machine learning \& knowledge extraction

# Interpretable Topic Extraction and Word Embedding Learning Using Non-Negative Tensor DEDICOM 

Lars Hillebrand ${ }^{1,2, *,+(\mathbb{D}}$, David Biesner ${ }^{1,2, *,+(\mathbb{D}}$, Christian Bauckhage ${ }^{1,2}$ and Rafet Sifa ${ }^{1}$<br>1 Fraunhofer IAIS, 53757 Sankt Augustin, Germany; christian.bauckhage@iais.fraunhofer.de (C.B.); rafet.sifa@iais.fraunhofer.de (R.S.)<br>2 Department of Computer Science, University of Bonn, 53113 Bonn, Germany<br>* Correspondence: lars.patrick.hillebrand@iais.fraunhofer.de (L.H.); david.biesner@iais.fraunhofer.de (D.B.)<br>$\dagger$ These authors contributed equally to this work.

Citation: Hillebrand, L.; Biesner, D.; Bauckhage, C.; Sifa, R. Interpretable Topic Extraction and Word Embedding Learning Using Non-Negative Tensor DEDICOM. Mach. Learn. Knowl. Extr. 2021, 3, 123-167. https://doi.org/10.3390 /make3010007

Received: 30 November 2020
Accepted: 13 January 2021
Published: 19 January 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/).


#### Abstract

Unsupervised topic extraction is a vital step in automatically extracting concise contentual information from large text corpora. Existing topic extraction methods lack the capability of linking relations between these topics which would further help text understanding. Therefore we propose utilizing the Decomposition into Directional Components (DEDICOM) algorithm which provides a uniquely interpretable matrix factorization for symmetric and asymmetric square matrices and tensors. We constrain DEDICOM to row-stochasticity and non-negativity in order to factorize pointwise mutual information matrices and tensors of text corpora. We identify latent topic clusters and their relations within the vocabulary and simultaneously learn interpretable word embeddings. Further, we introduce multiple methods based on alternating gradient descent to efficiently train constrained DEDICOM algorithms. We evaluate the qualitative topic modeling and word embedding performance of our proposed methods on several datasets, including a novel New York Times news dataset, and demonstrate how the DEDICOM algorithm provides deeper text analysis than competing matrix factorization approaches.


Keywords: matrix factorization; tensor factorization; word embeddings; topic modeling; NLP

## 1. Introduction

Matrix factorization methods have always been a staple in many natural language processing (NLP) tasks. Factorizing a matrix of word co-occurrences can create both lowdimensional representations of the vocabulary, so-called word embeddings [1,2], that carry semantic and topical meaning within them, as well as representations of meaning that go beyond single words to latent topics.

Decomposition into Directional Components (DEDICOM) is a matrix factorization technique that factorizes a square, possibly asymmetric, matrix of relationships between items into a loading matrix of low-dimensional representations of each item and an affinity matrix describing the relationships between the dimensions of the latent representation (see Figure 1 for an illustration).

We introduce a modified row-stochastic variation of DEDICOM, which allows for interpretable loading vectors and apply it to different matrices of word co-occurrence statistics created from Wikipedia based semi-artificial text documents. Our algorithm produces low-dimensional word embeddings, where one can interpret each latent factor as a topic that clusters words into meaningful categories. Hence, we show that row-stochastic DEDICOM successfully combines the task of learning interpretable word embeddings and extracting representative topics.

We further derive a similar model for factorization of three-dimensional data tensors, which represent word co-occurrence statistics for text corpora with intrinsic structure that allows for some separation of the corpus into subsets (e.g., a news corpus structered by time).


Figure 1. (a) The DEDICOM algorithm factorizes a square matrix $S \in \mathbb{R}^{n \times n}$ into a loading matrix $A \in \mathbb{R}^{n \times k}$ and an affinity matrix $R \in \mathbb{R}^{k \times k}$. (b) The tensor DEDICOM algorithm factorizes a three dimensional tensor $\underline{S} \in \mathbb{R}^{t \times n \times n}$ into a loading matrix $A \in \mathbb{R}^{n \times k}$ and a three dimensional affinity tensor $\underline{\boldsymbol{R}} \in \mathbb{R}^{t \times k \times k}$.

An interesting aspect of this type of factorization is the interpretability of the affinity matrix. An entry in the matrix directly describes the relationship between the topics of the respective row and column and one can therefore use this tool to extract topics that a certain text corpus deals with and analyze how these topics are connected in the given text.

In this work we first describe the aforementioned DEDICOM algorithm and provide details on the modified row-stochasticity constraint and on optimization. We further expand our model to factorize three-dimensional tensors and introduce a multiplicative update rule that facilitates the training procedure. We then present results of various experiments on both semi-artificial text documents (combinations of Wikipedia articles) and real text documents (movie reviews and news articles) that show how our approach is able to capture hidden latent topics within text corpora, cluster words in a meaningful way and find relationships between these topics within the documents.

This paper is an extension of previous work [3]. In addition to the algorithms and experiments described there, we here add the extension of the DEDICOM algorithm to three-dimensional tensors, introduce a multiplicative update rule to increase training stability and present new experiments on two additional text corpora (Amazon reviews and New York Times news articles).

## 2. Related Work

Matrix factorization describes the task of compressing the most relevant information from a high-dimensional input matrix into multiple low-dimensional factor matrices, with either approximate or exact input reconstruction (see for example [4] for a theoretical overview of common methods and their applications). In this work we consider the DEDICOM algorithm, which has a long history of providing an interpretable matrix or tensor factorization, mostly for rather low-dimensional tasks.

First described in [5], it since has been applied to analysis of social networks [6], email correspondence [7] and video game player behavior [8,9]. DEDICOM also has successfully been employed to NLP tasks such as part of speech tagging [10], however to the best of our knowledge we provide the first implementation of DEDICOM for simultaneous word embedding learning and topic modeling.

Many works deal with the task of putting constraints on the factor matrices of the DEDICOM algorithm. In [7,8], the authors constrain the affinity matrix $R$ to be nonnegative, which aids interpretability and improves convergence behavior if the matrix to be factorized is non-negative. However, their approach relies on the Kronecker product between matrices in the update step, solving a linear system of $n^{2} \times k^{2}$, where $n$ denotes the number of items in the input matrix and $k$ the number of latent factors. These dimensions make the application on text data, where $n$ describes the number of words in the vocabulary, a computationally futile task. Constraints on the loading matrix, $A$, include non-negativity as well (see [7]) or column-orthogonality as in [8].

In contrast, we propose a new modified row-stochasticity constraint on $A$, which is tailored to generate interpretable word embeddings that carry semantic meaning and represent a probability distribution over latent topics.

The DEDICOM algorithm has previously been applied to tensor data as well, for example in [11], in which the authors apply the algorithm on general multirelational data by computing an exact solution for the affinity matrix. Both [6,7] explore a slight variation of our tensor DEDICOM approach to analyze relations in email data and [12] apply a similar model on non-square input tensors.

Previous matrix factorization based methods in the NLP context mostly dealt with either word embedding learning or topic modeling, but not with both tasks combined.

For word embeddings, the GloVe [2] model factorizes an adjusted co-occurrence matrix into two matrices of the same dimension. The work is based on a large text corpus with a vocabulary of $n \approx 400,000$ and produces word embeddings of dimension $k=300$. In order to maximize performance on the word analogy task, the authors adjusted the cooccurrence matrix to the logarithmized co-occurrence matrix and added bias terms to the optimization objective.

A model conceived around the same time, word2vec [13], calculates word embeddings not from a co-occurrence matrix but directly from the text corpus using the skip-gram or continuous-bag-of-words approach. More recent work [1] has shown that this construction is equivalent to matrix factorization on the pointwise mutual information (PMI) matrix of the text corpus, which makes it very similar to the glove model described above.

Both models achieve impressive results on word embedding related tasks like word analogy, however the large dimensionality of the word embeddings makes interpreting the latent factors of the embeddings impossible.

On the topic modeling side, matrix factorization methods are routinely applied as well. Popular algorithms like non-negative matrix factorization (NMF) [14], singular value decomposition (SVD) [15,16] and principal component analysis (PCA) [17] compete against the probabilistic latent dirichlet allocation (LDA) [18] to cluster the vocabulary of a word co-occurrence or document-term matrix into latent topics. (More recent expansions of these methods can be found in $[19,20]$.) Yet, we empirically show that the implicitly learned word embeddings of these methods lack semantic meaning in terms of the cosine similarity measure.

We benchmark our approach qualitatively against these methods in Section 4.3 and the Appendixes A and B.

## 3. Constrained DEDICOM Models

In this section we provide a detailed theoretical view at different constrained DEDICOM algorithms utilized for factorizing word co-occurrence based positive pointwise mutual information matrices and tensors.

We first consider the case of a two-dimensional input matrix $S$ (see Figure 1a) in Section 3.1. We then present an extension of the algorithm for three-dimension input tensors $\underline{S}$ (see Figure 1b) in Section 3.2. Finally we derive a multiplicative update rule for non-negative tensor DEDICOM.

### 3.1. The Row-Stochastic DEDICOM Model for Matrices

For a given language corpus consisting of $n$ unique words $X=x_{1}, \ldots, x_{n}$ we calculate a co-occurrence matrix $\boldsymbol{W} \in \mathbb{R}^{n \times n}$ by iterating over the corpus on a word token level with a sliding context window of specified size. Then

$$
\begin{equation*}
\boldsymbol{W}_{i j}=\# \text { word } i \text { appears in context of word } j . \tag{1}
\end{equation*}
$$

Note that the word context window can be applied symmetrically or asymmetrically around each word. We choose a symmetric context window, which implies a symmetric co-occurrence matrix, $\boldsymbol{W}_{i j}=\boldsymbol{W}_{j i}$.

We then transform the co-occurrence matrix into the pointwise mutual information matrix (PMI), which normalizes the counts in order to extract meaningful co-occurrences from the matrix. Co-occurrences of words that occur regularly in the corpus are decreased since their appearance together might be nothing more than a statistical phenomenon, the
co-occurrence of words that appear less often in the corpus give us meaningful information about the relations between words and topics. We define the PMI matrix as

$$
\begin{equation*}
\mathbf{P M I}_{i j}:=\log W_{i j}+\log N-\log N_{i}-\log N_{j} \tag{2}
\end{equation*}
$$

where $N:=\sum_{i j=1}^{n} \boldsymbol{W}_{i j}$ is the sum of all co-occurrence counts of $\boldsymbol{W}, N_{i}:=\sum_{j=1}^{n} \boldsymbol{W}_{i j}$ the row sum and $N_{j}:=\sum_{i=1}^{n} W_{i j}$ the column sum.

Since the co-occurrence matrix $W$ is symmetrical, the transformed PMI matrix is symmetrical as well. Nevertheless, DEDICOM is able to factorize both symmetrical and non-symmetrical matrices. We expand details on symmetrical and non-symmetrical relationships in Section 3.4.

Additionally, we want all entries of the matrix to be non-negative, our final matrix to be factorized is therefore the positive PMI (PPMI)

$$
\begin{equation*}
S_{i j}=\mathbf{P P M I}_{i j}=\max \left\{0, \mathbf{P M I}_{i j}\right\} \tag{3}
\end{equation*}
$$

Our aim is to decompose this matrix using row-stochastic DEDICOM as

$$
\begin{equation*}
\boldsymbol{S} \approx \boldsymbol{A R} \boldsymbol{A}^{T}, \quad \text { with } \quad \boldsymbol{S}_{i j} \approx \sum_{b=1}^{k} \sum_{c=1}^{k} \boldsymbol{A}_{i b} \boldsymbol{R}_{b c} \boldsymbol{A}_{j c} \tag{4}
\end{equation*}
$$

where $A \in \mathbb{R}^{n \times k}, \boldsymbol{R} \in \mathbb{R}^{k \times k}, A^{T}$ denotes the transpose of $A$ and $k \ll n$. Literature often refers to $A$ as the loading matrix and $R$ as the affinity matrix. $A$ gives us for each word $i$ in the vocabulary a vector of size $k$, the number of latent topics we wish to extract. The square matrix $\boldsymbol{R}$ then provides possibility for interpretation of the relationships between these topics.

Empirical evidence has shown that the algorithm tends to favor columns unevenly, such that a single column receives a lot more weight in its entries than the other columns. We try to balance this behavior by applying a column-wise z-normalization on $A$, such that all columns have zero mean and unit variance.

In order to aid interpretability we wish each word embedding to be a distribution over all latent topics, i.e., entry $A_{i b}$ in the word-embedding matrix provides information on how much topic $b$ describes word $i$.

To implement these constraints we therefore apply a row-wise softmax operation over the column-wise z-normalized $A$ matrix by defining $A^{\prime} \in \mathbb{R}^{n \times k}$ as

$$
\begin{align*}
& A_{i b}^{\prime}:=\frac{\exp \left(\bar{A}_{i b}\right)}{\sum_{b^{\prime}=1}^{k} \exp \left(\bar{A}_{i b^{\prime}}\right)}, \quad \bar{A}_{i b}:=\frac{A_{i b}-\mu_{b}}{\sigma_{b}},  \tag{5}\\
& \mu_{b}:=\frac{1}{n} \sum_{i=1}^{n} A_{i b}, \quad \sigma_{b}:=\sqrt{\frac{1}{n} \sum_{i=1}^{n}\left(A_{i b}-\mu_{b}\right)^{2}}
\end{align*}
$$

and optimizing $A$ for the objective

$$
\begin{equation*}
\boldsymbol{S} \approx \boldsymbol{A}^{\prime} \boldsymbol{R}\left(\boldsymbol{A}^{\prime}\right)^{T} \tag{6}
\end{equation*}
$$

Note that after applying the row-wise softmax operation all entries of $A^{\prime}$ are nonnegative.

To judge the quality of the approximation (6) we apply the Frobenius norm, which measures the difference between $S$ and $\boldsymbol{A}^{\prime} \boldsymbol{R}\left(\boldsymbol{A}^{\prime}\right)^{T}$. The final loss function we optimize our model for is therefore given by

$$
\begin{align*}
\mathcal{L}(\boldsymbol{S}, \boldsymbol{A}, \boldsymbol{R}) & =\left\|\boldsymbol{S}-\boldsymbol{A}^{\prime} \boldsymbol{R}\left(\boldsymbol{A}^{\prime}\right)^{T}\right\|_{F}^{2}  \tag{7}\\
& =\sum_{i=1}^{n} \sum_{j=1}^{n}\left(\boldsymbol{s}_{i j}-\left(\boldsymbol{A}^{\prime} \boldsymbol{R}\left(\boldsymbol{A}^{\prime}\right)^{T}\right)_{i j}\right)^{2} \tag{8}
\end{align*}
$$

with

$$
\begin{equation*}
\left(\boldsymbol{A}^{\prime} \boldsymbol{R}\left(\boldsymbol{A}^{\prime}\right)^{T}\right)_{i j}=\sum_{b=1}^{k} \sum_{c=1}^{k} A_{i b}^{\prime} \boldsymbol{R}_{b c} \boldsymbol{A}_{j c}^{\prime} \tag{9}
\end{equation*}
$$

and $\boldsymbol{A}^{\prime}$ defined in (5).
To optimize the loss function we train both matrices using alternating gradient descent similar to [8]. Within each optimization step we apply

$$
\begin{array}{lll}
A \leftarrow \boldsymbol{A}-f_{\theta}\left(\nabla_{A}, \eta^{\boldsymbol{A}}\right), & \text { where } & \nabla_{A}=\frac{\partial \mathcal{L}(S, A, \boldsymbol{R})}{\partial A} \\
\boldsymbol{R} \leftarrow \boldsymbol{R}-f_{\theta}\left(\nabla_{\boldsymbol{R}}, \eta^{\boldsymbol{R}}\right), & \text { where } & \nabla_{\boldsymbol{R}}=\frac{\partial \mathcal{L}(\boldsymbol{S}, \boldsymbol{A}, \boldsymbol{R})}{\partial \boldsymbol{R}} \tag{11}
\end{array}
$$

with $\eta^{A}, \eta^{R}>0$ being individual learning rates for both matrices and $f_{\theta}(\cdot)$ representing an arbitrary gradient based update rule with additional hyperparameters $\theta$. For our experiments we employ automatic differentiation methods. For details on the implementation of the algorithm above refer to Section 4.2.

### 3.2. The Constrained DEDICOM Model for Tensors

In this section we extend the model described above to three-dimensional tensors as input data. As above, the input describes the co-occurrences of vocabulary items in a text corpus. However, we consider additionally structured text: Instead of one matrix describing the entire corpus we unite multiple $n \times n$ matrices of co-occurrences into one tensor $\underline{S} \in \mathbb{R}^{t \times n \times n}$. Each of the $t$ slices then consists of an adjusted PPMI matrix for a subset of the text corpus. This structure could originate for instance from different data (e.g., different Wikipedia articles), different topical subsets of the data source (e.g., reviews for different articles) or describe time-slices (e.g., news articles for certain time periods).

To construct the PPMI tensor we again take a vocabulary $X=x_{1}, \ldots, x_{n}$ over the entire corpus. For each subset $l$ we then calculate a co-occurrence matrix $\underline{W}_{l} \in \mathbb{R}^{n \times n}$ as described above. Stacking these matrices yields the co-occurrence tensor $\underline{\boldsymbol{W}} \in \mathbb{R}^{t \times n \times n}$.

When transforming slice $\underline{\boldsymbol{W}}_{l}$ into a PMI matrix we want to apply information from the entire corpus. We therefore do not only calculate the column, the row and the total sums on the corresponding subset but on the entire text corpus. Therefore

$$
\begin{equation*}
\underline{\mathbf{P M I}}_{i j}:=\log \underline{W}_{l i j}+\log N-\log N_{i}-\log N_{j} \tag{12}
\end{equation*}
$$

where $N:=\sum_{l=1}^{k} \sum_{i j=1}^{n} \underline{\boldsymbol{W}}_{l i j}$ is the sum of all co-occurrence counts of $\underline{\boldsymbol{W}}, N_{i}:=\sum_{l=1}^{k} \sum_{j=1}^{n} \underline{\boldsymbol{W}}_{l i j}$ the row sum and $N_{j}:=\sum_{l=1}^{k} \sum_{i=1}^{n} \underline{W}_{l i j}$ the column sum.

Finally we define the positive pointwise mutual information tensor as

$$
\begin{equation*}
\underline{S}_{l i j}=\underline{\mathbf{P P M I}}_{i j}=\max \left\{0, \underline{\mathbf{P M I}}_{i j}\right\} . \tag{13}
\end{equation*}
$$

We decompose this input tensor into a matrix $A \in \mathbb{R}^{n \times k}$ and a tensor $\underline{R} \in \mathbb{R}^{t \times k \times k}$, such that

$$
\begin{equation*}
\underline{S} \approx A \underline{R} A^{T} \tag{14}
\end{equation*}
$$

where we multiply each slice of $\underline{R}$ with $A$ and $A^{T}$ to reconstruct the corresponding slice of $\underline{S}$ :

$$
\begin{equation*}
\underline{\boldsymbol{S}}_{l i j} \approx \sum_{b=1}^{k} \sum_{c=1}^{k} \boldsymbol{A}_{i b} \underline{\boldsymbol{R}}_{l b c} \boldsymbol{A}_{j c} \tag{15}
\end{equation*}
$$

We keep our naming convention for $A$ as the loading matrix and $\underline{R}$ as the affinity tensor, since again $A$ gives us for each word $i$ in the vocabulary a vector of size $k$ and for each slice $l$ the square matrix $\underline{\boldsymbol{R}}_{l}:=\left(\underline{\boldsymbol{R}}_{l i j}\right)_{i, j=1}^{k}$ provides information on the relationships between the topics in the $l$-th input slice.

Analogous to (7) we construct a loss function

$$
\begin{align*}
\mathcal{L}(\underline{\boldsymbol{S}}, \boldsymbol{A}, \underline{\boldsymbol{R}}) & =\left\|\underline{\boldsymbol{S}}-\boldsymbol{A} \underline{\boldsymbol{R}}(\boldsymbol{A})^{T}\right\|_{F}^{2}  \tag{16}\\
& =\sum_{l=1}^{t} \sum_{i=1}^{n} \sum_{j=1}^{n}\left(\underline{\boldsymbol{S}}_{l i j}-\left(\boldsymbol{A} \underline{\boldsymbol{R}}_{l}(\boldsymbol{A})^{T}\right)_{i j}\right)^{2}  \tag{17}\\
& =\sum_{l=1}^{t} \mathcal{L}\left(\underline{\boldsymbol{S}}_{l}, A, \underline{\boldsymbol{R}}_{l}\right) \tag{18}
\end{align*}
$$

with

$$
\begin{equation*}
\left(\boldsymbol{A} \underline{\boldsymbol{R}}_{l}(\boldsymbol{A})^{T}\right)_{i j}=\sum_{b=1}^{k} \sum_{c=1}^{k} \boldsymbol{A}_{i b} \underline{\boldsymbol{R}}_{l b c} \boldsymbol{A}_{j c} \tag{19}
\end{equation*}
$$

Note that in this framework, the DEDICOM algorithm described in the previous section is equivalent to tensor DEDICOM with $t=1$.

