

Semi-Supervised Tensor Factorization for Node Classification in Complex Social Networks

Georgios Katsimbras and Georgios Paliouras
NCSR Demokritos, Ag. Paraskevi, Athens, Greece
gkatsibras@iit.demokritos.gr, paliourg@iit.demokritos.gr

Abstract

This paper proposes a method to guide tensor factorization, using class labels. Furthermore, it shows the advantages of using the proposed method in identifying nodes that play a special role in multi-relational networks, e.g. spammers. Most complex systems involve multiple types of relationships and interactions among entities. Combining information from different relationships may be crucial for various prediction tasks. Instead of creating distinct prediction models for each type of relationship, in this paper we present a tensor factorization approach based on RESCAL, which collectively exploits all existing relations. We extend RESCAL to produce a semi-supervised factorization method that combines a classification error term with the standard factor optimization process. The coupled optimization approach, models the tensorial data assimilating observed information from all the relations, while also taking into account classification performance. Our evaluation on real-world social network data shows that incorporating supervision, when available, leads to models that are more accurate.

1 Introduction

In recent years, we are experiencing an enormous growth of online social platforms where complex multi-relational networks take form. Interactions in a multi-relational network provide information that helps to identify the social position of participating parties [Ferrante, 2012]. Hence, identifying individuals of similar social position or role can be modeled as a task of identifying network nodes that have similar relationships [Tian Dai *et al.*, 2012]. In many cases there is also additional information about individuals apart from relationships and interactions.

Given the richness of information in multi-relational networks, new methods to analyze this information and produce prediction models are needed. In this direction, several approaches have been proposed to deal with the identification of key nodes in multi-relational networks, such as multilinear PageRank [Gleich *et al.*, 2015] and the multi-relational version of hubs and authorities [Li *et al.*, 2012]. However, meth-

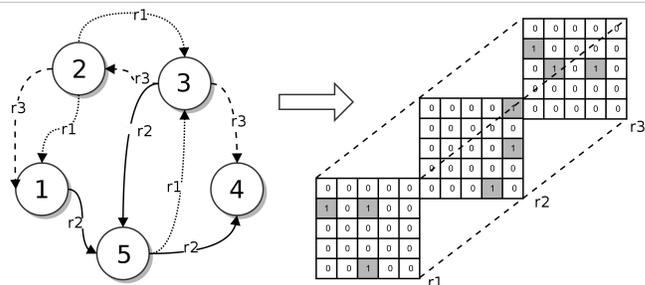


Figure 1: A multi-relational network can be easily mapped to a tensor. As we can see, the network at the left can be mapped to a $5 \times 5 \times 3$ tensor. Each frontal slice represents the adjacency matrix for a specific relation

ods that integrate interactions and properties of the nodes, utilizing tensor representations and tensor decomposition have not been studied extensively.

Tensors and their factorization have been used widely in machine learning and data mining. Due to the nature of tensors, relations expressed as tuples of the form (*object, relation, subject*), can be straightforwardly mapped to a tensor. For instance, the multi-relational network in Figure 1 can be represented as a $5 \times 5 \times 3$ tensor where an entry (i, j, r) is 1 only if the relation exists. Rows and columns represent the nodes and each slice/matrix represents the adjacency matrix for a distinct relation.

In this work, we approach the problem of identifying the position and role of individuals in multi-relational networks by representing the networks as tensors and performing tensor factorization. Tensors can provide a simple representation of multi-relational networks while tensor factorization approaches perform well in high dimensional and sparse data [Kolda and Bader, 2009]. Specifically, we propose a new method for ranking nodes according to their social position in multi-relational networks, combining tensor factorization and classification in a joint learning process. Similar to TripleRank [Franz *et al.*, 2009], our idea is to produce a ranking of users based on the factor scores that are computed during the factorization process. However, in our approach the factor matrices are a result of the proposed semi-supervised factorization process, which forces nodes that belong to the same class to obtain similar factor representation. Thus, ranking in

these biased factor matrices will take their class into account.

In order to achieve this, we propose an enhanced version of the RESCAL tensor factorization method that assimilates a-priori information about nodes. Additionally, the factorization process incorporates a component that minimizes classification error. In contrast to existing research that treat factorization and classification separately, we combine them in a single optimization process to obtain semi-supervised factorization. We evaluate the proposed method on a real-world dataset collected from the social network Tagged.com, that consists of a network of 5,317,649 users and seven types of relations [Fakhraei *et al.*, 2015].

