Unsupervised Content-Based Identification of Fake News Articles with Tensor Decomposition Ensembles

Conference Paper · February 2018

2 authors:

Seyedmehdi Hosseinimotlagh
University of California, Riverside

Evangelos E. Papalexakis
Carnegie Mellon University

Unsupervised Content-Based Identification of Fake News Articles with Tensor Decomposition Ensembles

Seyedmehdi Hosseinimotlagh
University of California Riverside
shoss007@ucr.edu

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

ABSTRACT

Social media provide a platform for quick and seamless access to information. However, the propagation of false information, especially during the last year, raises major concerns, especially given the fact that social media are the primary source of information for a large percentage of the population. False information may manipulate people's beliefs and have real-life consequences. Therefore, one major challenge is to automatically identify false information by categorizing it into different types and notify users about the credibility of different articles shared online.

Existing approaches primarily focus on feature generation and selection from various sources, including corpus-related features. However, so far, prior work has not paid considerable attention to the following question: how can we accurately distinguish different categories of false news, solely based on the content?

In this paper we work on answering this question. In particular, we propose a tensor modeling of the problem, where we capture latent relations between articles and terms, as well as spatial/contextual relations between terms, towards unlocking the full potential of the content. Furthermore, we propose an ensemble method which judiciously combines and consolidates results form different tensor decompositions into clean, coherent, and high-accuracy groups of articles that belong to different categories of false news. We extensively evaluate our proposed method on real data, for which we have labels, and demonstrate that the proposed algorithm was able to identify all different false news categories within the corpus, with average homogeneity per group of up to 80%.

KEYWORDS

Tensor decomposition, Ensemble method, Fake news

1 INTRODUCTION

The advent of social media services has facilitated production, sharing, and searching information at an unprecedented level. For instance, over 65% of the US adult population has access to Facebook and share information on a daily basis[1]. Social media allow people to express their feelings, state their opinions on news, and relay information from the media to their audience. Therefore, social media play a profound role in influencing the social, economic and political domains of everyday decision making. For instance, during 2016 US presidential election, candidates effectively used Twitter to send their messages and express their opinions directly to their supporters. Hillary Clinton's Twitter, for example, has reached 16 million followers, with her most popular tweet received more than 600,000 re-tweets [2] and one million likes.

Aside from all merits, the proliferation of false information raises major concerns. In the 2013 World Economic Forum, "the rapid spread of misinformation online" has been ranked among top ten trends for year 2014 [11]. False information being deliberately produced to manipulate readers is called disinformation for different purposes for example stating a biased belief or temping audience to clicking on the link. On the other hand, misinformation, is the unintentional relay of incorrect facts to an audience who may perceive the delivered information as true.

There exist several online fact-checking applications, like Hoaxy [30], which gather credibility scores of news from several online fact-checking websites, such as Snopes.com, PolitiFact.com, and FactCheck.org. Recent research shows that fake news spread fast through social media networks (for instance, Hurricane Sandy in 2012 [4] and the Boston Marathon blasts in 2013 [31]), and this can adversely amplify public anxiety. For instance, in 2013, $130 billion in "stock value was wiped out" abruptly after a rumor of Barack Obama being injured by an explosion at the White House [26]. Thus, early identification of fake news, before they find their way in online fact-checking repositories, potentially using solely their content, is imperative.

A popular line of work focuses on analyzing user connectivity, in order to understand how information is diffused on a social network. There exist several proposed models, such as independent cascade and linear threshold model [17], SIR [35] and tipping models [5], most of which seek to learn a parameterized probability of a user being infected, as a function of their friends and their behavior.

In addition to network-based analyses, fake news identification can be done based on the content of a particular article. For example, Facebook recently proposed a number of tips [3] for spotting fake news, such as being skeptical all-CAPS headlines and unusual formatting. Aside from using merely meta-data of an news article, such as its title or potential hashtags shared along with it, the very text of the article can reveal crucial information. In this paper, we focus on content-based unsupervised fake news identification, using factorization methods. Most unsupervised clustering methods in text retrieval with basis of *Non-negative Matrix Factorization* (NMF) [16, 34, 20, 32, 18, 12], *Singular Value Decomposition*

[1]http://www.pewinternet.org/2016/11/11/social-media-update-2016/

MIS2, Marina Del Rey, CA, USA

DOI: 10.475/123_4

[2]https://www.weforum.org/agenda/2016/08/hillary-clinton-or-donald-trump-winning-on-twitter/
[3]https://techcrunch.com/2017/04/06/facebook-puts-link-to-10-tips-for-spotting-false-news-atop-feed/

(SVD) [7] and *Independent Component Analysis* (ICA) [19] focus on frequencies of terms (words) in documents. To improve the results, tf-idf (term frequency-inverse document frequency) has been widely proposed as term-weighting schemes to enhance the effectiveness of words. The tf-idf value increases proportionally to the number of times a word appears in the document and reduce the weight of term appears more frequently in corpus generally. Although the frequencies of terms are considered in these algorithms, the affinity of terms are ignored that causes irrelevant news with same high keywords being selected in a same latent factor.

The aforementioned matrix techniques are able to capture correlations and similarities between different documents, however, they are not able to fully leverage the context of a document, since they are restricted to unigrams (or even n-grams, which still do not fully capture context). Context has been shown to be extremely useful in designing effective word vector representations, with prime example the Word2Vec model [22].

