other constraints, other than non-negative, for the C^3APTION are well suited for various application and have potential to offer more accurate results?
- How can we use coupling for incremental tensor data?

VI. Acknowledgments

Research was supported by the Department of the Navy, Naval Engineering Education Consortium under award no. N00174-17-1-0005, the National Science Foundation Grant no. 1901379 and an Adobe Data Science Research Faculty Award. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the funding parties.

References

[1] Kdd data. http://gmap.cs.arizona.edu/datasets.

[2] Evrim Acar, Rasmus Bro, and Age K Smilde. Data fusion in metabolomics using coupled matrix and tensor factorizations. *Proceedings of the IEEE*, 103(9):1602–1620, 2015.

[3] Evrim Acar, Tamara G Kolda, and Daniel M Dunlavy. All-at-once optimization for coupled matrix and tensor factorizations. *arXiv preprint arXiv:1105.3422*, 2011.

[4] Evrim Acar, Evangelos E Papalexakis, Morten A Rasmussen, Anders J Lawaetz, Mathias Nilsson, and Rasmus Bro. Structure-revealing data fusion. *BMC bioinformatics*, 15(1):239, 2014.

[5] Ardavan Afshar, Ioakeim Perros, Haesun Park, Christopher deFilippi, Xiaowei Yan, Walter Stewart, Joyce Ho, and Jimeng Sun. Taste: temporal and static tensor factorization for phenotyping electronic health records. In *Proceedings of the ACM Conference on Health, Inference, and Learning*, pages 193–203, 2020.

[6] Brett W Bader, Tamara G Kolda, et al. Matlab tensor toolbox version 2.6, available online, february 2015, 2015.

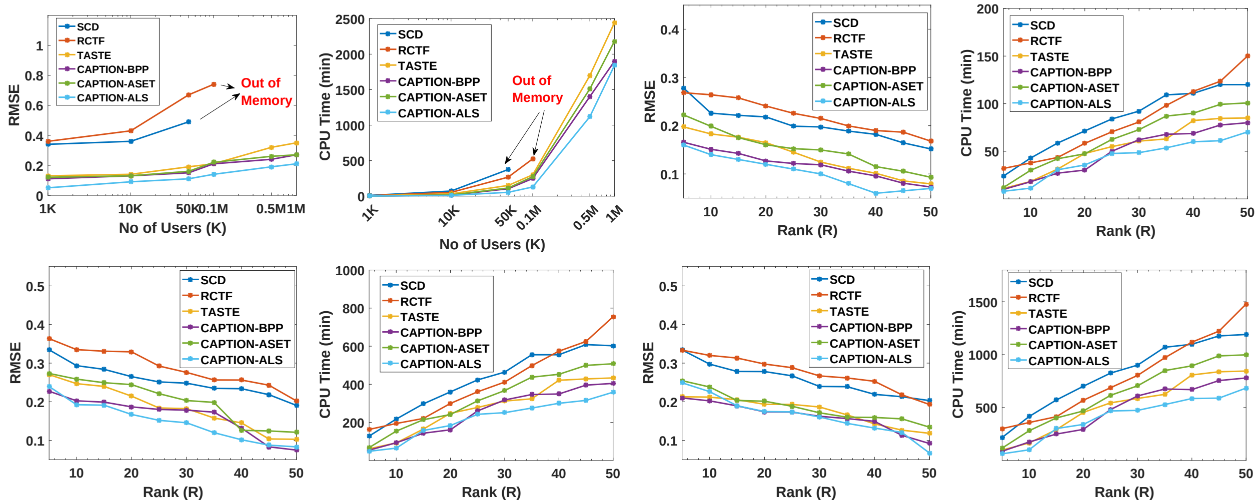

Figure. 5: Scalability analysis of C^3APTION method using synthetic and three real world datasets. (a-b): Scalability analysis of synthetic data with respect to varying number of users K, where K ranges $10^3 - 10^6$. Stable performance (RMSE) in the range $1K - 50K$, for most methods. Baseline method SCD and RCTF runs out of memory. (c-d): Scalability analysis with respect to Movielens dataset (c-d), Adobe dataset (e-f) and CMS heath record data (g-h). C^3APTION-ALS significantly outperforms the other methods even when data is very sparse.

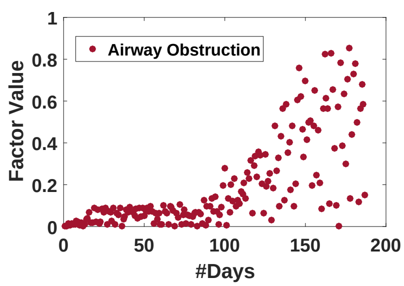

Figure. 6: Visualization of time-frame captured of the patient no. 11426 created by C^3APTION-ALS on CMS dataset.