In summary, the main contributions of the paper are:

- Combination of tensor factorization with classification, achieving class-driven modeling of tensorial data.
- Demonstration of the benefits of using the proposed method to rank social network users according to their social position, using a real-world dataset

2 Background

2.1 Notation

Throughout the paper we will use the following notation. The uppercase calligraphic letters denote a tensor $\mathcal{T} \in R^{N \times N \times M}$, as we assume the data is given in such dimensions. Here N is the number of entities (e.g. individuals) and M is the number of relations. Matrices are represented by uppercase italic letters like A . Lowercase bold letters like \mathbf{v} denote a vector. 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} . Also, \mathcal{T}_k represents the k frontal slice of tensor \mathcal{T} . Additionally, $vec(A)$ is the vectorization of A and the operator \otimes is the Kronecker product.

For our relational modeling of triples of the form (*object, relation, subject*), we use a binary representation:

$$\mathcal{T}_{ijk} = \begin{cases} 1 & \text{if the relation exists} \\ 0 & \text{otherwise} \end{cases}$$

This representation is also known as an adjacency tensor. As an example, the relation (*u1, r2, u5*) in Figure 1 exists, causes the corresponding entry in the tensor to be 1, i.e. $\mathcal{T}_{1,5,2} = 1$. In this sense, each frontal slice of \mathcal{T} represents an adjacency matrix for all entities with respect to a particular relation.

2.2 Related Work

As multi-relational data can be efficiently represented by tensors, many approaches have been proposed in the literature that employ tensor factorizations. Common tensor factorization approaches such as CANDECOMP/PARAFAC [Harshman and Lundy, 1994], Tucker factorization [Tucker, 1966] and DEDICOM [Bader *et al.*, 2007] have been utilized. Most of these studies use a common latent representation for both entities and relations. CANDECOMP/PARAFAC decomposes a tensor into a sum of rank-one tensors. Tucker decomposes a tensor into a core tensor and separate factor matrices for each mode of the tensor.

A recent approach to tensor factorization is RESCAL [Nickel *et al.*, 2011] which achieves high predictive performance in the task of link prediction. RESCAL, which we will describe in more detail in Section 3, uses a unique latent representation for entities.

Recently, several studies have combined matrix and tensor factorization in a joint model. These techniques achieve good performance when there is additional features about the nodes or when combining different datasets together. A CANDECOMP/PARAFAC based method for jointly factorizing a matrix and a tensor was proposed in [Acar *et al.*, 2013]. The matrix was used to hold additional node properties. Similar approaches appear in [Narita *et al.*, 2011; Erdos and Miettinen, 2013; Nimishakavi *et al.*, 2016].

Identifying Key Nodes in Multi-relational Data. The problem of identifying key nodes in a network has been extensively studied in the literature. While most of the existing methods apply to single relation networks, there is also great interest in multi-relational networks. The study in [Gleich *et al.*, 2015] is an extension of PageRank to multi-relational networks. In [Liu *et al.*, 2010] they propose a graphical model which utilizes heterogeneous link information and the textual information associated to each node. In [Zhaoyun *et al.*, 2013] a method for performing combined random walks exploiting multi-relational influence networks is presented. MultiRank [Ng *et al.*, 2011] is a method to simultaneously determine the importance of both objects and relations based on a probability distribution computed from multi-relational data. HAR [Li *et al.*, 2012] calculates hub, authority and relevance scores of nodes in multi-relational networks. In [Jingjing *et al.*, 2013] a three-stage process is proposed which a) collects features of the users, b) performs a tensor factorization on user relations and c) applies a ranking scheme to identify the most influential users.

As multi-relational data can be efficiently represented by tensors, TripleRank [Franz *et al.*, 2009] employs tensor factorization in order to rank entities in the context of linked data. TripleRank applies a common tensor factorization CANDECOMP/PARAFAC [Harshman and Lundy, 1994] to obtain two factor matrices which correspond to hub and authority scores. Another approach [Rendle *et al.*, 2009], utilizes tensor factorization for ranking tags in order to provide tag recommendations, using multi-target networks which involves more than one entity types.

In contrast to these approaches, our method incorporates classification error in the optimization process. In this direction, there are only few approaches that obtain the factor matrices in a supervised manner, that is, by utilizing the label information of the samples. Most of these approaches build a classifier using the factor matrices that are produced by the decomposition [Li and Ngom, 2010]. Differently from these methods, our approach performs decomposition by estimating the classification error with respect to the factor matrices in each update. Similar to our approach, the method in [Wu *et al.*, 2013] pursues discriminative tensor decomposition by coupling non-negative Tucker tensor factorization and a maximum margin classifier. In contrast to this method, our approach (a) utilizes RESCAL which does not require the construction of core tensor and (b) produces factor matrices by

employing Alternating Least Squares (ALS) which has been proved to be more efficient than gradient descent.

