Filtered Tensor Construction and Decomposition for Drug Repositioning

Dimitri Leggas, Muthu Baskaran, James Ezick, Brendan von Hofe

Reservoir Labs, Inc.

New York, NY

{leggas, baskaran, ezick, vonhofe}@reservoir.com

Abstract—Drug repositioning (also called "drug repurposing") is a drug development strategy that saves time and money by finding new uses for existing drugs. While a variety of computational approaches to drug repositioning exist, recent work has shown that tensor decomposition, an unsupervised learning technique for finding latent structure in multidimensional data, is a useful tool for drug repositioning. The known relationships between drugs, targets, and diseases can easily be encoded as a tensor, and by learning a low-rank representation of this tensor, decompositions can complete missing entries and therefore predict novel drug-disease associations. Multiple recent works, in the context of cancer and COVID-19 drug discovery, have used joint tensor decompositions to suggest drug repositioning candidates. While these methods make high-quality predictions, they rely on specialized decompositions formulated for specific problems. In this work, we use ENSIGN, a suite of tensor decomposition tools, to show that CP tensor decompositions of a single tensor encoding drug-target-disease associations are capable of predicting verifiable drug repositioning candidates. Because the tensors generated by drug repositioning problems are sparse, we introduce a filtered tensor construction to limit the span of the tensor without losing information needed to learn the relevant associations. We show that our method predicts verifiable novel drug-disease associations in cancer and COVID-19 data. The simplicity of our approach makes it an attractive tool for biomedical researchers looking for out-of-the-box solutions, and ENSIGN brings an added level of usability and scalability.

Index Terms—Drug Repositioning, COVID-19, Cancer, Tensor, Tensor Decompositions, Tensor Completion

I. INTRODUCTION

Drug repositioning is the process of finding alternative uses for drugs outside of their original medical indication. As a drug development strategy, repositioning has multiple benefits over *de novo* drug development including a better risk-versus-reward trade-off and reduced time and costs. Although speeding up and reducing the cost of drug development has often resulted in increased risk, this is not the case for drug repositioning as the compounds in question, having been developed for other indications, have often been targeted, optimized, screened *in vitro* and *in vivo*, manufactured, and possibly used in clinic [1]. Repositioned drugs have therefore cleared many of the regulatory hurdles that result in a 2% success rate in *de novo* drug development [2]. While licensing, further trials, and registration are still required before a

repositioned drug can go on the market, a key bottleneck is the identification of compounds that may treat diseases they were not designed to treat. Serendipitous observation of the side effects of known drugs is one approach to finding candidates for repositioning, but computational methods offer principled ways to examine the known relationships between treatments and diseases and predict novel relationships.

In this work, we consider the application of CANDE-COMP/PARAFAC (CP) tensor decomposition to drug repositioning. CP tensor decomposition is an unsupervised machine learning tool for decomposing high-dimensional arrays, or *tensors*, into a sum of rank-1 tensors, or *components*, analogous to the singular value decomposition for matrices. It is useful for studying correlations between different attributes of data [3], and it has been successfully applied in diverse fields and data, such as spatiotemporal analysis [4], cybersecurity [5], [6], chemistry [7], [8], machine learning [9]–[11], precision healthcare [12], [13], genomics [14], neuroscience [15]–[17], and others. Tensors and their decomposition provide an ideal framework for formulating and solving drug repositioning problems because known relationships are easily encoded as tensors, and decompositions provide a low-rank summary of the known relationships and, as we will show, directly predict novel relationships. A particularly useful set of relationships to consider are those between drugs, targets (genes or the proteins they express), and diseases.

Encoding known drug-target-disease triplets is straightforward: construct a 3-dimensional tensor in which the dimensions, or *modes*, correspond to drugs, targets, and diseases, respectively, and the indices in a given mode correspond to specific instances of that category. Each tensor entry therefore corresponds to one drug-target-disease triplet, and known relationships are quantified with nonzero values at the corresponding entries while unknown relationships correspond to zeroes in the tensor. Not only does a low-rank decomposition of a tensor summarize the data as a sum of rank-1 components, which are represented by *factor matrices*, but when reconstructed into a tensor, it includes nonzero entries not present in the original tensor that are implied by the multi-linear relationships discovered by the decomposition. Ranking these by magnitude provides a means of presenting top predictions of novel drug-target-disease relationships and thus identifying candidates for drug repositioning.

This is not the first work to use tensor decomposition for

drug repositioning. Wang et al. [18] create tensors encoding drug-protein-disease relationships according to the method above and decompose them with the additional constraints that the factor matrices also reproduce known drug-drug and protein-protein interactions. They apply their method to reposition drugs for cancer and show that their top predicted drug-disease relationships are validated in the literature. Kanatsoulis et al. [19] propose a coupled tensor-matrix knowledge graph embedding that encodes every pairwise relationship (e.g. drug-target, target-disease, drug-drug) as a tensor. By decomposing this coupled system of tensors so that entities share factor matrices across decompositions, they perform a completion of the knowledge graph. The factor matrices relating to drugs and diseases can then be used to predict new relationships, and this method finds 10 COVID-19 trial therapies in its top 100 predictions. Both of these works, in their reliance on coupled methods, make the implicit assumption that the sparse drug-target-protein tensor is insufficient to learn the underlying relationships in order to predict repositioning candidates.