[7] Riccardo Bellazzi, Marianna Diomidous, Indra Neil Sarkar, Katsuhiko Takabayashi, Andreas Ziegler, and Alexa T McCray. Data analysis and data mining: current issues in biomedical informatics. *Methods of information in medicine*, 50(6):536, 2011.

[8] Rasmus Bro and Henk AL Kiers. A new efficient method for determining the number of components in parafac models. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 17(5):274–286, 2003.

[9] J Douglas Carroll and Jih-Jie Chang. Analysis of individual differences in multidimensional scaling via an n-way generalization of "eckart-young" decomposition. *Psychometrika*, 35(3):283–319, 1970.

[10] Christos Chatzichristos, Mike Davies, Javier Escudero, Eleftherios Kofidis, and Sergios Theodoridis. Fusion of eeg and fmri via soft coupled tensor decompositions. In *2018 26th EUSIPCO*, pages 56–60. IEEE, 2018.

[11] CMS. Data. https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/SynPUFs, 2008. [Online].

[12] Georgios Theocharous ekta Gujral and Evangelos E. Papalexakis. Constrainted coupled cp and parafac2 tensor decomposition. 2020.

[13] Matthieu Genicot, P-A Absil, Renaud Lambiotte, and Saber Sami. Coupled tensor decomposition: a step towards robust components. In *2016 24th EUSIPCO*, pages 1308–1312. IEEE, 2016.

[14] Ekta Gujral and Evangelos E Papalexakis. Smacd: semi-supervised multi-aspect community detection. In *Proceedings of the 2018 SIAM SDM*, pages 702–710. SIAM, 2018.

[15] Ekta Gujral, Ravdeep Pasricha, and Evangelos Papalexakis. Beyond rank-1: Discovering rich community structure in multi-aspect graphs. In *Proceedings of The Web Conference 2020*, 2020.

[16] F Maxwell Harper and Joseph A Konstan. The movielens datasets: History and context. *Acm transactions on interactive intelligent systems*, 5:19, 2016.

[17] R.A. Harshman. Foundations of the parafac procedure: Models and conditions for an" explanatory" multimodal factor analysis. 1970.

[18] Richard A Harshman. Parafac2: Mathematical and technical notes. *UCLA working papers in phonetics*, 22(3044):122215, 1972.

[19] ByungSoo Jeon, Inah Jeon, Lee Sael, and U Kang. Scout: Scalable coupled matrix-tensor factorization-algorithm and discoveries. In *2016 IEEE 32nd International Conference on Data Engineering (ICDE)*, pages 811–822. IEEE, 2016.

[20] Hyunsoo Kim and Haesun Park. Nonnegative matrix factorization based on alternating nonnegativity constrained least squares and active set method. *SIAM journal on matrix analysis and applications*, 30(2):713–730, 2008.

[21] Jingu Kim and Haesun Park. Fast nonnegative tensor factorization with an active-set-like method. In *High-Performance Scientific Computing*, pages 311–326. Springer, 2012.

[22] T.G. Kolda and B.W. Bader. Tensor decompositions and applications. *SIAM review*, 51(3), 2009.

[23] Shuangzhe Liu and Gõtz Trenkler. Hadamard, khatri-rao, kronecker and other matrix products. *International Journal of Information and Systems Sciences*, 4(1):160–177, 2008.

[24] Anant Madabhushi and George Lee. Image analysis and machine learning in digital pathology: Challenges and opportunities, 2016.

[25] Olivera Novović, Sanja Brdar, and Vladimir Crnojević. Evolving connectivity graphs in mobile phone data. In *NetMob, The main conference on the scientific analysis of mobile phone datasets*, pages 73–75, 2017.

[26] Evangelos E Papalexakis. Automatic unsupervised tensor mining with quality assessment. In *Proceedings of the 2016 SIAM Int. Conf. on Data Mining*. SIAM, 2016.

[27] Evangelos E Papalexakis, Tom M Mitchell, Nicholas D Sidiropoulos, Christos Faloutsos, Partha Pratim Talukdar, and Brian Murphy. Scoup-smt: Scalable coupled sparse matrix-tensor factorization. *arXiv preprint arXiv:1302.7043*, 2013.

[28] Tara Safavi, Chandra Sripada, and Danai Koutra. Fast network discovery on sequence data via time-aware hashing. *Knowledge and Information Systems*, 61(2):987–1017, 2019.