

# CTD: Fast, Accurate, and Interpretable Method for Static and Dynamic Tensor Decompositions

Jungwoo Lee  
Seoul National University  
muon9401@gmail.com

Dongjin Choi  
Seoul National University  
skywalker5@snu.ac.kr

Lee Sael  
The State University of New York  
(SUNY) Korea  
sael@sunykorea.ac.kr

## ABSTRACT

How can we find patterns and anomalies in a tensor, or multi-dimensional array, in an efficient and directly interpretable way? How can we do this in an online environment, where a new tensor arrives each time step? Finding patterns and anomalies in a tensor is a crucial problem with many applications, including building safety monitoring, patient health monitoring, cyber security, terrorist detection, and fake user detection in social networks. Standard PARAFAC and Tucker decomposition results are not directly interpretable. Although a few sampling-based methods have previously been proposed towards better interpretability, they need to be made faster, more memory efficient, and more accurate.

In this paper, we propose CTD, a fast, accurate, and directly interpretable tensor decomposition method based on sampling. CTD-S, the static version of CTD, provably guarantees a high accuracy that is 17~83× more accurate than that of the state-of-the-art method. Also, CTD-S is made 5~86× faster, and 7~12× more memory-efficient than the state-of-the-art method by removing redundancy. CTD-D, the dynamic version of CTD, is the first interpretable dynamic tensor decomposition method ever proposed. Also, it is made 2~3× faster than already fast CTD-S by exploiting factors at previous time step and by reordering operations. With CTD, we demonstrate how the results can be effectively interpreted in the online distributed denial of service (DDoS) attack detection.

## CCS CONCEPTS

• **Information systems** → *Data mining*;

## 1 INTRODUCTION

Given a tensor, or multi-dimensional array, how can we find patterns and anomalies in an efficient and directly interpretable way? How can we do this in an online environment, where a new tensor arrives each time step? Many real-world data are multi-dimensional and can be modeled as sparse tensors. Examples include network traffic data (source IP - destination IP - time), movie rating data (user - movie - time), IoT sensor data, and healthcare data. Finding patterns and anomalies in those tensor data is a very important problem with many applications such as building safety monitoring [10], patient health monitoring [5, 13, 15, 22], cyber security [19], terrorist detection [1, 2, 14], and fake user detection in social networks [4, 11]. Tensor decomposition method, a widely-used tool in tensor analysis, has been used for this task. However, the standard tensor decomposition methods such as PARAFAC [9] and Tucker [21] do not provide interpretability and are not applicable for real-time analysis in environments with high-velocity data.

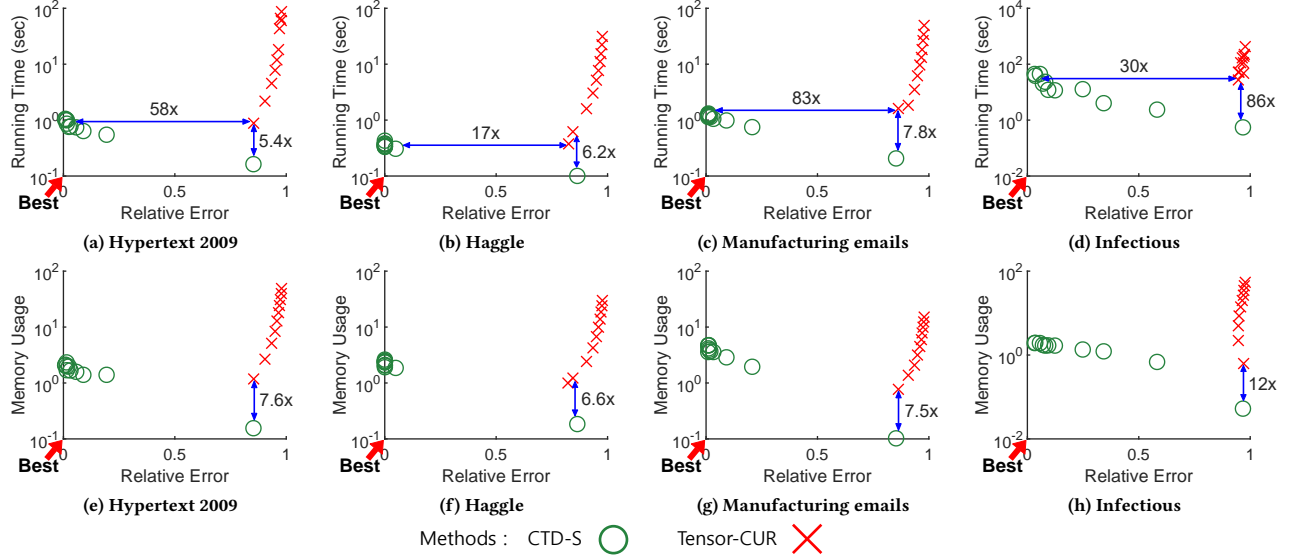
**Table 1: Comparison of our proposed CTD and the existing tensor-CUR. The static method CTD-S outperforms the state-of-the-art tensor-CUR in terms of time, memory usage, and accuracy. The dynamic method CTD-D is the fastest and the most accurate.**

	Existing	[Proposed]	
	Tensor-CUR [12]	CTD-S	CTD-D
<b>Interpretability</b>	✓	✓	✓
<b>Time</b>	fast	faster	<b>fastest</b>
<b>Memory usage</b>	low	<b>lower</b>	low
<b>Accuracy</b>	low	high	<b>highest</b>
<b>Online</b>			✓

Sampling-based tensor decomposition methods [3, 7, 12] arose as an alternative to due to their direct interpretability. The direct interpretability not only reduces time and effort involved in finding patterns and anomalies from the decomposed tensors but also provides clarity in interpreting the result. A sampling-based decomposition method for sparse tensors is also memory-efficient since it preserves the sparsity of the original tensors on the sampled factor matrices. However, existing sampling-based tensor decomposition methods are slow, have high memory usage, and produce low accuracy. For example, tensor-CUR [12], the state-of-the-art sampling-based static tensor decomposition method, has many redundant fibers including duplicates in its factors. These redundant fibers cause higher memory usage and longer running time. Tensor-CUR is also not accurate enough for real-world tensor analysis.

In addition to interpretability, demands for online method applicable in a dynamic environment, where multi-dimensional data are generated continuously at a fast rate, are also increasing. A real-time analysis is not feasible with static methods since all the data, i.e., historical and incoming tensors, need to be decomposed over again at each time step. There are a few dynamic tensor decomposition methods proposed [16, 17, 23]. However, proposed methods are not directly interpretable and do not preserve sparsity. To the best of our knowledge, there has been no sampling-based dynamic tensor decomposition method proposed.

In this paper, we propose CTD (Compact Tensor Decomposition), a fast, accurate, and interpretable sampling-based tensor decomposition method. CTD has two versions: CTD-S for static tensors, and CTD-D for dynamic tensors. CTD-S is optimal after sampling, and results in a compact tensor decomposition through careful sampling and redundancy elimination, thereby providing much better running time and memory efficiency than previous methods. CTD-D, the first sampling-based dynamic tensor decomposition



**Figure 1: Error, running time, and memory usage of CTD-S compared to those of tensor-CUR. CTD-S is more accurate, faster and more memory-efficient than tensor-CUR.**

method in literature, updates and modifies minimally on the components altered by the incoming data, making the method applicable for real-time analysis on a dynamic environment. Table 1 shows the comparison of CTD and the existing method, tensor-CUR. Our main contributions are as follows:

- **Method.** We propose CTD, a fast, accurate, and directly interpretable tensor decomposition method. We prove the optimality of the static method CTD-S which makes it more accurate than the state-of-the-art method. Also, to the best of our knowledge, the dynamic method CTD-D is the first sampling-based dynamic tensor decomposition method.
- **Performance.** CTD-S is 17~83x more accurate, 5~86x faster, and 7~12x more memory-efficient compared to tensor-CUR, the state-of-the-art competitor as shown in Figure 1. CTD-D is up to 3x faster than CTD-S as shown in Figure 5.
- **Interpretable Analysis.** We show how CTD results are directly interpreted to successfully detect DDoS attacks in network traffic data.

The code for CTD and datasets are available at <https://datalab.snu.ac.kr/CTD/>. The rest of this paper is organized as follows. Section 2 describes preliminaries and related works for tensor and sampling-based decomposition. Section 3 describes our proposed method CTD. Section 4 presents the experimental results. After presenting CTD at work in Section 5, we conclude in Section 6.