We show that out-of-the-box tensor decompositions applied to an appropriately-chosen subset of the drug-target-disease tensor are sufficient to predict novel drug-disease relationships. The smaller tensor is constructed by providing a list of entities (drugs, targets, or diseases) that are known to be relevant *a priori* and only encoding drug-target-disease relationships that include at least one of these entities. Assuming that the filter criteria are well-chosen, excluding information improves the quality of the predictions because the low-rank decomposition does not have to capture irrelevant relationships. We leverage the tensor construction tools, sparse decomposition methods, and post-processing tools in ENSIGN [20]–[25], a highly parallelized tensor decomposition package written in ANSI C, to perform our decompositions. We use the Python bindings for ENSIGN to perform our experiments in a Jupyter notebook [26], and we make the Jupyter notebook for the analysis of the cancer data available on the ENSIGN website [27]. To summarize our contributions:

- We propose a modified method for constructing drug-target-disease tensors that filters out relationships that are believed irrelevant
- We show that applying alternating tensor decompositions to our filtered tensors is sufficient for finding novel drug-disease relationships
- We demonstrate our approach on cancer data and show that our method is capable of suggesting candidates for repositioning drugs between 229 types of cancer
- We show that our method recommends 5 drugs for COVID-19 that underwent clinical trials and that 2 of these are not predicted Kanatsoulis et al. [19]

The organization of the rest of the paper is as follows. In Section II we discuss previous work on drug repositioning, tensor completion, and their intersection. In Section III we provide tensor preliminaries and an exposition of our method. In Section IV we show the results of our method on cancer and COVID-19 data. A final discussion is in Section V.

II. PREVIOUS WORK

A. Computational Drug Repositioning

For a recent survey of computational approaches to drug repositioning, see the survey by Jarada et al. [28]. We summarize the major approaches in comparison to our proposed method. There is a large body of work on using network analysis for drug repositioning due to the fact that relationships between drugs, targets, and diseases can naturally be expressed as graphs in which nodes are entities and edges are known relationships. Network analysis approaches can be categorized by what kinds of networks are considered. Possible data choices are any combination of the involved entity types: protein-protein interaction networks, drug-target interaction networks, drug-drug interaction networks, drug-disease association networks, drug-side effect association networks, and disease–disease interaction networks, and other heterogeneous networks.

The analysis of networks with one interaction type has been performed using bipartite graph models. Bleakley et al. [29] propose a supervised bipartite graph to learn novel drug-target relationships and classify drug-target interaction types. Kinnings et al. [30] study drug-drug interaction networks in order to learn drug communities, which enable the identification of drugs that may behave similarly. Finally, Hu et al. [31] build disease-drug networks to directly identify drug repositioning candidates. Our proposed method, like these, relies on networks of relationships between drugs, targets, and diseases, but it considers more than a single interaction class.

Others have developed non-bipartite graph approaches for studying more heterogeneous networks using clustering and network centrality measures. Using disease-gene associations and drug-target interactions, Wu et al. [32] build a network in which nodes represent genes and diseases and edges represent target interactions and perform clustering in order to predict important drug-disease links, which are presented as drug repositioning candidates. Rashkit et al. [33] build disease-target and drug-target networks and use a variety of network centrality measures as an ensemble for predicting important drugs and diseases across the two networks. While these and other works have leveraged more classes of interaction that are relevant to drug repositioning, we believe a tensor decomposition approach is superior for its ability to simultaneously learn relationships between all classes of entities.

Outside of network analysis, deep learning provides another set of tools for predicting drug repositioning candidates. Aliper et al. [34] use a fully-connected network to predict drug indications and outperform support vector machines. This method could be used for performing drug repositioning by predicting application classes of a given drug. Hu et al. [35] design a convolutional neural network that learns chemical structure and protein sequences for predicting drug-target interactions. Segler et al. [36] build a recurrent neural network for generating compounds with desired biological properties. While the method performs well, the authors note that it replicates the *de novo* drug development process. Finally, Zeng

et al. [37] use deep autoencoders to learn embeddings of relevant interactions. Despite the promising results from deep learning approaches, we argue that the inherent interpretability of tensor methods is a reason to develop strong tensor completion approaches.

B. Tensor Completion

This work frames drug repositioning as a tensor completion problem: known relationships are nonzero entries in a tensor, and we would like to predict new interactions (missing tensor entries). Tensor completion is performed by decomposing a tensor with missing entries in order to find its latent structure so that the reconstructed tensor fills out missing values. Acar et al. [38] propose an algorithm called CP-WOPT (CP Weighted OPTimization) that successfully recovers tensor entries even with 99% missing data and in the presence of noise.

Recently, Liu and Moitra [39] offer a tensor completion algorithm that consists of performing an alternating tensor decomposition with pre- and post-processing steps. They are able to make provable claims about the completion rate of their method. Namely, they show that when using their pre- and post-processing algorithms, a logarithmic (in the mode sizes) number of iterations of alternating minimization are necessary to estimate the entries of the true tensor within inverse polynomial (in the mode sizes) accuracy. However, their experimental results indicate that it is possible for the alternating minimization (i.e. the decomposition) alone to produce a completed tensor.