Update steps can then be taken via alternating gradient descent on $A$ and $\underline{\boldsymbol{R}}$. As in the previous section, one can now add additional constraints to $A$ and $\underline{R}$ and calculate the gradients as in (10), using automatic differentiation methods. Taking update steps of size $\eta^{A}$ and $\eta^{\underline{R}}$ respectively leads to an eventual convergence to some local or global minimum of the loss (16) with respect to the original or constrained $A$ and $\underline{\boldsymbol{R}}$.

Alternatively, constraints can be added to $A$ and $\underline{R}$ by methods like projected gradient descent and the Frank-Wolfe algorithm [21] which either adjust the respective matrix or tensor to be constrained after the gradient step or apply the general gradient step such that the matrix or tensor never leaves the constrained area.

However, empirical results show that automatic differentiation methods lead to slow and unstable training convergence and worse qualitative results when applying the mentioned constraints on the factor matrices and tensors. We therefore derive an alternative method of applying alternating gradient descent to $A$ and $\underline{R}$ based on multiplicative update rules. This does not only improve training stability and convergence behavior but also lead to better qualitative results (see Section 4.3 and Figure 2).

We derive the gradients for $A$ and $\underline{R}$ analytically and set the learning rates $\eta^{A}$ and $\eta \underline{R}$ individually for each element $(i, j)$ as $\eta_{i j}^{A}$ for matrix $A$ and for each element $(l, i, j)$ as $\eta_{l i j}^{\frac{R}{i}}$ for tensor $\underline{\boldsymbol{R}}$, such that the resulting update step is an element-wise multiplication of the respective matrix or tensor.

We derive the updates for the matrix algorithm first and later extend them for the tensor case. For detailed derivations refer to Appendix B. For $A$ we derive the gradient analytically as

$$
\begin{equation*}
\frac{\partial \mathcal{L}(S, A, R)}{\partial A}=-2\left(S A R^{T}+S^{T} A R-A\left(R A^{T} A R^{T}+R^{T} A^{T} A R\right)\right) \tag{20}
\end{equation*}
$$

Therefore the update step is

$$
\begin{equation*}
\boldsymbol{A}_{i j} \leftarrow \boldsymbol{A}_{i j}+\eta_{i j}^{A} 2\left(\left[S \boldsymbol{A} \boldsymbol{R}^{T}+\boldsymbol{S}^{T} \boldsymbol{A} \boldsymbol{R}\right]_{i j}-\left[\boldsymbol{A}\left(\boldsymbol{R} \boldsymbol{A}^{T} \boldsymbol{A} \boldsymbol{R}^{T}+\boldsymbol{R}^{T} \boldsymbol{A}^{T} \boldsymbol{A R}\right)\right]_{i j}\right) \tag{21}
\end{equation*}
$$



Figure 2. Reconstruction loss development during tensor factorization training. The $x$-axis plots the number of epochs on a logarithmic scale, the $y$-axis plots the corresponding reconstruction error for each method.

If we now chose $\eta^{A}$ as

$$
\begin{equation*}
\eta_{i j}^{A}:=\frac{\boldsymbol{A}_{i j}}{2\left[\boldsymbol{A}\left(\boldsymbol{R}^{T} \boldsymbol{A}^{T} \boldsymbol{A R}+\boldsymbol{R} \boldsymbol{A}^{T} \boldsymbol{A} \boldsymbol{R}^{T}\right)\right]_{i j}} \tag{22}
\end{equation*}
$$

the update (21) becomes

$$
\begin{equation*}
\boldsymbol{A}_{i j} \leftarrow \boldsymbol{A}_{i j} \frac{\left[\boldsymbol{S}^{T} \boldsymbol{A R}+\boldsymbol{S} \boldsymbol{A} \boldsymbol{R}^{T}\right]_{i j}}{\left[\boldsymbol{A}\left(\boldsymbol{R}^{T} \boldsymbol{A}^{T} \boldsymbol{A R}+\boldsymbol{R A}^{T} \boldsymbol{A} \boldsymbol{R}^{T}\right)\right]_{i j}} . \tag{23}
\end{equation*}
$$

For $\boldsymbol{R}$ we derive the gradient analytically as

$$
\begin{equation*}
\frac{\partial \mathcal{L}(S, A, R)}{\partial R}=-2\left(A^{T} S A-A^{T} A R A^{T} A\right) \tag{24}
\end{equation*}
$$

Therefore the update step is

$$
\begin{equation*}
\boldsymbol{R}_{i j}=\boldsymbol{R}_{i j}+\eta_{i j}^{R} 2\left(\left[\boldsymbol{A}^{T} \boldsymbol{S} \boldsymbol{A}\right]_{i j}-\left[\boldsymbol{A}^{T} \boldsymbol{A R} \boldsymbol{A}^{T} \boldsymbol{A}\right]_{i j}\right) \tag{25}
\end{equation*}
$$

Choose

$$
\begin{equation*}
\eta_{i j}^{R}:=\frac{\boldsymbol{R}_{i j}}{2\left[A^{T} A R A^{T} A\right]_{i j}} \tag{26}
\end{equation*}
$$

and the update (25) becomes

$$
\begin{equation*}
\boldsymbol{R}_{i j} \leftarrow \boldsymbol{R}_{i j} \frac{\left[\boldsymbol{A}^{T} \boldsymbol{S} \boldsymbol{A}\right]_{i j}}{\left[\boldsymbol{A}^{T} \boldsymbol{A R} \boldsymbol{A}^{T} A\right]_{i j}} \tag{27}
\end{equation*}
$$

Since $S_{i j} \geq 0$ for all $i, j$, in both (23) and (27) each element of the multiplier matrix is positive if both $A \geq 0$ and $R \geq 0$ in all entries. Therefore, initializing both matrices with positive values results in an update step that keeps the elements of $A$ and $R$ positive.

To extend this rule to tensor DEDICOM, consider that the analytical derivatives hold for $\underline{R}$ and $A$ by considering each slice $\underline{S}_{l}$ and $\underline{R}_{l}$ individually:

$$
\begin{align*}
& \frac{\partial \mathcal{L}\left(\underline{\boldsymbol{S}}_{l}, A, \underline{\boldsymbol{R}}_{l}\right)}{\partial \underline{\boldsymbol{R}}_{l}}=-2\left(\boldsymbol{A}^{T} \underline{\boldsymbol{S}}_{l} A-A^{T} A \underline{\boldsymbol{R}}_{l} \boldsymbol{A}^{T} A\right)  \tag{28}\\
& \frac{\partial \mathcal{L}\left(\underline{\boldsymbol{S}}_{l}, A, \underline{\boldsymbol{R}}_{l}\right)}{\partial A}=-2\left(\underline{\boldsymbol{S}}_{l}^{T} A \underline{\boldsymbol{R}}_{l}+\underline{\boldsymbol{S}}_{l} A \underline{\boldsymbol{R}}_{l}^{T}-A\left(\underline{\boldsymbol{R}}_{l}^{T} \boldsymbol{A}^{T} A \underline{\boldsymbol{R}}_{l}+\underline{\boldsymbol{R}}_{l} \boldsymbol{A}^{T} A \underline{\boldsymbol{R}}_{l}^{T}\right)\right) \tag{29}
\end{align*}
$$

Since by (18) we have $\mathcal{L}(\underline{\boldsymbol{S}}, \boldsymbol{A}, \underline{\boldsymbol{R}})=\sum_{l=1}^{t} \mathcal{L}\left(\underline{\boldsymbol{S}}_{l}, A, \underline{\boldsymbol{R}}_{l}\right)$ we can derive the full gradients as

$$
\begin{align*}
& \frac{\partial \mathcal{L}(\underline{\boldsymbol{S}}, A, \underline{\boldsymbol{R}})}{\partial \underline{\boldsymbol{R}}}=\left(A^{T} \underline{\boldsymbol{S}} A-A^{T} A \underline{\boldsymbol{R}} A^{T} A\right)  \tag{30}\\
& \frac{\partial \mathcal{L}(\underline{\boldsymbol{S}}, A, \underline{\boldsymbol{R}})}{\partial A}=\sum_{l=1}^{t}-2\left(\underline{S}_{l}^{T} A \underline{\boldsymbol{R}}_{l}+\underline{\boldsymbol{S}}_{l} A \underline{\boldsymbol{R}}_{l}^{T}-A\left(\underline{\boldsymbol{R}}_{l}^{T} A^{T} A \underline{\boldsymbol{R}}_{l}+\underline{\boldsymbol{R}}_{l} A^{T} A \underline{\boldsymbol{R}}_{l}^{T}\right)\right) \tag{31}
\end{align*}
$$

For $A$ we set $\eta^{A}$ as

$$
\begin{equation*}
\eta_{i j}^{A}:=\frac{\boldsymbol{A}_{i j}}{2 \sum_{l=1}^{t}\left[\boldsymbol{A}\left(\underline{\boldsymbol{R}}_{l}^{T} \boldsymbol{A}^{T} \boldsymbol{A} \underline{\boldsymbol{R}}_{l}+\underline{\boldsymbol{R}}_{l} \boldsymbol{A}^{T} A \underline{\boldsymbol{R}}_{l}^{T}\right)\right]_{i j}} \tag{32}
\end{equation*}
$$

Then the update step is

$$
\begin{align*}
A_{i j} & \leftarrow A_{i j}-\eta_{i j}^{A} \frac{\partial \mathcal{L}(\underline{\boldsymbol{S}}, \boldsymbol{A}, \underline{\boldsymbol{R}})}{\partial A}  \tag{33}\\
& =A_{i j} \frac{\sum_{l=1}^{t}\left[\underline{\boldsymbol{S}}_{l}^{T} \boldsymbol{A} \underline{\boldsymbol{R}}_{l}+\underline{\boldsymbol{S}}_{l} A \underline{\boldsymbol{R}}_{l}^{T}\right]}{\sum_{l=1}^{t}\left[\boldsymbol{A}\left(\underline{\boldsymbol{R}}_{l}^{T} \boldsymbol{A}^{T} A \underline{\boldsymbol{R}}_{l}+\underline{\boldsymbol{R}}_{l} \boldsymbol{A}^{T} A \underline{\boldsymbol{R}}_{l}^{T}\right)\right]_{i j}} . \tag{34}
\end{align*}
$$

For $\underline{R}$ we again set

$$
\begin{equation*}
\eta_{l i j}^{\underline{R}}:=\frac{\underline{\boldsymbol{R}}_{l i j}}{2\left[A^{T} A \underline{\boldsymbol{R}}_{l} A^{T} A\right]_{i j}} \tag{35}
\end{equation*}
$$

and the update (25) becomes

$$
\begin{equation*}
\underline{\boldsymbol{R}}_{l i j} \leftarrow \underline{\boldsymbol{R}}_{l i j} \frac{\left[\boldsymbol{A}^{T} \underline{\boldsymbol{S}}_{l} A\right]_{i j}}{\left[\boldsymbol{A}^{T} A \underline{\boldsymbol{R}}_{l} \boldsymbol{A}^{T} \boldsymbol{A}\right]_{i j}} \tag{36}
\end{equation*}
$$

Equations (23) and (27) provide multiplicative update rules that ensure the nonnegativity of $\boldsymbol{A}$ and $\boldsymbol{R}$ without any additional constraints. Equations (33) and (36) provide the corresponding rules for matrix $A$ and tensor $\underline{R}$ in tensor DEDICOM.

### 3.3. On Symmetry

The DEDICOM algorithm is able to factorize both symmetrical and asymmetrical matrices $S$. For a given matrix $A$, the symmetry of $R$ dictates the symmetry of the product $A R A^{T}$, since

$$
\begin{align*}
\left(\boldsymbol{A R} \boldsymbol{A}^{T}\right)_{i j} & =\sum_{b=1}^{k} \sum_{c=1}^{k} \boldsymbol{A}_{i b} \boldsymbol{R}_{b c} \boldsymbol{A}_{j c}=\sum_{b=1}^{k} \sum_{c=1}^{k} \boldsymbol{A}_{i b} \boldsymbol{R}_{c b} \boldsymbol{A}_{j c}  \tag{37}\\
& =\sum_{c=1}^{k} \sum_{b=1}^{k} \boldsymbol{A}_{j c} \boldsymbol{R}_{c b} \boldsymbol{A}_{i b}=\left(\boldsymbol{A R} \boldsymbol{A}^{T}\right)_{j i} \tag{38}
\end{align*}
$$

if $\boldsymbol{R}_{c b}=\boldsymbol{R}_{b c}$ for all $b, c$. We therefore expect a symmetric matrix $S$ to be decomposed into $A R A^{T}$ with a symmetric $R$, which is confirmed by our experiments. Factorizing a non-symmetric matrix leads to a non-symmetric $\boldsymbol{R}$, the asymmetric relation between items leads to asymmetric relations between the latent factors. The same relations hold for each slice $\underline{S}_{l}$ and $\underline{R}_{l}$ in tensor DEDICOM.

### 3.4. On Interpretability

We have

$$
\begin{equation*}
\boldsymbol{S}_{i j} \approx \sum_{b=1}^{k} \sum_{c=1}^{k} \boldsymbol{A}_{i b} \boldsymbol{R}_{b c} \boldsymbol{A}_{j c} \tag{39}
\end{equation*}
$$

i.e., we can estimate the probability of co-occurrence of two words $w_{i}$ and $w_{j}$ from the word embeddings $A_{i}$ and $A_{j}$ and the matrix $R$, where $A_{i}$ denotes the $i$-th row of $A$.

If we want to predict the co-occurrence between words $w_{i}$ and $w_{j}$ we consider the latent topics that make up the word embeddings $A_{i}$ and $A_{j}$, and sum up each component from $A_{i}$ with each component $A_{j}$ with respect to the relationship weights given in $\boldsymbol{R}$.

Two words are likely to have a high co-occurrence if their word embeddings have larger weights in topics that are positively connected by the $\boldsymbol{R}$ matrix. Likewise a negative entry $\boldsymbol{R}_{b, c}$ makes it less likely for words with high weight in the topics $b$ and $c$ to occur in the same context. See Figure 3 for an illustrated example.


Figure 3. The affinity matrix $\boldsymbol{R}$ describes the relationships between the latent factors. Illustrated here are two word embeddings, corresponding to the words $w_{i}$ and $w_{j}$. Darker shades represent larger values. In this example we predict a large co-occurrence at $S_{i i}$ and $S_{j j}$ because of the large weight on the diagonal of the $\boldsymbol{R}$ matrix. We predict a low co-occurrence at $S_{i j}$ and $S_{j i}$ since the large weights on $\boldsymbol{A}_{i 1}$ and $\boldsymbol{A}_{j 3}$ interact with low weights on $\boldsymbol{R}_{13}$ and $\boldsymbol{R}_{31}$.

Having an interpretable embedding model provides value beyond analysis of the affinity matrix of a single document. The worth of word embeddings is generally measured in their usefulness for downstream tasks. Given a prediction model based on word embeddings as one of the inputs, further analysis of the model behavior is facilitated when latent input dimensions easily translate to semantic meaning.

In most word embedding models, the embedding vector of a single word is not particularly useful in itself. The information only lies in its relationship (i.e., closeness or cosine similarity) to other embedding vectors. For example, an analysis of the change of word embeddings and therefore the change of word meaning within a document corpus (for example a news article corpus) can only show how various words form different clusters or drift apart over time. Interpretabilty of latent dimensions would provide tools to also consider the development of single words within the given topics.

All considerations above hold for the three-dimensional tensor case, in which we analyze a slice $\underline{R}_{l}$ together with the common word embedding matrix $A$ to gain insight into the input data slice $\underline{S}_{l}$.

## 4. Experiments and Results

In the following section we describe our experimental setup in full detail (Our Python implementation to reproduce the results is available on https:/ / github.com/LarsHill/ text-dedicom-paper. Additionally, we provide a snapshot of our versions of the applied public datasets (Wikipedia articles and Amazon reviews). ) and present our results on the simultaneous topic (relation) extraction and word embedding learning task. We compare these results against competing matrix and tensor factorization methods for topic modeling, namely NMF (including a Tucker-2 variation compatible with tensors), LDA and SVD.

### 4.1. Data

We conducted our experiments on three orthogonal text datasets which cover different text domains and allow for a thorough empirical analysis of our proposed methods.

The first corpus leveraged triplets of individual Wikipedia articles. The articles were retrieved as raw text via the official Wikipedia API using the wikipedia-api library. We differentiated between thematically similar (e.g., "Dolphin" and "Whale") and thematically different articles (e.g., "Soccer" and "Donald Trump"). Each article triplet was categorized
into one of three classes: All underlying Wikipedia articles were thematically different, two articles were thematically similar and one was different, and all articles were thematically similar. The previous paper [3] contained an extensive evaluation over 12 triples of articles in the supplementary material. In this work we focused on the three triples described in the previous main paper, namely

1. "Soccer", "Bee", "Johnny Depp",
2. "Dolphin", "Shark", "Whale", and
3. "Soccer", "Tennis", "Rugby".

Depending on whether the article triplets were represented as input matrix or tensor they were processed differently. In the case of a matrix input all three articles got concatenated to form a new artificially generated document. In the case of a tensor input the articles remained individual documents later which later represented slices in the tensor representation.

To analyze the topic extraction capability of constrained DEDICOM also on text which was rather prone to grammatical and syntactical errors, we utilized a subset of the Amazon review dataset [22]. In particular, we restricted ourselves to the "movie" product category and created a corpus consisting of six text documents holding the concatenated reviews from the following films respectively, "Toy Story 1", "Toy Story 3", "Frozen", "Monsters, Inc.", "Kung Fu Panda" and "Kung Fu Panda 2". Grouping the reviews by movie affiliation enabled us to generate a tensor representation of the corpus which we factorized via nonnegative tensor DEDICOM to analyze topic relations across movies. Table 1 lists the number of reviews per movie and shows that based on review count "Kung Fu Panda 1" was the most popular among the six films.

The third corpus represented a complete collection of New York Times news articles ranging from 1st September 2019 to 31st August 2020. The articles were taken from the New York Times website and covered a wide range of sections (see Table 2).

Instead of grouping the articles by section we binned and concatenated them by month yielding 12 news documents containing monthly information (see Table 3 for details on the article count per month). Thereby, the factorization of tensor DEDICOM allowed for an analysis of topic relations and their changes over time.

Table 1. Amazon movie review corpus grouped by movie and number of reviews per slice of input tensor.

| Movie | \# Reviews |
| :---: | :---: |
| Toy Story 1 | 2491 |
| Monsters, Inc. | 3203 |
| Kung Fu Panda 1 | 6708 |
| Toy Story 3 | 1209 |
| Kung Fu Panda 2 | 1208 |
| Frozen | 1292 |

Before transforming the text documents to matrix or tensor representations we applied the following textual preprocessing steps. First, the whole text was made lower-cased. Second, we tokenized the text making use of the word-tokenizer from the nltk library and removed common English stop words, including contractions such as "you're" and "we'll". Lastly we cleared the text from all remaining punctuation and deleted digits, single characters and multi-spaces (see Table 4 for an overview of corpora statistics after preprocessing).