3 Our Approach: Semi-Supervised RESCAL (CLASS-RESCAL)

RESCAL tensor factorization. RESCAL is a state-of-the-art relational learning method based on a variant of DEDICOM tensor factorization. It achieves high predictive performance in various tasks such as link prediction and collective classification [Nickel *et al.*, 2011]. A basic aspect of RESCAL’s model is that an entity has a unique representation over all relations in the data.

Standard tensor factorization models such as CANDECOMP/PARAFAC and Tucker use a bipartite model of relational data, meaning that entities have different latent representations as subjects or objects in relations. Consequently, all direct and indirect relations have a determining influence on the calculation of the latent factors.

RESCAL factorization is computed for each frontal slice as follows:

$$\mathcal{T}_k = AR_kA^T, \text{ for } k = 1, \dots, M$$

where A is a $n \times r$ latent factor matrix that models entities and R_k is an asymmetric $r \times r$ latent factor matrix that models relations. More precisely, each entity is represented via a unique row in A and each relation via R_k .

To compute this factorization, a regularized least squares optimization problem must be solved:

$$\begin{aligned} \underset{A, R}{\text{minimize}} \quad & \sum_k \|\mathcal{T}_k - AR_kA^T\|^2 \\ & + \lambda_A \|A\|^2 + \lambda_R \sum_k \|R_k\|^2 \end{aligned} \quad (1)$$

where λ_A and λ_R are regularization hyperparameters. This function can be minimized via Alternating Least Squares (ALS) [Bader *et al.*, 2007], where the updates of A and R_k can be calculated as:

$$\begin{aligned} A \leftarrow & \left[\sum_{k=1}^M \mathcal{T}_k AR_k^T + \mathcal{T}_k^T AR_k \right] \\ & \left[\sum_{k=1}^M R_k A^T AR_k^T + R_k^T A^T AR_k + \lambda I \right]^{-1} \end{aligned} \quad (2)$$

and

$$R_k \leftarrow (Z^T Z + \lambda I)^{-1} Z \text{vec}(\mathcal{T}_k) \quad (3)$$

where $Z = A \otimes A$. Here, it is important to recall that during the optimization process, a unique latent representation for each entity is shared over all relations. This means that the resulting A takes into account all relations.

Introducing classification error. We now describe our joint model for tensor factorization and classification. Suppose we have a set of labels for our existing tensorial data. For example, we have a label for some nodes in Figure 1 which indicates whether the user is a spammer or not. We could assume that spammers interact and relate with each other [McPherson *et al.*, 2001]. In other words, we expect

similarly labeled entities to share similar factors. Based on this assumption, we propose CLASS-RESCAL, to introduce class-label information in the tensor factorization, in order to move entities of the same class closer in the latent space. CLASS-RESCAL models this problem as a joint optimization of tensor factors and classification.

Given an adjacency tensor $\mathcal{T} \in \mathbb{R}^{N \times N \times M}$, as presented in Section 2, a set of labels $Y \in \{-1, 1\}$ where $i = 1, \dots, K$ with $K < N$, we solve the optimization problem presented in Eq. 4.

$$\underset{A, R}{\text{minimize}} \quad f(A, R) + g(A) + h(A, R) \quad (4)$$

where

$$f(A, R) = \sum_k \|\mathcal{T}_k - AR_kA^T\|^2$$

is the tensor factorization least squares problem,

$$g(A) = \lambda_g \|Y - y(A)\|^2 \quad (5)$$

is the prediction error of the classifier and

$$h(A, R) = \lambda_A \|A\|^2 + \lambda_R \sum_k \|R_k\|^2$$

is the regularization term. λ_g is a hyperparameter to control the influence of the classification error in the optimization. Note that $y(A)$ is produced with respect to the current values of the latent factor matrix A in each iteration.

The Classifier. Each row \mathbf{a}_i of the factor matrix A represents a unique latent-component representation of the i -th entity. The similarity of entities is assumed to be reflected in their latent-component representation. Namely, entities that share similar relations will also share similar latent representations. So, if the i -th entity is similar to j -th entity then also their rows on the factor matrix A should be similar, e.g. $\mathbf{a}_i \approx \mathbf{a}_j$. In order to force similar entities of the same class to have similar latent representation we minimize the prediction error of the classifier, Eq. 5. To induce the similarities between entities into the decomposition we use a k -NN classifier. Nearest neighbor techniques use the similarity of the local neighborhood of a node to make a prediction, assuming that the social position of a node is close to the social positions of its neighbors. Under this assumption, we assign a label to each row of A based on its top k neighbors in A (see Figure 2).