C. Tensor Decomposition for Drug Repositioning

We describe two recent works that use tensor decomposition/completion for drug repositioning. In addition to serving as the basis for our method, they serve as points of comparison for ours, as they both use publicly available data in their experiments. Wang et al. [18] introduce tensor decomposition for drug repositioning. They create a tensor in which the dimensions correspond to drugs, targets, and diseases and the nonzero entries reflect known relationships. Implicit in this model is the notion that there are unknown relationships, the corresponding entries of which are zero. Additionally, they consider two matrices that encode drug-drug and target-target interactions. After finding a low-rank decomposition of the tensor jointly with two matrices, they reconstruct the tensor and rank new entries by their scores. Their top predictions are validated by a literature search. Our approach can be considered a simplification of theirs in that we decompose the drug-target-disease tensor without coupling it with additional matrices.

Kanatsoulis et al. [19] use tensor completion to predict therapies for COVID-19. They consider a knowledge graph containing relationships between drugs, genes, diseases, side effects, symptoms, cellular components, pathways, pharmacological classes, molecular functions, anatomy, and biological processes. For every pair of entity types, they encode the known relationships as a 3-dimensional tensor in which two dimensions correspond to the types and the third dimension

corresponds to the relationship types. As each entity appears in multiple tensors, this formulation leads to a large system of tensor decomposition equations where the latent representations of each entity type are shared across tensors. After performing a joint decomposition of these tensors, they derive scores for drug-disease relationships, in particular for SARS-CoV-2 variants, the viruses that cause COVID-19. In their top 100 predictions, they identify 10 therapies that underwent clinical trials. Again, our approach does not require a joint decomposition.

III. METHODS

A. Tensor Preliminaries

Before presenting our method, we introduce the basic concepts related to tensor decomposition. For a full treatment on the topic, see [3]. A *tensor* is a multidimensional array. The *order* of a tensor is the number of dimensions, also known as modes. Thus, for any $N \in \mathbb{N}$, the N-dimensional array $\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ is an N-*mode* tensor. Hence, a vector is a 1-mode tensor and a matrix is a 2-mode tensor. We denote the $(i_1, \ldots, i_N)$th entry of the N-mode tensor $\mathcal{X}$ as $x_{i_1, \ldots, i_N}$.

In this work, we fix $N = 3$ and work with tensors encoding drug-target-disease relationships: $\mathcal{X} \in \mathbb{R}^{I_C \times I_T \times I_D}$ where C, T, and D signify drugs (compounds), targets, and diseases, respectively.

A rank-R CP decomposition is an approximation of $\mathcal{X}$ as a sum of R outer products of 3 vectors:

$$\mathcal{X} \approx \sum_{r=1}^{R} \boldsymbol{c}_r \circ \boldsymbol{t}_r \circ \boldsymbol{d}_r \qquad (1)$$

It is convenient to express the decomposition as the outer product of its *factor matrices* $\boldsymbol{C} \in \mathbb{R}^{I_C \times R}, \boldsymbol{T} \in \mathbb{R}^{I_T \times R}$, and $\boldsymbol{D} \in \mathbb{R}^{I_D \times R}$ where the rth column of $\boldsymbol{C}$ is the vector $\boldsymbol{c}_r$, and so on. We use the notation

$$\mathcal{X} \approx \mathcal{T} = [\![\boldsymbol{\lambda}; \boldsymbol{C}, \boldsymbol{T}, \boldsymbol{D}]\!]. \qquad (2)$$

to rewrite Equation 1 compactly. Here we have column-normalized the factor matrices and absorbed the weights into $\boldsymbol{\lambda} \in \mathbb{R}^R$ such that λ_r is the *weight* corresponding to the rth component. On a per-entry basis, (2) is equivalent to

$$x_{i_C, i_T, i_D} \approx t_{i_C, i_T, i_D} = \sum_{r=1}^{R} \lambda_r \times c_{i_C, r} \times t_{i_T, r} \times d_{i_D, r}. \qquad (3)$$

If the decomposition rank R is equal to the rank of the tensor $\mathcal{X}$, then (2) is exact and $\mathcal{T} = \mathcal{X}$. In general, however, we seek a *low-rank* representation of $\mathcal{X}$, and $\mathcal{T}$ is an approximation to the data tensor $\mathcal{X}$. In order to find $\mathcal{T}$, we minimize the sum of the elementwise losses as measured by a function $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$,

$$\min_{\mathcal{T}} \sum_{i_C=1}^{I_C} \sum_{i_T=1}^{I_T} \sum_{i_D=1}^{I_D} f(x_{\mathbf{i}}, t_{\mathbf{i}}). \qquad (4)$$
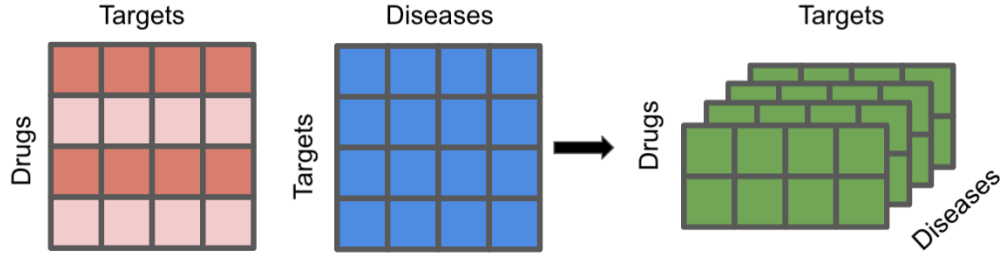
Fig. 1: We use the following construction criterion for building our tensors: A drug-target-disease relationship exists if the corresponding drug-target and target-disease relationships exist. The filtering criterion, our contribution to the tensor construction, is a set of entities of any type. During tensor construction, only drug-target-disease triplets with at least one entity in the specified set are considered. In the example above, there are four drugs, targets, and diseases. The highlighted rows in the drug-target adjacency matrix indicate that the filtering criterion is a subset of two drugs. As a result, the constructed tensor has dimensions $2 \times 4 \times 4$.