In this paper, we take this analysis a step further and explicitly model the context of words within a document via capturing the spatial vicinity of each word. In particular, we model the corpus as a third order tensor which simultaneously models article and term relations, as well as spatial/contextual relations between terms. Exploiting both aspects of the corpus, and in particular the spatial relations between words, is a determining factor for identifying coherent groups of articles that fall under different types of false news. To the best of our knowledge, this work is the first to simultaneously model spatial relations between terms, and latent relations between terms and articles, in the context of fake news identification. Our contributions are:

- **Exploiting higher-order context information**: We model the corpus as a three-mode (article, term, term) tensor which captures spatial term relations and article-term relations. We, subsequently use the CP/PARAFAC decomposition [14] to identify latent groups of articles that fall under coherent categories of false news.
- **Introducing an ensemble method**: We design an ensemble method which leverages multiple decompositions of the (article, term, term) tensor to further refine the latent groups discovered by the decomposition and produce a clean categorization of articles, which is more accurate than the state of the art. In particular, our proposed ensemble technique 1) discovers classes with higher homogeneity, 2) is able to identify "hard" categories of false news (where baselines fail), and 3) produces results with lower outlier diversity.
- **Evaluation on real-world data**: We apply our proposed algorithm on Kaggle fake news dataset [28] which contains several features including the body of news and also labels used only for evaluation. Our results show that the proposed algorithm finds most of fake news categories with homogeneity value up to 80% on overall, reduces the diversity of outliers to 2.5 and exploring all categories with high homogeneity whereas other algorithms are incapable of discovery all categories.

Indicatively, our proposed method was able to produce latent groups with 65% coherence overall and more 80% for most of categories, whereas NMF/SVD up to 50% on average of top 30 news for each factor.

The outline of the paper is as follows. In Section 2, we describe preliminaries and notations used in the paper. In Section 3 we define the problem and describe our two-tier proposed algorithm. We give the results of running numerous experiments with two different unsupervised clustering algorithms in Section 4. In Section 5 we provide a brief literature survey on false news. Finally, we describe our conclusions in Section 6 about the usefulness of our proposed method and how the two algorithms could be used together to detect and analyze false news categories.

2 PRELIMINARIES AND NOTATIONS

Before explaining our proposed algorithm, we provide a few necessary definitions and describe our notations on tensor decomposition and co-clustering which we use throughout this paper.

2.1 CP/PARAFAC Tensor Decomposition

Tensors are denoted by boldface underlined capital letters ($\underline{\mathbf{X}}$), matrices by boldface capital letters ($\mathbf{X}$), and vectors by boldface lower case letters ($\mathbf{x}$). Furthermore, $\underline{\mathbf{X}}_{i,:,:}$ denotes the ith horizontal slice in $\underline{\mathbf{X}}$. CP/PARAFAC decomposition [14] of a 3-way $\underline{\mathbf{X}}$, is written as $\underline{\mathbf{X}} \approx [\mathbf{A}, \mathbf{B}, \mathbf{C}] = \sum_{r=1}^{R} \lambda_r \ \mathbf{a}_r \ o \ \mathbf{b}_r \ o \ \mathbf{c}_r$ where the symbol o denotes the outer product and $\mathbf{a}_r$ (same for $\mathbf{b}_r$ and $\mathbf{c}_r$) denotes the normalized rth column (called *component*) of non-negative *factor matrix* $\mathbf{A}$ (same for $\mathbf{B}$ and $\mathbf{C}$. R is the *rank* of CP/PARAFAC decomposition is minimum number of columns needed that their outer products construct the original tensor. However, in most cases, a low-rank decomposition is preferred, because it is able to effectively capture hidden patterns and similarities in the data. In this paper, we employ the CP/PARAFAC decomposition with alternating optimization, given a Poisson distribution on the column of data, which has been shown to be tailored when dealing with sparse, count data [6]; hence $\{\mathbf{A}, \mathbf{B}, \mathbf{C}\} \in [0, 1]^{I_n \times R}$ and $||\mathbf{a}_r|| = 1$ (also for $\mathbf{b}_r$ and $\mathbf{c}_r$).

2.2 Soft Co-Clustering

Co-clustering searches for subsets of similar rows and columns in the data matrix with least squares similarity. In the Spare Matrix Regression algorithmin [23] the authors introduce a soft matrix co-clustering method, where co-clusters may overlap. In co-clustering since rows and columns are decomposed is called bilinear decomposition. So, $\mathbf{X}^{N \times M} \approx [R^{N \times K}, C^{M \times K}] = \sum_{i=1}^{K} \mathbf{r}_i \mathbf{c}_i{}^T$ where r_i and c_i^T is a latent factor of rows and columns, respectively and K is the rank of co-clustering. The authors of [23] further impose L1-norm regularization to promote sparsity on the latent factors. Therefore, the loss function is calculated as: $||\mathbf{X} - \mathbf{R}\mathbf{C}^T||_T^2 + \lambda \sum_{i,k} |\mathbf{R}(i,k)| + \lambda \sum_{j,k} |\mathbf{C}(j,k)||$.

3 PROBLEM DEFINITION AND PROPOSED METHOD

In this section, we explain our two-tier proposed method. First we explain how to extract spatial relations on terms and then we discus the co-clustering to decompose relevant documents.