Table 2. New York Times news corpus composition by section and number of articles.

| Section | \# Articles |
| :---: | :---: |
| Politics | 3204 |
| U.S. | 2610 |
| Business | 1624 |
| New York | 1528 |
| Europe | 988 |
| Asia Pacific | 839 |
| Health | 598 |
| Technology | 551 |
| Middle East | 443 |
| Science | 440 |
| Economy | 339 |
| Elections | 240 |
| Climate | 239 |
| World | 233 |
| Africa | 124 |
| Australia | 113 |
| Canada | 104 |

Table 3. New York Times news corpus grouped by month and number of articles. This corresponds to the number of articles per slice of input tensor.

| Month | \# Articles |
| :---: | :---: |
| September 2019 | 1586 |
| October 2019 | 1788 |
| November 2019 | 1623 |
| December 2019 | 1461 |
| January 2020 | 1725 |
| Febuary 2020 | 1602 |
| March 2020 | 1937 |
| April 2020 | 1712 |
| May 2020 | 1713 |
| June 2020 | 1828 |
| July 2020 | 1814 |
| August 2020 | 1886 |

Table 4. Overview of word count statistics after preprocessing for all datasets. Columns represent from left to right the total number of words per corpus, total number of unique words per corpus, average number of total words per article, average number of unique words per article and the cutoff frequency of the 10,000 th most common word. Wikipedia article combinations: Dolphin, Shark, Whale (DSW), Soccer, Bee, Johnny Depp (SBJ), Soccer, Tennis, Rugby (STR).

|  | Total | Unique | Avg. Total | Avg Unique | Cutoff |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Amazon Reviews | 252,400 | 15,560 | 15.2 | 13.4 | 1 |
| Wikipedia DSW | 14,500 | 4376 | 4833.3 | 2106.0 | 1 |
| Wikipedia SBJ | 10,435 | 4034 | 3478.3 | 1600.3 | 1 |
| Wikipedia STR | 11,501 | 3224 | 3833.7 | 1408.0 | 1 |
| New York Times | $12,043,205$ | 141,591 | 582.5 | 366.5 | 118 |

Next, we utilized all preprocessed documents in a corpus to extract a fixed size vocabulary of $n=10,000$ most frequent tokens. Since our dense input tensor was of dimensionality $t \times n \times n$, a larger vocabulary size led to a significant increase in memory consumption. Based on the total number of unique corpora words reported in Table 4, a maximum vocabulary size of $n=10,000$ was reasonable for the three Wikipedia corpora and the Amazon
reviews corpus. Only the New York Times dataset could potentially have benefited from a larger vocabulary size.

Based on this vocabulary a symmetric word co-occurrence matrix was calculated for each of the corpus documents. When generating the matrix we only considered context words within a symmetrical window around the base word. Analysis in [2,3] showed that the window size in the range of 6 to 10 had little impact on performance. Thus, following our implementation in [3], we chose a window size of 7, the default in the original glove implementation. Like in [2], each context word only contributed $1 / d$ to the total word pair count, given it was $d$ words apart from the base word. To avoid any bias or prior information from the structure and order of the concatenated Wikipedia articles, reviews or news articles, we randomly shuffled the vocabulary before creating the co-occurrence matrix. As described in Section 3 we then transformed the co-occurrence matrix to a positive PMI matrix. If the corpus consisted of just one document the generated PPMI matrix functions as input for the row-stochastic DEDICOM algorithm. If the corpus consisted of several documents (e.g., one news document per month) the individual PPMI matrices were stacked to a tensor which in turn represented the input for the non-negative tensor DEDICOM algorithm.

The next section sheds light upon the training process of row-stochastic DEDICOM, non-negative tensor DEDICOM and the above mentioned competing matrix and tensor factorization methods, which will be benchmarked against our results in Section 4.3 and in the Appendixes A and B.

### 4.2. Training

As thoroughly outlined in Section 3 we trained both the row-stochastic DEDICOM and non-negative tensor DEDICOM with the alternating gradient descent paradigm.

In the case of a matrix input and a row-stochasticity constraint on $A$ we utilize automatic differentiation from the PyTorch library to perform update steps on $A$ and $\boldsymbol{R}$. First, we initialized the factor matrices $\boldsymbol{A} \in \mathbb{R}^{n \times k}$ and $\boldsymbol{R} \in \mathbb{R}^{k \times k}$, by randomly sampling all elements from a uniform distribution centered around $1, \mathcal{U}(0,2)$. Note that after applying the softmax operation on $A$ all rows of $A$ were stochastic. Therefore, scaling $R$ by

$$
\begin{equation*}
\bar{s}:=\frac{1}{n^{2}} \sum_{i j}^{n} s_{i j}, \tag{40}
\end{equation*}
$$

would result in the initial decomposition $\boldsymbol{A}^{\prime} \boldsymbol{R}\left(\boldsymbol{A}^{\prime}\right)^{T}$ yielding reconstructed elements in the range of $\bar{s}$, the element mean of the PPMI matrix $S$, and thus, speeding up convergence. Second, $\boldsymbol{A}$ and $\boldsymbol{R}$ were iteratively updated employing the Adam optimizer [23] with constant individual learning rates of $\eta^{A}=0.001$ and $\eta^{R}=0.01$ and hyperparameters $\beta_{1}=0.9, \beta_{2}=0.999$ and $\epsilon=1 \times 10^{-8}$. Both learning rates were identified through an exhaustive grid search. We trained for num_epochs $=15,000$ until convergence, where each epoch consisted of an alternating gradient update with respect to $A$ and $\boldsymbol{R}$. Algorithm 1 illustrates the just described training procedure.

In the case of a tensor input and an additional non-negativity constraint on $\underline{R}$ we noticed an inferior training performance with automatic differentiation methods. Hence, due to faster and more stable training convergence and improved qualitative results, we updated $A$ and $\underline{R}$ iteratively via derived multiplicative update rules enforcing nonnegativity. Again, we initialized $A \in \mathbb{R}^{n \times k}$ and $\underline{R} \in \mathbb{R}^{t \times k \times k}$, by randomly sampling all elements from a uniform distribution centered around $1, \mathcal{U}(0,2)$. In order to ensure that the initialized components yielded a reconstructed tensor whose elements were in the same range of the input, we calculated an appropriate scaling factor for each tensor slice $\underline{S}_{l}$ as

$$
\begin{equation*}
\alpha_{l}:=\left(\frac{\bar{s}_{l}}{k^{2}}\right)^{\frac{1}{3}}, \quad \text { where } \quad \bar{s}_{l}:=\frac{1}{n^{2}} \sum_{i j}^{n} \underline{s}_{l i j} . \tag{41}
\end{equation*}
$$

```
Algorithm 1 The row-stochastic DEDICOM algorithm
    initialize \(A, R \leftarrow U(0,2) \cdot \bar{s}\)
        \(\triangleright\) See Equation (40) for the definition of \(\bar{s}\)
    initialize \(\beta_{1}, \beta_{2}, \epsilon \quad \triangleright\) Adam algorithm hyperparameters
    initialize \(\eta^{A}, \eta^{R} \quad \triangleright\) Individual learning rates
    for \(i\) in \(1, \ldots\), num_epochs do
        Calculate \(\operatorname{loss} \mathcal{L}=\mathcal{L}(S, A, R) \quad \triangleright\) See Equation (7)
        \(A \leftarrow A-\operatorname{Adam}_{\beta_{1}, \beta_{2}, \epsilon}\left(\nabla_{A}, \eta^{A}\right), \quad\) where \(\quad \nabla_{A}=\frac{\partial \mathcal{L}}{\partial A}\)
        \(\boldsymbol{R} \leftarrow \boldsymbol{R}-\operatorname{Adam}_{\beta_{1}, \beta_{2}, \epsilon}\left(\nabla_{\boldsymbol{R}}, \eta^{\boldsymbol{R}}\right), \quad\) where \(\quad \nabla_{\boldsymbol{R}}=\frac{\partial \mathcal{L}}{\partial \boldsymbol{R}}\)
    return \(A^{\prime}\) and \(\boldsymbol{R}\), where \(\boldsymbol{A}^{\prime}=\) row_softmax \((\operatorname{col}\) _norm \((\boldsymbol{A})) \quad \triangleright\) See Equation (5)
```

Next, we scaled $A$ by $\bar{\alpha}=\frac{1}{t} \sum_{l=1}^{t} \alpha_{l}$ and each slice $\underline{\boldsymbol{R}}_{l}$ by $\alpha_{l}$ before starting the alternating multiplicative update steps for num_epochs $=300$. The detailed derivation of the update rules is found in Section 3.2 and their iterative application in the training process is described in Algorithm 2.

```
Algorithm 2 The non-negative tensor DEDICOM algorithm
    initialize \(\boldsymbol{A}, \underline{\boldsymbol{R}} \leftarrow U(0,2)\)
    scale \(A\) by \(\bar{\alpha}\) and \(\underline{\boldsymbol{R}}_{l}\) by \(\alpha_{l} \quad \triangleright\) See Equation (41) for the definitions of \(\bar{\alpha}\) and \(\alpha_{l}\)
    for \(i\) in 1,..., num_epochs do
        Calculate loss \(\mathcal{L}=\mathcal{L}(\underline{\boldsymbol{S}}, \boldsymbol{A}, \underline{\boldsymbol{R}}) \quad \triangleright\) See Equation (17)
        \(\boldsymbol{A}_{i j} \hookleftarrow \boldsymbol{A}_{i j} \frac{\left[\sum_{l=1}^{t}\left(\underline{\boldsymbol{S}}_{l} \boldsymbol{A} \underline{\boldsymbol{R}}_{l}^{T}+\underline{\boldsymbol{S}}_{l}^{T} A \underline{\boldsymbol{R}}_{l}\right)\right]_{i j}}{\left[\boldsymbol{A} \sum_{l=1}^{t}\left(\underline{\boldsymbol{R}}_{l} \boldsymbol{A}^{T} A \underline{\boldsymbol{R}}_{l}^{T}+\underline{\boldsymbol{R}}_{l}^{T} \boldsymbol{A}^{T} A \underline{\boldsymbol{R}}_{l}\right)\right]_{i j}}\)
        \(\underline{\boldsymbol{R}}_{l i j} \leftrightarrow \underline{\boldsymbol{R}}_{l i j} \frac{\left[\boldsymbol{A}^{T} \underline{\boldsymbol{S}}_{l} \boldsymbol{A}\right]_{i j}}{\left[\boldsymbol{A}^{T} \boldsymbol{A} \underline{\boldsymbol{R}}_{l} \boldsymbol{A}^{T} \boldsymbol{A}\right]_{i j}}\)
    return \(A\) and \(\underline{R}\)
```

We implemented NMF, LDA and SVD using the sklearn library. In all cases the learnable factor matrices were initialized randomly and default hyperparameters were applied during training. For NMF the multiplicative update rule from [14] was utilized.

Figure 4 shows the convergence behavior of the row-stochastic matrix DEDICOM training process and the final loss of NMF and SVD. Note that LDA optimized a different loss function, which is why the calculated loss was not comparable and therefore excluded. We see that the final loss of DEDICOM was located just above the other losses, which is reasonable when considering the row stochasticity contraint on $A$ and the reduced parameter amount of $n k+k^{2}$ compared to NMF $(2 n k)$ and SVD $\left(2 n k+k^{2}\right)$.


Figure 4. Reconstruction loss development during matrix factorization training. The $x$-axis plots the number of epochs, the $y$-axis plots the corresponding reconstruction error for each method.

To also have a benchmark model for our constrained tensor DEDICOM methods to compare against, we implemented a Tucker-2 variation of NMF, named tensor NMF (TNMF), which factorized the input tensor $\underline{S}$ as

$$
\begin{equation*}
\underline{\boldsymbol{S}}_{l} \approx \underline{\boldsymbol{W}} \boldsymbol{H}_{l} . \tag{42}
\end{equation*}
$$

Its training procedure closely followed the above described alternating gradient descent approach for non-negative tensor DEDICOM. However, due to the two-way factorization (three-way for DEDICOM) the scaling factor $\alpha_{l}$ to properly initialize $\underline{W}$ and $\boldsymbol{H}$ had to be adapted to

$$
\begin{equation*}
\alpha_{l}:=\left(\frac{\bar{s}_{l}}{k}\right)^{\frac{1}{2}}, \quad \text { where } \quad \bar{s}_{l}:=\frac{1}{n^{2}} \sum_{i j}^{n} \underline{S}_{l i j} . \tag{43}
\end{equation*}
$$

Analogous to Figure 4, we compared the training stability and convergence speed of our implemented tensor factorization methods. In particular, Figure 2 visualizes the reconstruction loss development for non-negative tensor DEDICOM trained via multiplicative update rules, row-stochastic tensor DEDICOM trained with automatic differentiation and the Adam optimizer and tensor NMF. It could be clearly observed that row-stochastic tensor DEDICOM converged much slower than the other two models trained with multiplicative update rules (learning rates were implicit here and did not have to be tuned).

### 4.3. Results

In the following, we present our results of training the above mentioned constrained DEDICOM factorizations on different text corpora to simultaneously learn interpretable word embeddings and meaningful topic clusters and their relations.

First, we focused our analysis on row-stochastic matrix DEDICOM applied to the synthetic Wikipedia text documents described in Section 4.1. For compactness reasons we primarily considered the document "Soccer, Bee and Johnny Depp", set the number of topics to $k=6$ and refer to Appendix A. 1 for the other article combinations and competing matrix factorization results. Second, we extended our evaluation to the tensor representation of the Wikipedia documents ( $t=3$, one article per tensor slice) and compared the performance of non-negative (multiplicative updates) and row-stochastic (Adam updates) tensor DEDICOM. Lastly, we applied non-negative tensor DEDICOM to the binned Amazon movie and New York Times news corpora to investigate topic relations across movies and over time. We again point the interested reader to Appendix A for additional results and the comparison to tensor NMF.

In the first step, we evaluated the quality of the learned latent topics by assigning each word embedding $A_{i}^{\prime} \in \mathbb{R}^{1 \times k}$ to the latent topic dimension that represents the maximum value in $A_{i}^{\prime}$, e.g.,

$$
A_{i}^{\prime}=\left[\begin{array}{llllll}
0.05 & 0.03 & 0.02 & 0.14 & 0.70 & 0.06
\end{array}\right], \quad \operatorname{argmax}\left(A_{i}^{\prime}\right)=5,
$$

and thus, $A_{i}^{\prime}$ was matched to Topic 5 . Next, we decreasingly sorted the words within each topic based on their matched topic probability. Table 5 shows the overall number of allocated words and the resulting top 10 words per topic together with each matched probability.

Table 5. The top 10 representative words per dimension of the basis matrix $\boldsymbol{A}^{\prime}$, trained on the Wikipedia data as input matrix using automatic gradient methods.

|  | Topic 1 |  |  |  |  |  |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: |
| \#619 | Topic 2 | \#1238 | Topic 3 |  |  |  |
| \#628 | Topic 4 |  |  |  |  |  |
| \#595 | Topic 5 <br> \#612 | Topic 6 <br> \#389 |  |  |  |  |
|  | ball | film | salazar | cup | bees | heard |
|  | $(0.77)$ | $(0.857)$ | $(0.201)$ | $(0.792)$ | $(0.851)$ | $(0.738)$ |
|  | penalty | starred | geoffrey | football | species | court |
|  | $(0.708)$ | $(0.613)$ | $(0.2)$ | $(0.745)$ | $(0.771)$ | $(0.512)$ |
| 3 | may | role | rush | fifa | bee | depp |
|  | $(0.703)$ | $(0.577)$ | $(0.2)$ | $(0.731)$ | $(0.753)$ | $(0.505)$ |
| 4 | referee | series | brenton | world | pollen | divorce |
|  | $(0.667)$ | $(0.504)$ | $(0.199)$ | $(0.713)$ | $(0.658)$ | $(0.454)$ |
| 5 | goal | burton | hardwicke | national | honey | alcohol |
|  | $(0.66)$ | $(0.492)$ | $(0.198)$ | $(0.639)$ | $(0.602)$ | $(0.435)$ |
| 6 | team | character | thwaites | uefa | insects | paradis |
|  | $(0.651)$ | $(0.465)$ | $(0.198)$ | $(0.623)$ | $(0.576)$ | $(0.42)$ |
| 7 | players | played | catherine | continental | food | relationship |
|  | $(0.643)$ | $(0.451)$ | $(0.198)$ | $(0.582)$ | $(0.536)$ | $(0.419)$ |
| 8 | player | director | kaya | teams | nests | abuse |
|  | $(0.639)$ | $(0.45)$ | $(0.198)$ | $(0.576)$ | $(0.529)$ | $(0.41)$ |
| 9 | play | success | melfi | european | solitary | stating |
|  | $(0.606)$ | $(0.438)$ | $(0.198)$ | $(0.57)$ | $(0.513)$ | $(0.408)$ |
| 10 | game | jack | raimi | association | eusocial | stated |
|  | $(0.591)$ | $(0.434)$ | $(0.198)$ | $(0.563)$ | $(0.505)$ | $(0.402)$ |

Indicated by the high assignment probabilities, one can see that columns 1, 2, 4, 5 and 6 represent distinct topics, which can easily be interpreted. Topic 1 and 4 were related to soccer, where 1 focused on the game mechanics and 4 on the organizational and professional aspect of the game. Topic 2 and 6 clearly referred to Johnny Depp, where 2 focused on his acting career and 6 on his difficult relationship to Amber Heard. The fifth topic obviously related to the insect "bee". In contrast, Topic 3 did not allow for any interpretation and all assignment probabilities were significantly lower than for the other topics.

Further, we analyzed the relations between the topics by visualizing the trained $\boldsymbol{R}$ matrix as a heatmap (see Figure 5c).


Figure 5. (a) 2-dimensional representation of word embeddings $A^{\prime}$ colored by topic assignment. (b) 2-dimensional representation of word embeddings $A^{\prime}$ colored by original Wikipedia article assignment (words that occur in more than one article are excluded). (c) Colored heatmap of affinity matrix $\boldsymbol{R}$.

One thing to note was the symmetry of $R$ which was a first indicator of a successful reconstruction, $A^{\prime} \boldsymbol{R}\left(A^{\prime}\right)^{T}$, (see Section 3.3). In addition, the main diagonal elements were consistently blue (positive), which suggested a high distinction between the topics. Although not very strong one could still see a connection between Topic 2 and 6 indicated by the light blue entry $\boldsymbol{R}_{26}=\boldsymbol{R}_{62}$. While the suggested relation between Topic 1 and 4 was not clearly visible, element $\boldsymbol{R}_{14}=\boldsymbol{R}_{41}$ was the least negative one for Topic 1 . In order to visualize the topic cluster quality we utilized Uniform Manifold Approximation ad Projection (UMAP) [24] to map the $k$-dimensional word embeddings to a 2-dimensional space. Figure 5a illustrates this low-dimensional representation of $A^{\prime}$, where each word is colored based on the above described word to topic assignment. In conjunction with Table 5 one could nicely see that Topic 2 and 6 (Johnny Depp) and Topic 1 and 4 (Soccer) were close to each other. Hence, Figure 5a implicitly shows the learned topic relations as well.

As an additional benchmark, Figure 5b plots the same 2-dimensional representation, but now each word is colored based on the original Wikipedia article it belonged to. Words that occurred in more than one article were not considered in this plot.

Directly comparing Figure 5a,b shows that row-stochastic DEDICOM did not only recover the original articles but also found entirely new topics, which in this case represented subtopics of the articles. Let us emphasize that for all thematically similar article combinations, the found topics were usually not subtopics of a single article, but rather novel topics that might span across multiple Wikipedia articles (see for example Table A2 in the Appendix A). As mentioned at the top of this section, we are not only interested in learning meaningful topic clusters, but also in training interpretable word embeddings that capture semantic meaning.

Hence, we selected within each topic the two most representative words and calculated the cosine similarity between their word embeddings and all other word embeddings stored in $A^{\prime}$. Table 6 shows the four nearest neighbors based on cosine similarity for the top two words in each topic. We observed a high thematical similarity between words with large cosine similarity, indicating the usefulness of the rows of $A^{\prime}$ as word embeddings.

