

Tensor Factorization with Label Information for Fake News Detection

Frosso Papanastasiou*
School of Electrical and Computer
Engineering, National Technical
University of Athens
Athens, Greece
efpanastasiou@gmail.com

Georgios Katsimpras
NCSR Demokritos
Athens, Greece
gkatsibras@iit.demokritos.gr

Georgios Paliouras
NCSR Demokritos
Athens, Greece
paliourg@iit.demokritos.gr

ABSTRACT

The buzz over the so-called “fake news” has created concerns about a degenerated media environment and led to the need for technological solutions. As the detection of fake news is increasingly considered a technological problem, it has attracted considerable research. Most of these studies primarily focus on utilizing information extracted from textual news content. In contrast, we focus on detecting fake news solely based on structural information of social networks. We suggest that the underlying network connections of users that share fake news are discriminative enough to support the detection of fake news. Thereupon, we model each post as a network of friendship interactions and represent a collection of posts as a multidimensional tensor. Taking into account the available labeled data, we propose a tensor factorization method which associates the class labels of data samples with their latent representations. Specifically, we combine a classification error term with the standard factorization in a unified optimization process. Results on real-world datasets demonstrate that our proposed method is competitive against state-of-the-art methods by implementing an arguably simpler approach.

KEYWORDS

tensor factorization, fake news, social networks, joint optimization

1 INTRODUCTION

Many recent reports and studies suggest that the public is concerned about the impact of fake news. Specifically, a European-wide survey of 26,000 adults conducted during 2018, shows that fake news are widely spread across Europe and 85% of the participants believe that fake news is a serious problem [5]. The same study concludes that online social media is perceived as the least trusted source of news. Therefore, the need for a solution to stop the spread of fake news is emerging.

A great number of approaches to detecting fake news have been proposed in the literature. We can distinguish three main categories of such methods: i) content-based, ii) network-based and iii) hybrid. Content-based approaches emphasize on analysing the news

content and trying to associate language patterns with deception [10, 13, 22]. Also called linguistic, these methods rely on Natural Language Processing (NLP) tasks such as syntactic, semantic and sentiment analysis. While these methods show good results, this is only achieved in closed domains where the context is already known and static [6]. In contrast, network-based approaches focus on network information, such as friendship networks and propagation paths. These methods aim to extract information by analysing the connectivity of the networks [9, 14]. As such, network-based methods can perform well in dynamic environments and consequently in the task of fake news detection, which goes across topics and domains [29]. Incorporating both content and network information, in a complementary way, is the goal of the hybrid approaches. Although this is a promising direction, combining together heterogeneous information is a hard and time-consuming task [26].

Based on the facts that i) users play an important role in the dissemination of fake news [25], ii) users share information with similar users (friends) [30] and iii) network properties have been shown to be important for the classification of fake news [23], we investigate relationship networks of users and how they facilitate the spreading of fake news.

Therefore, in this work, we focus on the information that can be extracted from structured networks and specifically from relationship networks between users. In particular, we model each post as a network of friendship interactions and represent a collection of posts as a multidimensional tensor. Using this representation we can employ tensor factorization methods to produce compact representations of posts that capture the underlying network information. However, the standard factorization methods have limited effectiveness due to their unsupervised nature. In cases, where labeled data are available, incorporating class information could assist the factorization process to produce discriminatory representations and thus to better identify fake news.

In this paper, we propose CLASS-CP, a tensor-based semi-supervised approach for classifying fake news posts using network information and the available labeled data. Specifically, our approach combines tensor factorization and classification in a joint learning process. We model the data as a third-order tensor which represents the friendship relations that are present. At the same time, we use the labeled data we have in hand to influence the resulting factor matrices. To achieve this, we propose a tensor factorization method that assimilates class information about posts. While, commonly, factorization and classification are employed separately, we combine them in a single optimization process to

*Also with NCSR Demokritos.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

ROME 2019, July 25, 2019, Paris, France

© 2019 Copyright held by the owner/author(s).

obtain semi-supervised factorization. The main contributions of the paper are:

- a tensor-based approach for detecting fake news that use network information and labeled data,
- combination of tensor factorization with classification, achieving class-driven modeling of tensorial data,
- evaluation of the performance of the method on real-world datasets, in order to demonstrate its effectiveness.