3.1 Problem description

Given a corpus of fake news $\mathbf{d} = \{\mathbf{n}_1, \mathbf{n}_2, \mathbf{n}_3, ...\}$ with size of N where each document $\vec{n_i}$ is a vector of terms in a dictionary, $\sum = \{t_1, t_2, t_3, ...\}$ with size of $T = |\sum|$. The problem is clustering of documents based on their terms into homogeneous classes with respect to fake news categories. To this end, we first cluster documents based on appearance positions of each term in an article and its correlations with other terms (Spatial relation extraction) following by designing an automatic ensemble co-clustering to cluster documents according to their positions in different factors among various low-rank decompositions.

3.2 Tier-1: Spatial relation extraction

In this section, we propose a method to decompose the documents based on spatial relations on their terms. The intuition of this stage is that a subset of terms which are in affinity to other represents a meaning. In this paper, we call the affinities of terms in a document *spatial relations* between terms and we propose a tensor constructed based on these relations. Low-rank tensor decomposition leads to mine the most important relations which represents a category of false news.

In most of previous content-based decomposition algorithms, a document-term matrix (for example, $\mathbf{A}^{T \times N}$ has been constructed such that $\mathbf{A}(i, j)$ represents the number of occurrence of ith term of the dictionary in the document j. Documents that have similarities in frequencies of words have the similar coefficient in each factor of decomposed matrix. Therefore selecting top values of each factor in decomposed document matrix leads to a coherent number of documents.

In contrast to previous research that only existence and frequency of words in documents does matter, we also take the spatial relation of words into account. To this end, we propose a tensor of $N \times T \times T$. each horizontal slice of $\underline{\mathbf{X}}_{i,:,:}$ represents the spatial relations in a document. For a horizontal slice like $\mathbf{S}$, $\mathbf{S}(i, j)$ is the number of times that ith term and jth term of the dictionary appear in an affinity of each other. Therefore, each slice of the tensor contains the co-occurrence of terms within δ which is the size of affinity window.

The problem is exatracting a densely correlated set of documents and words, usually manifesting in the data as dense blocks. To this end we use CP/PARAFAC decomposition. We compute non-negative CP/PARAFAC with alternating Poisson regression (CP_APR) that uses Kullback-Leibler divergence because of very high sparsity in the tensor [6].

3.3 Tier 2: Tensor Ensemble Co-clustering

The second tier of our proposed method is a clustering algorithm for categorizing news articles based on their cluster membership among a set of different tensor decompositions. Intuitively, the proposed clustering seeks to find a subset of news articles that frequently cluster together in different configurations of the tensor decomposition of the first tier (where 'configuration" refers to the rank of the decomposition).

The intuition behind our proposed ensemble method is that news articles that tend to frequently appear surrounded each other among different rank configurations, while having the same ranking within their latent factors are more likely to ultimately belong to the same category. The ranking of a news article with respect to a latent factor is derived by simply sorting the coefficients of each latent factor, corresponding to the clustering membership of a news article to a latent factor.

Before we proceed with our proposed method, below we describe a number of different scenarios that for two different news articles in our results, both when dealing with a single decomposition, and when dealing with an ensemble of configurations:

3.3.1 Scenario 1: Within factors of CP/PARAFAC. Suppose that two news articles of n_1 and n_2 belong to a same category. Besides, they have much similarity with respect to content. Therefore, one can expect that after a tensor decomposition they should appear in top of some columns close to each other (see Fig. 1a).

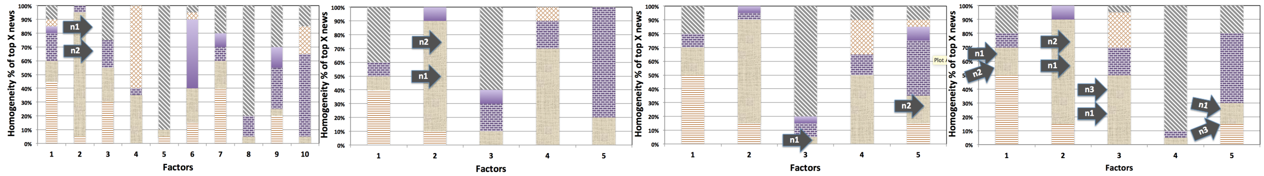
3.3.2 Scenario 2: Different configurations. If n_1 and n_2 are most of times in affinity of each other in different configurations of low-rank decomposition, one can expect that they belong to the same category (n_1 appears near n_2 in both Fig. 1a and Fig. 1b configurations).

3.3.3 Scenario 3: Different columns. To the contrary, if n_1 and n_2 appear at top of different columns (see fig. 1c), but in other factors they are usually near each other, it is more than likely that they are outliers to those columns.

3.3.4 Scenario 4: Transitivity . If there exists another news like n_3 such that n_1 and n_3 appear nearby in some configurations and n_1 and n_2 appear nearby in some other configurations, it's likely that n_1, n_2 and n_3 belong to a same category (see Fig. 1d).

In order to extract similar news articles from the ensemble of tensor decompositions with different configurations, we apply co-clustering. In particular, we propose to combine the clustering results of each individual tensor decomposition into a collective (news-article by latent-factor) matrix, from which we are going to extract co-clusters of news articles and the corresponding latent factors (coming from our ensemble of decompositions), indicating that news articles within the same co-cluster tend to appear (with high membership value) frequently in the same latent tensor decomposition factor.