## 2 PRELIMINARIES AND RELATED WORKS

In this section, we describe preliminaries and related works for tensor and sampling-based decompositions. Table 2 lists the definitions of symbols used in this paper.

### 2.1 Tensor

A tensor is a multi-dimensional array and is denoted by the boldface Euler script, e.g.  $\mathcal{X} \in \mathbb{R}^{I_1 \times \dots \times I_N}$  where  $N$  denotes the order (the number of axes) of  $\mathcal{X}$ . Each axis of a tensor is also known as mode or way. A fiber is a vector (1-mode tensor) which has fixed indices except one. Every index of a mode- $n$  fiber is fixed except  $n$ -th index. A fiber can be regarded as a higher-order version of a matrix row and column. A matrix column and row each correspond to mode-1 fiber and mode-2 fiber, respectively. A slab is an  $(N - 1)$ -mode tensor which has only one fixed index.  $\mathbf{X}_{(\alpha)} \in \mathbb{R}^{I_\alpha \times N_\alpha}$  denotes a mode- $\alpha$  matricization of  $\mathcal{X}$ , where  $N_\alpha = \prod_{n \neq \alpha} I_n$ .  $\mathbf{X}_{(\alpha)}$  is made by rearranging mode- $\alpha$  fibers of  $\mathcal{X}$  to be the columns of  $\mathbf{X}_{(\alpha)}$ .  $\|\mathcal{X}\|_F$  is the Frobenius norm of  $\mathcal{X}$  and is defined by Equation 1.

$$\|\mathcal{X}\|_F^2 = \sum_{i_1, i_2, \dots, i_N} x_{i_1 i_2 \dots i_N}^2 \quad (1)$$

$\mathcal{X} \times_n \mathbf{U} \in \mathbb{R}^{I_1 \times \dots \times I_{n-1} \times J \times I_{n+1} \times \dots \times I_N}$  denotes the  $n$ -mode product of a tensor  $\mathcal{X} \in \mathbb{R}^{I_1 \times \dots \times I_N}$  with a matrix  $\mathbf{U} \in \mathbb{R}^{J \times I_n}$ . Elementwise,

$$(\mathcal{X} \times_n \mathbf{U})_{i_1 \dots i_{n-1} j i_{n+1} \dots i_N} = \sum_{i_n=1}^{I_n} x_{i_1 \dots i_{n-1} i_n i_{n+1} \dots i_N} u_{j i_n} \quad (2)$$

$\mathcal{X} \times_n \mathbf{U}$  has a property shown in Equation 3.

$$\mathcal{Y} = \mathcal{X} \times_n \mathbf{U} \Leftrightarrow \mathbf{Y}_{(n)} = \mathbf{U} \mathbf{X}_{(n)} \quad (3)$$

We assume that a matrix or tensor is stored in a sparse-unordered representation (i.e. only nonzero entries are stored in a form of pair of indices and the corresponding value).  $\text{nnz}(\mathcal{X})$  denotes the number of nonzero elements in  $\mathcal{X}$ .

We describe existing sampling-based matrix and tensor decomposition methods in the following subsections.

Table 2: Table of symbols.

Symbol	Definition
$\mathcal{X}$	tensor (Euler script, bold letter)
$\mathbf{X}$	matrix (uppercase, bold letter)
$\mathbf{x}$	column vector (lower case, bold letter)
$x$	scalar (lower case, italic letter)
$\mathbf{X}_{(n)}$	mode- $n$ matricization of a tensor $\mathcal{X}$
$\mathbf{X}^\dagger$	pseudoinverse of $\mathbf{X}$
$N$	order (number of modes) of a tensor
$\times_n$	$n$ -mode product
$\ \bullet\ _F$	Frobenius norm
$\text{nnz}(\mathcal{X})$	number of nonzero elements in $\mathcal{X}$

## 2.2 Sampling Based Matrix Decomposition

Sampling-based matrix decomposition methods sample columns or rows from a given matrix and use them to make their factors. They produce directly interpretable factors which preserve sparsity since those factors directly reflect the sparsity of the original data. In contrast, a singular value decomposition (SVD) generates factors which are hard to understand and dense because the factors are in a form of linear combination of columns or rows from the given matrix. Definition 2.1 shows the definition for CX matrix decomposition [8], a kind of sampling-based matrix decomposition.

*Definition 2.1.* Given a matrix  $\mathbf{A} \in \mathbb{R}^{m \times n}$ , the matrix  $\tilde{\mathbf{A}} = \mathbf{C}\mathbf{X}$  is a CX matrix decomposition of  $\mathbf{A}$ , where a matrix  $\mathbf{C} \in \mathbb{R}^{m \times c}$  consists of actual columns of  $\mathbf{A}$  and a matrix  $\mathbf{X}$  is any matrix of size  $c \times n$ .

We introduce well-known CX matrix decomposition methods: LinearTimeCUR, CMD, and Colibri.

*LinearTimeCUR and CMD.* Drineas et al. [6] proposed LinearTimeCUR and Sun et al. [18] proposed CMD. In the initial step, LinearTimeCUR and CMD sample columns from an original matrix  $\mathbf{A}$  according to the probabilities proportional to the norm of each column with replacement. Drineas et al. [6] has proven that this biased sampling provides an optimal approximation. Then, they project  $\mathbf{A}$  into the column space spanned by those sampled columns and use the projection as the low-rank approximation of  $\mathbf{A}$ . LinearTimeCUR has many duplicates in its factors because a column or row with a higher norm is likely to be selected multiple times. These duplicates make LinearTimeCUR slow and require a large amount of memory. CMD handles the duplication issue by removing duplicate columns and rows in the factors of LinearTimeCUR, thereby reducing running time and memory significantly.

*Colibri.* Tong et al. [20] proposed Colibri-S which improves CMD by removing all types of linear dependencies including duplicates. Colibri-S is much faster and memory-efficient compared to LinearTimeCUR and CMD because the dimension of factors is much smaller than that of LinearTimeCUR and CMD. Tong et al. [20] also proposed the dynamic version Colibri-D. Although Colibri-D can update its factors incrementally, it fixes the indices of the initially sampled columns which need to be updated over time. Our CTD-D not only handles general dynamic tensors but also does not have to fix those indices.

## 2.3 Sampling Based Tensor Decomposition

Sampling-based tensor decomposition method samples actual fibers or slabs from an original tensor. In contrast to PARAFAC and Tucker, the most famous tensor decomposition methods, the resulting factors of sampling-based tensor decomposition method are easy to understand and usually sparse. There are two types of sampling based tensor decomposition: one based on Tucker and the other based on LR tensor decomposition which is defined in Definition 2.2. In Tucker-type sampling based tensor decomposition (e.g., Approx-TensorSVD [7] and FBTD (fiber-based tensor decomposition) [3]), factor matrices for all modes are either sampled or generated; the overhead of generating a factor matrix for each mode makes these methods too slow for applications to real-time analysis. We focus on sampling methods based on LR tensor decomposition which is faster than those based on Tucker decomposition.

*Definition 2.2.* (LR tensor decomposition) Given a tensor  $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ ,  $\tilde{\mathcal{X}} = \mathcal{L} \times_\alpha \mathbf{R}$  is a mode- $\alpha$  LR tensor decomposition of  $\mathcal{X}$ , where a matrix  $\mathbf{R} \in \mathbb{R}^{I_\alpha \times c}$  consists of actual mode- $\alpha$  fibers of  $\mathcal{X}$  and a tensor  $\mathcal{L}$  is any tensor of size  $I_1 \times \dots \times I_{\alpha-1} \times c \times I_{\alpha+1} \times \dots \times I_N$ .

*Tensor-CUR.* Mahoney et al. [12] proposed tensor-CUR, a mode- $\alpha$  LR tensor decomposition method. Tensor-CUR is an  $n$ -dimensional extension of LinearTimeCUR. Tensor-CUR samples fibers and slabs from an original tensor and builds its factors using the sampled ones. The only difference between LinearTimeCUR and tensor-CUR is that tensor-CUR exploits fibers and slabs instead of columns and rows. Thus, tensor-CUR has drawbacks similar to those of LinearTimeCUR. Tensor-CUR has many redundant fibers in its factors and these fibers make tensor-CUR slow and use a large amount of memory.

## 3 PROPOSED METHOD

In this section, we describe our proposed CTD (Compact Tensor Decomposition), an efficient and interpretable sampling-based tensor decomposition method. We first describe the static version CTD-S, and then the dynamic version CTD-D of CTD.