Table 6. For the most significant two words per topic, the four nearest neighbors based on cosine similarity are listed. Matrix $A^{\prime}$ trained on the wikipedia data as input matrix using automatic gradient methods.

|  | Topic 1 | Topic 2 | Topic 3 | Topic 4 | Topic 5 | Topic 6 |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: |
| 0 | ball | film | salazar | cup | bees | heard |
|  | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ |
| 1 | penalty | starred | geoffrey | fifa | bee | court |
|  | $(0.994)$ | $(0.978)$ | $(1.0)$ | $(0.995)$ | $(0.996)$ | $(0.966)$ |
| 2 | referee | role | rush | national | species | divorce |
|  | $(0.992)$ | $(0.964)$ | $(1.0)$ | $(0.991)$ | $(0.995)$ | $(0.944)$ |
| 3 | may | burton | bardem | world | pollen | alcohol |
|  | $(0.989)$ | $(0.937)$ | $(1.0)$ | $(0.988)$ | $(0.986)$ | $(0.933)$ |
| 4 | goal | series | brenton | uefa | honey | abuse |
|  | $(0.986)$ | $(0.935)$ | $(1.0)$ | $(0.987)$ | $(0.971)$ | $(0.914)$ |
|  | penalty | starred | geoffrey | football | species | court |
| 0 | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ |
|  | referee | role | rush | fifa | bees | divorce |
| 1 | $(0.999)$ | $(0.994)$ | $(1.0)$ | $(0.994)$ | $(0.995)$ | $(0.995)$ |
|  | goal | series | salazar | national | bee | alcohol |
| 2 | $(0.998)$ | $(0.985)$ | $(1.0)$ | $(0.983)$ | $(0.99)$ | $(0.987)$ |
|  | player | burton | brenton | cup | pollen | abuse |
| 3 | $(0.997)$ | $(0.981)$ | $(1.0)$ | $(0.983)$ | $(0.99)$ | $(0.982)$ |
| 4 | ball | film | thwaites | world | insects | settlement |
|  | $(0.994)$ | $(0.978)$ | $(1.0)$ | $(0.982)$ | $(0.977)$ | $(0.978)$ |

In comparison to DEDICOM, other matrix factorization methods also provided a useful clustering of words into topics, with varying degree of granularity and clarity. However, the application of these methods as word embedding algorithms mostly failed on the word similarity task, with words close in cosine similarity seldom sharing the same thematical similarity we have seen in DEDICOM. This can be seen in Table A1, which shows for each method, NMF, LDA and SVD, the resulting word to topic clustering and the cosine nearest neighbors of the top two word embeddings per topic. While the individual topics extracted by NMF looked very reasonable, its word embeddings did not seem to carry any semantic meaning based on cosine similarity; e.g., the four nearest neighbors of "ball" were "invoke", "replaced", "scores" and "subdivided". A similar nonsensical picture can be observed for the other main topic words. LDA and SVD performed slightly better on the similar word task, although not all similar words appeared to be sensible, e.g., "children", "detective", "crime", "magazine" and "barber". In addition, some topics could not be clearly defined due to mixed word assignments, e.g., Topic 4 for LDA and Topic 1 for SVD.

Before shifting our analysis to the Amazon movie review and the New York Times news corpus we investigated factorizing the tensor representation of the "Soccer, Bee and Johnny Depp" Wikipedia documents. In particular, we compared the qualitative factorization results of row-stochastic and non-negative tensor DEDICOM trained with automatic differentiation and multiplicative update rules, respectively. Tables 7 and 8 in conjunction with Figures 6 and 7 show the extracted topics and their relations for both methods.


Figure 6. Colored heatmap of affinity tensor $\underline{\boldsymbol{R}}$, trained on the Wikipedia data represented as input tensor using automatic gradient methods.

Table 7. Top 10 representative words per dimension of the basis matrix $\boldsymbol{A}^{\prime}$, trained on the wikipedia data as input tensor using automatic gradient methods.

|  | Topic 1 <br> \#481 | Topic 2 <br> \#661 | Topic 3 <br> \#414 | Topic 4 <br> \#457 | Topic 5 <br> \#316 | Topic 6 <br> \#1711 |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: |
| 1 | hind | game | film | heard | bees | disorder |
|  | $(0.646)$ | $(0.83)$ | $(0.941)$ | $(0.844)$ | $(0.922)$ | $(0.291)$ |
|  | segments | football | starred | court | bee | collapse |
|  | $(0.572)$ | $(0.828)$ | $(0.684)$ | $(0.566)$ | $(0.868)$ | $(0.29)$ |
| 3 | bacteria | players | role | divorce | honey | attrition |
|  | $(0.563)$ | $(0.782)$ | $(0.624)$ | $(0.51)$ | $(0.756)$ | $(0.285)$ |
| 4 | legs | ball | series | depp | insects | losses |
|  | $(0.562)$ | $(0.777)$ | $(0.562)$ | $(0.508)$ | $(0.68)$ | $(0.284)$ |
| 5 | antennae | team | burton | sued | food | invertebrate |
|  | $(0.555)$ | $(0.771)$ | $(0.547)$ | $(0.48)$ | $(0.634)$ | $(0.283)$ |
| 6 | females | may | character | stating | species | rate |
|  | $(0.549)$ | $(0.696)$ | $(0.499)$ | $(0.45)$ | $(0.599)$ | $(0.283)$ |
| 7 | wings | play | success | alcohol | nests | businesses |
|  | $(0.547)$ | $(0.692)$ | $(0.494)$ | $(0.449)$ | $(0.596)$ | $(0.282)$ |
| 8 | small | competitions | played | paradis | flowers | virgil |
|  | $(0.538)$ | $(0.677)$ | $(0.483)$ | $(0.446)$ | $(0.571)$ | $(0.282)$ |
| 9 | groups | match | films | alleged | pollen | iridescent |
|  | $(0.527)$ | $(0.672)$ | $(0.482)$ | $(0.445)$ | $(0.56)$ | $(0.282)$ |
| 10 | males | penalty | box | stated | larvae | detail |
|  | $(0.518)$ | $(0.664)$ | $(0.465)$ | $(0.444)$ | $(0.529)$ | $(0.281)$ |



Figure 7. Colored heatmap of affinity tensor $\underline{\boldsymbol{R}}$, trained on the Wikipedia data represented as input tensor using multiplicative update rules.

Table 8. Top 10 representative words per dimension of the basis matrix $A$, trained on the wikipedia data as input tensor using multiplicative update rules.

|  | Topic 1 \#521 | $\text { Topic } 2$ \#249 | Topic 3 \#485 | Topic 4 \#871 | Topic 5 \#445 | $\text { Topic } 6$ \#1469 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 1 | species | game | honey | allow | insects | depp |
|  | (3.105) | (3.26) | (2.946) | (0.668) | (2.794) | (2.419) |
| 2 | eusocial | football | bee | organised | pollen | film |
|  | (2.524) | (3.05) | (2.01) | (0.662) | (2.239) | (2.115) |
| 3 | solitary | players | beekeeping | winner | flowers | role |
|  | (2.279) | (2.699) | (1.933) | (0.632) | (2.019) | (1.32) |
| 4 | nest | ball | bees | officially | nectar | starred |
|  | (2.118) | (2.475) | (1.704) | (0.626) | (1.656) | (1.3) |
| 5 | females | may | increased | wins | wasps | actor |
|  | (1.993) | (2.447) | (1.589) | (0.617) | (1.602) | (1.155) |
| 6 | workers | team | humans | level | wings | series |
|  | (1.797) | (2.424) | (1.515) | (0.613) | (1.588) | (1.126) |
| 7 | nests | association | wild | free | many | burton |
|  | (1.787) | (1.92) | (1.415) | (0.6) | (1.588) | (1.112) |
| 8 | colonies | play | mites | constitute | hind | played |
|  | (1.722) | (1.834) | (1.4) | (0.596) | (1.577) | (1.068) |
| 9 | egg | referee | colony | regulation | hairs | heard |
|  | (1.692) | (1.809) | (1.35) | (0.595) | (1.484) | (1.005) |
| 10 | males | laws | beekeepers | prestigious | pollinating | success |
|  | (1.664) | (1.792) | (1.332) | (0.594) | (1.467) | (0.981) |

It could be seen that non-negative tensor DEDICOM yielded a more interpretable affinity tensor $\underline{\boldsymbol{R}}$ (Figure 7) due to its enforced non-negativity. For example, it clearly highlighted the bee related Topics 1,3 and 5 in the affinity tensor slice corresponding to the article "Bee". Moreover, all extracted topics in Table 8 were distinct and their relations were well represented in the individual slices of $\underline{\boldsymbol{R}}$. In contrast, Topic 6 in Table 7 did not represent a meaningful topic, which was also indicated by the low probability scores of the ranked topic words. Although the results of the similar word evaluation were arguably better for row-stochastic tensor DEDICOM (see Tables 9 and 10) we prioritized topic extraction and relation quality. That is why in the further analysis of the Amazon review and New York Times news corpus we restricedt our evaluation to non-negative tensor DEDICOM.

As described in Section 4.1 our Amazon movie review corpus comprised human written reviews for six famous animation films. Factorizing its PPMI tensor representation with non-negative tensor DEDICOM and the number of topics set to $k=10$ revealed not only movie-specific subtopics but also general topics that spanned over several movies. For example, Topics 1, 9 and 10 in Table 11 could uniquely be related to the films "Frozen", "Toy Story 1" and "Kung Fu Panda 1", respectively, whereas Topic 5 constituted bonus material on a DVD which held true for all films. The latter could also be seen in Figure 8 where Topic 5 was highlighted in each movie slice (strongly in the top and lightly in the bottom row). In the same sense one could observe that Topic 3 was present in both "Kung Fu Panda 1" and "Kung Fu Panda 2", which is reasonable considering the topic depicted the general notion of a fearsome warrior.

Table 9. For the most significant two words per topic, the four nearest neighbors based on cosine similarity are listed. Matrix $A^{\prime}$, trained on the wikipedia data as input tensor using automatic gradient methods.

|  | Topic 1 | Topic 2 | Topic 3 | Topic 4 | Topic 5 | Topic 6 |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: |
| 0 | hind | game | film | heard | bees | disorder |
|  | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ |
| 1 | segments | football | starred | court | bee | collapse |
|  | $(0.995)$ | $(1.0)$ | $(0.968)$ | $(0.954)$ | $(0.999)$ | $(1.0)$ |
| 2 | legs | players | role | divorce | honey | losses |
|  | $(0.994)$ | $(0.999)$ | $(0.954)$ | $(0.925)$ | $(0.99)$ | $(1.0)$ |
| 3 | antennae | ball | series | sued | insects | attrition |
|  | $(0.993)$ | $(0.999)$ | $(0.951)$ | $(0.907)$ | $(0.976)$ | $(0.999)$ |
| 4 | wings | team | burton | alleged | food | businesses |
|  | $(0.992)$ | $(0.998)$ | $(0.945)$ | $(0.897)$ | $(0.97)$ | $(0.999)$ |
|  | segments | football | starred | court | bee | collapse |
| 0 | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ |
|  | antennae | game | role | divorce | bees | disorder |
| 1 | $(1.0)$ | $(1.0)$ | $(0.993)$ | $(0.996)$ | $(0.999)$ | $(1.0)$ |
|  | wings | players | series | sued | honey | losses |
| 2 | $(0.999)$ | $(0.999)$ | $(0.978)$ | $(0.991)$ | $(0.995)$ | $(0.999)$ |
|  | bacteria | ball | burton | alleged | insects | pesticide |
| 3 | $(0.999)$ | $(0.999)$ | $(0.975)$ | $(0.981)$ | $(0.984)$ | $(0.998)$ |
|  | legs | team | film | alcohol | food | businesses |
| 4 | $(0.998)$ | $(0.999)$ | $(0.968)$ | $(0.981)$ | $(0.976)$ | $(0.998)$ |

Table 10. For the most significant two words per topic, the four nearest neighbors based on cosine similarity are listed. Matrix $A$, trained on the wikipedia data as input tensor using multiplicative update rules.

|  | Topic 1 | Topic $\mathbf{2}$ | Topic 3 | Topic 4 | Topic 5 | Topic 6 |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: |
| 0 | species | game | honey | allow | insects | depp |
|  | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ |
| 1 | easier | football | boatwrights | emancipation | ultraviolet | charlie |
|  | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ |
| 2 | tiny | players | glade | broadly | mechanics | infiltrate |
|  | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ |
| 3 | halictidae | association | tutankhamun | disabilities | exploit | thenwife |
|  | $(0.999)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ |
| 4 | provision | team | oracle | total | swallows | tourist |
|  | $(0.999)$ | $(0.997)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ |
|  | eusocial | football | bee | organised | pollen | film |
| 0 | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ |
|  | oligocene | players | subfamilies | comes | honeybees | starred |
| 1 | $(1.0)$ | $(1.0)$ | $(0.995)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ |
|  | architecture | game | internal | shows | enlarged | smoking |
| 2 | $(1.0)$ | $(1.0)$ | $(0.994)$ | $(1.0)$ | $(0.998)$ | $(0.999)$ |
|  | uncommon | association | studied | attention | simple | dislocated |
| 3 | $(1.0)$ | $(1.0)$ | $(0.994)$ | $(1.0)$ | $(0.998)$ | $(0.999)$ |
|  | termed | team | cladogram | deductions | drove | injuries |
| 4 | $(1.0)$ | $(0.997)$ | $(0.99)$ | $(1.0)$ | $(0.998)$ | $(0.999)$ |



Figure 8. Colored heatmap of affinity tensor $\underline{\boldsymbol{R}}$, trained on the Amazon review data represented as input tensor using multiplicative update rules.

Table 11. Top 10 representative words per dimension of the basis matrix $A$, trained on the Amazon review data as input tensor using multiplicative update rules.

|  | $\begin{gathered} \text { Topic } 1 \\ \# 528 \end{gathered}$ | Topic 2 <br> \#445 | Topic 3 \#1477 | Topic 4 | Topic 5 \#1917 | Topic 6 \#597 | Topic 7 <br>  <br>  | Topic 8 \#670 | Topic 9 \#1599 | $\begin{gathered} \text { Topic } 10 \\ \# 188 \end{gathered}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 1 | na | shen | legendary | lasseter | disc | screams | code | mike | woody | po |
|  | (4.215) | (4.21) | (1.459) | (2.887) | (1.367) | (3.779) | (3.292) | (4.325) | (6.12) | (5.737) |
| 2 | elsa | peacock | valley | director | birds | energy | email | crystal | buzz | master |
| 2 | (4.087) | (2.668) | (1.448) | (2.392) | (1.343) | (3.315) | (2.781) | (4.055) | (5.484) | (5.276) |
| 3 | olaf | oldman | temple | andrew | widescreen | monstropolis | promo | billy | andy | shifu |
| 3 | (2.315) | (2.627) | (1.31) | (2.158) | (1.327) | (3.13) | (2.645) | (3.911) | (4.355) | (4.707) |
| 4 | trolls | gary | kim | stanton | outtakes | world | free | goodman | toys | dragon |
|  | (2.241) | (2.423) | (1.307) | (2.119) | (1.238) | (3.109) | (2.343) | (3.812) | (4.119) | (4.344) |
| 5 | frozen | lord | fearsome | special | extras | monsters | promotion | sully | lightyear | warrior |
|  | (2.196) | (2.201) | (1.288) | (1.892) | (1.185) | (3.047) | (2.279) | (3.728) | (3.334) | (4.274) |
| 6 | kristoff | weapon | teacher | pete | dvd | city | promotional | wazowski | allen | tai |
|  | (2.155) | (1.469) | (1.288) | (1.612) | (1.142) | (2.994) | (2.266) | (3.513) | (2.752) | (4.082) |
| 7 | queen | wolf | battle | ranft | included | monster | amazon | randall | tim | lung |
|  | (2.055) | (1.405) | (1.264) | (1.564) | (1.13) | (2.978) | (2.129) | (3.404) | (2.609) | (3.993) |
| 8 | hans | inner | duk | joe | short | power | click | sulley | hanks | five |
|  | (2.054) | (1.38) | (1.257) | (1.564) | (1.101) | (2.919) | (2.024) | (3.396) | (2.471) | (2.952) |
| 9 | sister | yeoh | train | feature | games | scare | download | buscemi | cowboy | oogway |
|  | (1.904) | (1.359) | (1.221) | (1.555) | (1.1) | (2.614) | (1.889) | (3.266) | (2.407) | (2.918) |
| 10 | ice | michelle | warriors | ralph | tour | closet | instructions | james | space | furious |
|  | (1.839) | (1.354) | (1.22) | (1.518) | (1.031) | (2.555) | (1.877) | (3.264) | (2.375) | (2.806) |

Figure 9 and Table 12 refer to our experimental results on the dataset of New York Times news articles. We saw a diverse array of topics extracted from the text corpus, ranging from US-politics (Topics 4, 6, 7) to natural disasters (Topic 8), Hollywood sexual assault allegations (Topic 10) and the COVID epidemic both from a medical view (Topic 3) and a view on resulting restrictions to businesses (Topic 9).

The corresponding heatmap allowed us to infer when certain topics were most relevant in the last year. While the entries relating to the COVID pandemic remain light blue for the
first half of the heatmap we sawa the articles picking up on the topic around March 2020, when the effects of the Coronavirus started hitting the US. Even comparatively smaller events like the conviction of Harvey Weinstein and the death of George Floyd triggering the racism debate in the US could be recognized in the heatmap, with a large deviation of Topic 10 around February 2020 and Topic 4 around June 2020.


Figure 9. Colored heatmap of affinity tensor $\underline{\boldsymbol{R}}$, trained on the New York Times news article data represented as input tensor using multiplicative update rules.

Further empirical results on the Amazon review and New York Times news corpora, such as two-dimensional UMAP representations of the embedding matrix $A$ and extracted topics from tensor NMF, can be found in Appendixes A. 3 and A.4, respectively. For example, Table A21 shows that the tensor NMF factorization also extracted high quality topics but lacked the interpretable affinity tensor $\underline{R}$ which was crucial in order to properly comprehend a topic development over time.

Table 12. Top 10 representative words per dimension of the basis matrix $A$, trained on the New York Times news article data as input tensor using multiplicative update rules.

|  | $\begin{gathered} \text { Topic } 1 \\ \# 454 \end{gathered}$ | Topic 2 \#5984 | Topic 3 \#567 | $\begin{gathered} \text { Topic } 4 \\ \# 562 \end{gathered}$ | Topic 5 <br> \#424 | Topic 6 | Topic 7 $\# 515$ | Topic 8 <br> \#297 | $\underset{\# 431}{ } 9$ \#431 | $\begin{gathered} \text { Topic } 10 \\ \# 436 \end{gathered}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 1 | suleimani | loans | masks | floyd | contributed | confederate | ukraine | storm | restaurants | weinstein |
|  | (2.812) | (0.618) | (3.261) | (3.376) | (4.565) | (3.226) | (3.191) | (2.76) | (2.948) | (3.442) |
| 2 | iran | university | protective | minneapolis | reporting | statue | sondland | hurricane | bars | sexual |
|  | (2.593) | (0.551) | (2.823) | (2.551) | (2.788) | (2.649) | (2.881) | (2.622) | (2.021) | (2.71) |
| 3 | iraq | oil | gloves | police | michael | statues | zelensky | winds | reopen | rape |
|  | (2.453) | (0.549) | (2.516) | (2.255) | (2.707) | (2.416) | (2.133) | (1.715) | (1.684) | (2.102) |
| 4 | iranian | billion | ventilators | george | katie | monuments | ambassador | tropical | gyms | assault |
|  | (2.408) | (0.54) | (2.22) | (2.088) | (2.324) | (1.815) | (1.976) | (1.606) | (1.654) | (1.861) |
| 5 | iraqi | loan | surgical | protests | alan | monument | ukrainian | storms | stores | jury |
|  | (1.799) | (0.468) | (2.032) | (1.936) | (2.292) | (1.375) | (1.789) | (1.439) | (1.638) | (1.513) |
| 6 | baghdad | bonds | gowns | brutality | emily | flag | giuliani | coast | theaters | charges |
|  | (1.604) | (0.466) | (1.965) | (1.765) | (2.165) | (1.206) | (1.755) | (1.415) | (1.627) | (1.409) |
| 7 | qassim | payments | equipment | racism | nicholas | richmond | volker | laura | salons | predatory |
|  | (1.599) | (0.456) | (1.86) | (1.579) | (2.096) | (1.109) | (1.754) | (1.259) | (1.438) | (1.387) |
| 8 | strike | edited | supplies | knee | cochrane | symbols | investigations |  |  | harvey |
|  | (1.597) | (0.452) | (1.816) | (1.435) | (1.934) | (1.089) | (1.602) | (1.217) | (1.424) | (1.35) |
| 9 | gen | trillion | gear | killing | rappeport | remove | testified | category | shops | guilty |
|  | (1.513) | (0.451) | (1.742) | (1.429) | (1.613) | (1.058) | (1.584) | (1.192) | (1.325) | (1.312) |
| 10 | $\underset{(1.504)}{ }$ | graduated <br> (0.449) | mask <br> (1.502) | officers (1.405) | maggie <br> (1.529) | removal <br> (1.003) | testimony (1.558) | landfall <br> (1.106) | indoor (1.247) | sex (1.3) |

## 5. Conclusions and Outlook

We propose a constrained version of the DEDICOM algorithm that is able to factorize the pointwise mutual information matrices of text documents into meaningful topic clusters all the while providing interpretable word embeddings for each vocabulary item. Our study on semi-artificial data from Wikipedia articles has shown that this method recovers the underlying structure of the text corpus and provides topics with thematic granularity, meaning the extracted latent topics are more specific than a simple clustering of articles. A comparison to related matrix factorization methods has shown that the combination of relation aware topic modeling and interpretable word embedding learning given by our algorithm is unique in its class.