Here, we use the shorthand $x_{\mathbf{i}} = x_{i_C, i_T, i_D}$ and $t_{\mathbf{i}} = t_{i_C, i_T, i_D}$. The elementwise loss function used is $f(x_{\mathbf{i}}, t_{\mathbf{i}}) = (x_{\mathbf{i}} - t_{\mathbf{i}})^2$.

In this work, in addition to finding a low-rank approximation of the data that are known at the time of decomposition, our goal is to find additional entries that are found in the approximation $\mathcal{T}$ and not in $\mathcal{X}$. These are used to find repositioning candidates.

B. Filtered Tensor Construction

We encode our known drug-target-disease relationships as a 3-mode tensor in which each mode corresponds to one of the entity types and each index within a mode corresponds to a specific entity. Following Wang et al. [18], the tensor is a Boolean tensor in which the value at an index is 1 if the corresponding drug-target-disease triplet has a known relationship and 0 otherwise. In the use cases we consider, the available data consist of lists of known pairwise entity relationships, in particular known drug-target and target-disease relationships. Given these data, in order to construct the tensors representing the known information, we require a *construction criterion* for determining the existence of drug-target-disease relationships and a *filtering criterion* for removing irrelevant entities. Again following Wang et al. [18], we use the following criterion: A drug-target-disease relationship exists if the corresponding drug-target and target-disease relationships exist.

The filtering criterion, our contribution to the tensor construction, is a set of entities of any type. During tensor construction, only drug-target-disease triplets with at least one entity in the specified set are considered. This very simple filter enables us to obtain predictions we did not otherwise find. The tensors we consider are very sparse, and much of the sparsity comes from the large mode sizes. By removing entities that we know have no relevance to the drug repositioning problem in question, we are able to decrease the span of the tensor and improve the qualities of the decomposition.

Let R_{CT} and R_{TD} be the sets of known drug-target and target-disease relationships. In our use cases, the data provide these explicitly in the form of adjacency matrices. Let F be the set provided as the filtering criterion. Then the drug-target-disease tensor is given by

$$x_{i_C, i_T, i_D} = \begin{cases} 1 & (i_C, i_T) \in R_{CT} \text{ and} \\ & (i_T, i_D) \in R_{TD} \text{ and} \\ & (i_C \in F \text{ or } i_T \in F \text{ or } i_D \in F) \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

C. Tensor Completion

We do not perform a specialized tensor completion algorithm a la Acar et al. [38] or Liu et al. [39]. As stated, our objective is to test how basic tensor decomposition performs for the task of tensor completion in the context of drug repositioning. We use the alternating least-squares CP decomposition (CP-ALS) [44], [45], which assumes Gaussian-distributed data and solves least-squares problems to update the factor matrices in an alternating fashion. We use the scalable implementation provided in ENSIGN, and perform decompositions of the filtered drug-target-disease tensors. The factor matrices are then used to reconstruct the tensor, and any nonzero entries in the reconstruction that are not present in the original tensor are considered predictions. Because the original tensor was Boolean and the reconstruction is not perfect, the reconstructed values are between 0 and 1, which we interpret as the confidence in the prediction. This allows the user to set a threshold for which predictions should be considered as drug repositioning candidates.

IV. RESULTS

A. Identifying Cancer Treatments for Repositioning

We consider the data compiled by Luo et al. [46] and used by Wang et al. [18] available on GitHub [47]. The heterogeneous network consists of 708 drugs from Drug-Bank [48], 1,512 targets from Human Protein Reference Database (HPRD) [49], and 5,603 diseases from Comparative Toxicogenomics Database (CTD) [50]. There are adjacency matrices for every pair of entity types, and we consider

Drug	Original Indication	Novel Indication	Literature Validation
Verapamil	high blood pressure, heart arrhythmias, and angina	neoplasm invasiveness	Yohem et al. [40]
Halothane	anesthetic	colonic neoplasms / neoplasm invasiveness	Rudnick et al. [41]
Verapamil	high blood pressure, heart arrhythmias, and angina	renal cell carcinoma	Yu et al. [42]
Nitrazepam	hypnotic	prostatic neoplasms	None
Theophylline	chronic asthma and chronic lung disease	leukemia	Makower et al. [43]

TABLE I: We embed drug-target and target-disease adjacency matrices as drug-target-disease relationships in a 3-mode Boolean tensor. The non-zero entries represent known relationships, and we complete the tensor in order to discover novel relationships. To perform the completion process we decompose the tensor using CP-ALS and reconstruct the decomposition to find values that did not appear in the original tensor. Our top 5 predicted drug-target-disease relationships have scores above 0.9. From these predicted relationships, we derive the drug repositioning candidates found in this table. Four of our predictions are validated by a paper in the medical literature.

drug-target associations (collected from DrugBank), target-disease associations (collected from CTD), and drug-disease associations (collected from CTD).