### 3.1 CTD-S for Static Tensors

*Overview.* How can we design an efficient sampling-based static tensor decomposition method? Tensor-CUR, the existing state-of-the-art, has many redundant fibers in its factors and these fibers make tensor-CUR slow and use large memory. Our proposed CTD-S method removes all dependencies from the sampled fibers and maintains only independent fibers; thus, CTD-S is faster and more memory-efficient than tensor-CUR.

*Algorithm.* Figure 2 shows the scheme for CTD-S. CTD-S first samples fibers biased toward a norm of each fiber. Three different fibers (red, blue, green) are sampled in Figure 2. There are many duplicates after biased sampling process since CTD-S samples fibers multiple times with replacement and a fiber with a higher norm is likely to be sampled many times. There also exist linearly dependent fibers such as the green fiber which can be expressed as a linear combination of the red one and the blue one. Those linearly dependent fibers including duplicates are redundant in that they do not give new information when interpreting the result. CTD-S removes those redundant fibers and stores only the independent

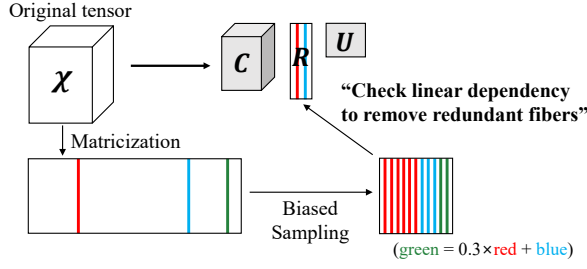


Figure 2: The scheme for CTD-S.

fibers in its factor  $\mathbf{R}$  to keep result compact. CTD-S only keeps one red fiber and one blue fiber in  $\mathbf{R}$  in Figure 2.

CTD-S decomposes a tensor  $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$  into one tensor  $\mathcal{C} \in \mathbb{R}^{I_1 \times \dots \times I_{\alpha-1} \times \tilde{s} \times I_{\alpha+1} \times \dots \times I_N}$ , and two matrices  $\mathbf{U} \in \mathbb{R}^{\tilde{s} \times \tilde{s}}$  and  $\mathbf{R} \in \mathbb{R}^{I_\alpha \times \tilde{s}}$  such that  $\mathcal{X} \approx \mathcal{C} \times_\alpha \mathbf{R} \mathbf{U}$ . CTD-S is a mode- $\alpha$  LR tensor decomposition method and is interpretable since  $\mathbf{R}$  consists of independent fibers sampled from  $\mathcal{X}$ .

---

**Algorithm 1** CTD-S for Static Tensor

---

**Input:** Tensor  $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ , mode  $\alpha \in \{1, \dots, N\}$ , sample size  $s \in \{1, \dots, N_\alpha\}$ , and tolerance  $\epsilon$   
**Output:**  $\mathcal{C} \in \mathbb{R}^{I_1 \times \dots \times I_{\alpha-1} \times \tilde{s} \times I_{\alpha+1} \times \dots \times I_N}$ ,  $\mathbf{U} \in \mathbb{R}^{\tilde{s} \times \tilde{s}}$ ,  $\mathbf{R} \in \mathbb{R}^{I_\alpha \times \tilde{s}}$

- 1: Let  $\mathbf{X}_{(\alpha)}$  be the mode- $\alpha$  matricization of  $\mathcal{X}$
- 2: Compute column distribution for  $i = 1, \dots, N_\alpha$ :  

$$P(i) \leftarrow \frac{|\mathbf{X}_{(\alpha)}(:, i)|^2}{\|\mathbf{X}_{(\alpha)}\|_F^2}$$
- 3: Sample  $s$  columns from  $\mathbf{X}_{(\alpha)}$  based on  $P(i)$ . Let  $I = \{i_1, \dots, i_s\}$
- 4: Let  $I' = \{i'_1, \dots, i'_s\}$  be a set consisting of unique elements in  $I$
- 5: Initialize  $\mathbf{R} \leftarrow [\mathbf{X}_{(\alpha)}(:, i'_1)]$  and  $\mathbf{U} \leftarrow 1/(\mathbf{X}_{(\alpha)}(:, i'_1)^T \mathbf{X}_{(\alpha)}(:, i'_1))$
- 6: **for**  $k = 2 : s' : s$  **do**
- 7:   Compute the residual:  

$$\vec{r}e\vec{s} \leftarrow (\mathbf{X}_{(\alpha)}(:, i'_k) - \mathbf{R} \mathbf{U}^T \mathbf{X}_{(\alpha)}(:, i'_k))$$
- 8:   **if**  $\|\vec{r}e\vec{s}\| \leq \epsilon \|\mathbf{X}_{(\alpha)}(:, i'_k)\|$  **then**
- 9:     continue
- 10:   **else**
- 11:     Compute:  $\delta \leftarrow \|\vec{r}e\vec{s}\|^2$  and  $\vec{y} \leftarrow \mathbf{U}^T \mathbf{X}_{(\alpha)}(:, i'_k)$
- 12:     Update  $\mathbf{U}$ :  

$$\mathbf{U} \leftarrow \begin{pmatrix} \mathbf{U} + \vec{y} \vec{y}^T / \delta & -\vec{y} / \delta \\ -\vec{y}^T / \delta & 1 / \delta \end{pmatrix}$$
- 13:     Expand  $\mathbf{R} : \mathbf{R} \leftarrow [\mathbf{R}, \mathbf{X}_{(\alpha)}(:, i'_k)]$
- 14:   **end if**
- 15: **end for**
- 16: Compute  $\mathcal{C} \leftarrow \mathcal{X} \times_\alpha \mathbf{R}^T$
- 17: **return**  $\mathcal{C}, \mathbf{U}, \mathbf{R}$

---

Algorithm 1 shows the procedure of CTD-S. First, CTD-S computes the probabilities of mode- $\alpha$  fibers of  $\mathcal{X}$ , which are proportional to the norm of each fiber, and then samples  $s$  fibers from  $\mathcal{X}$  according to the probabilities with replacement, in lines 1-3. Redundant fibers exist in the sampled fibers in this step. CTD-S selects unique fibers from the initially sampled  $s$  fibers in line 4 where  $s'$  denotes the number of those unique fibers. This step reduces the number of iterations in lines 6-15 from  $s - 1$  to  $s' - 1$ .  $\mathbf{R}$  is initialized by the first sampled fiber in line 5. In lines 6-15, CTD-S removes redundant mode- $\alpha$  fibers in the sampled fibers. The matrices  $\mathbf{U}$  and  $\mathbf{R}$  are computed incrementally in this step. The columns of  $\mathbf{R}$

always consist of independent mode- $\alpha$  fibers through the loop. In each iteration, CTD-S checks whether one of the sampled fibers is linearly independent of the column space spanned by  $\mathbf{R}$  or not in lines 7-8. If the fiber is independent, CTD-S updates  $\mathbf{U}$  and expands  $\mathbf{R}$  with the fiber in lines 10-13. Finally, CTD-S computes  $\mathcal{C}$  with  $\mathcal{X}$  and  $\mathbf{R}$  in line 16.

Lemma 3.1 shows the computational cost of CTD-S.

**LEMMA 3.1.** *The computational complexity of CTD-S is  $O((\tilde{s}I_\alpha + s)N_\alpha + s'(\tilde{s}^2 + \text{nnz}(\mathbf{R})) + s \log s + \text{nnz}(\mathcal{X}))$ , where  $N_\alpha$  is  $\prod_{n \neq \alpha} I_n$  and  $\tilde{s} \ll s' \leq s$ .*

**PROOF.** The mode- $\alpha$  matricization of  $\mathcal{X}$  in line 1 needs  $O(\text{nnz}(\mathcal{X}))$  operations. Computing column distribution in line 2 takes  $O(\text{nnz}(\mathcal{X}) + N_\alpha)$  and sampling  $s$  columns in line 3 takes  $O(sN_\alpha)$ .  $O(s \log s)$  operation is required in computing unique elements in  $I$  in line 4. Computing  $\mathbf{R}$  and  $\mathbf{U}$  in lines 5-15 takes  $O(s'(\tilde{s}^2 + \text{nnz}(\mathbf{R})))$  as proved in Lemma 1 in [20]. Computing  $\mathcal{C}$  in line 16 takes  $O(\tilde{s}I_\alpha N_\alpha)$ . Overall, CTD-S needs  $O((\tilde{s}I_\alpha + s)N_\alpha + s'(\tilde{s}^2 + \text{nnz}(\mathbf{R})) + s \log s + \text{nnz}(\mathcal{X}))$  operations.  $\square$

Lemma 3.2 shows that CTD-S has the optimal accuracy for given sampled fibers, and thus is more accurate than tensor-CUR.