Extending this algorithm to factorize three-dimensional input tensors allows for the study of changes in the relations between topics across subsets of a structured text corpus, e.g., news articles grouped by time period. Algorithmically, this can be solved via alternating gradient descent by either automatic gradient methods or by applying multiplicative update rules which decrease training time drastically and enhance training stability.

Due to memory constraints from matrix multiplications of high dimensional dense tensors our proposed approach is currently limited in vocabulary size or time dimension.

In further work we aim for developing algorithms capable of leveraging sparse matrix multiplications to avoid the above mentioned memory constraints. In addition, we plan to expand on the possibilities of constraining the factor matrices and tensors when applying a multiplicative update rule and further analyze the behavior of the factor tensors, for example by utilizing time series analysis to discover temporal relations between extracted topics and to potentially identify trends. Finally, further analysis may include additional quantitative evaluations of our proposed methods' topic modeling performance with competing approaches.

Author Contributions: Conceptualization, D.B.; Methodology, L.H.; Project administration, L.H. and D.B.; Supervision, C.B. and R.S.; Writing-original draft, L.H. and D.B. All authors have read and agreed to the published version of the manuscript.
Funding: The authors of this work were supported by the Competence Center for Machine Learning Rhine Ruhr (ML2R) which is funded by the Federal Ministry of Education and Research of Germany (grant no. 01 IS18038C). We gratefully acknowledge this support.
Data Availability Statement: Publicly available datasets (Amazon reviews, Wikipedia articles) were analyzed in this study. This data can be found here: https://github.com/LarsHill/text-dedicom-paper. The NYT news article data presented in this study are available on request from the corresponding author. The data are not publicly available due to potential copyright concerns.

Conflicts of Interest: The authors declare no conflict of interest.

## Appendix A. Additional Results

Appendix A.1. Additional Results on Wikipedia Data as Matrix Input
Articles: "Soccer", "Bee", "Johnny Depp".

Table A1. For each evaluated matrix factorization method we display the top 10 words for each topic and the five most similar words based on cosine similarity for the two top words from each topic.

|  |  | Topic 1 | Topic 2 | Topic 3 | Topic 4 | Topic 5 | Topic 6 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| NMF | $\begin{gathered} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \end{gathered}$ | \#619 ball may penalty referee players team goal game player play | \#1238 bees species bee pollen honey insects food nests solitary eusocial | \#628 film starred role series burton character films box office jack | \#595 football cup world fifa national association international women teams uefa | \#612 <br> heard depp court alcohol relationship stated divorce abuse paradis stating | \#389 <br> album band guitar vampires rock <br> hollywood song released perry debut |
|  | $\begin{aligned} & 0 \\ & 1 \\ & 2 \\ & 3 \\ & 4 \end{aligned}$ | ball invoke replaced scores subdivided | bees odors tufts colour affected | film burtondirected tone landau brother | football athenaeus paralympic governing varieties | heard <br> crew alleging oped asserted | album jones marilyn roots drums |
|  | $\begin{aligned} & \hline 0 \\ & 1 \\ & 2 \\ & 3 \\ & 4 \end{aligned}$ | may yd ineffectiveness tactical slower | ```species niko commercially microbiota strategies``` | starred <br> shared <br> whitaker <br> eccentric <br> befriends | cup inaugurated confederation gold headquarters | depp refer york leaders nonindian | band <br> heroes <br> bowie <br> debut <br> solo |
| LDA | $\begin{gathered} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \end{gathered}$ | $\begin{aligned} & \hline \# 577 \\ & \text { film } \end{aligned}$ series <br> man <br> played pirates character along cast also hollow | \#728 football women association fifa teams games world cup game international | \#692 <br> depp children life role starred alongside actor stated burton playing | \#607 penalty heard ball direct referee red time goal scored player | \#663 <br> bees <br> flowers bee honey pollen food increased pollination times larvae | \#814 <br> species workers solitary players colonies eusocial nest may size egg |
|  | $\begin{aligned} & 0 \\ & 1 \\ & 2 \\ & 3 \\ & 4 \end{aligned}$ | film charlie near thinking shadows | football cup canada zealand activities | depp critical february script song | penalty <br> extra <br> kicks <br> inner <br> moving | bees bee insects authors hives | species social chosen females subspecies |
|  | $\begin{aligned} & 0 \\ & 1 \\ & 2 \\ & 3 \\ & 4 \end{aligned}$ | series crybaby waters sang cast | women fifa opera exceeding cuju | children detective crime magazine barber | heard allison serious allergic cost | flowers always eusociality varroa wing | workers carcases lived provisioned cuckoo |
| SVD | $\begin{gathered} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \end{gathered}$ | \#1228 <br> bees also bee species played time one first two pollen | \#797 <br> depp film starred role series burton character actor released release | \#628 <br> game ball team players penalty play may referee competitions laws | \#369 cup football fifa world european uefa national europe continental confederation | \#622 <br> heard court divorce stating alcohol paradis documents abuse settlement sued | \#437 <br> beekeeping increased honey described use wild varroa mites colony flowers |
| SVD | $\begin{aligned} & \hline 0 \\ & 1 \\ & 2 \\ & 3 \\ & 4 \end{aligned}$ | bees bee develops studied crops | deppiiiracismappropriation <br> march | game correct abandoned maximum clear | cup continental contested confederations conmebol | heard alleging attempting finalized submitted | $\begin{gathered} \hline \text { beekeeping } \\ \text { varroa } \\ \text { animals } \\ \text { mites } \\ \text { plato } \end{gathered}$ |
|  | $\begin{aligned} & 0 \\ & 1 \\ & 2 \\ & 3 \\ & 4 \end{aligned}$ | also although told chosen stars | film waters robinson scott costars | ball finely poised worn manner | football er suffix word appended | court <br> declaration issued restraining verbally | increased usage farmers mentioned aeneid |

Articles: "Dolphin", "Shark", "Whale".

Table A2. Top half lists the top 10 representative words per dimension of the basis matrix $A$, bottom half lists the five most similar words based on cosine similarity for the two top words from each topic.

|  | $\begin{gathered} \text { Topic } 1 \\ \# 460 \end{gathered}$ | $\begin{gathered} \text { Topic } 2 \\ \# 665 \end{gathered}$ | Topic 3 \#801 | Topic 4 <br> \#753 | $\begin{gathered} \text { Topic } 5 \\ \# 854 \end{gathered}$ | Topic 6 \#721 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 1 | shark | calf | ship | conservation | water | dolphin |
|  | (0.665) | (0.428) | (0.459) | (0.334) | (0.416) | (0.691) |
| 2 | sharks | months | became | countries | similar | dolphins |
|  | (0.645) | (0.407) | (0.448) | (0.312) | (0.374) | (0.655) |
| 3 | fins | calves | poseidon | government | tissue | captivity |
|  | (0.487) | (0.407) | (0.44) | (0.309) | (0.373) | (0.549) |
| 4 | killed | females | riding | wales | body | wild |
|  | (0.454) | (0.399) | (0.426) | (0.304) | (0.365) | (0.467) |
| 5 | million | blubber | dionysus | bycatch | swimming | behavior |
|  | (0.451) | (0.374) | (0.422) | (0.29) | (0.357) | (0.461) |
| 6 | fish | young | ancient | cancelled | blood | bottlenose |
|  | (0.448) | (0.37) | (0.42) | (0.288) | (0.346) | (0.453) |
| 7 | international | sperm | deity | eastern | surface | sometimes |
|  | (0.442) | (0.356) | (0.412) | (0.287) | (0.344) | (0.449) |
| 8 | fin | born | ago | policy | oxygen | human |
|  | (0.421) | (0.355) | (0.398) | (0.286) | (0.34) | (0.421) |
| 9 | fishing | feed | melicertes | control | system | less |
|  | (0.405) | (0.349) | (0.395) | (0.285) | (0.336) | (0.42) |
| 10 | teeth | mysticetes | greeks | imminent | swim | various |
|  | (0.398) | (0.341) | (0.394) | (0.282) | (0.336) | (0.418) |
| 0 | shark | calf | ship | conservation | water | dolphin |
|  | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) |
| 2 | sharks | calves | dionysus | south | prey | dolphins |
|  | (0.981) | (0.978) | (0.995) | (0.981) | (0.964) | (0.925) |
| 3 | fins | females | riding | states | swimming | sometimes |
|  | (0.958) | (0.976) | (0.992) | (0.981) | (0.959) | (0.909) |
| 4 | killed | months | deity | united | allows | another |
|  | (0.929) | (0.955) | (0.992) | (0.978) | (0.957) | (0.904) |
| 5 | fishing | young | poseidon | endangered | swim | bottlenose |
| 5 | (0.916) | (0.948) | (0.987) | (0.976) | (0.947) | (0.903) |
| 0 | sharks | months | became | countries | similar | dolphins |
|  | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) |
| 2 | shark | born | old | eastern | surface | behavior |
|  | (0.981) | (0.992) | (0.953) | (0.991) | (0.992) | (0.956) |
| 3 | fins | young | later | united | brain | sometimes |
| 3 | (0.936) | (0.992) | (0.946) | (0.989) | $(0.97)$ | $(0.945)$ |
| 4 | tiger | sperm | ago | caught | sound | various |
|  | (0.894) | (0.985) | (0.939) | (0.987) | (0.968) | (0.943) |
| 5 | killed | calves | modern | south | object | less |
|  | (0.887) | (0.984) | (0.937) | (0.979) | (0.965) | (0.937) |

Articles: "Dolphin", "Shark", "Whale'.

Table A3. For each evaluated matrix factorization method we display the top 10 words for each topic and the five most similar words based on cosine similarity for the two top words from each topic.

|  |  | Topic 1 | Topic 2 | Topic 3 | Topic 4 | Topic 5 | Topic 6 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| NMF |  | \#492 | \#907 | \#452 | \#854 | \#911 | \#638 |
|  | 1 | blood | international | evidence | sonar | ago | calf |
|  | 2 | body | killed | selfawareness | may | teeth | young |
|  | 3 | heart | states | ship | surface | million | females |
|  | 4 | gills | conservation | dionysus | clicks | mysticetes | captivity |
|  | 5 | bony | new | came | prey | whales | calves |
|  | 6 | oxygen | united | another | use | years | months |
|  | 7 | organs | shark | important | underwater | baleen | born |
|  | 8 | tissue | world | poseidon | sounds | cetaceans | species |
|  | 9 | water | endangered | mark | known | modern | male |
|  | 10 | via | islands | riding | similar | extinct | female |
|  | 0 | blood | international | evidence | sonar | ago | calf |
|  | 1 | travels | proposal | flaws | poisoned | consist | uninformed |
|  | 2 | enters | fipotidae | methodological | signals | specialize | primary |
|  | 3 | vibration | banned | nictating | $\approx-$ | legs | born |
|  | 4 | tolerant | iniidae | wake | emitted | closest | leaner |
|  | 0 | body | killed | selfawareness | may | teeth | young |
|  | 1 | crystal | law | legendary | individuals | fuel | brood |
|  | 2 | blocks | consumers | humankind | helping | lamp | lacking |
|  | 3 | modified | pontoporiidae | helpers | waste | filterfeeding | accurate |
|  | 4 | slits | org | performing | depression | krill | consistency |
| LDA |  | \#650 | \#785 | \#695 | \#815 | \#635 | \#674 |
|  | 1 | killed | teeth | head | species | meat | air |
|  | 2 | system | baleen | fish | male | whale | using |
|  | 3 | endangered | mysticetes | dolphin | females | $\mathrm{ft}^{\text {d }}$ | causing |
|  | 4 | often | ago | fin | whales | fisheries | currents |
|  | 5 | close | jaw | eyes | sometimes | also | sounds |
|  | 6 | sharks | family | fat | captivity | ocean | groups |
|  | 7 | countries | water | navy | young | threats | sound |
|  | 8 | since | includes | popular | shark | children | research |
|  | 9 | called | allow | tissue | female | population | clicks |
|  | 10 | vessels | greater | tail | wild | bottom | burst |
|  | 0 | killed | teeth | head | species | meat | air |
|  | 1 | postures | dense | underside | along | porbeagle | australis |
|  | 2 | dolphinariums |  | grooves |  |  |  |
|  | 3 | town | tourism | eyesight | long | activities | melbourne |
|  | 4 | onethird | planktonfeeders | osmoregulation | sleep | comparable | spear |
|  | 0 | system | baleen |  |  | whale | using |
|  | 1 | dominate | mysticetes | mostly | females | live | communication |
|  | 2 | close | distinguishing | swim | aorta | human | become |
|  | 3 | controversy | unique | due | female | cold | associated |
|  | 4 | agree | remove | whole | position | parts | mirror |
| SVD |  | \#1486 | \#544 | \#605 | \#469 | \#539 | \#611 |
|  | 1 | dolphins | water | shark | million | poseidon | dolphin |
|  | 2 | species | body | sharks | years | became | meat |
|  | 3 | whales | tail | fins | ago | ship | family |
|  | 4 | fish | teeth | international | whale | riding | river |
|  | 5 | also | flippers | killed | two | evidence | similar |
|  | 6 | large | tissue | fishing | calf | melicertes | extinct |
|  | 7 | may | allows | fin | mya | deity | called |
|  | 8 | one | air | law | later | ino | used |
|  | 9 | animals | feed | new | months | came | islands |
|  | 10 | use | bony | conservation | mysticetes | made | genus |
|  |  |  |  |  |  |  |  |
|  | 1 | various | vertical | corpse | approximately | games | depicted |
|  | 2 | finding | unlike | stocks | assigned | phalanthus | makara |
|  | 3 | military | chew | galea | hybodonts | statue | capensis |
|  | 4 | selfmade | lack | galeomorphii | appeared | isthmian | goddess |
|  | 0 | species | body | sharks | years | became | meat |
|  | 1 | herd | heart | mostly | acanthodians | pirates | contaminated |
|  | 2 | reproduction | resisting | fda | spent | elder | harpoon |
|  | 3 | afford | fit | lists | stretching | mistook | practitioner |
|  | 4 | maturity | posterior | carcharias | informal | wealthy | pcbs |

Articles: "Soccer", "Tennis", "Rugby".
Table A4. Top half lists the top 10 representative words per dimension of the basis matrix $A$, bottom half lists the five most similar words based on cosine similarity for the two top words from each topic.

|  | Topic 1 \#539 | $\text { Topic } 2$ \#302 | Topic 3 \#563 | Topic 4 \#635 | Topic 5 \#650 | Topic 6 \#530 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 1 | $\begin{gathered} \text { may } \\ (0.599) \end{gathered}$ | $\begin{gathered} \hline \text { leads } \\ (0.212) \end{gathered}$ | tournaments (0.588) | greatest <br> (0.572) | $\begin{aligned} & \text { football } \\ & (0.553) \end{aligned}$ | $\begin{gathered} \text { net } \\ (0.644) \end{gathered}$ |
| 2 | penalty <br> (0.576) | $\begin{gathered} \text { sole } \\ (0.205) \end{gathered}$ | tournament (0.517) | $\begin{aligned} & \text { tennis } \\ & (0.497) \end{aligned}$ | $\begin{aligned} & \text { rugby } \\ & (0.542) \end{aligned}$ | $\begin{aligned} & \text { shot } \\ & (0.629) \end{aligned}$ |
| 3 | referee $(0.564)$ | competes (0.205) | $\begin{aligned} & \text { events } \\ & (0.509) \end{aligned}$ | female <br> (0.44) | $\begin{aligned} & \text { south } \\ & (0.484) \end{aligned}$ | $\begin{aligned} & \text { stance } \\ & (0.553) \end{aligned}$ |
| 4 | $\begin{gathered} \text { team } \\ (0.517) \end{gathered}$ | extending (0.204) | $\begin{aligned} & \text { prize } \\ & (0.501) \end{aligned}$ | $\begin{gathered} \text { ever } \\ (0.433) \end{gathered}$ | $\begin{aligned} & \text { union } \\ & \text { (0.47) } \end{aligned}$ | $\begin{aligned} & \text { stroke } \\ & (0.543) \end{aligned}$ |
| 5 | $\begin{gathered} \text { goal } \\ (0.502) \end{gathered}$ | fixing <br> (0.203) | $\begin{aligned} & \text { tour } \\ & (0.497) \end{aligned}$ | navratilova (0.405) | wales (0.459) | $\begin{gathered} \text { serve } \\ (0.537) \end{gathered}$ |
| 6 | $\begin{gathered} \text { kick } \\ (0.459) \end{gathered}$ | triggered (0.203) | money (0.488) | modern $(0.401)$ | national $(0.446)$ | rotation $(0.513)$ |
| 7 | $\begin{gathered} \text { play } \\ (0.455) \end{gathered}$ | bleeding $(0.202)$ | $\operatorname{cup}_{(0.486)}$ | $\begin{aligned} & \text { best } \\ & (0.4) \end{aligned}$ | england $(0.438)$ | backhand (0.508) |
| 8 | $\begin{gathered} \text { ball } \\ (0.452) \end{gathered}$ | $\begin{aligned} & \text { fraud } \\ & (0.202) \end{aligned}$ | $\begin{gathered} \text { world } \\ (0.467) \end{gathered}$ | wingfield (0.394) | $\begin{gathered} \text { new } \\ (0.416) \end{gathered}$ | $\begin{gathered} \text { hit } \\ (0.507) \end{gathered}$ |
| 9 | offence (0.444) | inflammation (0.202) | $\begin{gathered} \text { atp } \\ (0.464) \end{gathered}$ | sports <br> (0.39) | europe <br> (0.406) | forehand (0.499) |
| 10 | $\begin{gathered} \text { foul } \\ (0.443) \end{gathered}$ | $\begin{aligned} & \text { conditions } \\ & (0.201) \end{aligned}$ | $\begin{gathered} \text { men } \\ (0.463) \end{gathered}$ | williams (0.389) | states (0.404) | $\begin{aligned} & \text { torso } \\ & (0.487) \end{aligned}$ |
| 0 | may (1.0) | leads (1.0) | tournaments (1.0) | greatest (1.0) | football (1.0) | $\begin{gathered} \text { net } \\ (1.0) \end{gathered}$ |
| 2 | $\begin{gathered} \text { goal } \\ (0.98) \end{gathered}$ | tiredness (1.0) | events $(0.992)$ | female (0.98) | union (0.98) | $\begin{gathered} \text { shot } \\ (0.994) \end{gathered}$ |
| 3 | $\begin{gathered} \text { play } \\ (0.959) \end{gathered}$ | ineffectiveness (1.0) | $\begin{aligned} & \text { tour } \\ & (0.989) \end{aligned}$ | $\begin{aligned} & \text { ever } \\ & (0.971) \end{aligned}$ | $\begin{aligned} & \text { rugby } \\ & (0.979) \end{aligned}$ | $\begin{gathered} \text { serve } \\ (0.987) \end{gathered}$ |
| 4 | penalty <br> (0.954) | $\begin{aligned} & \text { recommences } \\ & (1.0) \end{aligned}$ | money (0.986) | navratilova (0.967) | $\begin{aligned} & \text { association } \\ & (0.96) \end{aligned}$ | $\begin{aligned} & \text { hit } \\ & (0.984) \end{aligned}$ |
| 5 | $\begin{gathered} \text { team } \\ (0.953) \end{gathered}$ | mandated (1.0) | $\begin{gathered} \text { prize } \\ (0.985) \end{gathered}$ | $\begin{aligned} & \text { tennis } \\ & (0.962) \end{aligned}$ | england $(0.958)$ | $\begin{aligned} & \text { stance } \\ & (0.955) \end{aligned}$ |
| 0 | penalty (1.0) | sole <br> (1.0) | tournament (1.0) | tennis (1.0) | rugby (1.0) | shot <br> (1.0) |
| 2 | referee <br> (0.985) | discretion (1.0) | events (0.98) | greatest <br> (0.962) | $\begin{aligned} & \text { football } \\ & (0.979) \end{aligned}$ | $\begin{gathered} \text { net } \\ (0.994) \end{gathered}$ |
| 3 | $\begin{gathered} \text { kick } \\ (0.985) \end{gathered}$ | synonym <br> (1.0) | $\begin{aligned} & \text { event } \\ & (0.978) \end{aligned}$ | $\begin{aligned} & \text { female } \\ & (0.953) \end{aligned}$ | union <br> (0.975) | $\begin{gathered} \text { serve } \\ (0.987) \end{gathered}$ |
| 4 | offence <br> (0.982) | violated (1.0) | $\begin{gathered} \text { atp } \\ (0.974) \end{gathered}$ | $\begin{gathered} \text { year } \\ (0.951) \end{gathered}$ | england $(0.961)$ | $\begin{aligned} & \text { hit } \\ & (0.983) \end{aligned}$ |
| 5 | $\begin{gathered} \text { foul } \\ (0.982) \end{gathered}$ | layout <br> (1.0) | money $(0.966)$ | navratilova (0.949) | wales (0.949) | stance (0.98) |



Figure A1. (a) 2-dimensional representation of word embeddings $A^{\prime}$ colored by topic assignment. (b) 2-dimensional representation of word embeddings $A^{\prime}$ colored by original Wikipedia article assignment (words that occur in more than one article are excluded). (c) Colored heatmap of affinity matrix $\boldsymbol{R}$.


