

# TeX-Graph: Coupled tensor-matrix knowledge-graph embedding for COVID-19 drug repurposing

Charilaos I. Kanatsoulis \*

Nicholas D. Sidiropoulos †

## Abstract

Knowledge graphs (KGs) are powerful tools that codify relational behaviour between entities in knowledge bases. KGs can simultaneously model many different types of subject-predicate-object and higher-order relations. As such, they offer a flexible modeling framework that has been applied to many areas, including biology and pharmacology – most recently, in the fight against COVID-19. The flexibility of KG modeling is both a blessing and a challenge from the learning point of view. In this paper we propose a novel coupled tensor-matrix framework for KG embedding. We leverage tensor factorization tools to learn concise representations of entities and relations in knowledge bases and employ these representations to perform drug repurposing for COVID-19. Our proposed framework is principled, elegant, and achieves 100% improvement over the best baseline in the COVID-19 drug repurposing task using a recently developed biological KG.

**Keywords**— knowledge graphs, tensor, drug repurposing, COVID-19, embedding, network

## 1 Introduction

How does COVID-19 relate to better-studied viral infections and biological mechanisms? Can we use existing drugs to effectively treat COVID-19 symptoms? Since the COVID-19 pandemic has disrupted our lives, there is a pressing need to answer such questions, and COVID-19 research has swiftly risen to the top of the scientific agenda, worldwide. While these questions will ultimately be answered by medical experts, data-driven methods can help to cut-down the immense search space, thus helping to accelerate progress and optimize the allocation of precious research resources. In this paper, our goal is to derive such a method by using network science and multi-view analysis tools.

Networks are powerful abstractions that model interactions between the entities of a system [3]. Networks and network science offer concise and informative modeling, analysis and processing for various biological, engineering and social systems, to name a few [11, 21]. Networks are usually rep-

resented by graphs, that are defined by a set of nodes and a set of edges connecting pairs of nodes. The entities of a system are usually represented by the nodes of the graph, and the interactions by the edges.

A knowledge graph (KG) models the relational behavior of various entities in knowledge bases. A KG is heterogeneous in the sense that it models interactions between entities of different type, e.g., drugs and diseases, and is also a multidimensional network (edge-labeled multi-graph) [4], since the edges (interactions) that connect the nodes (entities) can be multiple and also of different type. Knowledge graphs (KGs) have recently attracted significant attention due to their applicability to various science and engineering tasks. For instance, popular knowledge graphs are YAGO [31], DBpedia [1], NELL [8], Freebase [5], and the Google KG [28]. A recent trend codifies knowledge bases of biomedical components and processes, such as genes, diseases and drugs into KGs e.g., [14, 15, 17]. KGs can model any relations of the form subject-predicate-object, as well as higher-order generalizations. However, this broad modeling freedom can sometimes be a challenge, as the entities can be very diverse and the dimensions of the KG can turn prohibitively large.

A common way to exploit KGs for data mining and machine learning applications is via *knowledge graph embedding*. KG embedding aims to extract meaningful low dimensional representations of the entities and the relations present in a KG. A plethora of methods has been developed to perform KG embedding [2, 6, 12, 18–20, 22, 24, 25, 29, 32, 33]. The most popular among them adopt a single-layer perceptron or neural network approach e.g., [6, 20, 29, 32, 33]. Various tensor factorization models have also been proposed, e.g., [2, 12, 19, 22, 24]. Matrix factorization is also a tool that has been utilized for KG embedding, e.g., [18, 25].

In this paper we propose **TeX-Graph**, a novel coupled tensor-matrix framework to perform KG embedding. The proposed KG coupled tensor-matrix modeling extracts meaningful information from a set of diverse entities with multi-modal interactions in a principled and concise manner. **TeX-Graph** avoids modeling inefficiencies in previously proposed tensor models, and relative to neural network approaches it offers a principled and effective way to produce unique KG representations. The proposed framework is used for drug repurposing, a pivotal tool in the fight against COVID-19 and other diseases. Learning concise representations for drug compounds, diseases, and the relations between them, allows for link prediction between drug com-

\*Electrical and Computer Engineering Department, University of Minnesota. Email: kanat003@umn.edu, † Electrical and Computer Engineering Department, University of Virginia. Email: nikos@virginia.edu

pounds and COVID-19 or other diseases. The impact is critical. First, compound repurposing enables drug design that drastically reduces the design exploration cycle and the failure rate. Second, it markedly reduces drug development cost, as developing new therapeutic drugs is tremendously expensive.

The contributions of our work can be summarized as follows:

- **Novel KG modeling:** We propose a principled coupled tensor-matrix model tailored to KG needs for efficient and parsimonious representations.
- **Analysis:** The **TeX-Graph** embeddings are unique and permutation invariant, a property which is important for consistency and necessary for interpretability.
- **Algorithm:** We design a scalable algorithmic framework with lightweight updates, that can effectively handle very large KGs.
- **Application:** The proposed framework is developed to perform drug repurposing, a pivotal task in the fight against COVID-19.
- **Performance:** **TeX-Graph** achieves 100% performance improvement compared to the best available baseline for COVID-19 drug repurposing using a recently developed COVID-19 KG.

**Reproducibility:** The DRKG dataset used in the experiments is publicly available; The **TeX-Graph** code<sup>1</sup> is also publicly available.

**Notation:** Our notation is summarized in Table 1.

## 2 Preliminaries

In order to facilitate the upcoming discussion we now discuss some tensor algebra preliminaries. For more background on tensors the reader is referred to [27].

A third-order tensor  $\underline{\mathbf{X}} \in \mathbb{R}^{I \times J \times K}$  is a three-way array indexed by  $i, j, k$  with elements  $\underline{\mathbf{X}}(i, j, k)$ . It has three mode types – columns  $\underline{\mathbf{X}}(i, :, k)$  ( $:$  is used to denote all relevant index values for that mode, i.e., from beginning to end), rows  $\underline{\mathbf{X}}(:, j, k)$ , and fibers  $\underline{\mathbf{X}}(i, j, :)$  – see Fig. 1. A third order tensor can also be described by a set of matrices (slabs), i.e., horizontal  $\underline{\mathbf{X}}(i, :, :)$ , vertical  $\underline{\mathbf{X}}(:, j, :)$  and frontal slabs  $\underline{\mathbf{X}}(:, :, k)$  – see Fig. 2. A rank-one third order tensor  $\underline{\mathbf{Z}} \in \mathbb{R}^{I \times J \times K}$  is defined as the outer product of three vectors. Recall that a rank one matrix is the outer product of two vectors. Any third order tensor can be decomposed into a sum of three-way outer products (rank one tensors) as:

$$(2.1) \quad \underline{\mathbf{X}} = \sum_{f=1}^F \mathbf{A}(:, f) \circ \mathbf{B}(:, f) \circ \mathbf{C}(:, f),$$

where  $\mathbf{A} \in \mathbb{R}^{I \times F}$ ,  $\mathbf{B} \in \mathbb{R}^{J \times F}$ ,  $\mathbf{C} \in \mathbb{R}^{K \times F}$  are matrices collecting the respective mode factors, i.e., hold in their columns the vectors involved in the  $F$  three-way outer products. The above expression is known as the polyadic

Table 1: Overview of notation.