Following the tensor construction method in Section III-B, we consider a drug-target-disease interaction to exist when the drug-target and target-disease relationships exist. To evaluate our method for cancer drug repositioning, we define the filtering criterion F as the set of cancer-related diseases that have at least 300 target-disease associations and at least 100 drug-disease associations. The connectivity requirements help to ensure that the tensor is not too sparse. The resulting tensor contains 473 drugs, 397 targets, and 229 diseases. There are 51,694 entries, meaning that many drug-target-disease relationships are inferred from the adjacency matrices, for an overall sparsity of 99.88%.

We perform a rank-250 decomposition of the tensor using CP-ALS, then reconstruct the tensor entries from the factor matrices. After ranking the novel drug-target-disease triplets by score, there are only 5 drug-disease pairs derived from triplets with scores greater than 0.9. These drug-disease relationships are presented in Table I. While we predict relatively few novel associations with high confidence, owing to the sparsity of the tensor, 4 of those that we predict are validated by the literature as having been tenable hypotheses.

B. Proposing COVID-19 Therapies

In the context of COVID-19 drug development, we consider the Drug Repurposing Knowledge Graph for Covid-19 (DRKG) created by Ioannidis et al. [51] and used by Kanatsoulis et al. [19] in their tensor completion work. The DRKG is a heterogeneous network consisting of relationships between drugs, genes, diseases, side effects, symptoms, cellular components, pathways, pharmacological classes, molecular functions, anatomy, and biological processes. Moreover, the associations themselves are separated into classes (e.g., a disease-gene relationship may be a "blocker" or an "inhibitor"). The drug-target and drug-disease relationships are pulled from Drug-Bank [48], the Global Network of Biomedical Relationships (GNBR) [52], and Hetionet [53], while the target-disease relationships are pulled from the last two networks.

Following the tensor construction method in Section III-B, we consider a drug-target-disease interaction to exist when the drug-target and target disease relationships exist. To evaluate our method for COVID-19 drug repositioning, we defined the filtering criterion F as the set of genes that have an association with a COVID-19 variant or genes that are associated with one of these genes. This filtering drastically reduces the number of associations the tensor decomposition must capture in its low-rank representation while retains the most relevant data for finding associations related to COVID-19. The resulting tensor encodes 3,626 drugs (20 of which are known to have undergone trials for treating COVID-19), 4,215 genes, and 1,753 diseases (of which 27 are SARS-CoV-2 variants, the viruses that cause COVID-19).

We perform a rank-400 decomposition of the tensor using CP-ALS, then reconstruct the tensor entries from the factor matrices. As stated in Section III-C, we take reconstructed entries not seen in the original tensor with scores greater than a user-defined confidence threshold as repositioning candidates. Specifically, we set the confidence threshold to 0.99. Amongst the predictions surpassing the threshold are 5 drug-disease associations between known trial COVID-19 drugs and the SARS-CoV-2 variants. These are presented in Table II. There are 304 false positives

We did not report all of the drugs found by Kanatsoulis et al. [19], but we found 2 that they did not report: Ibuprofen and tranexamic acid. Both methods rank Dexamethasone as the top choice, and as they point out, there were originally contradictory recommendations for and against it, but our results align with the updated evidence that Dexamethasone reduces the 28-day mortality rate in COVID-19 patients [54].

While our method did not outperform that of Kanatsoulis et al. [19], there are advantages to our approach. First, rather than jointly decomposing many tensors and matrices, we decompose a single tensor with a common algorithm that is implemented in tensor decomposition software. Next, because we complete a single tensor, our scores are simply the nonzero unseen entries, which take values between 0 and 1. This allows the scores to be interpreted as confidences in the prediction. The practical and conceptual advantages of this approach warrant further work toward reducing the false positives.

V. Discussion

We use ENSIGN in order to demonstrate that out-of-the-box CP tensor decompositions are capable of performing tensor completion for drug repositioning. When we apply our method

Drug	Indication	Kanatsoulis et al. Rank [19]	Our Score
Dexamethasone	Steroid	1	0.999
Methylprednisolone	Steroid	6	0.996
Azithromycin	Antibiotic	13	-
Thalidomide	Immunomodulator	18	-
Losartan	Anti-hypertensive	41	0.995
Hydroxychloroquine	Immunosuppressive	47	-
Colchicine	Anti-inflammatory	48	-
Oseltamivir	Anti-viral	60	-
Chloroquine	Immunosuppressive	68	-
Deferoxamine	Iron reducer	88	-
Ibuprofen	Anti-inflammatory	-	0.997
Tranexamic acid	Clotting promoter	-	0.995

TABLE II: We embed drug-target and target-disease relationships found in a COVID-19 knowledge graph [51] as a drug-target-disease tensor encoding known associations. We use CP-ALS to decompose the tensor and reconstruct the decomposition to perform tensor completion. Any reconstructed value not seen in the original tensor with a value above the threshold of 0.99 is a predicted drug-target-disease pair, from which we derive our COVID-19 treatment predictions. We find 3 predictions made by Kanatsoulis et al. [19] in addition to 2 not noted in their paper.

of tensor construction and decomposition to a tensor encoding known drug-target-cancer relationships, we predict few novel drug-disease associations, yet all but one of them are validated by the literature. When applied to the COVID-19 knowledge graph, our method identifies 5 of the known trial therapies for treating COVID-19. In addition to being able to identify viable drug repositioning candidates, our method relies on constructing a single tensor and performing a straightforward decomposition. No joint or other specialized decompositions are necessary. The simplicity and efficacy of our approach make it attractive for use in rapid experimentation by the biomedical research community and drug developers.