When constructing the (news-article by latent-factor) matrix, we have to decide whether a particular news article belongs to one of the r factors of each decomposition. We, thus, use the value of the factor corresponding to each article, and assign it accordingly to one of the r factors. To that end, we divide each factor into several partitions, based on the Empirical Cumulative Distribution Function (ECDF) of their values, so that we do not create arbitrary partitions or make arbitrary assignment decisions. After computing the partitions, each column is reordered, as shown in Figure 2 so



(a) (b) (c) (d)

Figure 1: Examples of different conditions for news a) two news in their affinity in a factor b) two news in their affinity in a factor of different configuration c) two same label news appear as outliers of different factors d) transitive relation of news

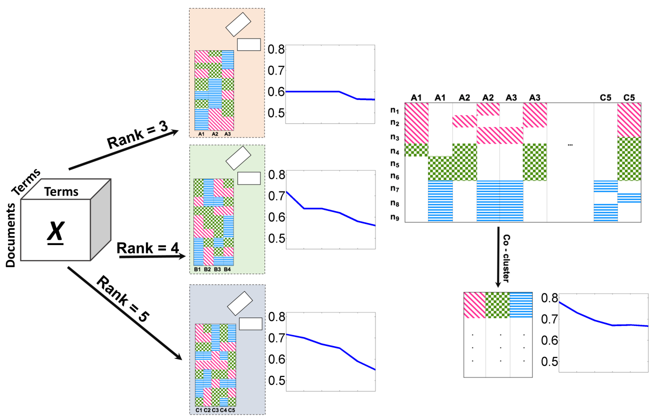


Figure 2: Overview of our proposed algorithm, demonstrating the construction of the S matrix out of an ensemble of tensor decompositions, and the extraction of high-quality clusters for fake news articles. Next to each decomposition of the ensemble we also plot the homogeneity (with respect to the actual labels) of each factor and we demonstrate how the homogeneity improves by combining the decomposition ensembles via co-clustering.

that all news articles that are assigned to the same latent factor (i.e., latent cluster) are located in contiguous positions in the vector.

After constructing the matrix, we co-cluster the rows (corresponding to news articles) and columns (corresponding to latent tensor factors) of that matrix. The result will be set of news articles that frequently appear in the same latent factor partition (computed by the ECDF of each factor) with similar factor-membership value. Intuitively, the proposed ensemble method decreases the likelihood of outliers, i.e., news articles which, by mere noise, would appear with very high value in one of the decompositions, but not in the entire ensemble.

4 EXPERIMENTAL EVALUATION

In this section, we evaluate our proposed method on a real dataset. We use the fake news dataset provide by Kaggle [28] which contain text and metadata of over 12000 articles. Table 1 describes the categories of fake news labeled by BS detector [8]. We discard *fake* category because of few instances and *bs* news since they are missing labels. We implement our proposed algorithms in Matlab. For tensor decomposition and SMR co-clustering, we use the tensor toolbox library [3] and the publicly available implmentation in [9], respectively. We only choose the news which contain more than 100 words after stemming and removing stop words. For balancing the dataset, we randomly select the equal number of instances for each category. Therefore, it results to have 75 instances for each category. The default kernel size to construct the tensor is 10 unless it is explicitly stated. We conduct experiments a number of times on different shuffled instances and report the median results.

Table 1: Fake news categories.

Labels from BS detector.	
class name [short form]	descriptions
[Satire]	Sources that provide humorous commentary on current events in the form of fake news.
Extreme [Bias]	Sources that traffic in political propaganda and gross distortions of fact.
[Conspiracy] Theory	Sources that are well-known promoters of kooky conspiracy theories.
[Junk Sci]ence	Sources that promote pseudoscience, metaphysics, naturalistic fallacies, and other scientifically dubious claims.
[Hate] Group	Sources that actively promote racism, misogyny, homophobia, and other forms of discrimination.
[State] News	Sources in repressive states operating under government sanction.