Figure A2. (a) 2-dimensional representation of word embeddings $A^{\prime}$ colored by topic assignment. (b) 2-dimensional representation of word embeddings $A^{\prime}$ colored by original Wikipedia article assignment (words that occur in more than one article are excluded). (c) Colored heatmap of affinity matrix $\boldsymbol{R}$.

Articles: "Soccer", "Tennis", "Rugby".

Table A5. For each evaluated matrix factorization method we display the top 10 words for each topic and the 5 most similar words based on cosine similarity for the 2 top words from each topic.

|  |  | Topic 1 | Topic 2 | Topic 3 | Topic 4 | Topic 5 | Topic 6 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| NMF |  | \#511 | \#453 | \#575 | \#657 | \#402 | \#621 |
|  | 1 | net | referee | national | tournaments | rackets | rules |
|  | 2 | shot | penalty | south | doubles | balls | wingfield |
|  | 3 | serve | may | football | singles | made | december |
|  | 4 | hit | kick | cup | events | size | game |
|  | 5 | stance | card | europe | tour | must | sports |
|  | 6 | stroke | listed | fifa | prize | strings | lawn |
|  | 7 | backhand | foul | union | money | standard | modern |
|  | 8 | ball | misconduct | wales | atp | synthetic | greek |
|  | 9 | server | red | africa | men | leather | fa |
|  | 10 | service | offence | new | grand | width | first |
|  | 0 | net | referee | national | tournaments | rackets | rules |
|  | 1 | defensive | retaken | serbia | bruno | pressurisation | collection |
|  | 2 | closer | interference | gold | woodies | become | hourglass |
|  | 3 | somewhere | dismissed | north | eliminated | equivalents | unhappy |
|  | 4 | center | fully | headquarters | soares | size | originated |
|  | 0 | shot | penalty | south | doubles | balls |  |
|  | 1 | rotated | prior | asian | combining | express | experimenting |
|  | 2 | execute | yellow | argentina | becker | OZ | llanelidan |
|  | 3 | strive | duration | la | exclusively | bladder | attended |
|  | 4 | curve | primary | kong | woodbridge | length | antiphanes |
| LDA |  | \#413 | \#518 | \#395 | \#776 | \#616 | \#501 |
|  |  | used | net | wimbledon | world | penalty | clubs |
|  | 2 | forehand | ball | episkyros | cup | score | rugby |
|  | 3 | use | serve | occurs | tournaments | goal | schools |
|  | 4 | large | shot | grass | football | team | navratilova |
|  | 5 | notable | opponent | roman | fifa | end | forms |
|  | 6 | also | hit | bc | national | players | playing |
|  | 7 | western | lines | occur | international | match | sport |
|  | 8 | twohanded | server | ad | europe | goals | greatest |
|  | 9 | doubles | service | island | tournament | time | union |
|  | 10 | injury | may | believed | states | scored | war |
|  |  |  |  |  |  |  | clubs |
|  | 1 | seconds | mistaken | result | british | measure | sees |
|  | 2 | restrictions | diagonal | determined | cancelled | crossed | papua |
|  | 3 | although | hollow | exists | combined | requiring | admittance |
|  | 4 | use | perpendicular | win |  |  | forces |
|  | 0 |  |  |  |  |  |  |
|  | 1 | twohanded | long | roman | multiple | penalty | union |
|  | 2 | grips | deuce | bc | inline | bar | public |
|  | 3 | facetiously | position | island | fifa | fouled | took |
|  | 4 | woodbridge | allows | believed | manufactured | hour | published |
| SVD |  | \#1310 | \#371 | \#423 | \#293 | \#451 | \#371 |
|  | 1 | players | net | tournaments | stroke | greatest | balls |
|  | 2 | player | ball | singles | forehand | ever | rackets |
|  | 3 | tennis | shot | doubles | stance | female | size |
|  | 4 | also | serve | tour | power | wingfield | square |
|  | 5 | play | opponent | slam | backhand | williams | made |
|  | 6 | football | may | prize | torso | navratilova | leather |
|  | 7 | team | hit | money | grip | game | weight |
|  | 8 | first | service | grand | rotation | said | standard |
|  | 9 | one | hitting | events | twohanded | serena | width |
|  | 10 | rugby | line | ranking | used | sports | past |
|  |  |  | net | tournaments | stroke | greatest | balls |
|  | 1 | breaking | pace | masters | rotates | lived | panels |
|  | 2 | one | reach | lowest | achieve | female | sewn |
|  | 3 | running | underhand | events | face | biggest | entire |
|  | 4 | often | air | tour | adds | potential | leather |
|  | 0 |  |  |  |  |  |  |
|  | 1 | utilize | keep | indian | twohanded | autobiography | meanwhile |
|  | 2 | give | hands | doubles | begins | jack | laminated |
|  | 3 | converted | pass | pro | backhand | consistent | wood |
|  | 4 | touch | either | rankings |  | gonzales | strings |

## Appendix A.2. Additional Results on Wikipedia Data as Tensor Input

Wikipedia Articles "Soccer", "Bee", "Johnny Depp"—DEDICOM Automatic gradient method.


Figure A3. (a) 2-dimensional representation of word embeddings $A^{\prime}$ colored by topic assignment. (b) 2-dimensional representation of word embeddings $A^{\prime}$ colored by original Wikipedia article assignment (words that occur in more than one article are excluded).

Wikipedia Articles "Soccer", "Bee", "Johnny Depp"—DEDICOM Multiplicative Update Rules.


Figure A4. (a) 2-dimensional representation of word embeddings $A^{\prime}$ colored by topic assignment. (b) 2-dimensional representation of word embeddings $A^{\prime}$ colored by original Wikipedia article assignment (words that occur in more than one article are excluded).

Wikipedia Articles "Dolphin", "Shark", "Whale"—DEDICOM Multiplicative Update Rules.

Table A6. Each column lists the top 10 representative words per dimension of the basis matrix $A$.

|  | Topic 1 <br> \#226 | Topic 2 <br> \#628 | Topic 3 <br> \#1048 | Topic 4 <br> \#571 | Topic 5 <br> \#1267 | Topic 6 <br> \#554 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 1 | cells | mysticetes | shark | bony | dolphin | whaling |
|  | $(1.785)$ | $(1.808)$ | $(3.019)$ | $(1.621)$ | $(3.114)$ | $(3.801)$ |
| 2 | brain | whales | sharks | blood | dolphins | iwc |
|  | $(1.624)$ | $(1.791)$ | $(2.737)$ | $(1.452)$ | $(2.908)$ | (2.159) |
| 3 | light | feed | fins | fish | bottlenose | aboriginal |
|  | $(1.561)$ | $(1.427)$ | $(1.442)$ | $(1.438)$ | $(1.629)$ | $(2.098)$ |
| 4 | cone | baleen | killed | gills | meat | canada |
|  | $(1.448)$ | $(1.33)$ | $(1.407)$ | $(1.206)$ | $(1.403)$ | $(1.912)$ |
| 5 | allow | odontocetes | endangered | teeth | behavior | moratorium |
|  | $(1.32)$ | $(1.278)$ | $(1.377)$ | $(1.088)$ | $(1.399)$ | $(1.867)$ |
|  | greater | consist | hammerhead | body | captivity | industry |
| 6 | $(1.292)$ | $(1.162)$ | $(1.269)$ | $(1.043)$ | $(1.298)$ | $(1.855)$ |
|  | slightly | water | conservation | system | river | us |
| 7 | $(1.269)$ | $(1.096)$ | $(1.227)$ | $(1.027)$ | $(1.281)$ | $(1.838)$ |
|  | ear | krill | trade | skeleton | common | belugas |
| 8 | $(1.219)$ | $(1.05)$ | $(1.226)$ | $(1.008)$ | $(1.275)$ | $(1.585)$ |
|  | cornea | toothed | whitetip | called | selfawareness | whale |
| 9 | $(1.158)$ | $(1.003)$ | $(1.203)$ | $(0.99)$ | $(1.248)$ | $(1.542)$ |
|  | rod | sperm | finning | tissue | often | gb£ |
| 10 | $(1.128)$ | $(0.991)$ | $(1.184)$ | $(0.875)$ | $(1.218)$ | $(1.528)$ |

Table A7. For the most significant two words per topic, the four nearest neighbors based on cosine similarity are listed.

|  | Topic 1 | Topic 2 | Topic 3 | Topic 4 | Topic 5 | Topic 6 |
| :--- | :---: | :---: | :---: | :---: | :---: | :---: |
| 0 | cells | mysticetes | shark | bony | dolphin | whaling |
|  | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ |
| 1 | sensitive | unborn | native | edges | hybrid | māori |
|  | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ |
| 2 | cone | grind | tl | mirabile | hybridization | trips |
|  | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ |
| 3 | rod | counterparts | predators-organisms | matches | yangtze | predominantly |
|  | $(0.998)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ |
| 4 | corneas | threechambered | cretaceous | turbulence | grampus | revenue |
|  | $(0.998)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ |
|  | brain | whales | sharks | blood | dolphins | iwc |
| 0 | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ |
|  | receive | extended | reminiscent | hydrodynamic | superpod | distinction |
| 1 | $(0.998)$ | $(0.996)$ | $(1.0)$ | $(0.998)$ | $(1.0)$ | $(1.0)$ |
| 2 | equalizer | bryde | electrical | scattering | masturbation | billion |
|  | $(0.998)$ | $(0.996)$ | $(1.0)$ | $(0.998)$ | $(1.0)$ | $(1.0)$ |
| 3 | lobes | closes | induced | reminder | interaction | spain |
|  | $(0.997)$ | $(0.996)$ | $(1.0)$ | $(0.998)$ | $(1.0)$ | $(1.0)$ |
| 4 | clear | effects | coarsely | $(1.0)$ | flows | stressful |
|  | $(0.997)$ | $(0.996)$ | $(0.998)$ | $(1.0)$ | competition | $(1.0)$ |



Figure A5. Colored heatmap of affinity tensor $\boldsymbol{R}$.

(a)
(b)

Figure A6. (a) 2-dimensional representation of word embeddings $A$ colored by topic assignment. (b) 2-dimensional representation of word embeddings $A$ colored by original Wikipedia article assignment (words that occur in more than one article are excluded).

Wikipedia Articles "Soccer", "Tennis", "Rugby"—DEDICOM Multiplicative Update Rules.


Figure A7. Colored heatmap of affinity tensor $\boldsymbol{R}$.

Table A8. Each column lists the top 10 representative words per dimension of the basis matrix $A$.

|  | Topic 1 <br> \#441 | Topic 2 \#861 | Topic 3 \#412 | Topic 4 \#482 | Topic 5 \#968 | Topic 6 \#57 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 1 | rugby | titles | rackets | net | penalty | doubles |
|  | (2.55) | (1.236) | (2.176) | (2.767) | (1.721) | (2.335) |
| 2 | union | wta | wingfield | shot | football | singles |
|  | (2.227) | (1.196) | (1.536) | (2.586) | (1.701) | (2.321) |
| 3 | wales | circuit | modern | serve | team | tournaments |
|  | (1.822) | (1.123) | (1.513) | (2.393) | (1.507) | (2.245) |
| 4 | georgia | futures | racket | hit | laws | tennis |
|  | (1.682) | (1.122) | (1.43) | (1.978) | (1.462) | (1.752) |
| 5 | fiji | earn | th | stance | referee | grand |
|  | (1.557) | (1.104) | (1.355) | (1.945) | (1.449) | (1.662) |
| 6 | samoa | offer | lawn | service | fifa | events |
|  | (1.474) | (1.096) | (1.316) | (1.83) | (1.439) | (1.648) |
| 7 | zealand | mixed | century | stroke | may | slam |
|  | (1.458) | (1.089) | (1.236) | (1.797) | (1.435) | (1.623) |
| 8 | new | draws | strings | server | goal | player |
|  | (1.414) | (1.085) | (1.179) | (1.761) | (1.353) | (1.344) |
| 9 | tonga | atp | yielded | backhand | competitions | professional |
|  | (1.374) | (1.072) | (1.121) | (1.692) | (1.345) | (1.328) |
| 10 | south | challenger | balls | forehand | associations | players |
|  | (1.369) | (1.07) | (1.101) | (1.554) | (1.288) | (1.316) |

Table A9. For the most significant two words per topic, the four nearest neighbors based on cosine similarity are listed.

|  | Topic 1 | Topic 2 | Topic 3 | Topic 4 | Topic 5 | Topic 6 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 0 | rugby | titles | rackets | n | penalty | doubles |
|  | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) |
| 1 | ireland | hopman | proximal | hit | organisers | singles |
| 2 | union | dress | interlaced | formally | elapsed | tournaments |
| 2 | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (0.985) |
| 3 | backfired | tennischannel | harry | offensive | polite | grand |
|  | (1.0) | (0.998) | (1.0) | (1.0) | (1.0) | (0.975) |
| 4 | kilopascals <br> (1.0) | $\begin{gathered} \text { seoul } \\ (0.998) \end{gathered}$ | deserves (1.0) | deeply (1.0) | modest (1.0) | $\begin{gathered} \text { slam } \\ (0.971) \end{gathered}$ |
| 0 | union | wta | wingfield | shot | football | singles |
| 0 | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) |
| 1 | rugby | helps | proximal | requires | circumference | doubles |
| 1 | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) |
| 2 | ireland <br> (1.0) | hamilton (1.0) | interlaced (1.0) | backwards (1.0) | touchline (1.0) | tournaments (0.985) |
| 3 | backfired | weeks |  |  | sanctions |  |
| 3 | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (0.975) |
| 4 | zealand (1.0) | couple <br> (1.0) | deserves (1.0) | torso (1.0) | $\begin{gathered} \text { home } \\ (0.999) \end{gathered}$ | $\begin{gathered} \text { slam } \\ (0.971) \end{gathered}$ |



Figure A8. (a) 2-dimensional representation of word embeddings $A$ colored by topic assignment. (b) 2-dimensional representation of word embeddings $A$ colored by original Wikipedia article assignment (words that occur in more than one article are excluded).

> Wikipedia Articles "Soccer", "Tennis", "Rugby"—TNMF.

Table A10. Each column lists the top 10 representative words per dimension of the basis matrix $\boldsymbol{A}^{\prime}$.

|  | Topic 1 <br> \#275 | Topic 2 <br> \#505 | Topic 3 <br> \#607 | Topic 4 <br> \#459 | Topic 5 <br> \#816 | Topic 6 <br> \#559 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 1 | greatest | rackets | tournaments | net | football | penalty |
|  | $(39.36)$ | $(33.707)$ | $(29.126)$ | $(29.789)$ | $(27.534)$ | $(27.793)$ |
| 2 | ever | modern | events | shot | rugby | referee |
|  | $(26.587)$ | $(24.281)$ | $(25.327)$ | $(27.947)$ | $(24.037)$ | $(23.632)$ |
| 3 | female | balls | tour | serve | union | goal |
|  | $(25.52)$ | $(22.016)$ | $(23.488)$ | $(25.722)$ | $(21.397)$ | $(23.072)$ |
| 4 | navratilova | wingfield | prize | hit | south | may |
|  | $(24.348)$ | $(20.923)$ | $(21.823)$ | $(21.344)$ | $(20.761)$ | $(22.978)$ |
| 5 | best | tennis | atp | stance | national | team |
|  | $(24.114)$ | $(19.863)$ | $(21.124)$ | $(20.75)$ | $(19.586)$ | $(21.258)$ |
| 6 | williams | strings | money | service | fifa | kick |
|  | $(22.207)$ | $(18.602)$ | $(20.667)$ | $(19.7)$ | $(19.331)$ | $(21.052)$ |
| 7 | serena | racket | doubles | server | wales | foul |
|  | $(21.256)$ | $(18.369)$ | $(19.919)$ | $(19.051)$ | $(18.627)$ | $(19.018)$ |
| 8 | said | made | ranking | stroke | league | listed |
|  | $(20.666)$ | $(17.622)$ | $(19.736)$ | $(18.781)$ | $(18.31)$ | $(17.736)$ |
| 9 | martina | yielded | us | backhand | cup | free |
|  | $(20.153)$ | $(17.284)$ | $(19.431)$ | $(17.809)$ | $(17.015)$ | $(17.702)$ |
| 10 | budge | th | masters | ball | association | goals |
|  | $(20.111)$ | $(16.992)$ | $(18.596)$ | $(17.2)$ | $(16.721)$ | $(17.209)$ |
|  |  |  |  |  |  |  |

Table A11. For the most significant two words per topic, the four nearest neighbors based on cosine similarity are listed.

|  | Topic 1 | Topic 2 | Topic 3 | Topic 4 | Topic 5 | Topic 6 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 0 | greatest | rackets | tournaments | net | football | penalty |
|  | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ |
| 1 | illustrated | garden | us | lob | midlothian | whole |
|  | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ |
| 2 | johansson | construction | earned | receiving | alcock | corner |
|  | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ |
| 3 | wilton | yielded | partipating | rotates | capital | offender |
|  | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ |
| 4 | jonathan | energy | receives | adds | representatives | stoke |
|  | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ |
|  | ever | modern | events | shot | rugby | referee |
| 0 | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ |
|  | deserved | design | juniors | lobber | slang | dismissed |
| 1 | $(1.0)$ | $(0.999)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ |
|  | stated | version | bowl | unable | colonists | showing |
| 2 | $(1.0)$ | $(0.999)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ |
|  | female | shape | comprised | alter | sevenaside | stoppage |
| 3 | $(1.0)$ | $(0.998)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ |
|  | contemporaries | stitched | carlo | applying | seldom | layout |
| 4 | $(1.0)$ | $(0.998)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ | $(1.0)$ |
|  |  |  |  |  |  |  |


(a)
(b)

Figure A9. (a) 2-dimensional representation of word embeddings $H$ colored by topic assignment. (b) 2-dimensional representation of word embeddings $H$ colored by original Wikipedia article assignment (words that occur in more than one article are excluded).

Wikipedia Articles "Dolphin", "Shark", "Whale"—TNMF.