|  |              |   |
|--|--------------|---|
| $x, y, z$  | $\triangleq$ | scalars   |
| $(m, n), (h, r, t)$  | $\triangleq$ | ordered tuple   |
| $\mathbf{x}, \mathbf{y}, \mathbf{z}$                                     | $\triangleq$ | vectors   |
| $\mathbf{A}, \mathbf{B}, \mathbf{C}$                                     | $\triangleq$ | matrices  |
| $\underline{\mathbf{X}}, \underline{\mathbf{Y}}, \underline{\mathbf{Z}}$ | $\triangleq$ | tensors   |
| $\mathcal{S}, \mathcal{S}_n^+, \mathcal{S}_n^-$                          | $\triangleq$ | sets  |
| $\mathbf{A}(:, f)$   | $\triangleq$ | $f$ -th column of matrix $\mathbf{A}$                   |
| $\mathbf{A}(i, :)$   | $\triangleq$ | $i$ -th row of matrix $\mathbf{A}$                      |
| $\underline{\mathbf{X}}^k$   | $\triangleq$ | $k$ -th frontal slab of tensor $\underline{\mathbf{X}}$ |
| $\mathbf{A}^T$   | $\triangleq$ | transpose of matrix $\mathbf{A}$                        |
| $\ \mathbf{A}\ _F$   | $\triangleq$ | Frobenius norm of matrix $\mathbf{A}$                   |
| $\odot$  | $\triangleq$ | Khatri-Rao (columnwise Kronecker) product               |
| $\circ$  | $\triangleq$ | outer product   |
| $*$  | $\triangleq$ | Hadamard product  |
| $\text{diag}(\mathbf{x})$  | $\triangleq$ | diagonal matrix of vector $\mathbf{x}$                  |
| $\lfloor x \rfloor$  | $\triangleq$ | largest integer that is less than or equal to $x$       |
| $\text{nnz}$   | $\triangleq$ | number of non-zeros                                     |

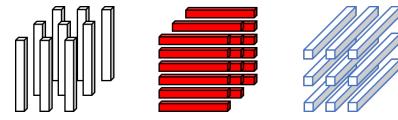


Figure 1: The columns, rows and fibers of a third-order tensor.

decomposition (PD) of a third-order tensor. If  $F$  is the minimum number of outer products required to synthesize  $\underline{\mathbf{X}}$ , then  $F$  is the *rank* of tensor  $\underline{\mathbf{X}}$  and the decomposition is known as the *canonical polyadic decomposition* (CPD) or *parallel factor analysis* (PARAFAC) [13]. For the rest of the paper we use the notation  $\underline{\mathbf{X}} = [\mathbf{A}, \mathbf{B}, \mathbf{C}]$  to denote the CPD of the tensor.

A striking property of the CPD is that it is essentially unique under mild conditions, even if the rank is higher than the dimensions – see [9] for a generic result.

A tensor can be represented in a matrix form by employing the *matricization* operation. There are three common ways to matricize (or unfold) a third-order tensor,

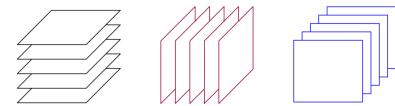


Figure 2: The horizontal, vertical, and frontal slabs of a third-order tensor.

<sup>1</sup>[https://github.com/marhar19/TeX-Graph\\_KG\\_embedding](https://github.com/marhar19/TeX-Graph_KG_embedding)

by stacking columns, rows, or fibers of the tensor to form a matrix. To be more precise let:

$$(2.2) \quad \underline{\mathbf{X}}(:,:,k) = \mathbf{X}^k \in \mathbb{R}^{I \times J},$$

where  $\mathbf{X}^k$  are the frontal slabs of tensor  $\underline{\mathbf{X}}$  and in the context of this paper they model adjacency matrices. Then the mode-1, mode-2 and mode-3 unfoldings of  $\underline{\mathbf{X}}$  are:

$$(2.3) \quad \mathbf{X}^{(1)} = [\mathbf{X}^1, \dots, \mathbf{X}^K]^T = (\mathbf{C} \odot \mathbf{B}) \mathbf{A}^T \in \mathbb{R}^{JK \times I},$$

$$(2.4) \quad \mathbf{X}^{(2)} = [\mathbf{X}^{1^T}, \dots, \mathbf{X}^{K^T}] = (\mathbf{C} \odot \mathbf{A}) \mathbf{B}^T \in \mathbb{R}^{IK \times J},$$

$$(2.5) \quad \mathbf{X}^{(3)} = \begin{bmatrix} \mathbf{X}^1(:,1), \dots, \mathbf{X}^K(:,1) \\ \vdots \\ \mathbf{X}^1(:,J), \dots, \mathbf{X}^K(:,J) \end{bmatrix} = (\mathbf{B} \odot \mathbf{A}) \mathbf{C}^T \in \mathbb{R}^{IJ \times K},$$

Another important tensor model is the coupled CPD. In coupled CPD we are interested in decomposing an array of tensors that share at least one common latent factor. In particular, consider a collection of  $N$  tensors  $\underline{\mathbf{X}}_n \in \mathbb{R}^{I \times J_n \times K_n}$ ,  $n \in \{1, \dots, N\}$ . The rank- $F$  coupled CPD of  $\{\underline{\mathbf{X}}_n\}$  can be expressed as:

$$(2.6) \quad \underline{\mathbf{X}}_n = [\mathbf{A}, \mathbf{B}_n, \mathbf{C}_n], \quad n \in \{1, \dots, N\},$$

where  $\mathbf{A} \in \mathbb{R}^{I \times F}$  is the common factor and  $\mathbf{B}_n \in \mathbb{R}^{J_n \times F}$ ,  $\mathbf{C}_n \in \mathbb{R}^{K_n \times F}$  are unshared factors. The coupled CPD is also unique under certain conditions, even if individual CPDs of  $\underline{\mathbf{X}}_n$  are not unique. In this work we will use the following uniqueness theorem for coupled CPD:

**THEOREM 1.** [30, p. 510] Consider the coupled CPD as expressed in (2.6). The decomposition is essentially unique if:

1.  $\mathbf{A}$  has full column rank,
2.  $\mathbf{G}$  has full column rank,

where  $\mathbf{G}$  is defined as:

$$(2.7) \quad \mathbf{G} = \begin{bmatrix} \mathbf{C}_2(\mathbf{C}_1) \odot \mathbf{C}_2(\mathbf{B}_1) \\ \vdots \\ \mathbf{C}_2(\mathbf{C}_N) \odot \mathbf{C}_2(\mathbf{B}_N) \end{bmatrix} \in \mathbb{R}^{\frac{1}{4} \sum_{n=1}^N J_n(K_n-1)K_n \times \frac{1}{2} F(F-1)},$$

and  $\mathbf{C}_2(\mathbf{C}_n)$  is the second compound matrix of  $\mathbf{C}_n$  – see [10, p. 861-862]. In the context of coupled CPD, essential uniqueness corresponds to  $\mathbf{A}$  being unique and  $\{\mathbf{B}_n, \mathbf{C}_n\}$  being identifiable up to column scaling and counter-scaling.

### 3 Problem Statement

As mentioned in the introduction knowledge graphs (KGs) have attracted significant attention over the past decade due to their tremendous modeling capabilities. In particular, KGs model triplets of subject-predicate-object, denoted as

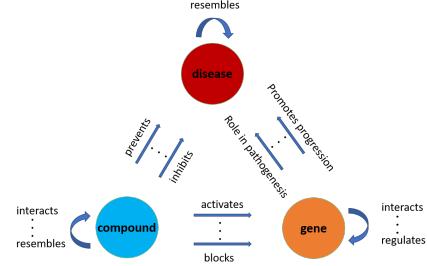


Figure 3: Schematic representation of biological KG.

(head, relation, tail) or (h, r, t). Subjects (heads) and objects (tails) are entities that are represented as graph nodes and predicates (relations) define the type of edge according to which the subject is connected to the object. A schematic representation of a KG, which models relations between genes, compounds and diseases is presented in Fig. 3.