The code for the proposed method is publicly available and can be found at <https://github.com/FrossoPap/class-cp>.

2 RELATED WORK

When it comes to fake news detection on social media, most approaches focus on finding patterns from news content, such as vocabulary, syntax, writing styles and images. For example, Hosseinotlagh et al. [13] proposed an unsupervised tensor modeling of the problem, based on term frequency and spatial relations between terms and articles. Guacho et al. [7] introduced a semi-supervised model via tensor embeddings that uses spatial/contextual information about news articles. Gupta et al. [8] proposed a classifier to estimate tweet credibility from features such as number of words, URLs, hashtags, emojis, presence of swear words, pronouns. Horne et al. [12] employed an SVM-based algorithm using stylistic, complexity and psychological features of both title and body text to classify real, fake and satire news. However, most of these content-based methods require closed domains with predefined context in order to perform well.

Recent work investigates network-based properties and features generated by the users’ social profiles and interaction with the news. For example, Shu et al. [28] proposed to compare explicit and implicit user profile features, in order to measure their potential to identify fake news. Yang et al. [32] developed an unsupervised fake news detection algorithm to understand the dependencies among the truths of news, the users’ opinions, and their credibility. Jin et al. [15] exploited conflicting viewpoints in news tweets with a topic model method and created a credibility propagation network of tweets that generates the final result. Castillo et al. [3] exploited features from users’ posting and re-posting behavior, text and citations to external sources. In another work, Shu et al. [27] proposed a tri-relationship embedding framework called TriFN, that models both publisher-news relations and user-news interactions for fake news classification. Existing state-of-the-art methods that consider network information, such as the TriFN framework, require the simultaneous modeling of many parameters that come from heterogeneous sources of information. This can be a hard and time consuming task.

In regard to the joint factorization framework that is proposed in this work, we can find several similar techniques in the literature. Context-aware methods [19, 21] combine contextual information with factorization to solve recommendation tasks. On the other hand, class-aware approaches [2, 16, 31] integrate class information into the factorization process but all these methods are optimized to perform in domains different from fake news detection.

Our method builds on top of the class-aware approaches and aims to develop a less complex but still effective fake news detection system that exploits the network of friendship interactions and

the class information of posts. The proposed method introduces an alternative way for modeling news posts through structural information and thus, it can be effective in numerous domains where content information is limited.

3 PRELIMINARIES AND NOTATION

Throughout the paper we use the following notation. The uppercase calligraphic letters denote tensors, e.g. $\mathcal{T} \in \mathbb{R}^{p \times u \times u}$, where p is the number of posts and u is the number of nodes (e.g. users). Matrices are represented by uppercase italic letters like A . Lowercase bold letters, like \mathbf{v} , denote vectors. The (i, j) element of a matrix A is denoted by a_{ij} . To refer to the i -th row of a matrix A we use \mathbf{a}_i . Similarly, an element (i, j, k) of a tensor \mathcal{T} will be denoted as \mathcal{T}_{ijk} . Additionally, $vec(A)$ is the vectorization of A , the operator \otimes is the Kronecker product and the operator \odot is the Khatri-Rao product. For our tensor modeling we use a binary representation:

$$\mathcal{T}_{ijk} = \begin{cases} 1, & \text{if user } k \text{ has engaged with post } i \\ & \text{and user } j \text{ follows user } k \\ 0, & \text{otherwise} \end{cases}$$

The *order* of a tensor, also known as *ways* or *modes*, is the number of its dimensions, therefore, \mathcal{T} is called a third-order tensor. Also, a r -order tensor is of rank-one if it can be strictly decomposed into the outer product of r vectors. In addition, to simplify the notation, we follow the same notation as in [17] for Kruskal operators, to express the factorization models in the next sections.

4 PROBLEM DEFINITION AND PROPOSED METHOD

4.1 Problem definition

Given a set of posts $P = (p_1, p_2, \dots, p_p)$, a set of friendship networks $G = (g_1, g_2, \dots, g_p)$ and assuming that a set of labels $Y = (y_1, y_2, \dots, y_l)$ is available for some posts (e.g. $l < p$), we want to efficiently predict the labels of the remaining $p - l$ posts. Solving this task raises the following research questions:

- (Q1) How can we combine the available friendship networks to facilitate the classification of nodes?
- (Q2) How can we incorporate the label information of the data in hand to enhance the analysis?