We have laid the foundations on which to build an elegant and robust tensor completion method for drug repositioning. Despite the ability of our method to predict novel drug-disease associations that can be validated by the literature, a remaining challenge is to reduce the number of false positives produced by our method. If the goal is that these methods should play a large role in the drug development process, then it is not sufficient that they produce correct predictions, but that they only produce correct predictions. Future work will include improving the robustness of our simplified approach and adding advancements that help to increase confidence in the predictions.

It is also important to note that our method is easily extended to incorporate additional information into new dimensions of the tensor. We consider a 3-dimensional drug-target-disease tensor, but genotypic, phenotypic, chemical, pathway, or symptom information could be incorporated in additional modes analogous to how other methods incorporate these features through coupled tensors. This approach would retain the simplicity and ease-of-use of building and working with a single tensor.

ACKNOWLEDGMENTS

REFERENCES

[1] T. T. Ashburn and K. B. Thor, "Drug repositioning: identifying and developing new uses for existing drugs," *Nature Reviews Drug Discovery*, vol. 3, no. 8, pp. 673–683, 2004. [Online]. Available: https://doi.org/10.1038/nrd1468

[2] H. Xue, J. Li, H. Xie, and Y. Wang, "Review of drug repositioning approaches and resources," *International journal of biological sciences*, vol. 14, no. 10, pp. 1232–1244, 07 2018. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/30123072

[3] T. G. Kolda and B. W. Bader, "Tensor Decompositions and Applications," *SIAM Review*, vol. 51, no. 3, pp. 455–500, September 2009.

[4] T. Henretty, M. Baskaran, J. Ezick, D. Bruns-Smith, and T. A. Simon, "A quantitative and qualitative analysis of tensor decompositions on spatiotemporal data," in *High Performance Extreme Computing Conference (HPEC), 2017 IEEE*. IEEE, 2017, pp. 1–7.

[5] M. Baskaran, T. Henretty, J. Ezick, R. Lethin, and D. Bruns-Smith, "Enhancing network visibility and security through tensor analysis," *Future Generations Computer Systems*, vol. 96, 2 2019.

[6] T. M. Ranadive and C. J. Coley, "Ieee bigdata 2019 cup: Binary classification via tensor completion," in *2019 IEEE International Conference on Big Data (Big Data)*, 2019, pp. 5908–5914.

[7] K. Murphy, C. Stedmon, D. Graeber, and R. Bro, "Fluorescence spectroscopy and multi-way techniques. parafac," *Analytical methods*, vol. 5, 09 2013.

[8] R. Jaffé, K. Cawley, and Y. Yamashita, "Applications of excitation emission matrix fluorescence with parallel factor analysis (eem-parafac) in assessing environmental dynamics of natural dissolved organic matter (dom) in aquatic environments: A review," *ACS Symposium Series*, pp. 27–73, 06 2014.

[9] E. Acar and B. Yener, "Unsupervised multiway data analysis: A literature survey," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 1, pp. 6–20, 2009.

[10] G. Beylkin, J. Garcke, and M. J. Mohlenkamp, "Multivariate regression and machine learning with sums of separable functions," *SIAM Journal on Scientific Computing*, vol. 31, no. 3, pp. 1840–1857, 2009.

[11] S. Rendle, "Factorization machines with libfm," *ACM Trans. Intell. Syst. Technol.*, vol. 3, no. 3, pp. 57:1–57:22, May 2012.

[12] Y. Wang, R. Chen, J. Ghosh, J. C. Denny, A. Kho, Y. Chen, B. A. Malin, and J. Sun, "Rubik: Knowledge guided tensor factorization and completion for health data analytics," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '15. New York, NY, USA: Association for Computing Machinery, 2015, p. 1265–1274.

[13] J. Henderson, H. He, B. A. Malin, J. C. Denny, A. N. Kho, J. Ghosh, and J. C. Ho, "Phenotyping through semi-supervised tensor factorization (psst)," in *AMIA Annual Symposium Proceedings 2018*, 2018, p. 564–573.

[14] V. Hore, A. Viñuela, A. Buil, J. Knight, M. I. McCarthy, K. Small, and J. Marchini, "Tensor decomposition for multiple-tissue gene expression experiments," *Nature genetics*, vol. 48, no. 9, p. 1094, 2016.

[15] E. Acar, C. Aykut-Bingol, H. Bingol, R. Bro, and B. Yener, "Multiway analysis of epilepsy tensors," *Bioinformatics*, vol. 23, no. 13, pp. i10–i18, 07 2007.