In this paper, we focus our attention on a biological KG that models relational triplets between biological entities. For example, (compound 1, interacts with, compound 2), (compound 1, activates, gene 1), (gene 1, regulates, gene 2), (compound 1, prevents, disease 1), (gene 2, is linked with, disease 2) are common triplets in numerous recently developed knowledge bases [14, 15, 17]. Modeling these types of relations as a KG enables embedding entities and relations in a Euclidean space which can further facilitate any type of processing and analysis. For instance, obtaining a low dimensional representation of compounds, diseases and the ‘prevents’ relation allows measuring similarity, and thus predicting and testing hypotheses regarding (compound, prevents, disease) interactions. Drug repurposing can be performed by predicting candidate compounds for new and existing target diseases. Note that the proposed framework to be introduced shortly is not limited to biological KGs – it can be applied to a wide variety of interesting KGs.

**3.1 Prior Art** Several methods have been proposed to learn low dimensional representations of KGs. To properly describe them we need to define the score function  $f(\cdot)$  and the loss function  $\mathcal{L}(\cdot)$ .

Let  $(h_n, r_n, t_n)$  be an available triplet and  $\mathbf{h}_n \in \mathbb{R}^F$ ,  $\mathbf{t}_n \in \mathbb{R}^F$  and  $\mathbf{r}_n \in \mathbb{R}^d$  be the low dimensional embeddings we aim to learn. Note that entity and relation embeddings need not be of the same dimension. The score function determines the relation model between the head (subject) and the tail (object). In simple words, high values of the score function  $f(\mathbf{h}_n, \mathbf{r}_n, \mathbf{t}_n)$  are desirable for existing triplets  $(h_n, r_n, t_n)$  and low values of  $f(\mathbf{h}_n, \mathbf{r}_n, \mathbf{t}_n)$  for non-existing ones.

In order to produce the entity and relational embeddings we define the following forward model for each triplet  $(h_n, r_n, t_n)$ :

$$(3.8a) \quad \mathbf{h}_n = \gamma \left( \mathbf{W}_e^T \mathbf{o}_n^h \right) \in \mathbb{R}^F,$$

Table 2: Knowledge Graph models

| Model         | score function $f(\mathbf{h}, \mathbf{r}, \mathbf{t})$   |
|---------------|--|
| TransE [6]    | $1 - \ \mathbf{h} + \mathbf{r} - \mathbf{t}\ _2$ or $1 - \ \mathbf{h} + \mathbf{r} - \mathbf{t}\ _1$ |
| TransR [20]   | $1 - \ \mathbf{M}_r \mathbf{h} + \mathbf{r} - \mathbf{M}_t \mathbf{t}\ _2$                           |
| DistMult [33] | $\mathbf{h}^T \text{diag}(\mathbf{r}) \mathbf{t}$  |
| RESCAL [22]   | $\mathbf{h}^T \mathbf{R} \mathbf{t}$   |
| RotatE [32]   | $1 - \ \mathbf{h} * \mathbf{r} - \mathbf{t}\ $   |

$$(3.8b) \quad \mathbf{t}_n = \gamma \left( \mathbf{W}_e^T \mathbf{o}_n^t \right) \in \mathbb{R}^F,$$

$$(3.8c) \quad \mathbf{r}_n = \delta \left( \mathbf{W}_r^T \mathbf{o}_n^r \right) \in \mathbb{R}^d,$$

where  $\mathbf{o}_n^h \in \{0, 1\}^{L_e}$ ,  $\mathbf{o}_n^t \in \{0, 1\}^{L_e}$ ,  $\mathbf{o}_n^r \in \{0, 1\}^{K_r}$  are one-hot input vectors corresponding to the head, tail and relation index of the triplet  $(h_n, r_n, t_n)$  respectively, with  $L_e$ ,  $K_r$  being the total number of entities (nodes) and types of relations;  $\gamma(\cdot)$  and  $\delta(\cdot)$  are element-wise functions and  $\mathbf{W}_e \in \mathbb{R}^{L_e \times F}$ ,  $\mathbf{W}_r \in \mathbb{R}^{K_r \times d}$  are matrices that contain the model parameters to be learned.

Popular choices for  $\gamma(\cdot)$  and  $\delta(\cdot)$  are the identity function and hyperbolic tangent. If  $\gamma(\cdot)$  or  $\delta(\cdot)$  are identity functions then the rows of  $\mathbf{W}_e$  or  $\mathbf{W}_r$  are the learned embeddings for entities and relations respectively. For TransE, DistMult and RotatE  $F = d$ , whereas for TransR and RESCAL  $d \neq F$ . In the TransR model  $\mathbf{M}_r \in \mathbb{R}^{d \times F}$  is a projection matrix associated with relation  $r$  and in RESCAL  $\mathbf{R} \in \mathbb{R}^{F \times F}$ .

In order to learn the embeddings, state-of-the-art methods attempt to minimize the following risk:

$$(3.9) \quad \min_{\mathbf{W}_e, \mathbf{W}_r} \frac{1}{N} \sum_{n=1}^N \mathcal{L}(y_n - f(\mathbf{h}_n, \mathbf{r}_n, \mathbf{t}_n))$$

where  $N$  is the number of data points (triplets or non-triplets),  $y_n = 1$  if the triplet  $(h_n, r_n, t_n)$  exists, else  $y_n = 0$ . Typical loss functions include the logistic loss, square loss, pairwise ranking loss, margin-based ranking loss and variants of them. In order to tackle the problem in (3.9) the most popular approach is stochastic gradient descent (SGD) or batch SGD [7].

**3.2 The 3-way model** Modeling a KG using a third order tensor has been considered in [2, 12, 19, 22, 24]. In these works, the first and second mode of the tensor is the concatenation of all the available entities, regardless of their type, whereas the third mode represents the different type of relations – i.e., each frontal slab of the third order tensor represents a certain interaction type between the entities of the KG. The methods in [2, 24] work with incomplete tensors, whereas [12, 19, 22] model each frontal slab as an adjacency matrix. To be more precise, let  $\underline{\mathbf{Z}} \in \{0, 1\}^{L_e \times L_e \times K_r}$  be the third order tensor in [12, 19, 22]. Then  $\underline{\mathbf{Z}}(i, j, k) = 1$  if entity  $i$  interacts with entity  $j$  through relation  $k$  and  $\underline{\mathbf{Z}}(i, j, k) = 0$  if there is no interaction between entities  $i$  and  $j$  via the  $k$  relation.

An important observation is that although the first and second mode of tensor  $\underline{\mathbf{Z}}$  represent the same entities, each frontal slab  $\mathbf{Z}^k$  is not necessarily symmetric. The reason is that subject-predicate-object does not necessarily imply object-predicate-subject. The works in [12, 19] compute the CPD of  $\underline{\mathbf{Z}}$  (or scaled versions of  $\underline{\mathbf{Z}}$ ) and produce two embeddings for each entity, one as a subject and another as an object. Although this is not always a drawback, it can result in an overparametrized model because in many applications entities usually act *either* as a subject *or* as an object, but not both. Furthermore, a single unified representation is usually preferable. In order to overcome this issue, RESCAL [22] proposed the following model for each frontal slab:

$$(3.10) \quad \mathbf{Z}^k = \mathbf{A} \mathbf{R}^k \mathbf{A}^T, \quad k = 1, \dots, K_r,$$

where  $\mathbf{R}^k \in \mathbb{R}^{F \times F}$  is square matrix holding the relation embeddings associated with relation  $k$ . Note that the RESCAL model is different than the traditional CPD (symmetric in mode 1 and 2) in the sense that  $\mathbf{R}^k$  is not constrained to be diagonal. Relaxing the diagonal constraints allows matrix  $\mathbf{R}^k$  to absorb in the relation embedding the direction in which different entities interact. On the downside, this type of relaxation forfeits the parsimony and uniqueness properties of the CPD. This is an important point, since uniqueness is a prerequisite for model interpretability when we are interested in exploratory / explanatory analysis (and not simply in making ‘black box’ predictions).