**LEMMA 3.2.** *CTD-S is more accurate than tensor-CUR. For a given  $\mathbf{R}_0$  consisting of initially sampled fibers, CTD-S has the minimum error.*

**PROOF.** CTD-S and tensor-CUR are both mode- $\alpha$  LR tensor decomposition methods and have errors in the form of Equation 4.

$$\|\mathcal{X} - \mathcal{L} \times_\alpha \mathbf{R}\|_F = \|\mathbf{X}_{(\alpha)} - \mathbf{R} \mathbf{L}_{(\alpha)}\|_F \quad (4)$$

where the equality comes from Equation 3. They both sample fibers from  $\mathcal{X}$  in the same way in their initial step. Let  $\mathbf{R}_0$  be the matrix consisting of those initially sampled fibers. Assume the same  $\mathbf{R}_0$  is given for both methods. Then, the error is now a function of  $\mathbf{L}_{(\alpha)}$  as shown in Equation 5. Equation 6 shows the  $\mathbf{L}_{(\alpha)}$  which minimizes the error.

$$f(\mathbf{L}_{(\alpha)}) = \|\mathbf{X}_{(\alpha)} - \mathbf{R}_0 \mathbf{L}_{(\alpha)}\|_F \quad (5)$$

$$\arg \min_{\mathbf{L}_{(\alpha)}} f(\mathbf{L}_{(\alpha)}) = \mathbf{R}_0^\dagger \mathbf{X}_{(\alpha)} \quad (6)$$

We show CTD-S has the minimum error. Let  $\mathbf{R}$  consists of the independent columns of  $\mathbf{R}_0$ .

$$\begin{aligned} \|\mathbf{X}_{(\alpha)} - \mathbf{R}_0 \mathbf{R}_0^\dagger \mathbf{X}_{(\alpha)}\|_F &= \|\mathbf{X}_{(\alpha)} - \mathbf{R} \mathbf{R}^\dagger \mathbf{X}_{(\alpha)}\|_F \\ &= \|\mathbf{X}_{(\alpha)} - \mathbf{R} (\mathbf{R}^T \mathbf{R})^{-1} \mathbf{R}^T \mathbf{X}_{(\alpha)}\|_F \\ &= \|\mathbf{X}_{(\alpha)} - \mathbf{R} \mathbf{U} \mathbf{C}\|_F \\ &= \text{Error of CTD-S} \end{aligned} \quad (7)$$

The first equality in Equation 7 holds because  $\mathbf{R}_0 \mathbf{R}_0^\dagger \mathbf{X}_{(\alpha)}$  means the projection of  $\mathbf{X}_{(\alpha)}$  onto the column space of  $\mathbf{R}_0$ , and  $\mathbf{R}$  and  $\mathbf{R}_0$  have the same column space. The third equality holds because CTD-S uses  $(\mathbf{R}^T \mathbf{R})^{-1}$  for its factor  $\mathbf{U}$  (theorem 1 in [20]), and  $\mathbf{R}^T \mathbf{X}_{(\alpha)}$  for its factor  $\mathbf{C}$ . In contrast, tensor-CUR does not have the minimum error because  $f(\mathbf{L}_{(\alpha)})$  in the Equation 5 is convex and tensor-CUR has  $\mathbf{L}_{(\alpha)}$  which is different from  $\mathbf{R}_0^\dagger \mathbf{X}_{(\alpha)}$ . Specifically, tensor-CUR further samples rows (called slabs) from  $\mathbf{X}_{(\alpha)}$  to construct its  $\mathbf{L}_{(\alpha)}$ .  $\square$

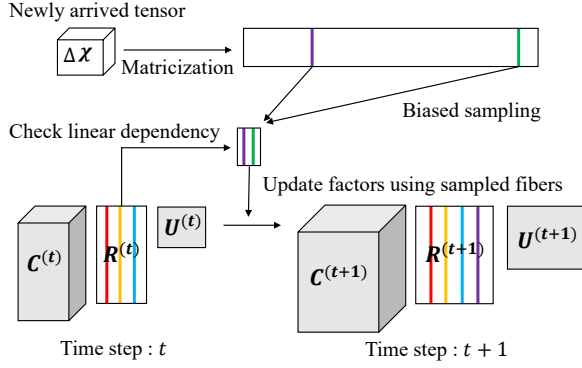


Figure 3: The scheme for CTD-D.

### 3.2 CTD-D for Dynamic Tensors

*Overview.* How can we design an efficient sampling-based dynamic tensor decomposition method? In a dynamic setting, a new tensor arrives at every time step and we want to keep track of sampling-based tensor decomposition. The main challenge is to update factors quickly while preserving accuracy. Note that there has been no sampling-based dynamic tensor decomposition method in the literature. Applying CTD-S at every time step is not a feasible option since it starts from scratch to update its factors, and thus running time increases rapidly as tensor grows. We propose CTD-D, the first sampling-based dynamic tensor decomposition method. CTD-D samples mode- $\alpha$  fibers only from the newly arrived tensor, and then updates the factors appropriately using those sampled ones. The main idea of CTD-D is to update the factors of CTD-S incrementally by (1) exploiting factors at previous time step and (2) reordering operations.

*Algorithm.* Figure 3 shows the scheme for CTD-D. At each time step, CTD-D samples fibers from newly arrived tensor and updates factors by checking linear dependency of sampled fibers with the factor at previous time step. Purple and green fiber are sampled from newly arrived tensor in Figure 3. Note that the purple fiber is added to the factor  $\mathbf{R}$  since it is linearly independent of the fibers in the factor at the previous time step, while the linearly dependent green fiber is ignored.

For any time step  $t$ , CTD-D maintains its factors  $\mathcal{C}^{(t)} \in \mathbb{R}^{I_1 \times \dots \times I_{\alpha-1} \times \tilde{d}_t \times I_{\alpha+1} \times \dots \times I_{N-1} \times t}$ ,  $\mathbf{U}^{(t)} \in \mathbb{R}^{\tilde{d}_t \times \tilde{d}_t}$ , and  $\mathbf{R}^{(t)} \in \mathbb{R}^{I_\alpha \times \tilde{d}_t}$  such that  $\mathcal{X}^{(t)} \approx \mathcal{C}^{(t)} \times_\alpha \mathbf{R}^{(t)} \mathbf{U}^{(t)}$ , where the upper subscript  $(t)$  indicates that the factor is at time step  $t$ .  $\mathcal{X}^{(t)}$  grows along the time mode and we assume that  $N$ -th mode is the time mode in a dynamic setting, where  $N$  denotes the order of  $\mathcal{X}^{(t)}$ . At the next time step  $t+1$ , CTD-D receives newly arrived tensor  $\Delta\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_{\alpha-1} \times I_{\alpha+1} \times \dots \times I_{N-1} \times 1}$  and updates  $\mathcal{C}^{(t)}$ ,  $\mathbf{U}^{(t)}$ , and  $\mathbf{R}^{(t)}$  into  $\mathcal{C}^{(t+1)} \in \mathbb{R}^{I_1 \times \dots \times I_{\alpha-1} \times \tilde{d}_{t+1} \times I_{\alpha+1} \times \dots \times I_{N-1} \times (t+1)}$ ,  $\mathbf{U}^{(t+1)} \in \mathbb{R}^{\tilde{d}_{t+1} \times \tilde{d}_{t+1}}$ , and  $\mathbf{R}^{(t+1)} \in \mathbb{R}^{I_\alpha \times \tilde{d}_{t+1}}$ , respectively such that  $\mathcal{X}^{(t+1)} \approx \mathcal{C}^{(t+1)} \times_\alpha \mathbf{R}^{(t+1)} \mathbf{U}^{(t+1)}$ .