4.2 The CLASS-CP approach

Our approach combines two basic mechanisms to solve questions (Q1) and (Q2): the collective learning achieved through tensor factorization and the class-aware optimization in relation to the latent factors, respectively. For the former task, we employ the Canonical/Parafac (CP) factorization [1]. CP is a widely used tensor factorization method, which decomposes a tensor into a sum of rank-one tensors. The first step in our approach is to build a tensor from the friendship networks, i.e. from social media data¹. The result is a 3rd-order tensor, $\mathcal{T} \in \mathbb{R}^{p \times u \times u}$. The CP decomposition of such a tensor is computed by the following least-squares loss:

$$\min_{A, B, C} \|\mathcal{T} - \llbracket A, B, C \rrbracket\|^2,$$

¹It is worth noticing that the method can be readily applied to other domains as well (e.g. biology and medicine)

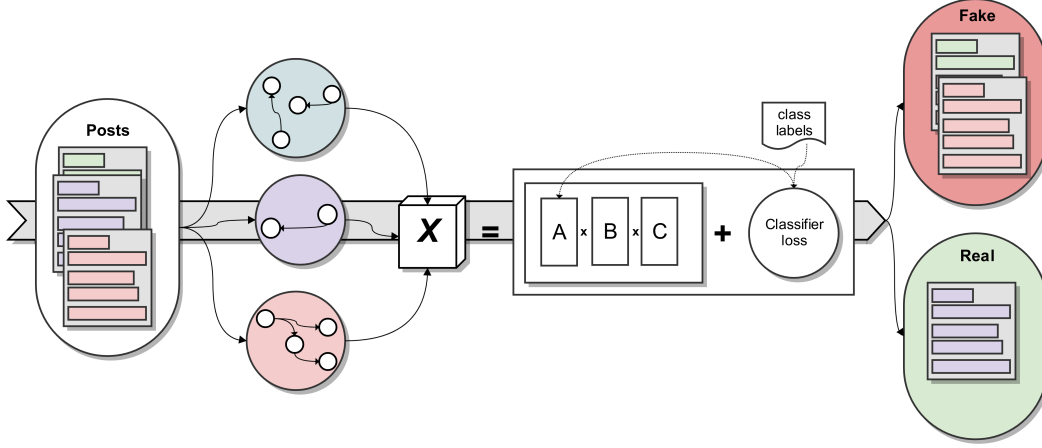


Figure 1: The conceptual diagram of CLASS-CP.

where $A \in \mathbb{R}^{p \times r}$ represents the posts-factor matrix, $B \in \mathbb{R}^{u \times r}$ and $C \in \mathbb{R}^{u \times r}$ represent the users-factor matrices, and r is the rank of the factorization.

To capture class-aware information we add a classification error term, based on the latent factor A . Thus, CLASS-CP extends the CP factorization using supervision. That is, the 3rd-order tensor \mathcal{T} is approximated, while taking into account the available labeled data. In particular, the post vectors in A , are linked to class labels.

The conceptual diagram of CLASS-CP is depicted in Figure 1. Firstly, we model posts as a network of friendship interactions and we represent a collection of them as a three-dimensional tensor, though, CLASS-CP is able to accept higher modes as well. As we will show, in the learned latent space, the posts-factor matrix A is biased by a classifier. To achieve this, we employ a joint optimization process that learns to predict fake news posts.

4.2.1 Joint optimization in CLASS-CP. Assuming a set of labels for some of the posts, we expect similarly labeled posts to share similar factors[20]. As mentioned above, we introduce class-label information into the tensor factorization, in order to move posts of the same class closer in the latent space. In particular, we introduce a matrix of coefficients $W \in \mathbb{R}^{r \times c}$, where r is the number of factors and c is the number of class labels. This matrix assigns labels to posts based on the factor matrix A as in a common regression problem $Y = AW$.