[16] A. H. Williams, T. H. Kim, F. Wang, S. Vyas, S. I. Ryu, K. V. Shenoy, M. Schnitzer, T. G. Kolda, and S. Ganguli, "Unsupervised discovery of demixed, low-dimensional neural dynamics across multiple timescales through tensor component analysis," *Neuron*, vol. 98, no. 6, pp. 1099–1115.e8, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0896627318303878

[17] F. Cong, Q.-H. Lin, L.-D. Kuang, X.-F. Gong, P. Astikainen, and T. Ristaniemi, "Tensor decomposition of eeg signals: A brief review," *Journal of Neuroscience Methods*, vol. 248, pp. 59–69, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165027015001016

[18] R. Wang, S. Li, L. Cheng, M. H. Wong, and K. S. Leung, "Predicting associations among drugs, targets and diseases by tensor decomposition for drug repositioning," *BMC Bioinformatics*, vol. 20, no. 26, p. 628, 2019. [Online]. Available: https://doi.org/10.1186/s12859-019-3283-6

[19] C. I. Kanatsoulis and N. D. Sidiropoulos, "Tex-graph: Coupled tensor-matrix knowledge-graph embedding for COVID-19 drug repurposing," *CoRR*, vol. abs/2010.11367, 2020. [Online]. Available: https://arxiv.org/abs/2010.11367

[20] R. Labs, "ENSIGN Tensor Toolbox." [Online]. Available: https://www.reservoir.com/support/ensign/

[21] M. Baskaran, B. Meister, N. Vasilache, and R. Lethin, "Efficient and Scalable Computations with Sparse Tensors," in *IEEE High Performance Extreme Computing Conference*, Waltham, MA, Sep. 2012.

[22] M. Baskaran, B. Meister, and R. Lethin, "Low-overhead Load-balanced Scheduling for Sparse Tensor Computations," in *IEEE High Performance Extreme Computing Conference*, Waltham, MA, Sep. 2014.

[23] M. M. Baskaran, B. Meister, and R. Lethin, "Parallelizing and Optimizing Sparse Tensor Computations," in *Proceedings of the 28th ACM International Conference on Supercomputing*, ser. ICS '14, 2014, pp. 179–179.

[24] J. Cai, M. Baskaran, B. Meister, and R. Lethin, "Optimization of Symmetric Tensor Computations," in *IEEE High Performance Extreme Computing Conference*, Waltham, MA, Sep. 2015.

[25] M. Baskaran, T. Henretty, B. Pradelle, M. H. Langston, D. Bruns-Smith, J. Ezick, and R. Lethin, "Memory-efficient parallel tensor decompositions," in *High Performance Extreme Computing Conference (HPEC), 2017 IEEE*. IEEE, 2017, pp. 1–7.

[26] P. Jupyter, "Jupyter." [Online]. Available: https://jupyter.org

[27] R. Labs, "Tensor Completion for Cancer Drug Repositioning." [Online]. Available: https://reservoir-ensign.github.io/usecases/bioinformatics.html

[28] T. N. Jarada, J. G. Rokne, and R. Alhajj, "A review of computational drug repositioning: strategies, approaches, opportunities, challenges, and directions," *Journal of Cheminformatics*, vol. 12, no. 1, p. 46, 2020. [Online]. Available: https://doi.org/10.1186/s13321-020-00450-7

[29] K. Bleakley and Y. Yamanishi, "Supervised prediction of drug-target interactions using bipartite local models," *Bioinformatics (Oxford, England)*, vol. 25, no. 18, pp. 2397–2403, 09 2009. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/19605421

[30] S. L. Kinnings, N. Liu, N. Buchmeier, P. J. Tonge, L. Xie, and P. E. Bourne, "Drug discovery using chemical systems biology: repositioning the safe medicine comtan to treat multi-drug and extensively drug resistant tuberculosis," *PLoS computational biology*, vol. 5, no. 7, pp. e1 000 423–e1 000 423, 07 2009. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/19578428

[31] G. Hu and P. Agarwal, "Human disease-drug network based on genomic expression profiles," *PLoS one*, vol. 4, no. 8, pp. e6536–e6536, 08 2009. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/19657382

[32] C. Wu, R. C. Gudivada, B. J. Aronow, and A. G. Jegga, "Computational drug repositioning through heterogeneous network clustering," *BMC systems biology*, vol. 7 Suppl 5, no. Suppl 5, pp. S6–S6, 2013. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/24564976

[33] H. Rakshit, P. Chatterjee, and D. Roy, "A bidirectional drug repositioning approach for parkinson's disease through network-based inference," *Biochemical and biophysical research communications*, vol. 457, 01 2015.

[35] S. Hu, C. Zhang, P. Chen, P. Gu, J. Zhang, and B. Wang, "Predicting drug-target interactions from drug structure and protein sequence using

[34] A. Aliper, S. Plis, A. Artemov, A. Ulloa, P. Mamoshina, and A. Zhavoronkov, "Deep learning applications for predicting pharmacological properties of drugs and drug repurposing using transcriptomic data," *Molecular Pharmaceutics*, vol. 13, 05 2016.

novel convolutional neural networks," *BMC Bioinformatics*, vol. 20, 12 2019.