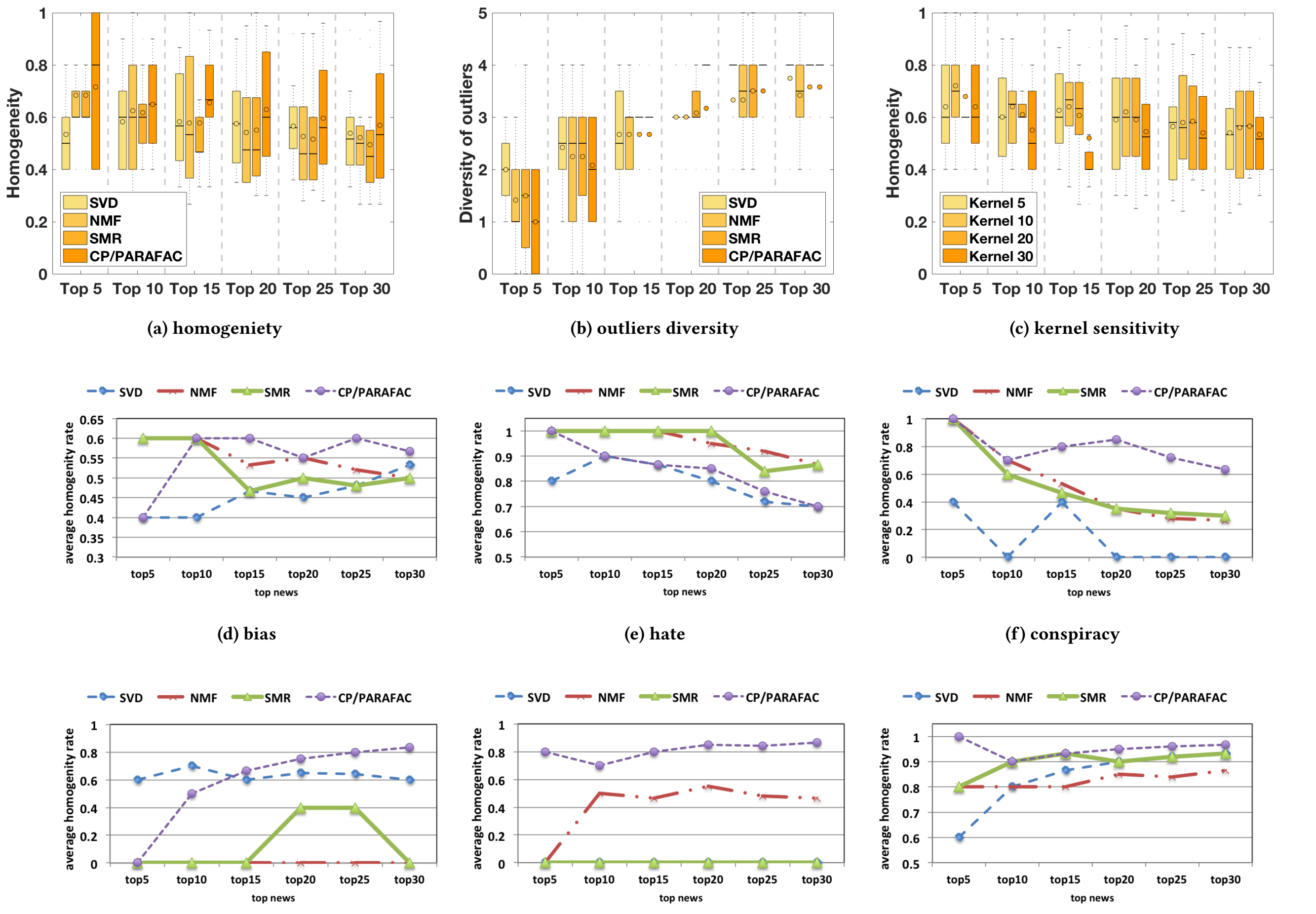
4.1 Evaluation of Tier-1

First, we show our results for the first tier of our approach, evaluating it in different aspects. More specifically, we analyze the results of 1) homogeneity and 2) variety of outliers in different algorithms, and 3) investigate the sensitivity of word kernel.

4.1.1 Homogeneity. Our primary objective is to identify high-quality clusters of fake news in an unsupervised manner. To this end, we show the homogeneity of factors in different decomposition algorithms, indicating how *pure* each latent cluster is. Figure 3a depicts this metric for different thresholds of top news articles. We observe that the homogenity of CP/PARAFAC falls slightly and remains level to almost 56% on average when selecting top 30 news of each factor while that of other algorithms drops significantly in other decomposition methods.

4.1.2 Outlier variety. Another important evaluation metric that we measure in this paper is the variety of outliers. As shown in Fig. 3b, although CP/PARAFAC reduces the number of outliers substantially, outliers belong to different news categories. During our experiments, we noticed that the number of outliers increases when increasing the number of top news articles to 25 or larger. Overall, comparing the results of homogeneity with the variety of outliers indicates that even though homogeneity is less sensitive to the increase of the number of news articles per cluster, the number of outliers increases substantially, warranting the second tier of our proposed method.

4.1.3 Categories identified. The third dimension on which we evaluate our proposed method is its ability to discover as many categories of fake news as possible. As seen in Fig. 3d-f, *bias*, *hate* and *junksci* are relatively easy for all of the decomposition algorithms.



(a) homogeniety (b) outliers diversity (c) kernel sensitivity

(d) bias (e) hate (f) conspiracy

(g) satire (h) state (i) junksci

Figure 3: 1tier decomposition algorithm a) homogeneity b) diversity of outliers c) kernel sensitivity d-f) average homogeneity in different categories.