Given a tensor $\mathcal{T} \in \mathbb{R}^{p \times u \times u}$ and a set of labels $Y \in \mathbb{R}^{l \times c}$, where l is the number of labeled posts with $l < p$, and c is the number of class labels, we solve the optimization problem presented in eq. 1.

$$\underset{A, B, C, W}{\text{minimize}} \quad f(A, B, C) + g(A, W), \quad (1)$$

where

$$f(A, B, C) = \|\mathcal{T} - \llbracket A, B, C \rrbracket\|^2$$

$$g(A, W) = \lambda_g \|Y - DAW\|^2 \quad (2)$$

is the tensor factorization least squares problem,

is the prediction error of the classifier, where $D = [I^{l \times l}, 0^{l \times (p-l)}] \in \mathbb{R}^{l \times p}$ and λ_g is a hyperparameter to control the influence of the

classification error in the optimization. Note that $g(A, W)$ is produced with respect to the current values of the post factor matrix A .

We solve the minimization problem in eq. 1, using the efficient alternating least squares method (ALS). This approach alternately fixes and solves factor matrices following update rules. The update rules are derived by setting the gradient of eq. 1 with respect to each factor matrix to zero.

For CP, the ALS method fixes every factor matrix except one and solves for it. The matricized form of f (one per mode) can be written as:

$$\begin{aligned} \mathcal{T}_{(1)} &\approx A(C \odot B)^T, \\ \mathcal{T}_{(2)} &\approx B(C \odot A)^T, \\ \mathcal{T}_{(3)} &\approx C(B \odot A)^T. \end{aligned} \quad (3)$$

Recall that the \odot operator denotes the Khatri-Rao product.

4.2.2 Updating factor matrices B and C . To find the updates for B and C , we ignore the second term of eq.1 and we solve the respective equations as presented in eq. 3. The ALS approach fixes A and C to solve for B , and A, B to solve for C . Let $Z_{(2)} = (C \odot A)$ and $Z_{(3)} = (B \odot A)$, then by combining eq. 1 and eq. 3 we can write:

$$\min_B \quad \|\mathcal{T}_{(2)} - BZ_{(2)}^T\|^2 \quad (4)$$

and

$$\min_C \quad \|\mathcal{T}_{(3)} - CZ_{(3)}^T\|^2. \quad (5)$$

The solution to eq. 4 and eq. 5 is given by [18] and so the updates of B and C can be written as:

$$B = (Z_{(2)}^T Z_{(2)})^{-1} Z_{(2)}^T \mathcal{T}_{(2)} \quad (6)$$

and

$$C = (Z_{(3)}^T Z_{(3)})^{-1} Z_{(3)}^T \mathcal{T}_{(3)} \quad (7)$$

4.2.3 *Updating factor matrix A.* Let $Z_{(1)} = (C \odot B)$. To find the update for factor matrix A , we can combine eq. 1 and eq. 3 and write:

$$\min_A \left\| \mathcal{T}_{(1)} - AZ_{(1)}^T \right\|^2 + \|Y - DAW\|^2. \quad (8)$$

Equation 8 can be solved through its vectorized form:

$$\min_A \left\| \text{vec}(\mathcal{T}_{(1)}) - (Z_{(1)} \otimes I_p) \text{vec}(A) \right\|^2 + \left\| \text{vec}(Y) - (W^T \otimes D) \text{vec}(A) \right\|^2. \quad (9)$$

Letting $G = (Z_{(1)} \otimes I_p)$ and $L = (W^T \otimes D)$, the solution to eq. 9 can be calculated and the update for A is given by:

$$\text{vec}(A) = (G^T G + L^T L)^{-1} (G^T \text{vec}(\mathcal{T}_{(1)}) + L^T \text{vec}(Y)) \quad (10)$$

4.2.4 *Updating factor matrix W.* To find the update of W , we ignore the first term of eq. 1 and so we solve:

$$\min_W \|Y - DAW\|^2. \quad (11)$$

The solution of eq. 11 with respect to W gives the update of W :

$$W = (A^T D^T D A)^{-1} A^T D^T Y \quad (12)$$

Algorithm 1 CLASS-CP: Given a tensor \mathcal{T} and a set of labels Y , approximate A, B, C and W

Require: tensor \mathcal{T} , labels Y

Ensure: factor matrices A, B, C and coefficients W

- 1: Initialize A, B, C, W and hyperparameter λ_g
 - 2: **repeat**
 - 3: update A using Eq. 10
 - 4: update B using Eq. 6
 - 5: update C using Eq. 7
 - 6: update W using Eq. 12
 - 7: **until** convergence
-