Algorithm 2 shows the procedure of CTD-D. First, CTD-D computes the probabilities of mode- $\alpha$  fibers of  $\Delta\mathcal{X}$ , which are proportional to the norm of each fiber, and then samples  $d$  fibers according to the probabilities with replacement in lines 1-3. CTD-D selects

---

#### Algorithm 2 CTD-D for Dynamic Tensor

---

**Input:** Tensor  $\Delta\mathcal{X} \in \mathbb{R}^{I_1 \times \dots \times I_{N-1} \times 1}$ , mode  $\alpha \in \{1, \dots, N-1\}$ ,  $\mathcal{C}^{(t)}$ ,  $\mathbf{U}^{(t)}$ ,  $\mathbf{R}^{(t)}$ , sample size  $d \in \{1, \dots, \Delta N_\alpha\}$ , and tolerance  $\epsilon$   
**Output:**  $\mathcal{C}^{(t+1)}$ ,  $\mathbf{U}^{(t+1)}$ ,  $\mathbf{R}^{(t+1)}$

- 1: Let  $\Delta\mathbf{X}_{(\alpha)}$  be the mode- $\alpha$  matricization of  $\Delta\mathcal{X}$
- 2: Compute column distribution for  $i = 1, \dots, \Delta N_\alpha$ :  

$$P(i) \leftarrow \frac{|\Delta\mathbf{X}_{(\alpha)}(:, i)|^2}{\|\Delta\mathbf{X}_{(\alpha)}\|_F^2}$$
- 3: Sample  $d$  columns from  $\Delta\mathbf{X}_{(\alpha)}$  based on  $P(i)$ . Let  $I = \{i_1, \dots, i_d\}$
- 4: Let  $I' = \{i'_1, \dots, i'_{d'}\}$  be a set consisting of unique elements in  $I$
- 5: Initialize  $\mathbf{R}^{(t+1)} \leftarrow \mathbf{R}^{(t)}$ ,  $\mathbf{U}^{(t+1)} \leftarrow \mathbf{U}^{(t)}$ , and  $\Delta\mathbf{R} \leftarrow []$
- 6: **for**  $k = 1 : d'$  **do**
- 7:   Let  $\mathbf{x} \leftarrow \Delta\mathbf{X}_{(\alpha)}(:, i'_k)$
- 8:   Compute the residual:  

$$\overline{\mathbf{r}}\mathbf{e}\mathbf{s} \leftarrow (\mathbf{x} - \mathbf{R}^{(t+1)}\mathbf{U}^{(t+1)}(\mathbf{R}^{(t+1)})^T\mathbf{x})$$
- 9:   **if**  $\|\overline{\mathbf{r}}\mathbf{e}\mathbf{s}\| \leq \epsilon \|\mathbf{x}\|$  **then**
- 10:     **continue**
- 11:   **else**
- 12:     Compute:  $\delta \leftarrow \|\overline{\mathbf{r}}\mathbf{e}\mathbf{s}\|^2$  and  $\overline{\mathbf{y}} \leftarrow \mathbf{U}^{(t+1)}(\mathbf{R}^{(t+1)})^T\mathbf{x}$
- 13:     Update  $\mathbf{U}^{(t+1)}$ :  

$$\mathbf{U}^{(t+1)} \leftarrow \begin{pmatrix} \mathbf{U}^{(t+1)} + \overline{\mathbf{y}}\overline{\mathbf{y}}^T/\delta & -\overline{\mathbf{y}}/\delta \\ -\overline{\mathbf{y}}^T/\delta & 1/\delta \end{pmatrix}$$
- 14:     Expand  $\mathbf{R}^{(t+1)}$  and  $\Delta\mathbf{R} : \mathbf{R}^{(t+1)} \leftarrow [\mathbf{R}^{(t+1)}, \mathbf{x}]$  and  $\Delta\mathbf{R} \leftarrow [\Delta\mathbf{R}, \mathbf{x}]$
- 15:   **end if**
- 16: **end for**
- 17: Update  $\mathcal{C}_{(\alpha)}^{(t+1)}$ :
- 18: **if**  $\Delta\mathbf{R}$  is not empty **then**
- 19:   
$$\mathcal{C}_{(\alpha)}^{(t+1)} \leftarrow \begin{pmatrix} \mathcal{C}_{(\alpha)}^{(t)} & (\mathbf{R}^{(t)})^T \Delta\mathbf{X}_{(\alpha)} \\ (\Delta\mathbf{R})^T \mathbf{R}^{(t)} \mathbf{U}^{(t)} \mathcal{C}_{(\alpha)}^{(t)} & (\Delta\mathbf{R})^T \Delta\mathbf{X}_{(\alpha)} \end{pmatrix}$$
- 20: **else**
- 21:   
$$\mathcal{C}_{(\alpha)}^{(t+1)} \leftarrow \begin{pmatrix} \mathcal{C}_{(\alpha)}^{(t)} & (\mathbf{R}^{(t)})^T \Delta\mathbf{X}_{(\alpha)} \end{pmatrix}$$
- 22: **end if**
- 23: Fold  $\mathcal{C}_{(\alpha)}^{(t+1)}$  into  $\mathcal{C}^{(t+1)}$
- 24: **return**  $\mathcal{C}^{(t+1)}$ ,  $\mathbf{U}^{(t+1)}$ ,  $\mathbf{R}^{(t+1)}$

---

unique  $d'$  fibers in line 4 and initializes  $\mathbf{R}^{(t+1)}$ ,  $\mathbf{U}^{(t+1)}$ , and  $\Delta\mathbf{R}$  with  $\mathbf{R}^{(t)}$ ,  $\mathbf{U}^{(t)}$ , and an empty matrix respectively in line 5, where  $\Delta\mathbf{R}$  consists of differences between  $\mathbf{R}^{(t)}$  and  $\mathbf{R}^{(t+1)}$ . In lines 6-16, CTD-D expands  $\mathbf{R}^{(t+1)}$  with those sampled fibers by sequentially evaluating linear dependency of each fiber with the column space of  $\mathbf{R}^{(t+1)}$ .  $\mathbf{R}^{(t+1)}$  and  $\mathbf{U}^{(t+1)}$  are updated in this step. Finally,  $\mathcal{C}_{(\alpha)}^{(t+1)}$  is updated in lines 17-21.

In the following, we describe two main ideas of CTD-D to update  $\mathcal{C}_{(\alpha)}^{(t+1)}$ ,  $\mathbf{R}^{(t+1)}$ , and  $\mathbf{U}^{(t+1)}$  efficiently while preserving accuracy: exploiting factors at previous time step, and reordering operations. **(1) Exploiting factors at previous time step:** First, we explain how we update  $\mathbf{R}^{(t+1)}$  and  $\mathbf{U}^{(t+1)}$  using the idea. In line 5 of Algorithm 1, CTD-S initializes  $\mathbf{R}$  and  $\mathbf{U}$  using one of the sampled fibers. This is because CTD-S requires  $\mathbf{R}$  to consist of linearly independent columns and it is satisfied when  $\mathbf{R}$  has only one fiber. Since  $\mathbf{R}^{(t)}$  already consists of linearly independent columns, we initialize  $\mathbf{R}^{(t+1)}$  and  $\mathbf{U}^{(t+1)}$  with  $\mathbf{R}^{(t)}$  and  $\mathbf{U}^{(t)}$  respectively in line 5 of Algorithm 2. In lines 6-16, we check linear independence of each sampled fiber from  $\Delta\mathbf{X}_{(\alpha)}$  with  $\mathbf{R}^{(t+1)}$ . If the fiber is linearly independent, we expand  $\mathbf{R}^{(t+1)}$  and update  $\mathbf{U}^{(t+1)}$  as in the lines 11-13 of Algorithm 1.

Second, we describe how we update  $\mathbf{C}^{(t+1)}$  using the idea. We assume that  $\Delta\mathbf{R}$  is not empty after line 16 of Algorithm 2. At time step  $t$  and its successor step  $t + 1$ , CTD-S satisfies Equations 8 and 9, where  $\mathbf{C}_{(\alpha)}^{(t)}$  has the size  $\tilde{d}_t \times N_\alpha^{(t)}$  and  $\mathbf{C}_{(\alpha)}^{(t+1)}$  has the size  $\tilde{d}_{t+1} \times N_\alpha^{(t+1)}$ .

$$\mathbf{C}_{(\alpha)}^{(t)} \leftarrow (\mathbf{R}^{(t)})^T \mathbf{X}_{(\alpha)}^{(t)} \quad (8)$$

$$\mathbf{C}_{(\alpha)}^{(t+1)} \leftarrow (\mathbf{R}^{(t+1)})^T \mathbf{X}_{(\alpha)}^{(t+1)} \quad (9)$$

We can rewrite  $\mathbf{R}^{(t+1)}$  and  $\mathbf{X}_{(\alpha)}^{(t+1)}$  as Equations 10 and 11 respectively, where  $\Delta\mathbf{R}$  has the size  $I_\alpha \times \Delta\tilde{d}$  and  $\Delta\mathbf{X}_{(\alpha)}$  has the size  $I_\alpha \times \Delta N_\alpha$  such that  $N_\alpha^{(t+1)} = N_\alpha^{(t)} + \Delta N_\alpha$  and  $\tilde{d}_{t+1} = \tilde{d}_t + \Delta\tilde{d}$ .