Another important drawback of the tree-way model is that it models unnecessary interactions. To see this, consider a KG that describes interactions between genes and diseases. Suppose that the observed interactions are of gene-gene and gene-disease type but there are no available data for disease-disease interactions. The tree-way model involves disease-disease interactions in the learning process (as non-edges), even though there are no data to justify it. As we will see in the upcoming section our proposed coupled tensor-matrix modeling addresses all the aforementioned challenges.

#### 4 The TeX-Graph model

In this paper we leverage coupled tensor-matrix factorization to extract low dimensional representations of entities (head, tail) as well as representations for the interactions (relation). KGs can be naturally represented by a collection of tensors and matrices, as shown in Fig. 4. To see this, consider the previous example of gene, compound and disease entities. Gene-compound interactions, of a certain type, can be represented by an adjacency matrix. Since there are multiple types of interactions, multiple adjacency matrices are necessary to model every interaction, resulting in a tensor  $\underline{\mathbf{X}}_{g,c} \in \{0, 1\}^{L_g \times L_c \times K_{g,c}}$ , where  $L_g$ ,  $L_c$  are the number of genes and compounds respectively, and  $K_{g,c}$  is the number of different interactions between genes and compounds. The same idea can be applied to any (entity, interaction, entity) triplet.

To facilitate the discussion let  $\underline{\mathbf{X}}_{m,n} \in \{0, 1\}^{L_m \times L_n \times K_{m,n}}$  be the tensor of interactions be-

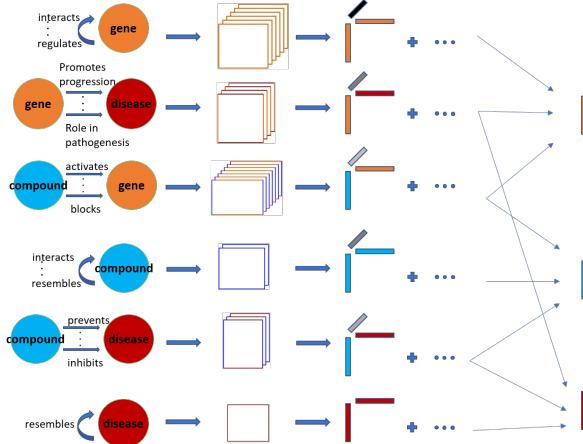


Figure 4: Schematic representation of TeX-Graph model.

tween entity of type- $m$  and type- $n$ , e.g.,  $m$  codifies genes and  $n$  codifies compounds. Also let  $L_T$  be the total number of different entity types, then  $m, n \in \{1, \dots, L_T\}$ .  $\underline{\mathbf{X}}_{m,n}(i, j, k) = 1$  if the  $i$ -th entity of type- $m$  interacts with the  $j$ -th entity of type- $n$  via relation  $k$  and  $\underline{\mathbf{X}}_{m,n}(i, j, k) = 0$  if there is no type- $k$  interaction between the  $i$ -th entity of type- $m$  and the  $j$ -th entity of type- $n$ . The KG is represented by a collection of tensors as:

$$(4.11) \quad \underline{\mathbf{X}}_{m,n} \in \{0, 1\}^{L_m \times L_n \times K_{m,n}}, \quad (m, n) \in \mathcal{S}$$

$$\mathcal{S} = \{(m, n) : m \leq n, \exists (h, r, t) \text{ with } (h, t) \in \text{type } (m, n) \text{ or } (n, m)\},$$

where  $\sum_{n=1}^{L_T} L_n = L_e$  and  $\sum_{(m,n) \in \mathcal{S}} K_{m,n} = K_r$ . Note that tensors  $\underline{\mathbf{X}}_{m,n}$  and  $\underline{\mathbf{X}}_{n,m}$  contain the same information since  $\underline{\mathbf{X}}_{m,n}^k = \underline{\mathbf{X}}_{n,m}^{k^T}$ . Therefore we only consider  $(m, n)$  tuples where  $m \leq n$ .

Each of the tensors in the array  $\{\underline{\mathbf{X}}_{m,n}, (m, n) \in \mathcal{S}\}$  admits a CPD and the overall model is cast as:

$$(4.12) \quad \underline{\mathbf{X}}_{m,n} = [\mathbf{A}_m, \mathbf{A}_n, \mathbf{C}_{m,n}], \quad (m, n) \in \mathcal{S},$$

where  $\mathbf{A}_n \in \mathbb{R}^{L_n \times F}, \mathbf{C}_{m,n} \in \mathbb{R}^{K_{m,n} \times F}$ . The  $i$ -th row of  $\mathbf{A}_n$  represents the  $F$ -dimensional embedding of the  $i$ -th type- $n$  entity and the  $k$ -th row of  $\mathbf{C}_{m,n}$  represents the  $F$ -dimensional embedding of the  $k$ -th type relation between type- $m$  and type- $n$  entities. Note that in the case where entities of type- $m$  interact with entities of type- $n$  via only one type of relation,  $\underline{\mathbf{X}}_{m,n} \in \{0, 1\}^{L_m \times L_n}$  is a matrix and can be factored as:

$$(4.13) \quad \underline{\mathbf{X}}_{m,n} = \mathbf{A}_m \text{diag}(\mathbf{c}_{m,n}) \mathbf{A}_n^T$$

The model in (4.12) is a coupled CPD as the factors  $\mathbf{A}_n$  appear in multiple tensors. For instance, type-1-type-1 interactions (gene-gene), type-1-type-2 interactions (gene-compound), type-1-type-3 interactions (gene-disease), result in the factor  $\mathbf{A}_1$  appearing in tensors

$$\underline{\mathbf{X}}_{1,1} = [\mathbf{A}_1, \mathbf{A}_1, \mathbf{C}_{1,1}], \underline{\mathbf{X}}_{1,2} = [\mathbf{A}_1, \mathbf{A}_2, \mathbf{C}_{1,1}] \text{ and } \underline{\mathbf{X}}_{1,3} = [\mathbf{A}_1, \mathbf{A}_3, \mathbf{C}_{1,3}].$$

The proposed TeX-Graph exhibits several favorable properties. First, the produced embeddings are unique, provided that they appear in more than one adjacency matrices.

**PROPOSITION 1.** (Uniqueness of the embeddings) *If the coupled tensor model in (4.12) is indeed low-rank,  $F$ , there exist entity and relation embedding vectors in  $F$  dimensional space that generate the given knowledge base. Then the  $F$ -dimensional TeX-Graph embeddings for type- $n$  entities and type- $(m, n)$  relations are unique and permutation invariant provided that  $\sum_{m \in \mathcal{S}_n^+} K_{m,n} + \sum_{p \in \mathcal{S}_n^-} K_{n,p} > 1$  and  $K_{m,n} > 1$  respectively, where  $\mathcal{S}_n^+, \mathcal{S}_n^-$  are defined in (4.16).*

The proof of Proposition 1 utilizes the uniqueness results of Theorem 1 and is relegated to the journal version due to space limitations. In the case where  $K_{m,n} = 1$  and type- $m$  entities appear in multiple tensors but type- $n$  entities only in one, the TeX-Graph model identifies  $\mathbf{A}_m$  and  $\mathbf{A}_n \text{diag}(\mathbf{c}_{m,n})$ , since there is rotational freedom between  $\mathbf{A}_n$  and  $\mathbf{c}_{m,n}$ .

Another important property of the proposed TeX-Graph is that it avoids modeling of spurious ‘cross-product’ relations that can never be observed. The coupled tensor-matrix model allows for a concise KG representation that eliminates such spurious relations from the start, contrary to the three-way model. To see this, consider the previous example of gene-disease KG that observes relational triplets between gene-gene and gene-disease type but not for disease-disease type. The proposed TeX-Graph does not model disease-disease interactions, whereas the three-way model treats them as non-edges.