The algorithm of CLASS-CP that brings together all of the above is depicted in Alg. 1. To compute the factor matrices A, B, C and the coefficient matrix W the algorithm performs alternating updates until it converges to a criterion or it reaches a maximum number of iterations. As a convergence criterion we use the *relative change* which can be calculated in each iteration as:

$$\frac{|(f_{new} + g_{new}) - (f_{old} + g_{old})|}{f_{old} + g_{old}} \quad (13)$$

5 EXPERIMENTAL EVALUATION

5.1 Datasets

To evaluate our approach we used two real-world public datasets which have been used previously in the literature [24–26]. The datasets were collected from two platforms, BuzzFeed and PolitiFact. They include both news content and social context features with fact-checked ground truth labels. News content features include meta-information such as the title and body text, and social context includes the related user’s profile information and activity (e.g., user sharing news on Twitter, user’s Twitter followers). For our

Table 1: Numbers of FakeNewsNet dataset

Media Platform	BuzzFeed	PolitiFact
# Users	1449	1697
# Engagements	8598	10249
# Social Links	6571	3093
# True news	91	120
# Fake news	91	120

evaluation, we use only social context features such as the friendship network which indicates the following/followee structure of users who post related posts.

In order to decrease the size and the sparsity of the data, we removed users with a node degree < 3 . As a result, we produced two tensors of size $182 \times 1449 \times 1449$ and $240 \times 1697 \times 1697$ for the BuzzFeed and PolitiFact datasets respectively. The basic statistics of the datasets after nodes deletion are shown in more detail in Table 1. User engagements include posts, re-posts and replies related with a news article.

5.2 Experimental Settings

To calculate the performance of CLASS-CP, we use Accuracy, Precision, Recall, and F1, that are commonly used as evaluation metrics in similar problems. We choose the first 80% of news posts as our training set and the remaining 20% for testing. Our data is balanced, meaning that the number of news posts labeled as true is equal to the number of fake posts. We perform the same process 10 times, independently for each dataset, and we report the average results.

Number of factors. Before comparing CLASS-CP against other approaches, we measure the effect of the number of factors on the results. We set the number of factors $r = 2, 3, 5, 10, 15, 20$. A larger r increases the computation time of the factorization. The best overall performance is achieved for $r = 5$ as shown in Figure 2 and we use it for the rest of the experiments.

Impact of training set size on performance. Additionally, we explore the behavior of CLASS-CP with respect to the size of the labeled data. In Figure 3, we show the F1-score learning curves in relation to the labeled data ratio on both datasets. From the plot, we can see that the more labeled instances are used in the training stage, the better performance.

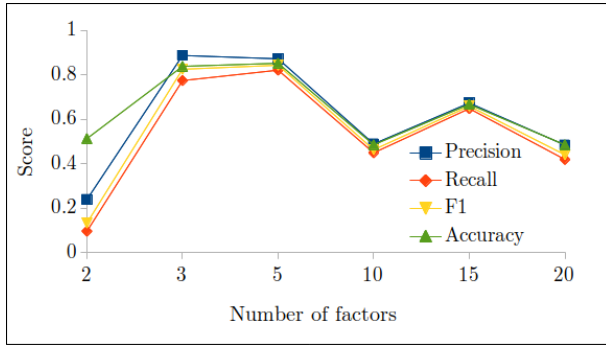
Hyperparameter λ_g . We set the value of the hyperparameter $\lambda_g = 1$ so as the classification loss can contribute equally to the optimization process, as the factorization.

Execution time. Each iteration takes less than a minute to finish and we need on average a few minutes (< 10) to reach convergence.

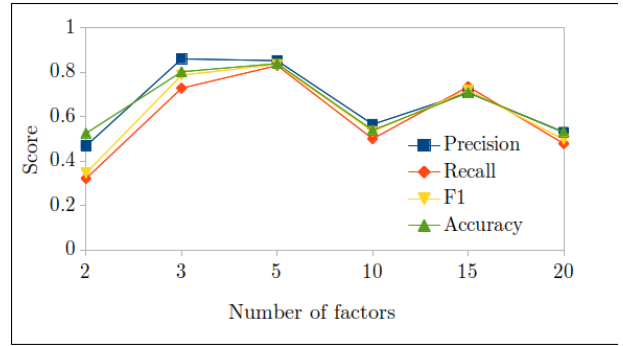
5.3 Classification of Fake News

We compare the proposed CLASS-CP algorithm with the following methods:

- **SVM:** A baseline linear SVM model that is built by using the rows of the tensor \mathcal{T} as input feature vectors.
- **CP+SVM:** The original CP factorization combined with a linear SVM classifier. This is a two-step process. In the first step we calculate the tensor embeddings of posts using CP in an unsupervised setting and in the second step we train a



(a) PolitiFact



(b) BuzzFeed

Figure 2: Behavior of CLASS-CP with different number of factors for each dataset.