$$\mathbf{R}^{(t+1)} = [\mathbf{R}^{(t)} \quad \Delta\mathbf{R}] \quad (10)$$

$$\mathbf{X}_{(\alpha)}^{(t+1)} = [\mathbf{X}_{(\alpha)}^{(t)} \quad \Delta\mathbf{X}_{(\alpha)}] \quad (11)$$

We replace  $\mathbf{R}^{(t+1)}$  and  $\mathbf{X}_{(\alpha)}^{(t+1)}$  in Equation 9 with those in Equations 10 and 11, respectively, to obtain the Equation 12.

$$\begin{aligned} \mathbf{C}_{(\alpha)}^{(t+1)} &\leftarrow \begin{bmatrix} (\mathbf{R}^{(t)})^T & (\Delta\mathbf{R})^T \end{bmatrix} \begin{bmatrix} \mathbf{X}_{(\alpha)}^{(t)} & \Delta\mathbf{X}_{(\alpha)} \end{bmatrix} \\ &= \begin{bmatrix} (\mathbf{R}^{(t)})^T \mathbf{X}_{(\alpha)}^{(t)} & (\mathbf{R}^{(t)})^T \Delta\mathbf{X}_{(\alpha)} \\ (\Delta\mathbf{R})^T \mathbf{X}_{(\alpha)}^{(t)} & (\Delta\mathbf{R})^T \Delta\mathbf{X}_{(\alpha)} \end{bmatrix} \end{aligned} \quad (12)$$

CTD-S computes all the 4 elements  $(\mathbf{R}^{(t)})^T \mathbf{X}_{(\alpha)}^{(t)}$ ,  $(\mathbf{R}^{(t)})^T \Delta\mathbf{X}_{(\alpha)}$ ,  $(\Delta\mathbf{R})^T \mathbf{X}_{(\alpha)}^{(t)}$ , and  $(\Delta\mathbf{R})^T \Delta\mathbf{X}_{(\alpha)}$  in Equation 12 from scratch, hence requires a lot of computations. To make computation of  $\mathbf{C}_{(\alpha)}^{(t+1)}$  incremental, we exploit existing factors at time step  $t$ :  $\mathbf{C}_{(\alpha)}^{(t)}$ ,  $\mathbf{R}^{(t)}$ , and  $\mathbf{U}^{(t)}$ . First, we use  $\mathbf{C}_{(\alpha)}^{(t)}$  instead of  $(\mathbf{R}^{(t)})^T \mathbf{X}_{(\alpha)}^{(t)}$  as in the Equation 8. Second, we should replace  $\mathbf{X}_{(\alpha)}^{(t)}$  in  $(\Delta\mathbf{R})^T \mathbf{X}_{(\alpha)}^{(t)}$  with the factors at time step  $t$ , since CTD-D does not have  $\mathbf{X}_{(\alpha)}^{(t)}$  as its input unlike CTD-S. We substitute  $\mathbf{R}^{(t)}\mathbf{U}^{(t)}\mathbf{C}_{(\alpha)}^{(t)}$  for  $\mathbf{X}_{(\alpha)}^{(t)}$ . This is because CTD-S ensures  $\mathbf{X}^{(t)} \approx \mathbf{C}^{(t)} \times_\alpha \mathbf{R}^{(t)}\mathbf{U}^{(t)}$  which can be rewritten as  $\mathbf{X}_{(\alpha)}^{(t)} \approx \mathbf{R}^{(t)}\mathbf{U}^{(t)}\mathbf{C}_{(\alpha)}^{(t)}$  by Equation 3. Equation 13 shows the final form of  $\mathbf{C}_{(\alpha)}^{(t+1)}$  which is the same as line 18 in Algorithm 2.

$$\mathbf{C}_{(\alpha)}^{(t+1)} \leftarrow \begin{bmatrix} \mathbf{C}_{(\alpha)}^{(t)} & (\mathbf{R}^{(t)})^T \Delta\mathbf{X}_{(\alpha)} \\ (\Delta\mathbf{R})^T \mathbf{R}^{(t)}\mathbf{U}^{(t)}\mathbf{C}_{(\alpha)}^{(t)} & (\Delta\mathbf{R})^T \Delta\mathbf{X}_{(\alpha)} \end{bmatrix} \quad (13)$$

$(\Delta\mathbf{R})^T \mathbf{R}^{(t)}\mathbf{U}^{(t)}\mathbf{C}_{(\alpha)}^{(t)}$  and  $(\Delta\mathbf{R})^T \Delta\mathbf{X}_{(\alpha)}$  are ignored when  $\Delta\mathbf{R}$  is empty as expressed in line 20 of Algorithm 2.

**(2) Reordering computations:** The computation order for the element  $(\Delta\mathbf{R})^T \mathbf{R}^{(t)}\mathbf{U}^{(t)}\mathbf{C}_{(\alpha)}^{(t)}$  is important since each order has a different computation cost. We want to determine the optimal parenthesization among possible parenthesizations. It can be shown that  $((\Delta\mathbf{R})^T \mathbf{R}^{(t)})\mathbf{U}^{(t)}\mathbf{C}_{(\alpha)}^{(t)}$  is the optimal one with  $O((\Delta\tilde{d})\tilde{d}_t(I_\alpha + \tilde{d}_t + N_\alpha^{(t)}))$  operations and can be done by parenthesizing from the left.

We prove that CTD-D is faster than CTD-S in Lemma 3.3.

**Table 3: Summary of the tensor data used.**

Name	$I_1$	$I_2$	$I_3$	Nonzeros
Hypertext 2009 <sup>1</sup>	112	113	5,246	20,818
Hagggle <sup>2</sup>	77	274	1,567	27,972
Manufacturing emails <sup>3</sup>	167	166	2,615	70,523
Infectious <sup>4</sup>	407	410	1,392	17,298
CAIDA (synthetic)	189	189	1,000	20,511

LEMMA 3.3. *CTD-D is faster than CTD-S. The computational complexity of CTD-D is  $O((\Delta\tilde{d})\tilde{d}_t(N_\alpha^{(t)} + I_\alpha) + (\tilde{d}_{t+1}I_\alpha + d)\Delta N_\alpha + d'(\tilde{d}_{t+1}^2 + \text{nnz}(\mathbf{R}^{(t+1)})) + d \log d + \text{nnz}(\Delta\mathbf{X}))$ .*

PROOF. The lines 1-4 of Algorithm 2 for CTD-D are similar to those of Algorithm 1 for CTD-S. The only difference is that CTD-D samples  $d$  columns from  $\Delta\mathbf{X}_{(\alpha)}$  while CTD-S samples  $s$  columns from  $\mathbf{X}_{(\alpha)}$ . Thus, lines 1-4 takes  $O(\text{nnz}(\Delta\mathbf{X}) + d\Delta N_\alpha + d \log d)$ . Updating  $\mathbf{R}^{(t+1)}$  and  $\mathbf{U}^{(t+1)}$  in lines 5-16 needs  $O(d'(\tilde{d}_{t+1}^2 + \text{nnz}(\mathbf{R}^{(t+1)})))$  operations as proved in Lemma 1 in [20]. In updating  $\mathbf{C}^{(t+1)}$  in lines 17-18,  $(\mathbf{R}^{(t)})^T \Delta\mathbf{X}_{(\alpha)}$  takes computational cost of  $O(\tilde{d}_t I_\alpha \Delta N_\alpha)$ .  $(\Delta\mathbf{R})^T \Delta\mathbf{X}_{(\alpha)}$  takes  $O(\Delta\tilde{d} I_\alpha \Delta N_\alpha)$  and  $(\Delta\mathbf{R})^T \mathbf{R}^{(t)}\mathbf{U}^{(t)}\mathbf{C}_{(\alpha)}^{(t)}$  takes  $O((\Delta\tilde{d})\tilde{d}_t(I_\alpha + \tilde{d}_t + N_\alpha^{(t)}))$ . Overall, CTD-D takes  $O((\Delta\tilde{d})\tilde{d}_t(N_\alpha^{(t)} + I_\alpha) + (\tilde{d}_{t+1}I_\alpha + d)\Delta N_\alpha + d'(\tilde{d}_{t+1}^2 + \text{nnz}(\mathbf{R}^{(t+1)})) + d \log d + \text{nnz}(\Delta\mathbf{X}))$ .

CTD-D is faster than CTD-S because CTD-S has  $\tilde{s}I_\alpha N_\alpha$  in its complexity, which is much larger than all the terms in the complexity of CTD-D.  $\square$

## 4 EXPERIMENTS

In this and the next sections, we perform experiments to answer the following questions.