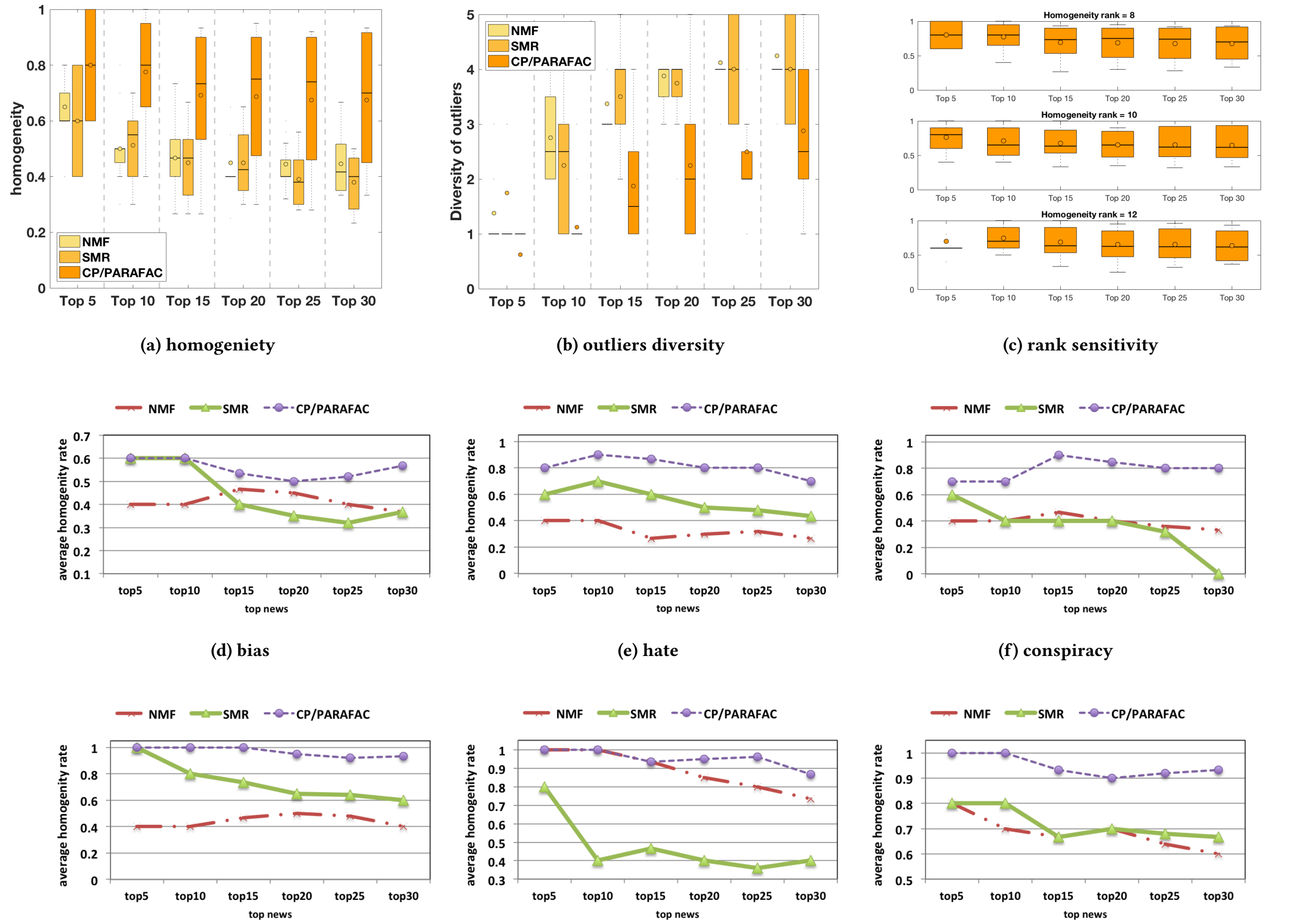
In contrast to other algorithms, our proposed CP/PARAFAC-based method is able to classify other, harder, categories with high homogeneity. More specifically, most other methods algorithms to even identify the category of *state* while our proposed method is identifying this category with 80% accuracy. Therefore, our proposed one-tier algorithm outperforms other algorithms in both overall homogeneity and also capability of finding all categories.

4.1.4 Spatial kernel sensitivity. We investigate the sensitivity of proposed CP/PARAFAC the size of the spatial kernel δ for the terms. As seen previously, considering the spatial relations in constructing the corresponding tensor lead to homogeneity improvements of factors. However, one may ask which kernel size of terms results in a better homogeneity, or how sensitive is the proposed decomposition to kernel size. Figure 3c illustrates the homogeneity of CP/PARAFAC with respect to kernel size. We observe than our method is not particularly sensitive to the exact choice of kernel δ.

4.2 Evaluation of Tier-2

In this section, we discuss that better homogeneity can be obtained after keeping track of relations between news. We will show that the proposed method boosts the homogeneity while decreases the variety of outliers. Furthermore, we show that homogeneity of each false news category individually. In this paper, we partition factors of SMR with their ecdf values of 90, 80 and 65% percentiles after excluding zeros. Then we add a column to for each corresponding CP/PARAFAC factor for zero values.

4.2.1 Homogeneity. In this section SMR is employed after applying Tier-1 decomposition method. Fig. 4a depicts the results of the proposed algorithm comparing with others. Our proposed algorithm outperforms other algorithms significantly. One merit of employing the proposed SMR on the Tier-1 decomposition is stabilization on homogeneity. We observe that there in less variation in homogeneity when selecting more top news from each factor. Furthermore, comparing to Tier-1 CP/PARAFAC shows significant improvement of the proposed CP/PARAFAC-SMR in



(a) homogeneity (b) outliers diversity (c) rank sensitivity

(d) bias (e) hate (f) conspiracy

(g) satire (h) state (i) junksci

Figure 4: The results of proposed Tier-2 decomposition with $\delta = 10$ a) homogeneity of ensemble co-cluster is almost 13% more than 1-tier CP/PARAFAC. b) diversity of outliers reduces by one on average comparing to 1-tier algorithms. in c) sensitivity to different SMR ranks d-f)kmeans on constructed ensemble matrix of CP/PARAFAC. *state*, *satire*, *junksci* and *biased* clustered with low homogeneity. g-l) average homogeneity in different categories. all categories are clustered with an average of 90% homogenity.

purity of some factors. Later, we will discuss the homogeneity of each new category separately.

4.2.2 Outlier variety. Fig. 4b illustrates the varieties of outliers using the proposed algorithm. As one can see, not only the number of outliers but also the varieties of outliers reduce significantly. In some factors, there exist only one type of outlier. Comparing the proposed method with tier-1 CP/PARAFAC reveals that outlier variety reduces by more than one on average. This indicates that the proposed algorithm effectively suppresses outliers.

Table 2 describes the percentage of categories existed as outliers in other categories. *Conspiracy* is the only outlier of *state* category and the dominant outliers of *bias*. Moreover, it seems similarity in *hate* and *satire* news although the class of *hate* news has a profound homogeneity.