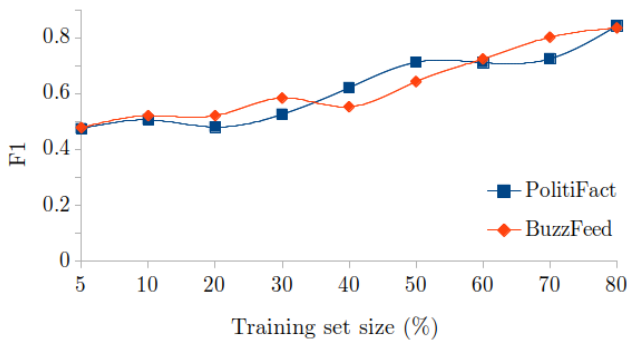


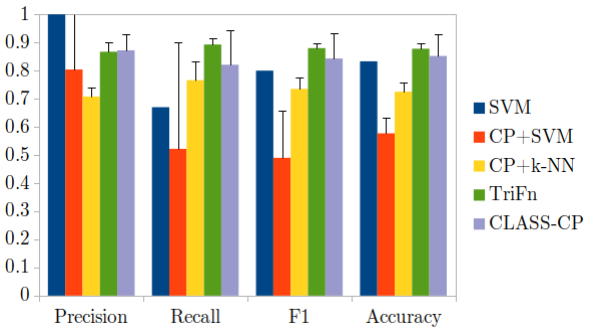
Figure 3: The learning curve on BuzzFeed and PolitiFact

SVM classifier using the embeddings produced in the first step.

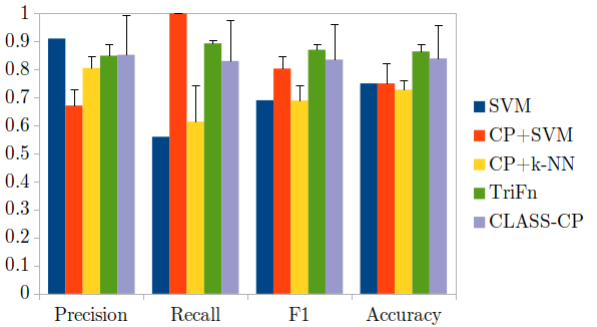
- **CP+k-NN**: The CP factorization combined with a k -NN classifier. We follow the same procedure as in CP+SVM but instead of a SVM classifier, we use a k -NN classifier. As suggested by [4, 11], we set k equal to the square root of the size of the training data i.e. $k = \sqrt{p}$.
- **TriFN**: TriFN is a state-of-the-art fake news detection method that extracts features separately from news publisher and user interactions and captures the interrelationship simultaneously. We use the same evaluation settings as proposed in [27] so as to compare our results with the ones reported in that study.

Experimental results. Figure 4 and Tables 2 and 3 show the comparison of CLASS-CP and the aforementioned methods for different metrics. From these, we can make the following observations:

- The SVM baseline approach gives the highest Precision results, but is weaker in terms of the other metrics.
- The use of CP does not improve the performance of the SVM.
- CLASS-CP significantly outperforms SVM, CP+SVM and CP+ k -NN both on the PolitiFact and the BuzzFeed datasets.
- CLASS-CP performs comparable to TriFN in both datasets.



(a) PolitiFact



(b) BuzzFeed

Figure 4: Performance comparison for different metrics

The above suggest that (i) using only network information and the available labeled data, (ii) by incorporating class information during and not after the factorization process and (iii) with a small number of factors that lead to short computation times, we can achieve an equally good performance as state-of-the-art frameworks, such as the TriFN, that require the fusion of many heterogeneous data sources and complex calculations. Therefore, we confirm our initial intuition that the use of the underlying network between users can reveal valuable information about fake news.