- Q1:** What is the performance of our static method CTD-S compared to the competing method tensor-CUR? (Section 4.2)
- Q2:** How do the performance of CTD-S and tensor-CUR change with regard to the sample size parameter? (Section 4.2)
- Q3:** What is the performance of our dynamic method CTD-D compared to the static method CTD-S? (Section 4.3)
- Q4:** What is the result of applying CTD-D for online DDoS attack detection? (Section 5)

### 4.1 Experimental Settings

*Measure.* We define three metrics (1. relative error, 2. memory usage, and 3. running time) as follows. First, a relative error is defined as Equation 14.  $\mathbf{X}$  denotes the original tensor and  $\tilde{\mathbf{X}}$  is the tensor reconstructed from the factors of  $\mathbf{X}$ . For example,  $\tilde{\mathbf{X}} = \mathbf{C} \times_\alpha \mathbf{R}\mathbf{U}$  in CTD-S.

$$\text{RelativeError} = \frac{\|\tilde{\mathbf{X}} - \mathbf{X}\|_F^2}{\|\mathbf{X}\|_F^2} \quad (14)$$

Second, memory usage is defined as Equation 15. Memory usage measures the relative amount of memory for storing the resulting

<sup>1</sup><http://konect.uni-koblenz.de/networks/sociopatterns-hypertext>

<sup>2</sup><http://konect.uni-koblenz.de/networks/contact>

<sup>3</sup>[http://konect.uni-koblenz.de/networks/radoslaw\\_email](http://konect.uni-koblenz.de/networks/radoslaw_email)

<sup>4</sup><http://konect.uni-koblenz.de/networks/sociopatterns-infectious>

factors. The denominator and numerator indicate the amount of memory for storing the original tensor and the resulting factors, respectively.

$$\text{MemoryUsage} = \frac{\text{nnz}(\mathcal{C}) + \text{nnz}(\mathbf{U}) + \text{nnz}(\mathbf{R})}{\text{nnz}(\mathcal{X})} \quad (15)$$

Finally, running time is measured in seconds.

*Data.* Table 3 shows the data we used in our experiments.

*Machine.* All the experiments are performed on a machine with a 10-core Intel 2.20 GHz CPU and 256 GB RAM.

*Competing method.* We compare our proposed method CTD with tensor-CUR [12], the state-of-the-art sampling-based tensor decomposition method. Both methods are implemented in MATLAB.

## 4.2 Performance of CTD-S

We measure the performance of CTD-S to answer Q1 and Q2. As a result, CTD-S is 17~83× more accurate for the same level of running time, 5~86× faster, and 7~12× more memory-efficient for the same level of error compared to tensor-CUR. CTD-S is more accurate over various sample sizes and its running time and memory usage are relatively constant compared to the tensor-CUR. The detail of the experiment is as follows.

Both CTD-S and tensor-CUR takes a given tensor  $\mathcal{X}$ , mode  $\alpha$ , and a sample size  $s$  as input because they are LR tensor decomposition methods. In each experiment, we give the same input and compare the performance. We set  $\alpha = 1$  and perform experiments under various sample sizes  $s$ . We set the number of slabs to sample  $r = s$  and the rank  $k = 10$  in tensor-CUR, and set  $\epsilon = 10^{-6}$  in CTD-S.

Figure 1 shows the running time vs. error and memory usage vs. error of CTD-S compared to tensor-CUR, which are the answers for Q1. We use sample sizes from 1 to 1000. The error of tensor-CUR is much larger than that of CTD-S. This phenomenon coincides with the Lemma 3.2, which guarantees that CTD-S is more accurate than tensor-CUR theoretically. We compare running time and memory usage under the same level of error, not under the same sample size, because there is a huge gap between the error of CTD-S and that of tensor-CUR under the same sample size.

Figure 4 shows the error, running time, and memory usage of CTD-S compared to those of tensor-CUR over increasing sample sizes  $s$  for the Huggle dataset, which are the answers for Q2. The error of CTD-S decreases as  $s$  increases because it gets more chance to sample important fibers which describe the original tensor well. The running time and memory usage of CTD-S are relatively constant compared to those of tensor-CUR. This is because CTD-S keeps only the linearly independent fibers, the number of which is bound to the rank of  $\mathbf{X}_{(\alpha)}$ . There are small fluctuations in the graphs since the sampling process of both CTD-S and tensor-CUR are based on randomness.

## 4.3 Performance of CTD-D

We compare the performance of CTD-D with those of CTD-S to answer Q3. As a result, CTD-D is 2~3× faster for the same level of error compared to CTD-S.

To simulate a dynamic environment, we divide a given dataset into two parts along the time mode. We use the first 80% of the

dataset as historical data and the later 20% as incoming data. We assume that historical data is already given and incoming data arrives sequentially at every time step, such that the whole data grows along the time mode. We measure the performance of CTD-D and CTD-S at each time step and calculate the average. We set the sample size  $d$  of CTD-D to be much smaller than that of CTD-S because CTD-D samples fibers only from the increment  $\Delta\mathcal{X}$  while CTD-S samples from the whole data  $\mathcal{X}$ . We set  $d$  of CTD-D to be 0.01 times of  $s$  of CTD-S,  $\alpha = 1$ , and  $\epsilon = 10^{-6}$ .

Figure 5 shows the error vs. running time and error vs. memory usage relation of CTD-D compared to those of CTD-S. Note that CTD-D is much faster than CTD-S though CTD-D uses the same or slightly more memory than CTD-S does. This is because multiplication between sparse matrices used in updating  $\mathcal{C}$  does not always produce sparse output, thus the number of nonzero entries in  $\mathcal{C}$  increases slightly over time steps.

## 5 CTD AT WORK

In this section, we use CTD-D for mining real-world network traffic tensor data. Our goal is to detect DDoS attacks in network traffic data efficiently in an online fashion; detecting DDoS attacks is a crucial task in network forensic. We propose a novel online DDoS attack detection method based on CTD-D's interpretability. We show that CTD-D is one of the feasible options for online DDoS attack detection and show how it detects DDoS attacks successfully. In contrast to the standard PARAFAC and Tucker decomposition methods, CTD-D can determine DDoS attacks from its decomposition result without expensive overhead. We aim to dynamically find a victim (destination host) and corresponding attackers (source hosts) of each DDoS attack in network traffic data. The victim receives a huge amount of traffic from a large number of attackers.

The online DDoS attack detection method based by CTD-D is as follows. First, we apply CTD-D on network traffic data which is a 3-mode tensor in the form of (source IP - destination IP - time). We assume an online environment where each slab of the network traffic data in the form of (source IP - destination IP) arrives sequentially at every time step. We use source IP mode as mode  $\alpha$ . Second, we inspect the factor  $\mathbf{R}$  of CTD-D, which consists of actual mode- $\alpha$  fibers from the original data.  $\mathbf{R}$  is composed of important mode- $\alpha$  fibers which signify major activities such as DDoS attack or heavy traffic to the main server. Thanks to CTD, we directly find out destination host and occurrence time of a major activity represented in a fiber in  $\mathbf{R}$ , by simply tracking the indices of fibers. We regard fibers with the same destination host index represent the same major activity, and consider the first fiber among those with the same destination host index to be the representative of each major activity. Then, we select fibers with the norm higher than the average among the first fibers and suggest them as candidates of DDoS attack. This is because DDoS attacks have much higher norms than normal traffic does.

We generate network traffic data by injecting DDoS attacks on the real-world CAIDA network traffic dataset. We assume that randomly selected 20% of source hosts participate in each DDoS attack. Table 4 shows the result of DDoS attack detection method of CTD-D. CTD-D achieves high F1 score for various number  $n$



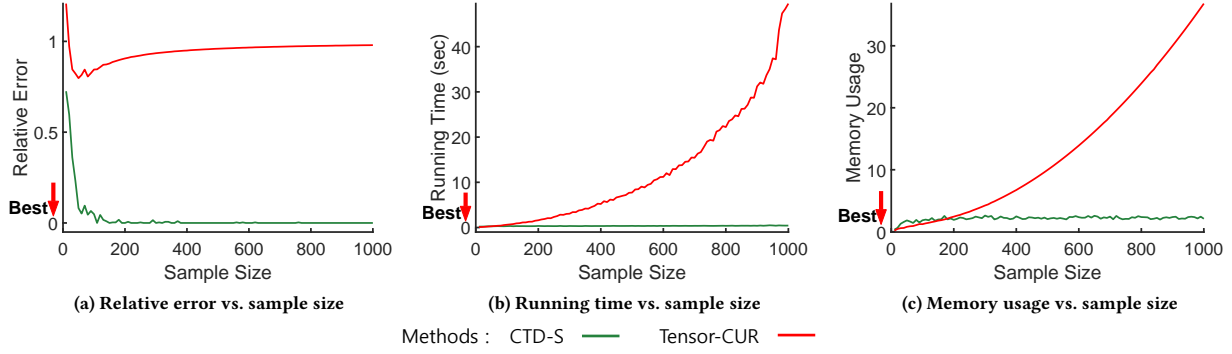


Figure 4: Error, running time, and memory usage of CTD-S compared to those of tensor-CUR over sample size  $s$  for Huggle dataset. CTD-S is more accurate over various sample sizes, and its running time and memory usage are relatively constant compared to the tensor-CUR.