It is worth noticing that TeX-Graph makes the implicit assumption that  $\underline{\mathbf{X}}_{n,n}$  are symmetric in the first and second mode. This is not always the case, since interactions between some entity types might be directed. To overcome this issue we assume that  $(h, r, t)$  implies  $(t, r, h)$  for  $(h, t)$  of the same type. Although this assumption ignores the direction in this type of interactions, it results in a more parsimonious model for the entity embeddings.

**4.1 Algorithmic framework** In order to learn the  $F$ -dimensional embeddings of all entities and relations we formulate the KG embedding problem as:

$$(4.14) \quad \min_{\{\mathbf{A}_m\}, \{\mathbf{C}_{m,n}\}} \sum_{(m,n) \in \mathcal{S}} \|\underline{\mathbf{X}}_{m,n} - [\mathbf{A}_m, \mathbf{A}_n, \mathbf{C}_{m,n}]\|_F^2,$$

The problem in (4.14) is non-convex and NP-hard in general. In order to tackle it we propose to fix all variables but one and update the remaining variable. This procedure is repeated in an alternating fashion. The update for  $\mathbf{A}_n$  is a system of linear equations and takes the form:

$$(4.15) \quad \sum_{m \in \mathcal{S}_n^+} (\mathbf{C}_{m,n} \odot \mathbf{A}_m)^T \underline{\mathbf{X}}_{m,n}^{(2)} + \sum_{p \in \mathcal{S}_n^-} (\mathbf{C}_{n,p} \odot \mathbf{A}_p)^T \underline{\mathbf{X}}_{n,p}^{(1)} = \left( \sum_{m \in \mathcal{S}_n^+} (\mathbf{C}_{m,n}^T \mathbf{C}_{m,n} * \mathbf{A}_m^T \mathbf{A}_m) + \sum_{p \in \mathcal{S}_n^-} (\mathbf{C}_{n,p}^T \mathbf{C}_{n,p} * \mathbf{A}_p^T \mathbf{A}_p) \right) \mathbf{A}_n^T,$$

---

**Algorithm 1** **TeX-Graph**


---

**Input:**  $\{(h_n, r_n, t_n)\}_{n=1}^N$ ,  $\{\mathbf{A}_m\}$ ,  $\{\mathbf{C}_{m,n}\}$ .  
**Output:**  $\{\mathbf{A}_n\}_{n=1}^{L_e}$ ,  $\{\mathbf{C}_{m,n}\}_{(m,n) \in \mathcal{S}}$ .  
Create  $\{\underline{\mathbf{X}}_{m,n}\}_{(m,n) \in \mathcal{S}}$  from  $\{(h_n, r_n, t_n)\}_{n=1}^N$ ;  
**repeat**  
    **for**  $n \in \{1, \dots, L_E\}$  **do**  
         $\mathbf{A}_n \leftarrow$  solve (4.15);  
    **end for**  
    **for**  $(m, n) \in \mathcal{S}$  **do**  
         $\mathbf{C}_{(m,n)} \leftarrow$  solve (4.17)  
    **end for**  
**until** criterion is met.

---

**Algorithm 2** **TeX-Graph-initialization**


---

**Input:**  $\{(h_n, r_n, t_n)\}_{n=1}^N$ .  
**Output:**  $\{\mathbf{A}_n\}_{n=1}^{L_e}$ ,  $\{\mathbf{C}_{m,n}\}_{(m,n) \in \mathcal{S}}$ .  
Create tensor  $\underline{\mathbf{Z}}$  from  $\{(h_n, r_n, t_n)\}_{n=1}^N$  as explained in section 3.2;  
Form  $\underline{\mathbf{Y}}$  as:  $\underline{\mathbf{Y}}(i, j, k) = \min\{1, \underline{\mathbf{Z}}(i, j, k) + \underline{\mathbf{Z}}(j, i, k)\}$ ;  
Solve  $\underline{\mathbf{Y}} = [\mathbf{A}, \mathbf{A}, \mathbf{C}]$  via sparse EVD;  
Form  $\{\mathbf{A}_n\}_{n=1}^{L_e}$ ,  $\{\mathbf{C}_{m,n}\}_{(m,n) \in \mathcal{S}}$  from  $\mathbf{A}$ ,  $\mathbf{C}$ .

---

where

$$(4.16) \quad \mathcal{S}_n^+ = \{m : (m, n) \in \mathcal{S}\}, \quad \mathcal{S}_n^- = \{p : (n, p) \in \mathcal{S}\}$$

The update for  $\mathbf{C}_{m,n}$  is the solution to the following system of linear equations:

$$(4.17) \quad (\mathbf{A}_n \odot \mathbf{A}_m)^T \mathbf{X}_{m,n}^{(3)} = (\mathbf{A}_n^T \mathbf{A}_n * \mathbf{A}_m^T \mathbf{A}_m) \mathbf{C}_{m,n}^T$$

The derivations for these updates as well as implementation details are presented in Appendix A.

The proposed **TeX-Graph** is presented in Algorithm 1. **TeX-Graph** is an iterative algorithm that tackles a non-convex problem and NP-hard in general. As a result different initial points might produce different results. Although we have observed that random initialization is sufficient most of the times we propose an alternative initialization procedure that yields consistent and reproducible results. To be more specific we form a symmetric version of tensor  $\underline{\mathbf{Z}}$  as:

$$(4.18) \quad \underline{\mathbf{Y}}(i, j, k) = \min\{1, \underline{\mathbf{Z}}(i, j, k) + \underline{\mathbf{Z}}(j, i, k)\}$$

Then we compute the semi-symmetric CPD of  $\underline{\mathbf{Y}} = [\mathbf{A}, \mathbf{A}, \mathbf{C}]$  using sparse eigenvalue decomposition (EVD) [26]. The proposed initialization procedure is presented in Algorithm 2.

**4.2 Computational complexity analysis** In terms of memory requirements and computational complexity, the main bottleneck of **TeX-Graph** lies in instantiating and computing the matricized tensor times Khatri-Rao product (MTTKRP) in the left hand side (LHS) of (4.15) and (4.17). The number of flops needed to compute the LHS of (4.15)

and (4.17) is  $\mathcal{O}\left(F \cdot \text{nnz}\left(\sum_{m \in \mathcal{S}_n^+} \underline{\mathbf{X}}_{m,n} + \sum_{p \in \mathcal{S}_n^-} \underline{\mathbf{X}}_{n,p}\right)\right)$  and  $\mathcal{O}\left(F \cdot \text{nnz}(\underline{\mathbf{X}}_{m,n})\right)$  respectively. For small values of  $F$ , which is the case in practice, the complexity is linear in the number of triplets participating in each update. Moreover, the Khatri-Rao products in the LHS of (4.15) and (4.17) are not being instantiated as shown in Appendix A.

## 5 Drug Repurposing for COVID-19

In this section we apply **TeXGraph** to a recently developed KG [17] in order to perform drug repurposing for COVID-19 disease. All algorithms were implemented in Matlab or Python, and executed on a Linux server comprising 32 cores at 2GHz and 128GB RAM.

**5.1 Data** The dataset used in the experiments is the Drug Repurposing Knowledge Graph (DRKG)<sup>2</sup> [17]. It codifies triplets of biological interactions between 97,238 different entities of 13 types, namely, genes, compounds, diseases, anatomy, tax, biological process, cellular component, pathway, molecular function, anatomical therapeutic chemical (Atc), side effect, pharmacological class, and symptom. The total number of triplets is 5,874,258 and there are 107 different types of interactions. The KG is organised in 6 adjacency tensors and 11 adjacency matrices. Detailed description of the dataset and the modeling can be found in Table 3. Each row denotes a different adjacency tensor or matrix and # type- $m$  entities, # type- $m$  entities, # relation types correspond to the dimension of mode 1, mode 2, and mode 3 respectively. The last column (sparsity) denotes the sparsity of each tensor, i.e.,  $\frac{\text{nnz}(\underline{\mathbf{X}}_{m,n})}{L_m L_n K_{m,n}}$ .