Table 2: Performance comparison on PoliFact

Metric	SVM	CP+SVM	CP+ k -NN	TriFN	CLASS-CP
Precision	1.000	0.804 \pm 0.253	0.708 \pm 0.033	0.867 \pm 0.034	0.872 \pm 0.058
Recall	0.670	0.522 \pm 0.380	0.766 \pm 0.066	0.893 \pm 0.023	0.821 \pm 0.122
F1	0.800	0.490 \pm 0.168	0.735 \pm 0.039	0.880 \pm 0.017	0.843 \pm 0.089
Accuracy	0.833	0.577 \pm 0.055	0.725 \pm 0.034	0.878 \pm 0.020	0.852 \pm 0.078

Table 3: Performance comparison on BuzzFeed

Metric	SVM	CP+SVM	CP+ k -NN	TriFN	CLASS-CP
Precision	0.910	0.671 \pm 0.058	0.805 \pm 0.041	0.849 \pm 0.040	0.852 \pm 0.143
Recall	0.560	1.000	0.614 \pm 0.137	0.893 \pm 0.013	0.830 \pm 0.146
F1	0.690	0.803 \pm 0.044	0.689 \pm 0.056	0.870 \pm 0.019	0.835 \pm 0.127
Accuracy	0.750	0.750 \pm 0.071	0.728 \pm 0.032	0.864 \pm 0.026	0.839 \pm 0.118

6 CONCLUSION

In this paper, instead of employing tensor factorization and classification separately, as is common, we propose a method that combines them in a joint learning process for detecting fake news posts in social media. This tensor-based approach uses the follower/followee structure of users that have engaged with the news, as well as a set of labeled news that are provided. In order to combine this information, we introduced an extension of the standard tensor factorization method CP that incorporates class-label information into the factorization itself. In this manner we obtain a class-aware semi-supervised tensor factorization method.

For the evaluation of the method, we conducted experiments with two real-world public datasets that confirm the importance of incorporating the class information in the factorization process, rather than using it in a separate step. The results also confirmed our initial intuition, that the way users are connected with each other can reveal the truthfulness of the news they share on social media.

As further work, we would like to explore the extent to which CLASS-CP improves when more features, both from the network and the content of the posts, are added. Additionally, we plan to assess the performance of our approach on more datasets and investigate the impact of data size on scalability. At a methodological level, it would be interesting to explore new ways of representing the available information with tensors, in order to incorporate it in the proposed framework.