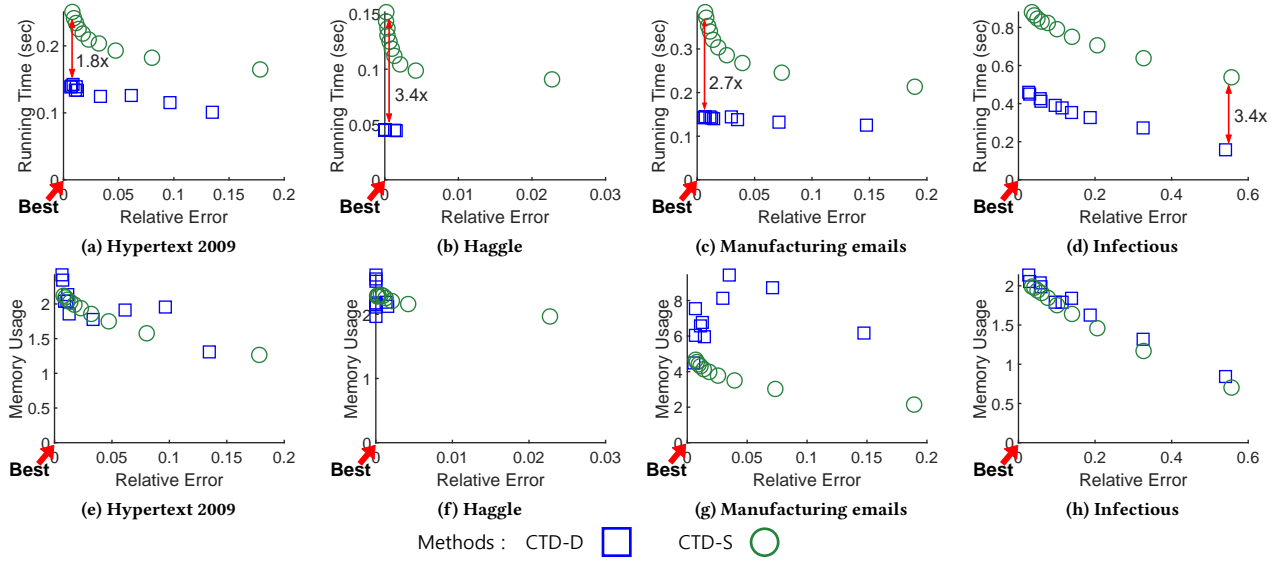


Figure 5: Error, running time, and memory usage relation of CTD-D compared to those of CTD-S. CTD-D is faster and has smaller error while using the same or slightly larger memory space compared to CTD-S.

Table 4: The result of online DDoS attack detection method based on CTD-D. CTD-D achieves high F1 score for various  $n$  with notable precision, where  $n$  denotes the number of injected DDoS attacks.

$n$	Recall	Precision	F1 score
1	1.000	1.000	1.000
3	1.000	1.000	1.000
5	0.880	1.000	0.931
7	0.857	1.000	0.921

of injected DDoS attacks with notable precision. We set  $d = 10$ ,  $\epsilon = 0.15$ .

## 6 CONCLUSION

We propose CTD, a fast, accurate, and directly interpretable tensor decomposition method based on sampling. The static version CTD-S is 17~83 $\times$  more accurate, 5~86 $\times$  faster, and 7~12 $\times$  more memory-efficient compared to tensor-CUR, the state-of-the-art method. The dynamic version CTD-D is 2~3 $\times$  faster than CTD-S for an online environment. CTD-D is the first method providing interpretable *dynamic* tensor decomposition. We show the effectiveness of CTD for online DDoS attack detection.

## REFERENCES

- [1] J. Allanach, H. Tu, S. Singh, P. Willett, and K. Pattipati. 2004. Detecting, tracking, and counteracting terrorist networks via hidden Markov models. IEEE Aerospace.



- [2] A. Arulselvan, C. W. Commander, L. Eleftheriadou, and P. M. Pardalos. 2009. Detecting Critical Nodes in Sparse Graphs. *Comput. Oper. Res.* (July 2009), 8.
- [3] C. F. Caiafa and A. Cichocki. 2010. Generalizing the column-row matrix decomposition to multi-way arrays. *Linear Algebra Appl.* (2010).
- [4] Q. Cao, M. Sirivianos, X. Yang, and T. Pregueiro. 2012. Aiding the Detection of Fake Accounts in Large Scale Social Online Services. NSDI, 15–15.
- [5] B. Cyganek and M. Woźniak. 2015. Tensor based representation and analysis of the electronic healthcare record data. BIBM.
- [6] P. Drineas, R. Kannan, and M. Mahoney. 2006. Fast Monte Carlo algorithms for matrices III: Computing a compressed approximate matrix decomposition. *SIAM J. Comput.* (2006).
- [7] P. Drineas and M. W. Mahoney. 2007. A randomized algorithm for a tensor-based generalization of the singular value decomposition. *Linear algebra and its applications* (2007).
- [8] P. Drineas, M. W. Mahoney, and S. Muthukrishnan. 2008. Relative-error CUR matrix decompositions. *SIAM J. Matrix Anal. Appl.* (2008).
- [9] R. A. Harshman. 1970. Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-modal factor analysis. (1970).
- [10] N. L. D. Khoa, B. Zhang, Y. Wang, W. Liu, F. Chen, S. Mustapha, and P. Runcie. 2015. On Damage Identification in Civil Structures Using Tensor Analysis. (2015).
- [11] G. Kontaxis, I. Polakis, S. Ioannidis, and E. P. Markatos. 2011. Detecting social network profile cloning. PERCOM.
- [12] M. W. Mahoney, M. Maggioni, and P. Drineas. 2008. Tensor-CUR decompositions for tensor-based data. *SIAM J. Matrix Anal. Appl.* (2008).
- [13] I. Perros, R. Chen, R. Vuduc, and J. Sun. 2015. Sparse Hierarchical Tucker Factorization and Its Application to Healthcare. ICDM.
- [14] C. Phua, V. C. S. Lee, K. Smith-Miles, and R. W. Gayler. 2010. A Comprehensive Survey of Data Mining-based Fraud Detection Research. *CoRR*.
- [15] M. A. Prada, J. Toivola, J. Kullaa, and J. Hollmén. 2012. Three-way Analysis of Structural Health Monitoring Data. *Neurocomput.* (2012).
- [16] J. Sun, S. Papadimitriou, and P. S. Yu. 2006. Window-based Tensor Analysis on High-dimensional and Multi-aspect Streams. ICDM.
- [17] J. Sun, D. Tao, and C. Faloutsos. 2006. Beyond streams and graphs: dynamic tensor analysis. KDD.
- [18] J. Sun, Y. Xie, H. Zhang, and C. Faloutsos. 2007. Less is More: Compact Matrix Decomposition for Large Sparse Graphs. SDM.
- [19] B. M. Thuraisingham. 2006. Data Mining for Security Applications. WISI.
- [20] H. Tong, S. Papadimitriou, J. Sun, P. S. Yu, and C. Faloutsos. 2008. Colibri: fast mining of large static and dynamic graphs. KDD.
- [21] L. R. Tucker. 1966. Some mathematical notes on three-mode factor analysis. *Psychometrika* (1966).
- [22] Y. Wang, R. Chen, J. Ghosh, J. C. Denny, A. Kho, Y. Chen, B. A. Malin, and J. Sun. 2015. Rubik: Knowledge Guided Tensor Factorization and Completion for Health Data Analytics. KDD.
- [23] S. Zhou, X. V. Nguyen, J. Bailey, Y. Jia, and I. Davidson. 2016. Accelerating Online CP Decompositions for Higher Order Tensors. KDD.