Table A12. Each column lists the top 10 representative words per dimension of the basis matrix $H$.

|  | $\begin{gathered} \hline \text { Topic } 1 \\ \# 675 \end{gathered}$ | $\underset{\# 996}{\text { Topic } 2}$ | $\begin{gathered} \text { Topic } 3 \\ \# 279 \end{gathered}$ | $\begin{gathered} \text { Topic } 4 \\ \# 491 \end{gathered}$ | Topic 5 \#1190 | $\underset{\# 663}{\text { Topic } 6}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 1 | whaling (34.584) | sharks (30.418) | $\begin{aligned} & \text { young } \\ & \text { (33.823) } \end{aligned}$ | $\begin{gathered} \text { killed } \\ (23.214) \end{gathered}$ | dolphin <br> (35.52) | mysticetes (24.404) |
| 2 | whale | fish | born | shark | dolphins | flippers |
|  | (25.891) | (23.648) | (27.62) | (22.6) | (31.881) | (22.059) |
| 3 | whales (21.653) | $\begin{gathered} \text { bony } \\ (19.689) \end{gathered}$ | oviduct (23.706) | states (21.24) | bottlenose (18.198) | odontocetes (21.621) |
| 4 | belugas | prey | viviparity | endangered | behavior | water |
|  | (20.933) | (18.785) | (23.694) | (20.976) | (18.003) | (21.087) |
| 5 | aboriginal (19.44) | $\begin{gathered} \text { teeth } \\ (18.242) \end{gathered}$ | embryos (22.966) | conservation (20.398) | selfawareness (16.48) | $\begin{gathered} \text { tail } \\ (18.268) \end{gathered}$ |
| 6 | iwc | blood | continue | fins | meat | mya |
| 6 | (19.226) | (16.521) | (21.752) | (18.641) | (16.02) | (17.79) |
| 7 | canada | gills | calves | new | often | baleen |
|  | (18.691) | (13.34) | (21.25) | (18.445) | (15.687) | (17.189) |
| 8 | $\stackrel{\text { arctic }}{ }$ | tissue | blubber | international | captivity | limbs |
|  | (17.406) | (12.927) | (21.094) | (18.4) | (15.452) | (16.56) |
| 9 | industry |  |  | drum | river | allow |
|  | (16.837) | (12.691) | (20.735) | (17.587) | (14.68) | (16.552) |
| 10 | $\begin{gathered} \text { right } \\ (16.766) \end{gathered}$ | skeleton (12.52) | $\begin{aligned} & \text { fluids } \\ & (20.662) \end{aligned}$ | $\begin{aligned} & \text { finning } \\ & (17.321) \end{aligned}$ | $\begin{gathered} \text { common } \\ (14.389) \end{gathered}$ | $\begin{aligned} & \text { toothed } \\ & \text { (16.489) } \end{aligned}$ |

Table A13. For the most significant two words per topic, the four nearest neighbors based on cosine similarity are listed.

|  | Topic 1 | Topic 2 | Topic 3 | Topic 4 | Topic 5 | Topic 6 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 0 | whaling | sharks | young | killed | dolphin | mysticetes |
| 0 | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) |
| 1 | antarctica | loan | getting | alzheimer | behaviors | digits |
|  | (1.0) | $\stackrel{(1.0)}{ }$ | $\stackrel{(1.0)}{\text { insulation }}$ |  |  | $\stackrel{(1.0)}{\text { streamlined }}$ |
| 2 | spain | leopard | insulation | queensland | familiar | streamlined |
| 2 | (1.0) | (1.0) | (1.0) | (1.0) |  |  |
| 3 | caro <br> (1.0) | dogfish | harsh <br> (1.0) | als <br> (1.0) | pantropical | archaeocete (1.0) |
| 4 | excluded | lifespans | primary | control | test | defines |
| 4 | (1.0) | (0.999) | (1.0) | (1.0) | (1.0) | (0.999) |
| 0 | whale | fish | born | shark | dolphins | flippers |
| 0 | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) |
| 1 | reason | like | getting | figure | levels | expel |
|  | (1.0) | (0.997) | (1.0) | (1.0) | (1.0) | (1.0) |
| 2 | respected | lifetime | young | sources | moderate | compress |
|  | (1.0) | (0.992) | (1.0) | (1.0) | (0.999) | (1.0) |
| 3 | divinity | content | leaner | video | injuries | protocetus |
|  | (0.999) | (0.992) | (1.0) | (0.998) | (0.999) | (1.0) |
| 4 | $\begin{aligned} & \text { taken } \\ & (0.998) \end{aligned}$ | hazardous (0.992) | insulation (1.0) | dogfishes <br> (0.997) | $\begin{aligned} & \text { seems } \\ & (0.998) \end{aligned}$ | nostrils <br> (1.0) |



Figure A10. (a) 2-dimensional representation of word embeddings $H$ colored by topic assignment. (b) 2-dimensional representation of word embeddings $\boldsymbol{H}$ colored by original Wikipedia article assignment (words that occur in more than one article are excluded).

Wikipedia Articles "Soccer", "Bee", "Johnny Depp"—TNMF.
Table A14. Each column lists the top 10 representative words per dimension of the basis matrix $\boldsymbol{H}$.

|  | Topic 1 <br> $\# 793$ | Topic 2 <br> \#554 | Topic 3 <br> \#736 | Topic 4 <br> \#601 | Topic 5 <br> \#740 | Topic 6 <br> \#616 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 1 | film | ball | honey | football | heard | species |
|  | $(37.29)$ | $(27.588)$ | $(29.778)$ | $(32.591)$ | $(37.167)$ | $(32.973)$ |
| 2 | starred | may | insects | fifa | depp | eusocial |
|  | $(23.821)$ | $(25.768)$ | $(27.784)$ | $(25.493)$ | $(30.275)$ | $(25.001)$ |
| 3 | role | penalty | bees | world | court | females |
|  | $(23.006)$ | $(25.04)$ | $(27.679)$ | $(25.414)$ | $(20.771)$ | $(24.24)$ |
| 4 | series | players | bee | cup | divorce | solitary |
|  | $(19.563)$ | $(24.063)$ | $(26.936)$ | $(24.925)$ | $(17.289)$ | $(21.173)$ |
| 5 | burton | referee | food | association | sued | nest |
|  | $(18.694)$ | $(23.649)$ | $(23.44)$ | $(22.331)$ | $(16.105)$ | $(20.198)$ |
| 6 | played | team | flowers | national | stated | males |
|  | $(17.583)$ | $(22.9)$ | $(22.374)$ | $(20.958)$ | $(15.984)$ | $(18.3)$ |
| 7 | character | goal | pollination | women | alcohol | workers |
|  | $(16.646)$ | $(22.859)$ | $(18.09)$ | $(20.668)$ | $(15.238)$ | $(17.16)$ |
| 8 | success | player | larvae | international | stating | typically |
|  | $(16.41)$ | $(22.054)$ | $(17.73)$ | $(20.16)$ | $(15.199)$ | $(16.886)$ |
| 9 | films | play | pollen | tournament | paradis | colonies |
|  | $(15.74)$ | $(21.774)$ | $(17.666)$ | $(18.26)$ | $(14.98)$ | $(16.528)$ |
| 10 | box | game | predators | uefa | alleged | queens |
|  | $(15.024)$ | $(20.471)$ | $(17.634)$ | $(18.029)$ | $(14.971)$ | $(16.427)$ |

Table A15. For the most significant two words per topic, the four nearest neighbors based on cosine similarity are listed.

|  | Topic 1 | Topic 2 | Topic 3 | Topic 4 | Topic 5 | Topic 6 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 0 | film | ball | honey | football | heard | species |
|  | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) |
| 1 | avril | officials | triangulum | entered | obtained | progressive |
|  | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) |
| 2 | $\begin{gathered} \text { office } \\ \text { (1.0) } \end{gathered}$ | invoke (1.0) | consumption (1.0) | $\begin{aligned} & \text { most } \\ & (1.0) \end{aligned}$ | countersued (1.0) | halictidae (1.0) |
| 3 | landau | heading | copper | excess | depths | temperate |
| 3 | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) |
| 4 | chamberlain (1.0) | twohalves (1.0) | might <br> (1.0) | $\begin{gathered} \text { uk } \\ (1.0) \end{gathered}$ | mismanagement (1.0) | spring <br> (1.0) |
|  | starred | may | insects | fifa | depp | eusocial |
| 0 | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) |
| 1 | raimi | noninternational | blooms | oceania | city | unfertilized |
| 1 | (1.0) | (0.992) | (1.0) | (1.0) | (1.0) | (1.0) |
| 2 | candidate | red | eats | sudamericana | tribute | females |
|  | (1.0) | (0.991) | (1.0) | ${ }^{(1.0)}$ | (1.0) | (1.0) |
| 3 | hardwicke | required | catching | widened | mick | paper |
|  | (1.0) | (0.991) | (1.0) | (1.0) | (1.0) | ${ }_{\text {chiber }}$ |
| 4 | peter (1.0) | $\begin{gathered} \mathrm{yd} \\ (0.989) \end{gathered}$ | disease (1.0) | oversee (1.0) | $\begin{aligned} & \text { elvis } \\ & \text { (1.0) } \end{aligned}$ | hibernate (1.0) |

Appendix A.3. Additional Results on Amazon Review Data as Tensor Input Amazon Reviews-DEDICOM Mulitplicative Update Rules.

(a)
(b)

Figure A11. (a) 2-dimensional representation of word embeddings $H$ colored by topic assignment. (b) 2-dimensional representation of word embeddings $\boldsymbol{H}$ colored by original Wikipedia article assignment (words that occur in more than one article are excluded).


Figure A12. (a) 2-dimensional representation of word embeddings $A$ colored by topic assignment. (b) 2-dimensional representation of word embeddings $A$ colored by original review article.


Figure A13. (a) 2-dimensional representation of word embeddings $H$ colored by topic assignment. (b) 2-dimensional representation of word embeddings $\boldsymbol{H}$ colored by original review article.

Table A16. For the most significant two words per topic, the four nearest neighbors based on cosine similarity are listed.

|  | Topic 1 | Topic 2 | Topic 3 | Topic 4 | Topic 5 | Topic 6 | Topic 7 | Topic 8 | Topic 9 | Topic 10 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 0 | anna | shen | legendary | lasseter | disc | screams | code | mike | woody | po |
|  | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) |
| 1 | christoph | canons | sacred <br> (0.999) | andrew | thxcertified | harvested | confirm | bogg | rips <br> (10) | panda <br> (0.985) |
|  | (1.0) | (1.0) | $(0.999)$ | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (0.985) |
| 2 | readiness (1.0) | yeoh <br> (1.0) | fulfill <br> (0.999) | stanton <br> (1.0) | presentation <br> (0.999) | speciallytrained <br> (1.0) | discount <br> (1.0) | chased (0.996) | spy | $\begin{gathered} \text { fu } \\ (0.983) \end{gathered}$ |
|  | (1.0) | (1.0) | (0.999) |  |  |  |  | (0.996) | (1.0) | (0.983) |
| 3 | $\begin{gathered} \text { carrots } \\ (1.0) \end{gathered}$ | $\begin{aligned} & \text { Wolf } \\ & \text { (1.0) } \end{aligned}$ | $\begin{aligned} & \text { roster } \\ & (0.999) \end{aligned}$ | (1.0) | $\underset{(0.999)}{\text { upgrade }}$ | screamprocessing <br> (1.0) | browser <br> (1.0) | $\begin{aligned} & \text { tlair } \\ & (0.996) \end{aligned}$ | $\begin{aligned} & \text { josie } \\ & (1.0) \end{aligned}$ | $\begin{gathered} \text { black } \\ (0.983) \end{gathered}$ |
| 4 | poverty <br> (1.0) | weapon (1.0) | megafan (0.999) | $\begin{gathered} \text { uncredited } \\ (1.0) \end{gathered}$ | featurettes (0.998) | corporation (1.0) | popup (1.0) | $\begin{gathered} \text { slot } \\ (0.995) \end{gathered}$ | supurb <br> (1.0) | $\begin{aligned} & \text { kung } \\ & (0.981) \end{aligned}$ |
| 0 | elsa | peacock | valley | director | birds | energy | email | crystal | buzz | master |
|  | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) |
| 1 | shipwreck | shen | praying | producer | pressed | powered | confirm | oz | hist | shifu |
|  | (1.0) | (1.0) | (0.999) | (0.998) | (1.0) | (0.998) | (1.0) | (1.0) | (1.0) | (0.999) |
| 2 |  | mcbride | kim | teaser | gadget | frightened | code | mae | wayne | warrior |
|  | (1.0) | (1.0) | (0.997) | (0.997) | (1.0) | (0.997) | (1.0) | $(1.0)$ | (1.0) | $(0.999)$ |
| 3 | idena | yeoh | chorgum | globes | classically | screams | fwiw | celia | reunited | dragon |
|  | (1.0) | (1.0) | (0.997) | (0.997) | (1.0) | (0.996) | (1.0) | (1.0) | (1.0) | (0.998) |
| 4 | prodding | michelle | preying | rousing | starz | scarry | android | cristal | hockey | martial |
|  | (1.0) | (1.0) | (0.997) | (0.997) | (1.0) | (0.996) | (1.0) | (1.0) | (1.0) | (0.994) |

Table A17. For the most significant two words per topic, the four nearest neighbors based on cosine similarity are listed.

|  | Topic 1 | Topic 2 | Topic 3 | Topic 4 | Topic 5 | Topic 6 | Topic 7 | Topic 8 | Topic 9 | Topic 10 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 0 | anna | shen | legendary | lasseter | disc | screams | code | mike | woody | (1.0) |
|  | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) |
| 1 | christoph | canons | sacred | andrew | thxcertified | harvested | confirm | bogg | rips | panda |
|  | $(1.0)^{\mathrm{r}}$ | (1.0) | (0.999) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (0.985) |
| 2 | readiness | yeoh | fulfill | stanton | presentation | speciallytrained | discount | chased | spy | fu |
|  | (1.0) | (1.0) | (0.999) | $(1.0)$ | (0.999) | (1.0) | (1.0) | (0.996) | (1.0) | $(0.983)$ |
| 3 | carrots | wolf | roster | eggleston | upgrade | screamprocessing | browser | flair | josie | black |
|  | (1.0) | (1.0) | (0.999) | (1.0) | (0.999) | (1.0) | (1.0) | (0.996) | (1.0) | (0.983) |
| 4 | poverty (1.0) | weapon (1.0) | megafan (0.999) | uncredited (1.0) | featurettes (0.998) | corporation (1.0) | popup $(1.0)$ | $\begin{gathered} \text { slot } \\ (0.995) \end{gathered}$ | supurb (1.0) | $\begin{gathered} \text { kung } \\ (0.981) \end{gathered}$ |
| 0 | elsa | peacock | valley | director | birds | energy | email | crystal | buzz | master |
|  | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) |
| 1 | shipwreck | shen | praying | producer | pressed | powered | confirm | OZ | hist | shifu |
|  | (1.0) | (1.0) | (0.999) | (0.998) | (1.0) | (0.998) | (1.0) | (1.0) | (1.0) | (0.999) |
| 2 | marriage | mcbride | kim | teaser |  |  |  | mae | wayne | warrior |
|  | (1.0) | (1.0) | (0.997) | (0.997) | (1.0) | (0.997) | (1.0) | (1.0) | (1.0) | (0.999) |
| 3 | idena |  | chorgum | globes | classically | screams | fwiw | celia | reunited | dragon |
|  | (1.0) | (1.0) | (0.997) | (0.997) | (1.0) | (0.996) | (1.0) | (1.0) | (1.0) | (0.998) |
| 4 | prodding | michelle | preying | rousing | starz | scarry | android | cristal | hockey | martial |
|  | (1.0) | (1.0) | (0.997) | (0.997) | (1.0) | (0.996) | (1.0) | (1.0) | (1.0) | (0.994) |

Amazon Reviews-TNMF.

Table A18. Each column lists the top 10 representative words per dimension of the basis matrix $\boldsymbol{H}$.

|  | Topic 1 $\# 590$ | Topic 2 $\# 1052$ | Topic 3 \#456 | $\begin{gathered} \text { Topic } 4 \\ \# 350 \end{gathered}$ | Topic 5 \#4069 | $\underset{\# 733}{\text { Topic } 6}$ | Topic 7 <br> \#1140 | Topic 8 \#582 | Topic 9 \#423 | $\begin{gathered} \text { Topic } 10 \\ \# 605 \end{gathered}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 1 | anna | woody | director | allen | widescreen | code | master | mike | film | screams |
|  | (109.81) | (134.366) | (100.622) | (83.875) | (34.484) | (88.94) | (89.686) | (88.628) | (58.514) | (87.782) |
| 2 | elsa | buzz | lasseter | hanks | outtakes | email | po | crystal | animation | energy |
|  | (106.148) | (120.93) | (93.134) | (77.313) | (30.724) | (73.645) | (85.79) | (82.472) | (53.628) | (78.113) |
| 3 | olaf | andy | andrew | tim | disc | promo | shifu | billy | characters | world |
|  | (59.353) | (105.728) | (81.452) | (75.511) | (30.688) | (67.483) | (82.465) | (79.244) | (46.861) | (73.888) |
| 4 | trolls | toys | stanton | rickles | extras | amazon | warrior | goodman | films | monstropolis |
|  | (58.811) | (98.523) | (80.132) | (74.217) | (30.09) | (64.978) | (75.609) | (76.831) | (44.937) | (73.484) |
| 5 | kristoff | lightyear | john | tom | versions | promotion | dragon | sully | pixar | monsters |
|  | (56.309) | (68.336) | (73.004) | (72.401) | (27.894) | (58.631) | (74.235) | (75.612) | (44.873) | (71.721) |
| 6 | hans | sid | pete |  | included |  |  | wazowski |  |  |
|  | (55.628) | (52.34) | (70.556) | (69.776) | (27.455) | (58.207) | (71.721) | (71.588) | (44.176) | (71.352) |
| 7 | frozen | cowboy | docter | varney | material | promotional | lung | randall | animated | power |
|  | (54.257) | (48.588) | (64.734) | (66.053) | (26.887) | (57.738) | (70.786) | (69.695) | (43.492) | (70.642) |
| 8 | queen | space | ralph | slinky | edition | click | furious | sulley | also | monster |
|  | (53.956) | (47.88) | (54.884) | (62.326) | (26.546) | (55.373) | (63.232) | (69.604) | (43.484) | (70.197) |
| 9 | sister | room | joe | potato | contains | download | oogway | james | dvd | closet |
|  | (52.749) | (42.655) | (53.7) | (62.237) | (25.386) | (50.788) | (60.879) | (68.574) | (42.736) | (61.451) |
| 10 | $\begin{gathered} \text { ice } \\ (4971) \end{gathered}$ | $\begin{gathered} \text { toy } \\ (42.042) \end{gathered}$ | ranft <br> (53.41) | $\begin{gathered} \mathrm{mr} \\ (61.801) \end{gathered}$ | extra <br> (25.144) | purchase <br> (50.327) | $\begin{aligned} & \text { five } \\ & \text { (59 259) } \end{aligned}$ | buscemi <br> (66.028) | well <br> (40.124) | scare <br> (61.243) |
|  | (49.71) | (42.042) | (53.41) | (61.801) | (25.144) | (50.327) | (59.259) | (66.028) | (40.124) | (61.243) |

Table A19. For the most significant two words per topic, the four nearest neighbors based on cosine similarity are listed.

|  | Topic 1 | Topic 2 | Topic 3 | Topic 4 | Topic 5 | Topic 6 | Topic 7 | Topic 8 | Topic 9 | Topic 10 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 0 | anna | woody | director | allen | widescreen | code | master | mike | film | screams |
|  | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) |
| 1 | marriage | acciently | producer | trustworthy | benefactors | card | furious | longtime | films | electrical |
|  | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (0.995) | (1.0) | (1.0) |
|  | trolls | limp | jackson | arguments | pioneers | confirmation | dragon | cyclops | first | screamprocessing |
| 2 | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (0.995) | (0.994) | (1.0) |
| 3 | flees | jealousey | rabson | hanks | keepcase | assuming | shifu | slot | animated | chlid |
|  | (1.0) | (0.999) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (0.995) | (0.993) | (1.0) |
| 4 | christian (1.0) | swells (0.999) | composer (1.0) | knowitall | redone (1.0) | android | warrior <br> (1.0) | humanlike (0.994) | animation (0.989) | shortage |
| 0 | elsa | buzz | lasseter | hanks | outtakes | email | po | crystal | animation | energy |
|  | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) |
|  | marriage | recive | nathan | trustworthy | storyboarding | promo | martial | billy | film | supply |
| 1 | (1.0) | (0.999) | (1.0) | (1.0) | (0.993) | (1.0) | (0.998) | (1.0) | (0.989) | (1.0) |
| 2 | heals | zorg | officer | allen | informative | avail | fight | talkative | story | powered |
|  | (1.0) | (0.999) | (1.0) | (1.0) | (0.991) | (1.0) | (0.997) | (0.999) | (0.987) | (1.0) |
| 3 | marrying | limp | cunningham | tom | contents | flixster | arts | competitor | scenes | collect |
| 3 | (1.0) | (0.997) | (1.0) | (1.0) | (0.99) | (1.0) | (0.996) | (0.999) | (0.987) | (1.0) |
| 4 | feminist (1.0) | acciently <br> (0.997) | derryberry <br> (1.0) | arguments <br> (1.0) | $\begin{gathered} \text { logo } \\ (0.99) \end{gathered}$ | confirming <br> (1.0) | adopted <br> (0.995) | devilishly (0.999) | also (0.984) | screams <br> (0.999) |

Appendix A.4. Additional Results on the New York Times News Article Data as Tensor Input New York Times News Articles - DEDICOM Mulitplicative Update Rules.