REFERENCES

- [1] Richard A. Harshman and Margaret E. Lundy. 1994. PARAFAC: Parallel factor analysis. *Computational Statistics Data Analysis* (1994).
- [2] Bokai Cao, Chun-Ta Lu, Xiaokai Wei, S Yu Philip, and Alex D. Leow. 2016. Semi-supervised tensor factorization for brain network analysis. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*.
- [3] Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. 2011. Information Credibility on Twitter. (2011).
- [4] Richard O. Duda, Peter E. Hart, and David G. Stork. 2001. *Pattern Classification* (2nd. ed.). Wiley.
- [5] EU. 2018. Flash Eurobarometer on Fake News and Online Disinformation. <https://ec.europa.eu/digital-single-market/en/news/final-results-eurobarometer-fake-news-and-online-disinformation>
- [6] Torstein Granskogen. 2018. *Automatic Detection of Fake News in Social Media using Contextual Information*. Master’s thesis.
- [7] Gisel Bastidas Guacho, Sara Abdali, Neil Shah, and Evangelos E. Papalexakis. 2018. Semi-supervised Content-based Detection of Misinformation via Tensor Embeddings. *arXiv:1804.09088v1* (2018).
- [8] Aditi Gupta and Ponnurangam Kumaraguru. 2012. Credibility ranking of tweets during high impact events.
- [9] Manish Gupta, Peixiang Zhao, and Jiawei Han. [n. d.]. Evaluating event credibility on twitter. In *Proceedings of the 2012 SIAM International Conference on Data Mining*.
- [10] Momchil Hardalov, Ivan Koychev, and Preslav Nakov. 2016. In search of credible news. In *International Conference on Artificial Intelligence: Methodology, Systems, and Applications*.
- [11] Ahmad Basheer Hassanat, Mohammad Ali Abbadi, Ghada Awad Altarawneh, and Ahmad Ali Alhasanat. 2014. Solving the Problem of the K Parameter in the KNN Classifier Using an Ensemble Learning Approach. (2014).
- [12] Benjamin D. Horne and Sibel Adalás. 2017. This Just In: Fake News Packs a Lot in Title, Uses Simpler, Repetitive Content in Text Body, More Similar to Satire than Real News. (2017).
- [13] Seyedmehdi Hosseinimotlagh and Evangelos E. Papalexakis. 2018. Unsupervised Content-Based Identification of Fake News Articles with Tensor Decomposition Ensembles.
- [14] Zhiwei Jin, Juan Cao, Yongdong Zhang, and Jiebo Luo. 2016. News verification by exploiting conflicting social viewpoints in microblogs. In *Thirtieth AAAI Conference on Artificial Intelligence*.
- [15] Zhiwei Jin, Juan Cao, Yongdong Zhang, and Jiebo Luo. 2016. News verification by exploiting conflicting social viewpoints in microblogs. (2016).
- [16] Georgios Katsimpras and Georgios Paliouras. 2018. Semi-Supervised Tensor Factorization for Identifying Nodes in Complex Social Networks. In *The Joint International Workshop on Social Influence Analysis and Mining Actionable Insights from Social Networks*.
- [17] Tamara Gibson Kolda. 2006. *Multilinear operators for higher-order decompositions*. Technical Report.
- [18] Tamara G Kolda and Brett W Bader. 2009. Tensor decompositions and applications. *SIAM review* (2009).
- [19] Yize Li, Jiazhong Nie, and Yi Zhang. 2010. Contextual Recommendation based on Text Mining. *Proceedings of the 23rd International Conference on Computational Linguistics* (2010).
- [20] Miller McPherson, Lynn Smith-Lovin, and James M Cook. 2001. Birds of a feather: Homophily in social networks. *Annual review of sociology* (2001).
- [21] Steffen Rendle, Zeno Gantner, Christoph Freudenthaler, and Lars Schmidt-Thieme. 2011. Fast context-aware recommendations with factorization machines. *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval (SIGIR '11)* (2011).
- [22] Victoria Rubin, Niall Conroy, Yimin Chen, and Sarah Cornwell. 2016. Fake news or truth? using satirical cues to detect potentially misleading news. In *Proceedings of the second workshop on computational approaches to deception detection*.
- [23] Kai Shu, H Russell Bernard, and Huan Liu. 2019. Studying fake news via network analysis: detection and mitigation. In *Emerging Research Challenges and Opportunities in Computational Social Network Analysis and Mining*.
- [24] Kai Shu, Deepak Mahudeswaran, Suhang Wang, Dongwon Lee, and Huan Liu. 2018. FakeNewsNet: A Data Repository with News Content, Social Context and Dynamic Information for Studying Fake News on Social Media. *arXiv preprint arXiv:1809.01286* (2018).
- [25] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations*

Newsletter (2017).

- [26] Kai Shu, Suhang Wang, and Huan Liu. 2017. Exploiting Tri-Relationship for Fake News Detection. *arXiv preprint arXiv:1712.07709* (2017).
- [27] Kai Shu, Suhang Wang, and Huan Liu. 2018. Beyond News Contents: The Role of Social Context for Fake News Detection. (2018).
- [28] Kai Shu, Suhang Wang, and Huan Liu. 2018. Understanding User Profiles on Social Media for Fake News Detection. (2018).
- [29] Eugenio Tacchini, Gabriele Ballarin, Marco L Della Vedova, Stefano Moret, and Luca de Alfaro. 2017. Some like it hoax: Automated fake news detection in social networks. *arXiv preprint arXiv:1704.07506* (2017).
- [30] Bing Tian Dai, Freddy Chong Tat Chua, and Ee-Peng Lim. 2012. Structural analysis in multi-relational social networks. In *Proceedings of the 2012 SIAM international conference on data mining*.
- [31] Yan-Hui Xiao, Zhen-Feng Zhu, Yao Zhao, and Yun-Chao Wei. 2013. Class-driven non-negative matrix factorization for image representation. *Journal of Computer Science and Technology* (2013).
- [32] Shuo Yang, Kai Shu, Suhang Wang, Renjie Gu, Fan Wu, and Huan Liu. 2019. Unsupervised Fake News Detection on Social Media:A Generative Approach. (2019).