**5.2 Procedure** Drug repurposing refers to the task of discovering existing drugs that can effectively manage certain diseases – COVID-19 in our study. DRKG encodes relational triplets of (compound,treats,disease) and (compound,inhibits,disease). Therefore drug repurposing in the context of DRKG boils down to predicting new ‘treats’ and ‘inhibits’ edges (links) between compounds and diseases of interest.

We follow the evaluation procedure proposed in [17]. In the training phase we learn low dimensional representations for the entities and relations, using all the edges in DRKG. In the testing phase, we assign a score to (compound,treats,disease) and (compound,inhibits,disease) triplets according to the scoring function used for training. For the proposed **TeX-Graph**, the scores assigned to the triplet (hyper-edge) (compound  $i$ , treats, disease  $j$ ) and (compound  $i$ , inhibits, disease  $j$ ) are:

$$\text{score}_{i,j,2} = \mathbf{A}_2(i, :) \text{diag}(\mathbf{C}_{2,3}(2, :)) \mathbf{A}_2(j, :)^T,$$

$$\text{score}_{i,j,9} = \mathbf{A}_2(i, :) \text{diag}(\mathbf{C}_{2,3}(9, :)) \mathbf{A}_2(j, :)^T,$$

since ‘treats’ and ‘inhibits’ relations correspond to the second and ninth frontal slab of  $\underline{\mathbf{X}}_{2,3}$ , respectively. The testing set consists of 34 corona-virus related diseases,

<sup>2</sup>[github.com/gnn4dr/DRKG](https://github.com/gnn4dr/DRKG)

Table 3: Coupled tensor-matrix DRKG modeling.

| entity type-m | entity type-n         | # type-m entities 1 | # type-n entities 2 | # relation types | tensor  | sparsity             |
|---------------|-----------------------|---------------------|---------------------|------------------|---|----------------------|
| Gene          | Gene                  | 39,220              | 39,220              | 32               | $\underline{\mathbf{X}}_{1,1} = [\mathbf{A}_1, \mathbf{A}_1, \mathbf{C}_{1,1}]$                 | $6.12 \cdot 10^{-5}$ |
|               | Compound              | 39,220              | 24,313              | 34               | $\underline{\mathbf{X}}_{1,2} = [\mathbf{A}_1, \mathbf{A}_2, \mathbf{C}_{1,2}]$                 | $6.50 \cdot 10^{-6}$ |
|               | Disease               | 39,220              | 5,103               | 15               | $\underline{\mathbf{X}}_{1,3} = [\mathbf{A}_1, \mathbf{A}_3, \mathbf{C}_{1,3}]$                 | $4.13 \cdot 10^{-5}$ |
|               | Anatomy               | 39,220              | 400                 | 3                | $\underline{\mathbf{X}}_{1,4} = [\mathbf{A}_1, \mathbf{A}_4, \mathbf{C}_{1,4}]$                 | 0.0154               |
|               | Tax                   | 39,220              | 215                 | 1                | $\underline{\mathbf{X}}_{1,5} = \mathbf{A}_1 \text{diag}(\mathbf{c}_{1,5}) \mathbf{A}_5^T$      | 0.0017               |
|               | Biological Process    | 39,220              | 11,381              | 1                | $\underline{\mathbf{X}}_{1,6} = \mathbf{A}_1 \text{diag}(\mathbf{c}_{1,6}) \mathbf{A}_6^T$      | 0.0013               |
|               | Cellular Component    | 39,220              | 1,391               | 1                | $\underline{\mathbf{X}}_{1,7} = \mathbf{A}_1 \text{diag}(\mathbf{c}_{1,7}) \mathbf{A}_7^T$      | 0.0013               |
|               | Pathway               | 39,220              | 1,822               | 1                | $\underline{\mathbf{X}}_{1,8} = \mathbf{A}_1 \text{diag}(\mathbf{c}_{1,8}) \mathbf{A}_8^T$      | 0.0012               |
|               | Molecular Function    | 39,220              | 2,884               | 1                | $\underline{\mathbf{X}}_{1,9} = \mathbf{A}_1 \text{diag}(\mathbf{c}_{1,9}) \mathbf{A}_9^T$      | $8.610^{-4}$         |
| Compound      | Compound              | 24,313              | 24,313              | 2                | $\underline{\mathbf{X}}_{2,2} = [\mathbf{A}_2, \mathbf{A}_2, \mathbf{C}_{2,2}]$                 | 0.0023               |
|               | Disease               | 24,313              | 5,103               | 10               | $\underline{\mathbf{X}}_{2,3} = [\mathbf{A}_2, \mathbf{A}_3, \mathbf{C}_{2,3}]$                 | $6.76 \cdot 10^{-5}$ |
|               | Atc                   | 24,313              | 4,048               | 1                | $\underline{\mathbf{X}}_{2,10} = \mathbf{A}_2 \text{diag}(\mathbf{c}_{2,10}) \mathbf{A}_{10}^T$ | $1.6 \cdot 10^{-4}$  |
|               | Side Effect           | 24,313              | 5,701               | 1                | $\underline{\mathbf{X}}_{2,11} = \mathbf{A}_2 \text{diag}(\mathbf{c}_{2,11}) \mathbf{A}_{11}^T$ | 0.0010               |
|               | Pharmacological Class | 24,313              | 345                 | 1                | $\underline{\mathbf{X}}_{2,12} = \mathbf{A}_2 \text{diag}(\mathbf{c}_{2,12}) \mathbf{A}_{12}^T$ | $1.22 \cdot 10^{-4}$ |
| Disease       | Disease               | 5,103               | 5,103               | 1                | $\underline{\mathbf{X}}_{3,3} = \mathbf{A}_3 \text{diag}(\mathbf{c}_{3,3}) \mathbf{A}_3^T$      | $4.17 \cdot 10^{-5}$ |
|               | Anatomy               | 5,103               | 400                 | 1                | $\underline{\mathbf{X}}_{3,4} = \mathbf{A}_3 \text{diag}(\mathbf{c}_{3,4}) \mathbf{A}_4^T$      | 0.0018               |
|               | Symptom               | 5,103               | 415                 | 1                | $\underline{\mathbf{X}}_{3,13} = \mathbf{A}_3 \text{diag}(\mathbf{c}_{3,13}) \mathbf{A}_{13}^T$ | 0.0016               |

including SARS, MERS and SARS-COV2 and 8,103 FDA-approved drugs in Drugbank. Drugs with molecule weight less than 250 daltons are excluded from testing. Ribavirin was also excluded from the testing set, since there exist a ‘treat’ edge in the training set between Ribavirin and a target disease. In order to evaluate the performance of the proposed **TeX-Graph** and the alternatives we retrieve the top-100 ranked drugs that appear in the highest testing scoring (hyper-)edges. These are the proposed candidate drugs for COVID-19. Then we assess how many of the 32 clinical trial drugs <sup>3</sup> (Ribavirin is excluded) appear in the proposed candidate top-100 drugs.