Table A20. For the most significant two words per topic, the four nearest neighbors based on cosine similarity are listed.

|  | Topic 1 | Topic 2 | Topic 3 | Topic 4 | Topic 5 | Topic 6 | Topic 7 | Topic 8 | Topic 9 | Topic 10 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 0 | suleimani | loans | masks | floyd | contributed | confederate | ukraine | storm | restaurants | weinstein |
|  | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | 1.0) | (1.0) | (1.0) | (1.0) |
| 1 | qassim | spend | sanitizer | brutality | alan | statue | lutsenko | storms | salons | raped |
|  | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) |
| 2 | iran (1.0) | smallbusiness (1.0) | wipes (1.0) | police <br> (1.0) | edmondson (1.0) | monuments (1.0) | ukrainians (1.0) | isaias <br> (1.0) | cafes (1.0) | predatory (1.0) |
| 3 | iranian | rent | cloth | systemic | mervosh | statues | yovanovitch | landfall | pubs | mann |
| 3 | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) |
| 4 | militias (1.0) | incentives (1.0) | homemade (1.0) | knee <br> (1.0) | emily <br> (1.0) | honoring (1.0) | burisma (1.0) | forecasters (1.0) | nightclubs (1.0) | sciorra <br> (1.0) |
| 0 | $\begin{aligned} & \text { iran } \\ & (1.0) \end{aligned}$ | university (1.0) | protective (1.0) | minneapolis (1.0) | reporting (1.0) | statue (1.0) | sondland (1.0) | hurricane (1.0) | bars (1.0) | sexual (1.0) |
| 1 | qassim | jerome | gowns | breonna | rabin | monuments | zelensky | bahamas | dining | rape |
|  | (1.0) | (0.999) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) |
| 2 | suleimani <br> (1.0) | $\begin{aligned} & \text { oxford } \\ & (0.998) \end{aligned}$ | ventilators <br> (1.0) | kueng (1.0) | contributed (1.0) | statues (1.0) | volker (1.0) | hurricanes (1.0) | theaters (1.0) | metoo $(1.0)$ |
| 3 | iranian | columbia | respirators | floyd | keith | confederate | giuliani | forecasters | venues | sexually |
|  | (1.0) | (0.998) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) |
| 4 | militias (1.0) | economics (0.998) | supplies (1.0) | police <br> (1.0) | chokshi <br> (1.0) | honoring (1.0) | quid (1.0) | landfall <br> (1.0) | malls <br> (1.0) | mann (1.0) |

## New York Times News Articles - TNMF.

Table A21. Each column lists the top 10 representative words per dimension of the basis matrix $A$.

|  | $\begin{gathered} \text { Topic } 1 \\ \# 977 \end{gathered}$ | $\begin{gathered} \text { Topic } 2 \\ \# 360 \end{gathered}$ | Topic 3 $\# 420$ | Topic 4 $\# 4192$ | $\begin{gathered} \text { Topic } 5 \\ \# 489 \end{gathered}$ | $\begin{gathered} \text { Topic } 6 \\ \# 405 \end{gathered}$ | Topic 7 $\# 1135$ | $\begin{gathered} \text { Topic } 8 \\ \# 748 \end{gathered}$ | Topic 9 $\# 108$ | $\begin{gathered} \hline \text { Topic } 10 \\ \# 1166 \end{gathered}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 1 | floyd (82.861) | contributed | iran 80.007) | masks <br> (40.463) | ship <br> (87.178) | syria (94.686) | senator <br> (77.285) | restaurants (93.511) | bloom (110.77) | ukraine <br> (79.611) |
| 2 | police | reporting | suleimani | patients | crew | syrian | storm | bars | julie | sondland |
| 2 | (64.649) | (84.889) | (78.78) | (34.77) | (70.694) | (82.565) | (43.439) | (64.889) | (103.282) | (61.099) |
| 3 | protesters | michael | iranian | ventilators | aboard | kurdish | hurricane | reopen | edited (100.159) | testimony |
|  | (63.588) | (76.156) | (72.581) | (34.299) | (67.464) | (82.013) | (42.213) | (57.541) | (100.159) | (49.982) |
| 4 | minneapolis (63.216) | $\begin{gathered} \text { katie } \\ (63.146) \end{gathered}$ | $\operatorname{ciraq}_{(63.27)}$ | protective (33.719) | passengers (65.895) | $\begin{aligned} & \text { turkey } \\ & (80.374) \end{aligned}$ | $\begin{gathered} \text { iowa } \\ (41.985) \end{gathered}$ | $\begin{aligned} & \text { stores } \\ & (55.435) \end{aligned}$ | $\begin{aligned} & \text { los } \\ & (95.747) \end{aligned}$ | $\begin{aligned} & \text { testified } \\ & (49.959) \end{aligned}$ |
| 5 | protests | emily |  |  |  |  | republican | gyms | graduated | zelensky |
|  | (61.585) | (60.696) | (50.966) | (28.178) | (63.535) | (75.912) | (40.993) | (50.487) | (93.51) | (48.427) |
| 6 | george | alan | strike | supplies | princess | kurds | gov | theaters | angeles | ambassador |
|  | (53.378) | (59.499) | (49.026) | (27.15) | (45.714) | (63.639) | (37.689) | (49.866) | (92.349) | (46.086) |
| 7 | brutality | nicholas | iraqi | gloves | flight | fighters | buttigieg | closed | berkeley | weinstein |
| 7 | (44.051) | (55.899) | (46.027) | (26.724) | (45.375) | (62.313) | (37.1) | (46.544) | (85.653) | (45.053) |
| 8 | officers | cochrane | qassim | equipment | nasa | forces | democrat | indoor | grew | ukrainian |
|  | (43.581) | (52.045) | (45.861) | (26.252) | (43.306) | (57.659) | (37.087) | (44.541) | (84.145) | (43.716) |
| 9 | racism | ben | maj | respiratory | navy | troops | representative | salons | today | giuliani |
|  | $\xrightarrow{(43.457)}$ | (41.414) | (44.921) | (25.447) | (40.396) | (54.045) | (35.807) | (41.99) | (41.818) | (42.674) |
| 10 | demonstrations (42.547) | $\begin{aligned} & \text { maggie } \\ & (41.328) \end{aligned}$ | baghdad (44.867) | $\begin{aligned} & \text { testing } \\ & \text { (24.179) } \end{aligned}$ | astronauts (37.073) | $\begin{gathered} \text { isis } \\ (53.743) \end{gathered}$ | $\begin{aligned} & \text { bernie } \\ & (35.228) \end{aligned}$ | shops <br> (40.311) | california (38.807) | $\begin{gathered} \text { sexual } \\ (39.966) \end{gathered}$ |

Table A22. For the most significant two words per topic, the four nearest neighbors based on cosine similarity are listed.

|  | Topic 1 | Topic 2 | Topic 3 | Topic 4 | Topic 5 | Topic 6 | Topic 7 | Topic 8 | Topic 9 | Topic 10 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 0 | floyd | contributed | iran | masks | ship | syria | senator | restaurants | bloom | ukraine |
| 0 | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) |
| 1 | demonstrations | shear | suleimani | providers | aboard | isis | wyden | shops | graduated | volker |
| 1 | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) |
| 2 | systemic | annie | retaliation | distressed | capsule | ceasefire | iowa | takeout | edited | inquiry |
| 2 | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (0.999) | (1.0) | (1.0) | (1.0) | $(1.0)$ |
| 3 | protests | mazzei | qassim | tobacco | diamond | fighters | steyer | nightclubs | berkeley | transcript |
| 3 | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (0.999) | (1.0) | (1.0) | (1.0) | (1.0) |
| 4 | defund <br> (1.0) | kitty <br> (1.0) | revenge <br> (1.0) | selfemployed (1.0) | dragon (1.0) | syrian <br> (0.999) | klobuchar <br> (1.0) | pubs <br> (1.0) | grew <br> (1.0) | investigations (1.0) |
|  | (1.0) | (1.0) | (1.0) | (1.0) |  | (0.999) | (1.0) | (1.0) | (1.0) | (1.0) |
| 0 | police | reporting | suleimani | patients | crew | syrian | storm | bars | julie | sondland |
| 0 | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) |
| 1 | systemic | luis | strike | treating | aboard | alassad | carolina | reopen | garcetti | testifying |
| 1 | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (0.993) | (1.0) |
| 2 | peaceful | beachy | maj | infection | capsule | recep | rubio | nonessential | graduated | mick |
| 2 | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (0.988) | (1.0) |
| 3 |  |  |  |  | princess |  | hampshire |  | edited | quid |
| 3 | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | (1.0) | $(0.988)$ | (1.0) |
| 4 | knee <br> (1.0) | glueck <br> (1.0) | retaliation (1.0) | repay (1.0) | cruise (1.0) | kurds <br> (1.0) | landfall (1.0) | takeout <br> (1.0) | berkeley (0.988) | impeachment (1.0) |




Figure A14. 2-dimensional representation of word embeddings $A$ colored by topic assignment.


Figure A15. 2-dimensional representation of word embeddings $A^{\prime}$ colored by topic assignment).

## Appendix B. Matrix Derivatives

In this section we derive the derivatives in (20) and (24) analytically. We write the loss in trace form by

$$
\begin{aligned}
\mathcal{L}(S, A, \boldsymbol{R}) & =\left\|S-\boldsymbol{A} \boldsymbol{R} \boldsymbol{A}^{T}\right\|_{F}^{2} \\
& =\operatorname{tr}\left[\left(S-\boldsymbol{A} \boldsymbol{R} \boldsymbol{A}^{T}\right)^{T}\left(S-\boldsymbol{A} \boldsymbol{R} \boldsymbol{A}^{T}\right)\right] \\
& =\operatorname{tr}\left[\boldsymbol{Q}^{T} \boldsymbol{Q}\right]
\end{aligned}
$$

Then

$$
\begin{aligned}
\mathrm{d} \mathcal{L} & =\mathrm{d} \operatorname{tr}\left[\boldsymbol{Q}^{T} \boldsymbol{Q}\right] \\
& =\operatorname{tr}\left[\mathrm{d}\left(\boldsymbol{Q}^{T} \boldsymbol{Q}\right)\right] \\
& =\operatorname{tr}\left[(\mathrm{d} \boldsymbol{Q})^{T} \boldsymbol{Q}+\boldsymbol{Q}^{T} \mathrm{~d} \boldsymbol{Q}\right] \\
& =\operatorname{tr}\left[\boldsymbol{Q}^{T} \mathrm{~d} \boldsymbol{Q}+\boldsymbol{Q}^{T} \mathrm{~d} \boldsymbol{Q}\right] \\
& =\operatorname{tr}\left[\left(\boldsymbol{Q}^{T}+\boldsymbol{Q}^{T}\right) \mathrm{d} \boldsymbol{Q}\right] \\
& =2 \operatorname{tr}\left[\boldsymbol{Q}^{T} \mathrm{~d} \boldsymbol{Q}\right] \\
& =2 \operatorname{tr}\left[\boldsymbol{Q}^{T} \mathrm{~d}\left(\boldsymbol{S}-\boldsymbol{A}^{\prime} \boldsymbol{R} \boldsymbol{A}^{T}\right)\right] \\
& =2 \operatorname{tr}\left[\boldsymbol{Q}^{T} \mathrm{~d} \boldsymbol{S}\right]-2 \operatorname{tr}\left[\boldsymbol{Q}^{T} \mathrm{~d}\left(\boldsymbol{A R} \boldsymbol{A}^{T}\right)\right] \\
& =-2 \operatorname{tr}\left[\boldsymbol{Q}^{T} \mathrm{~d}\left(\boldsymbol{A R} \boldsymbol{A}^{T}\right)\right]
\end{aligned}
$$

Differential in terms of d $\boldsymbol{R}$ :

$$
\begin{aligned}
\mathrm{d} \mathcal{L} & =-2 \operatorname{tr}\left[\boldsymbol{Q}^{T} \mathrm{~d}\left(A \boldsymbol{R} \boldsymbol{A}^{T}\right)\right] \\
& =-2 \operatorname{tr}\left[\boldsymbol{Q}^{T} \boldsymbol{A} \mathrm{~d} \boldsymbol{R} \boldsymbol{A}^{T}\right] \\
& =-2 \operatorname{tr}\left[\boldsymbol{A}^{T} \boldsymbol{Q}^{T} \boldsymbol{A d} \boldsymbol{R}\right]
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial R} & =-2\left(\boldsymbol{A}^{T} \boldsymbol{Q A}\right) \\
& =-2\left(\boldsymbol{A}^{T}\left(S-A R A^{T}\right) A\right) \\
& =-2\left(\boldsymbol{A}^{T} \boldsymbol{S} \boldsymbol{A}-\boldsymbol{A}^{T} \boldsymbol{A R} \boldsymbol{A}^{T} \boldsymbol{A}\right)
\end{aligned}
$$

Differential in terms of $\mathrm{d} A$ :

$$
\begin{aligned}
\mathrm{d} \mathcal{L} & =-2 \operatorname{tr}\left[\boldsymbol{Q}^{T} \mathrm{~d}\left(\boldsymbol{A R} \boldsymbol{A}^{T}\right)\right] \\
& =-2 \operatorname{tr}\left[\boldsymbol{Q}^{T} \mathrm{~d} \boldsymbol{A R} \boldsymbol{A}^{T}+\boldsymbol{Q}^{T} \boldsymbol{A R}(\mathrm{~d} \boldsymbol{A})^{T}\right] \\
& =-2 \operatorname{tr}\left[\boldsymbol{R} \boldsymbol{A}^{T} \boldsymbol{Q}^{T} \mathrm{~d} A+\boldsymbol{R}^{T} \boldsymbol{A}^{T} \boldsymbol{Q} \mathrm{~d} A\right] \\
& =-2 \operatorname{tr}\left[\left(\boldsymbol{R} \boldsymbol{A}^{T} \boldsymbol{Q}^{T}+\boldsymbol{R}^{T} \boldsymbol{A}^{T} \boldsymbol{Q}\right) \mathrm{d} \boldsymbol{A}\right]
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial A} & =-2\left(Q \boldsymbol{A} \boldsymbol{R}^{T}+Q^{T} \boldsymbol{A R}\right) \\
& =-2\left(\left(S-A \boldsymbol{R} \boldsymbol{A}^{T}\right) \boldsymbol{A} \boldsymbol{R}^{T}+\left(S-A \boldsymbol{R} \boldsymbol{A}^{T}\right)^{T} \boldsymbol{A R}\right) \\
& =-2\left(S \boldsymbol{A} \boldsymbol{R}^{T}-\boldsymbol{A R} \boldsymbol{A}^{T} \boldsymbol{A} \boldsymbol{R}^{T}+\boldsymbol{S}^{T} \boldsymbol{A R}-\boldsymbol{A} \boldsymbol{R}^{T} \boldsymbol{A}^{T} \boldsymbol{A R}\right) \\
& =-2\left(S \boldsymbol{A} \boldsymbol{R}^{T}+\boldsymbol{S}^{T} \boldsymbol{A} \boldsymbol{R}-\boldsymbol{A}\left(\boldsymbol{R} \boldsymbol{A}^{T} \boldsymbol{A} \boldsymbol{R}^{T}+\boldsymbol{R}^{T} \boldsymbol{A}^{T} \boldsymbol{A R}\right)\right)
\end{aligned}
$$

## References

1. Levy, O.; Goldberg, Y. Neural Word Embedding as Implicit Matrix Factorization. In Advances in Neural Information Processing Systems; NIPS'14; MIT Press: Cambridge, MA, USA, 2014; Volume 2, pp. 2177-2185.
2. Pennington, J.; Socher, R.; Manning, C. Glove: Global Vectors for Word Representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, Doha, Qatar, 25-29 October 2014; pp. 1532-1543.
3. Hillebrand, L.P.; Biesner, D.; Bauckhage, C.; Sifa, R. Interpretable Topic Extraction and Word Embedding Learning Using Row-Stochastic DEDICOM. In Machine Learning and Knowledge Extraction-4th International Cross-Domain Conference, CD-MAKE; Lecture Notes in Computer Science; Springer: Dublin, Ireland, 2020; Volume 12279, pp. 401-422.
4. Symeonidis, P.; Zioupos, A. Matrix and Tensor Factorization Techniques for Recommender Systems; Springer International Publishing: New York, NY, USA, 2016. [CrossRef]
5. Harshman, R.; Green, P.; Wind, Y.; Lundy, M. A Model for the Analysis of Asymmetric Data in Marketing Research. Market. Sci. 1982, 1, 205-242. [CrossRef]
6. Andrzej, A.H.; Cichocki, A.; Dinh, T.V. Nonnegative DEDICOM Based on Tensor Decompositions for Social Networks Exploration. Aust. J. Intell. Inf. Process. Syst. 2010, 12, 10-15.
7. Bader, B.W.; Harshman, R.A.; Kolda, T.G. Pattern Analysis of Directed Graphs Using DEDICOM: An Application to Enron Email; Office of Scientific \& Technical Information Technical Reports; Sandia National Laboratories: Albuquerque, NM, USA, 2006.
8. Sifa, R.; Ojeda, C.; Cvejoski, K.; Bauckhage, C. Interpretable Matrix Factorization with Stochasticity Constrained Nonnegative DEDICOM. In Proceedings of the KDML-LWDA, Rostock, Germany, 11-13 September 2017.
9. Sifa, R.; Ojeda, C.; Bauckage, C. User Churn Migration Analysis with DEDICOM. In Proceedings of the 9th ACM Conference on Recommender Systems; RecSys '15; Association for Computing Machinery: New York, NY, USA, 2015; pp. 321-324.
10. Chew, P.; Bader, B.; Rozovskaya, A. Using DEDICOM for Completely Unsupervised Part-of-Speech Tagging. In Proceedings of the Workshop on Unsupervised and Minimally Supervised Learning of Lexical Semantics, Boulder, CO, USA, 5 June 2009; Association for Computational Linguistics: Boulder, CO, USA, 2009; pp. 54-62.
11. Nickel, M.; Tresp, V.; Kriegel, H.P. A Three-Way Model for Collective Learning on Multi-Relational Data. In Proceedings of the 28th International Conference on Machine Learning, Bellevue, WA, USA, 28 June-2 July 2009; Association for Computing Machinery: Bellevue, WA, USA, 2011; pp. 809-816.
12. Sifa, R.; Yawar, R.; Ramamurthy, R.; Bauckhage, C.; Kersting, K. Matrix- and Tensor Factorization for Game Content Recommendation. KI-Künstl. Intell. 2019, 34, 57-67. [CrossRef]
13. Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G.; Dean, J. Distributed Representations of Words and Phrases and their Compositionality. Adv. Neural Inf. Proc. Syst. 2013, 26, 3111-3119.
14. Lee, D.D.; Seung, H.S. Algorithms for Non-Negative Matrix Factorization. In Proceedings of the 13th International Conference on Neural Information Processing Systems; NIPS’00; MIT Press: Cambridge, MA, USA, 2000; pp. 535-541.
15. Furnas, G.W.; Deerwester, S.; Dumais, S.T.; Landauer, T.K.; Harshman, R.A.; Streeter, L.A.; Lochbaum, K.E. Information Retrieval Using a Singular Value Decomposition Model of Latent Semantic Structure. In ACM SIGIR Forum; ACM: New York, NY, USA, 1988.
16. Wang, Y.; Zhu, L. Research and implementation of SVD in machine learning. In Proceedings of the 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS), Wuhan, China, 24-26 May 2017; pp. 471-475.
17. Jolliffe, I. Principal Component Analysis; John Wiley and Sons Ltd.: Hoboken, NJ, USA, 2005.
18. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent Dirichlet Allocation. J. Mach. Learn. Res. 2003, 3, 993-1022.
19. Lebret, R.; Collobert, R. Word Embeddings through Hellinger PCA. In Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, Gothenburg, Sweden, 26-30 April 2014; pp. 482-490.
20. Nguyen, D.Q.; Billingsley, R.; Du, L.; Johnson, M. Improving Topic Models with Latent Feature Word Representations. Trans. Assoc. Comput. Linguist. 2015, 3, 299-313. [CrossRef]
21. Frank, M.; Wolfe, P. An algorithm for quadratic programming. Naval Res. Logist. Q. 1956, 3, 95-110. [CrossRef]
22. Ni, J.; Li, J.; McAuley, J. Justifying recommendations using distantly-labeled reviews and fine-grained aspects. 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), Hong Kong, China, 3-7 November 2019; pp. 188-197.
23. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. arXiv 2014, arXiv:1412.6980.
24. McInnes, L.; Healy, J.; Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv 2018, arXiv:1802.03426.