[36] M. H. S. Segler, T. Kogej, C. Tyrchan, and M. P. Waller, "Generating focused molecule libraries for drug discovery with recurrent neural networks," *ACS central science*, vol. 4, no. 1, pp. 120–131, 01 2018. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/29392184

[37] X. Zeng, S. Zhu, X. Liu, Y. Zhou, R. Nussinov, and F. Cheng, "Deepdr: A network-based deep learning approach to in silico drug repositioning," *Bioinformatics (Oxford, England)*, vol. 35, 05 2019.

[38] E. Acar, D. M. Dunlavy, T. G. Kolda, and M. Mørup, "Scalable tensor factorizations for incomplete data," *Chemometrics and Intelligent Laboratory Systems*, vol. 106, no. 1, pp. 41–56, March 2011.

[39] A. Liu and A. Moitra, "Tensor completion made practical," *CoRR*, vol. abs/2006.03134, 2020. [Online]. Available: https://arxiv.org/abs/2006.03134

[40] K. Yohem, J. Clothier, S. L. Montague, R. J. Geary, A. Winters, M. Hendrix, and D. Welch, "Inhibition of tumor cell invasion by verapamil," *Pigment cell research*, vol. 4 5-6, pp. 225–33, 1991.

[41] S. Rudnick, G. Stevenson, S. Hall, I. Espinoza-Delgado, H. Stevenson, and D. Longo, "Halothane potentiates the antitumor activity of gamma-interferon and mimics calmodulin-blocking agents," *Anesthesiology*, vol. 74, pp. 115–9, 1991.

[42] D.-S. Yu, G.-H. Sun, C. Ma, and S. Chang, "Verapamil modulation of multidrug resistance in renal cell carcinoma," *Journal of the Formosan Medical Association = Taiwan yi zhi*, vol. 99, pp. 311–6, 05 2000.

[43] D. Makower, U. Malik, Y. Novik, and P. Wiernik, "Therapeutic efficacy of theophylline in chronic lymphocytic leukemia," *Medical Oncology*, vol. 16, pp. 69–71, 1999.

[44] J. D. Carroll and J.-J. Chang, "Analysis of individual differences in multidimensional scaling via an n-way generalization of "eckart-young" decomposition," *Psychometrika*, vol. 35, pp. 283–319, 1970.

[45] R. Harshman, "Foundations of the parafac procedure: Models and conditions for an "explanatory" multi-modal factor analysis," *UCLA Working Papers in Phonetics*, vol. 16, 1970.

[46] Y. Luo, X. Zhao, J. Zhou, J. Yang, Y. Zhang, W. Kuang, J. Peng, L. Chen, and J. Zeng, "A network integration approach for drug-target interaction prediction and computational drug repositioning from heterogeneous information," *Nature Communications*, vol. 8, no. 1, p. 573, 2017. [Online]. Available: https://doi.org/10.1038/s41467-017-00680-8

[47] Y. Luo, "Dtinet," https://github.com/luoyunan/DTINet.

[48] C. Knox, V. Law, T. Jewison, P. Liu, S. Ly, A. Frolkis, A. Pon, K. Banco, C. Mak, V. Neveu, Y. Djoumbou, R. Eisner, A. C. Guo, and D. S. Wishart, "Drugbank 3.0: a comprehensive resource for 'omics' research on drugs," *Nucleic acids research*, vol. 39, no. Database issue, pp. D1035–D1041, 01 2011. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/21059682

[49] T. S. Keshava Prasad, R. Goel, K. Kandasamy, S. Keerthikumar, S. Kumar, S. Mathivanan, D. Telikicherla, R. Raju, B. Shafreen, A. Venugopal, L. Balakrishnan, A. Marimuthu, S. Banerjee, D. S. Somanathan, A. Sebastian, S. Rani, S. Ray, C. J. Harrys Kishore, S. Kanth, M. Ahmed, M. K. Kashyap, R. Mohmood, Y. L. Ramachandra, V. Krishna, B. A. Rahiman, S. Mohan, P. Ranganathan, S. Ramabadran, R. Chaerkady, and A. Pandey, "Human protein reference database–2009 update," *Nucleic acids research*, vol. 37, no. Database issue, pp. D767–D772, 01 2009. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/18988627

[50] A. P. Davis, C. G. Murphy, R. Johnson, J. M. Lay, K. Lennon-Hopkins, C. Saraceni-Richards, D. Sciaky, B. L. King, M. C. Rosenstein, T. C. Wiegers, and C. J. Mattingly, "The comparative toxicogenomics database: update 2013," *Nucleic acids research*, vol. 41, no. Database issue, pp. D1104–D1114, 01 2013. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/23093600

[51] V. N. Ioannidis, X. Song, S. Manchanda, M. Li, X. Pan, D. Zheng, X. Ning, X. Zeng, and G. Karypis, "Drkg - drug repurposing knowledge graph for covid-19," https://github.com/gnn4dr/DRKG/, 2020.

[52] B. Percha and R. Altman, "A global network of biomedical relationships derived from text," *Bioinformatics (Oxford, England)*, vol. 34, 02 2018.

[53] "Hetionet," https:het.io, accessed: 2021-07-13.

[54] "Dexamethasone in hospitalized patients with covid-19," *New England Journal of Medicine*, vol. 384, no. 8, pp. 693–704, 2021/07/15 2020. [Online]. Available: https://doi.org/10.1056/NEJMoa2021436