**5.3 Methods** The methods used in the experiments are:

- **TeX-Graph.** The proposed **TeXGraph** algorithm initialized with Algorithm 2. The embedding dimension is set to  $F = 50$  and the algorithm runs for 10 iterations.
- **TransE-DRKG** [6,17]. **TransE** learns low dimensional KG embeddings using the score function shown in Table 2. For the the task of drug repurposing we use the specifications proposed in [17]. The  $l_2$  norm is chosen in the score function and training is performed using the deep graph library for knowledge graphs [34]. To evaluate the performance of TransE-DRKG on the drug repurposing task we used the 400-dimensional pretrained embeddings in [17], with which the drug repurposing results were better than the stand-alone code without pretraining.
- **3-way KG embeddings (3-way KGE).** We add as a baseline the embeddings produced by computing the CPD of tensor  $\underline{\mathbf{Y}}$  in (4.18). Recall that we use an

algebraic CPD of  $\underline{\mathbf{Y}}$  to initialize **TeX-Graph**. In 3-way KGE we initialize using the same procedure and also run 10 alternating least-squares iterations to compute the CPD of  $\underline{\mathbf{Y}}$ . 3-way KGE is tested with  $F = 50$ .

**5.4 Results** Table 4 shows the clinical trial drugs that appear in the top-100 recommendations along with their [rank-order]. The proposed approach retrieves 10 clinical trial drugs in the top-100 positions, and 7 in the top-50. Compared to **TransE-DRKG** that was the first proposed algorithm to perform drug-repurposing for COVID-19, **TeX-Graph** achieves 75% and 100% improvement in precision in the top-50 and top-100 respectively.

It is worth emphasizing that the proposed **TeX-Graph** retrieves approximately 1/3 of the COVID-19 clinical trial drugs, in the top-100, among a testing set of 8,103 drugs. This result is pretty remarkable and can essentially help cutting down the immense search space of medical research. For instance, consider the case of Dexamethasone, which is retrieved by **TeX-Graph** in the top ranked position (it admitted the highest score among all 8,103 drugs). At the onset of the pandemic, the initial guidance for Dexamethasone and other corticosteroids was indecisive. Guidelines from different sources issued either a weak recommendation to use Dexamethasone (with an asterisk that further evidence was required) or a weak recommendation against corticosteroids and Dexamethasone [23]. However, recent results indicate that treatment with Dexamethasone reduces mortality in patients with COVID-19 [16]. The results of **TeX-Graph** coaligned with the latest evidence and rank Dexamethasone as the top recommended drug. This suggests that our proposed data-driven approach could have essentially contributed in overturning the initial hesitancy to administrate Dexamethasone as a first line treatment.

<sup>3</sup>[www.covid19-trials.com](http://www.covid19-trials.com)

Table 4: Proposed candidate drugs for COVID-19

| TeX-Graph               | TransE-DRKG             | 3-way KGE        |
|-------------------------|-------------------------|------------------|
| F=50                    | F=400                   | F=50             |
| Dexamethasone [1]       | Dexamethasone [4]       | Oseltamivir [89] |
| Methylprednisolone [6]  | Colchicine [8]          |                  |
| Azithromycin [13]       | Methylprednisolone [16] |                  |
| Thalidomide [18]        | Oseltamivir [49]        |                  |
| Losartan [41]           | Deferoxamine [87]       |                  |
| Hydroxychloroquine [47] |                         |                  |
| Colchicine [48]         |                         |                  |
| Oseltamivir [60]        |                         |                  |
| Chloroquine [68]        |                         |                  |
| Deferoxamine [88]       |                         |                  |

## 6 Conclusion

In this paper we proposed a novel coupled tensor-matrix framework for knowledge graph embedding. The proposed model is principled and enjoys several favorable properties, including parsimony and uniqueness. The developed algorithmic framework admits lightweight updates and can handle very large graphs. Finally the proposed **TeX-Graph** showed very promising results in a timely application to drug repurposing, a task of paramount importance in the fight against COVID-19.

## 7 Acknowledgements

The work of the authors was supported in part by NSF IIS-1908070, ARO W911NF1910407. The authors also acknowledge Ioanna Papadatou, M.D., Ph.D, for contributing in the medical assessment of the produced results.

## A Appendix: TeX-Graph updates

**TeX-Graph** solves the following problem:

$$(A.1) \quad \min_{\{\mathbf{A}_m\}, \{\mathbf{C}_{m,n}\}} \sum_{(m,n) \in \mathcal{S}} \|\mathbf{X}_{m,n} - [\mathbf{A}_m, \mathbf{A}_n, \mathbf{C}_{m,n}]\|_F^2.$$

Then the update for  $\mathbf{A}_n$  is the solution of:

$$(A.2) \quad \min_{\mathbf{A}_n} \sum_{m \in \mathcal{S}_n^+} \|\mathbf{X}_{m,n} - [\mathbf{A}_m, \mathbf{A}_n, \mathbf{C}_{m,n}]\|_F^2 +$$

$$\sum_{p \in \mathcal{S}_n^-} \|\mathbf{X}_{n,p} - [\mathbf{A}_n, \mathbf{A}_p, \mathbf{C}_{n,p}]\|_F^2 \iff$$

$$\min_{\mathbf{A}_n} \sum_{m \in \mathcal{S}_n^+} \left\| \mathbf{X}_{m,n}^{(1)} - (\mathbf{C}_{m,n} \odot \mathbf{A}_m) \mathbf{A}_n^T \right\|_F^2 +$$

$$(A.3) \quad \sum_{p \in \mathcal{S}_n^-} \left\| \mathbf{X}_{n,p}^{(2)} - (\mathbf{C}_{n,p} \odot \mathbf{A}_p) \mathbf{A}_n^T \right\|_F^2.$$

where  $\mathcal{S}_n^+, \mathcal{S}_n^-$  are defined in (4.16). Taking the gradient of (A.3) with respect to  $\mathbf{A}_n$  and setting it to zero yields the equation in (4.15). The main bottleneck of (4.15) in terms of memory requirements and computational complexity is instantiating the Khatri-Rao products  $(\mathbf{C}_{n,p} \odot \mathbf{A}_p), (\mathbf{C}_{m,n} \odot \mathbf{A}_m)$  and computing the MTTKRP  $(\mathbf{C}_{n,p} \odot \mathbf{A}_p)^T \mathbf{X}_{n,p}^{(1)}, (\mathbf{C}_{m,n} \odot \mathbf{A}_m)^T \mathbf{X}_{m,n}^{(2)}$ . We focus on the computation of:

$$(A.4) \quad (\mathbf{C}_{n,p} \odot \mathbf{A}_p)^T \mathbf{X}_{n,p}^{(1)}.$$

Equation (A.4) can be equivalently written as:

$$(A.5) \quad \begin{bmatrix} \mathbf{A}_p \text{diag}(\mathbf{C}_{n,p}(1,:)) \\ \vdots \\ \mathbf{A}_p \text{diag}(\mathbf{C}_{n,p}(K_{n,p},:)) \end{bmatrix}^T \begin{bmatrix} \mathbf{X}_{n,p}^{1^T} \\ \vdots \\ \mathbf{X}_{n,p}^{K_{n,p}^T} \end{bmatrix} = \sum_{k=1}^{K_{n,p}} \text{diag}(\mathbf{C}_{n,p}(k,:)) \mathbf{A}_p \mathbf{X}_{n,p}^{k^T}.$$

It is clear from equation (A.5) that  $(\mathbf{C}_{n,p} \odot \mathbf{A}_p)$  need not be instantiated. The number of flops to compute (A.5) is  $\mathcal{O}(F \cdot \text{nnz}(\mathbf{X}_{n,p}))$ . Note that computing  $(\mathbf{C}_{m,n} \odot \mathbf{A}_m)^T \mathbf{X}_{m,n}^{(2)}$  is only different in the fact that the frontal slabs are not transposed, and is thus omitted.