4.2.3 Categories identified. We investigate the homogeneity and outliers of different news categories individually. As seen in Fig. 4 g-l, all of categories are clustered successfully. Although the results seems similar to tier-1 CP/PARAFAC, the improvement to each news category is because of eliminating the irrelevant news. It means that comparing the varieties of outliers in this stage and 1-tier CP/PARAFAC shows that relevant news is substituted with

Table 2: Outliers of each category.

category	outliers (in descending order) - percentage of outliers
Satire	Conspiracy (< 50%), JunkSci (< 50%)
Bias	Conspiracy (< 70%), State (< 20%), JunkSci (< 10%)
Conspiracy	Bias(< 50%), State (< 30%), hate (< 10%)
JunkSci	Satire (< 50%), Conspiracy (< 50%)
Hate	Satire (< 95%), Bias(< 5%)
State	Conspiracy (< 99%)

news of the category that there is no other news from this category exists in affinity.

4.2.4 Sensitivity to decomposition rank. Fig. 4c shows the sensitivity of SMR after employing CP/PARAFAC. As one can see, the homogeneity of the proposed method is robust with respect to different decomposition rank configurations of SMR. We observed that although the homogeneity of classes in rank 8 is slightly higher than that of other ranks, it finds fewer kinds of categories comparing to others.

5 RELATED WORK

In order to detect misinformation, supervised learning approaches have been widely used to detect false information. In [10] logistic regression classifier proposed to detect the mismatch between headline and content of articles. Gupta et al. [13] proposed a classifier to estimate the credibility of tweet with various features such as number of words, URLs, hashtags, emojis, presence of swear words, pronouns. Horne et al. [15] proposed a SVM algorithm on selected stylistic, complexity and psychological features of both title and body of articles to classify real, fake and satire news based on their contents. Qazvinian et al. [25] proposed a Bayes classifier on content-based, network-based, and microblog-specific features detect rumors on tweets' contents. In [27], Klatsch framework has been proposed with AdaBoost and SVM classifers on topological, crowdsourced and content-based (Hashtags, Mentions, URLs, and Phrases with sentiment extraction) features to detect political misinformation at the early stage on Twitter. In [36], author extracted several statistical features from tweets such as tweet lengths, the average number of hashtags, the entropy ratio of the word frequency distribution for training a decision tree model to rank the likelihood tweets. In all of aforementioned research, extracted features have been substituted for content.

NMF based algorithms have been widely used to clustering documents. Kuang et al. [20] proposed a 2NMF to cluster hierarchical documents based on contents. The NMF feature extraction of documents in [32] suffers from adjusting the number of keywords for categories. In [16], authors proposed an anchor-free topic identification based on word-word co-occurrence matrix. However, the location of terms and their affinity relations in documents have not been studied in this research.

There are some research that take the early labeled data in account to estimate the credibility of newly emerging data. In [33], authors proposed framework to clusters data into different rumor categories, select features, and train classifiers that can detect emerging rumors using prior labeled rumors. Sampson et al. [29] proposed to use implicit linking on hashtags and web address for helping the classification on set of content-based features [21] within detection deadline. In this paper, we focus our research on relation between news to cluster them into categories instead of utilizing the previous labeled information.

In [22], authors proposed several techniques for learning distributed representations of words such as Continuous Bag-of-Words (CBOW) Model, Continuous Skip-gram Model to train high dimensional word vectors. On contrary, we employ the co-occurance of terms within a sliding window for each document, accordingly constructing a tensor of documents. Unlike [22], we proposed a spatial CP/PARAFAC that considers the relations of terms among documents and representation of words within a documents without further training for classification.

6 CONCLUSION

In this paper we set out to cluster false news into different categories. We proposed a tensor based scheme which effectively leverages term context via capturing spatial relations between terms for each article. We further introduce an ensemble method which is able to consolidate and refine the results of multiple tensor decompositions into a single, high-quality, and high-coherence set of article clusters which achieves higher coherence than state-of-the-art baselines, and is able to identify all different categories of fake news within our dataset.

7 ACKNOWLEDGEMENTS

Research was supported by an Adobe Data Science Research Faculty Award. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the funding parties.

REFERENCES

[1] Fake news challenge dataset. Available online, http://www.fakenewschallenge.org, FEB 2017.

[2] Hunt Allcott and Matthew Gentzkow. Social media and fake news in the 2016 election. Technical report, National Bureau of Economic Research, 2017.

[3] Brett W. Bader, Tamara G. Kolda, et al. Matlab tensor toolbox version 2.6. Available online, February 2015.

[4] Farida Vis Burgess, Jean and Axel Bruns. Hurricane sandy: the most tweeted pictures.fi the guardian data blog. Available online, http://www.guardian.co.uk/news/datablog/gallery/2012/nov/06/hurricane-sandy-tweeted-pictures, November 2012.

[5] Damon Centola. The spread of behavior in an online social network experiment. *science*, 329(5996):1194–1197, 2010.

[6] Eric C Chi and Tamara G Kolda. On tensors, sparsity, and nonnegative factorizations. *SIAM Journal on Matrix Analysis and Applications*, 33(4):1272–1299, 2012.

[7] Scott Deerwester, Susan T Dumais, George W Furnas, Thomas K Landauer, and Richard Harshman. Indexing by latent semantic analysis. *Journal of the American society for information science*, 41(6):391, 1990.

[8] B.S. Detector. Fake news classifications. http://bsdetector.tech, 2017.

[9] N. D. Sidiropoulos E. E. Papalexakis and M. N. Garofalakis. Smr library. Available online, http://www.models.life.ku.dk/cocluster, February 2015.

[10] William Ferreira and Andreas Vlachos. Emergent: a novel data-set for stance classification. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. ACL, 2016.

[11] The Gaurdian. To tackle the spread of misinformation online we must first understand it. Available online, https://www.theguardian.com/commentisfree/2014/apr/24/tackle-spread-misinformation-online, 2014.

[12] Nicolas Gillis. Robustness analysis of hottopixx, a linear programming model for factoring nonnegative matrices. *SIAM Journal on Matrix Analysis and Applications*, 34(3):1189–1212, 2013.

[13] Aditi Gupta and Ponnurangam Kumaraguru. Credibility ranking of tweets during high impact events. In *Proceedings of the 1st Workshop on Privacy and Security in Online Social Media*, PSOSM '12, pages 2:2–2:8, New York, NY, USA, 2012. ACM.

[14] Richard A Harshman. Foundations of the parafac procedure: Models and conditions for an "explanatory" multi-modal factor analysis. *UCLA Working Papers in Phonetic*, 1970, Available at http://publish.uwo.ca/flharshman/wpppfac0.pdf.

[15] Benjamin Horne and Sibel Adali. This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. In *11th International AAAI Conference on Web and Social Media*, 2017.

[16] Kejun Huang, Xiao Fu, and Nikolaos D Sidiropoulos. Anchor-free correlated topic modeling: Identifiability and algorithm. In *Advances in Neural Information Processing Systems*, pages 1786–1794, 2016.

[17] David Kempe, Jon Kleinberg, and ftva Tardos. Maximizing the spread of influence through a social network. In *In Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, KDD fi03*, pages 137–146. ACM, 2003.

[18] Jingu Kim and Haesun Park. Toward faster nonnegative matrix factorization: A new algorithm and comparisons. In *Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on*, pages 353–362. IEEE, 2008.

[19] Thomas Kolenda, Lars Kai Hansen, and Sigurdur Sigurdsson. Independent components in text. In *Advances in independent component analysis*, pages 235–256. Springer, 2000.

[20] Da Kuang and Haesun Park. Fast rank-2 nonnegative matrix factorization for hierarchical document clustering. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '13, pages 739–747, New York, NY, USA, 2013. ACM.

[21] Jing Ma, Wei Gao, Zhongyu Wei, Yueming Lu, and Kam-Fai Wong. Detect rumors using time series of social context information on microblogging websites. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, CIKM '15, pages 1751–1754, New York, NY, USA, 2015. ACM.

[22] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

[23] Evangelos E Papalexakis and Nicholas D Sidiropoulos. Co-clustering as multilinear decomposition with sparse latent factors. In *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, pages 2064–2067. IEEE, 2011.

[24] Martin Porter. Porterstemming library. Available online, https://tartarus.org/martin/PorterStemmer/, 2017.

[25] Vahed Qazvinian, Emily Rosengren, Dragomir R Radev, and Qiaozhu Mei. Rumor has it: Identifying misinformation in microblogs. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1589–1599. Association for Computational Linguistics, 2011.

[26] Kenneth Rapoza. Can 'fake news' impact the stock market? Available online, https://www.forbes.com/sites/kenrapoza/2017/02/26/can-fake-news-impact-the-stock-market/#6584de9d2fac, Feb 2017.

[27] Jacob Ratkiewicz, Michael Conover, Mark R Meiss, Bruno Gonçalves, Alessandro Flammini, and Filippo Menczer. Detecting and tracking political abuse in social media. *ICWSM*, 11:297–304, 2011.

[28] Megan Risdal. Fake news dataset. https://www.kaggle.com/mrisdal/fake-news, 2017.

[29] Justin Sampson, Fred Morstatter, Liang Wu, and Huan Liu. Leveraging the implicit structure within social media for emergent rumor detection. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, CIKM '16, pages 2377–2382, New York, NY, USA, 2016. ACM.

[30] Chengcheng Shao, Giovanni Luca Ciampaglia, Alessandro Flammini, and Filippo Menczer. Hoaxy: A platform for tracking online misinformation. In *Proceedings of the 25th International Conference Companion on World Wide Web*, pages 745–750. International World Wide Web Conferences Steering Committee, 2016.

[31] The Independent Shih, Gerry. Boston marathon bombings: How twitter and reddit got it wrong. Available online, http://www.independent.co.uk/news/world/americas/boston-marathon-bombings-how-twitter-and-reddit-got\-it-wrong-8581167.html, April 2013.

[32] Hyun Ah Song and Soo-Young Lee. Hierarchical representation using nmf. In *International Conference on Neural Information Processing*, pages 466–473. Springer, 2013.

[33] Liang Wu, Jundong Li, Xia Hu, and Huan Liu. Gleaning wisdom from the past: Early detection of emerging rumors in social media. SDM, 2016.

[34] Wei Xu, Xin Liu, and Yihong Gong. Document clustering based on non-negative matrix factorization. In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, pages 267–273. ACM, 2003.

[35] Reza Zafarani, Mohammad Ali Abbasi, and Huan Liu. *Social media mining: an introduction*. Cambridge University Press, 2014.

[36] Zhe Zhao, Paul Resnick, and Qiaozhu Mei. Enquiring minds: Early detection of rumors in social media from enquiry posts. In *Proceedings of the 24th International Conference on World Wide Web*, WWW '15, pages 1395–1405, Republic and Canton of Geneva, Switzerland, 2015. International World Wide Web Conferences Steering Committee.