The update for  $\mathbf{C}_{m,n}$  is the solution of:

$$(A.6) \quad \min_{\mathbf{C}_{m,n}} \|\mathbf{X}_{m,n} - [\mathbf{A}_m, \mathbf{A}_n, \mathbf{C}_{m,n}]\|_F^2 \iff$$

$$(A.7) \quad \min_{\mathbf{C}_{m,n}} \left\| \mathbf{X}_{m,n}^{(3)} - (\mathbf{A}_m \odot \mathbf{A}_n) \mathbf{C}_{m,n}^T \right\|_F^2.$$

Taking the gradient of (A.7) with respect to  $\mathbf{C}_{m,n}$  and setting it to zero yields the equation in (4.17). The main memory and computation bottleneck of (4.17) is computing the MTTKRP. The formula in (A.5) can be utilized if  $\mathbf{C}_{n,p}$  is replaced by  $\mathbf{A}_n$ ,  $\mathbf{A}_p$  is replaced by  $\mathbf{A}_m$  and the transposed frontal slabs  $\mathbf{X}_{m,n}^{k^T}$  are replaced by vertical slabs.

## References

- [1] Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary Ives. Dbpedia: A nucleus for a web of open data. In *The semantic web*, pages 722–735. Springer, 2007.
- [2] Ivana Balazevic, Carl Allen, and Timothy Hospedales. Tucker: Tensor factorization for knowledge graph completion. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5188–5197, 2019.
- [3] Albert-László Barabási et al. *Network science*. Cambridge university press, 2016.
- [4] Michele Berlingario, Michele Coscia, Fosca Giannotti, Anna Monreale, and Dino Pedreschi. Foundations of multidimensional network analysis. In *2011 international conference on advances in social networks analysis and mining*, pages 485–489. IEEE, 2011.
- [5] Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. Freebase: a collaboratively created graph database for structuring human knowledge. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, pages 1247–1250, 2008.
- [6] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. In *Advances in neural information processing systems*, pages 2787–2795, 2013.

[7] Léon Bottou. Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT'2010*, pages 177–186. Springer, 2010.

[8] Andrew Carlson, Justin Betteridge, Bryan Kisiel, Burr Settles, Estevam R Hruschka, and Tom M Mitchell. Toward an architecture for never-ending language learning. In *Twenty-Fourth AAAI Conference on Artificial Intelligence*, 2010.

[9] Luca Chiantini and Giorgio Ottaviani. On generic identifiability of 3-tensors of small rank. *SIAM Journal on Matrix Analysis and Applications*, 33(3):1018–1037, 2012.

[10] Ignat Domanov and Lieven De Lathauwer. On the uniqueness of the canonical polyadic decomposition of third-order tensors — part ii: Uniqueness of the overall decomposition. *SIAM Journal on Matrix Analysis and Applications (SIMAX)*, 34(3):876–903, 2013.

[11] David Easley, Jon Kleinberg, et al. *Networks, crowds, and markets*, volume 8. Cambridge university press Cambridge, 2010.

[12] Thomas Franz, Antje Schultz, Sergej Sizov, and Steffen Staab. Triplerank: Ranking semantic web data by tensor decomposition. In *International semantic web conference*, pages 213–228. Springer, 2009.

[13] Richard A Harshman, Margaret E Lundy, et al. Parafac: Parallel factor analysis. *Computational Statistics and Data Analysis*, 18(1):39–72, 1994.

[14] Daniel S Himmelstein and Sergio E Baranzini. Heterogeneous network edge prediction: a data integration approach to prioritize disease-associated genes. *PLoS computational biology*, 11(7), 2015.

[15] Daniel Scott Himmelstein, Antoine Lizee, Christine Hessler, Leo Brueggeman, Sabrina L Chen, Dexter Hadley, Ari Green, Pouya Khankhanian, and Sergio E Baranzini. Systematic integration of biomedical knowledge prioritizes drugs for repurposing. *Elife*, 6:e26726, 2017.

[16] Peter Horby, Wei Shen Lim, Jonathan R Emberson, Marion Matham, Jennifer L Bell, Louise Linsell, Natalie Staplin, Christopher Brightling, Andrew Ustianowski, Einas Elmahi, et al. Dexamethasone in hospitalized patients with covid-19-preliminary report. *The New England journal of medicine*, 2020.

[17] Vassilis N. Ioannidis, Xiang Song, Saurav Manchanda, Mufei Li, Xiaoqin Pan, Da Zheng, Xia Ning, Xiangxiang Zeng, and George Karypis. Drkg - drug repurposing knowledge graph for covid-19. <https://github.com/gnn4dr/DRKG/>, 2020.

[18] Xueyan Jiang, Volker Tresp, Yi Huang, and Maximilian Nickel. Link prediction in multi-relational graphs using additive models. *SeRSy*, 919:1–12, 2012.

[19] Tamara G Kolda, Brett W Bader, and Joseph P Kenny. Higher-order web link analysis using multilinear algebra. In *Proceedings of Fifth IEEE International Conference on Data Mining*, pages 8–pp. IEEE, 2005.

[20] Hailun Lin, Yong Liu, Weiping Wang, Yinliang Yue, and Zheng Lin. Learning entity and relation embeddings for knowledge resolution. *Procedia Computer Science*, 108:345–354, 2017.

[21] Mark Newman. *Networks*. Oxford university press, 2018.

[22] Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. A three-way model for collective learning on multi-relational data. In *Icml*, volume 11, pages 809–816, 2011.

[23] Hallie C Prescott and Todd W Rice. Corticosteroids in covid-19 ards: evidence and hope during the pandemic. *Jama*, 324(13):1292–1295, 2020.

[24] Steffen Rendle and Lars Schmidt-Thieme. Pairwise interaction tensor factorization for personalized tag recommendation. In *Proceedings of the third ACM international conference on Web search and data mining*, pages 81–90, 2010.

[25] Sebastian Riedel, Limin Yao, Andrew McCallum, and Benjamin M Marlin. Relation extraction with matrix factorization and universal schemas. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 74–84, 2013.

[26] Eugenio Sanchez and Bruce R Kowalski. Tensorial resolution: a direct trilinear decomposition. *Journal of Chemometrics*, 4(1):29–45, 1990.

[27] Nicholas D Sidiropoulos, Lieven De Lathauwer, Xiao Fu, Kejun Huang, Evangelos E Papalexakis, and Christos Faloutsos. Tensor decomposition for signal processing and machine learning. *IEEE Transactions on Signal Processing*, 65(13):3551–3582, 2017.

[28] Amit Singhal. Introducing the knowledge graph: things, not strings. *Official google blog*, 16, 2012.

[29] Richard Socher, Danqi Chen, Christopher D Manning, and Andrew Ng. Reasoning with neural tensor networks for knowledge base completion. In *Advances in neural information processing systems*, pages 926–934, 2013.

[30] MIKAEL Sørensen, Ignat Domanov, and L De Lathauwer. Coupled canonical polyadic decompositions and (coupled) decompositions in multilinear rank-(lr, n, lr, n, 1) terms—part ii: Algorithms. *SIAM Journal on Matrix Analysis and Applications*, 36:1015–1045, 2015.

[31] Fabian M Suchanek, Gjergji Kasneci, and Gerhard Weikum. Yago: a core of semantic knowledge. In *Proceedings of the 16th international conference on World Wide Web*, pages 697–706, 2007.

[32] Zhiqing Sun, Zhi-Hong Deng, Jian-Yun Nie, and Jian Tang. Rotate: Knowledge graph embedding by relational rotation in complex space. In *International Conference on Learning Representations*, 2018.

[33] Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. Embedding entities and relations for learning and inference in knowledge bases. *arXiv preprint arXiv:1412.6575*, 2014.

[34] Da Zheng, Xiang Song, Chao Ma, Zeyuan Tan, Zihao Ye, Jin Dong, Hao Xiong, Zheng Zhang, and George Karypis. Dgl-ke: Training knowledge graph embeddings at scale. *arXiv preprint arXiv:2004.08532*, 2020.