For example, assume that Figure 2 shows the factor scores of A for the network in Figure 1. The latent representation of node 3 is similar to the latent representations of nodes 1 and 4. Assuming that both 1 and 4 are spammers according to supervision, the majority rule will identify user 3 as spammer.

Updates. We solve the minimization problem in Eq. 4 by using the efficient alternating least squares method (ALS), as in [Nickel *et al.*, 2011]. This approach fixes and solves A and R by update rules, that set the gradient of Eq. 4 with respect to each matrix to zero. The update rule for R_k will remain the same compared to the original RESCAL update rule, as in Eq. 3, while the update rule for A becomes:

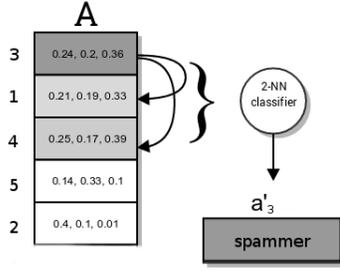


Figure 2: The process of k-NN classification. The top-k similar entities are considered for labeling. The labeling is based on majority voting. Here, top-2 nn of node 3 are nodes 1 and 4, who are labeled as spammers (shaded). As a result, node 3 will also be classified as spammer.

$$A \leftarrow \left[\sum_{k=1}^m \mathcal{T}_k A R_k^T + \mathcal{T}_k^T A R_k + 2(Y - y(A)) \right. \\ \left. \left[\sum_{k=1}^m R_k A^T A R_k^T + R_k^T A^T A R_k + \lambda I \right]^{-1} \right] \quad (6)$$

When $\frac{f(A,R)+g(A)+h(A,R)}{\|\mathcal{T}\|^2}$ converges to a predefined threshold ϵ or a maximum number of iterations is exceeded, the procedure stops. The details of the proposed method are summarized in Algorithm 1.

Algorithm 1 CLASS-RESCAL: Given a tensor \mathcal{T} and a set of labels Y , approximate A and R

Require: adjacency tensor \mathcal{T} , labels Y ,
Ensure: latent matrix A , latent matrices R_k

- 1: Initialize A , R and hyperparameters λ_g, λ_A
- 2: **repeat**
- 3: **procedure** UPDATE(A)
- 4: $y(A) \leftarrow (\text{CLASSIFIER}(A))$
- 5: update A using Eq. 6
- 6: **end procedure**
- 7: **procedure** UPDATE(R)
- 8: **for** each k in k -relations **do**
- 9: update R_k using Eq. 3
- 10: **end for**
- 11: **end procedure**
- 12: **until** convergence

Identifying entities. In RESCAL, the factor matrix A is a latent representation of users. Each row \mathbf{a}_i contains r scores, one for each latent factor. A ranking of users based on the factor scores will thus prioritize the users on how they are placed on the factor. However, in CLASS-RESCAL the factor matrix A is also influenced by the classifier which forces users of the same class to obtain similar factor representation. Therefore, to identify key users we rank the rows of the factor matrix A . For each row \mathbf{a}_i we compute a ranking score based

on the factor scores as follows:

$$\mathbf{a}_i^{\text{score}} = \frac{\sum_{j=1}^r a_{ij}}{r}$$

Folding-in a new instance. Given a new instance, we want to project it to the existing latent space of A and R . Figure 3, illustrates the folding-in process for CLASS-RESCAL. The shaded part of \mathcal{T} is the new instance which represents a new user, $\mathcal{T}_{(new)}$. R remains unchanged, while A will have to be updated by the following equation:

$$A_{(new)} = \left[\sum_{k=1}^m \mathcal{T}_{k(new)} A_{(old)} R_{k(ol)}^T + \mathcal{T}_{k(new)}^T A_{(old)} R_{k(ol)} \right] \\ \left[\sum_{k=1}^m R_{k(ol)} A_{(old)}^T A_{(old)} R_{k(ol)}^T \right. \\ \left. + R_{k(ol)}^T A_{(old)}^T A_{(old)} R_{k(ol)} + \lambda I \right]^{-1} \quad (7)$$

This procedure is fast requiring just a few simple matrix operations and hence it can be used for the classification of unseen entities.

Complexity. Following the analysis in [Nickel *et al.*, 2012], the complexity of updating A and R in RESCAL is $O(pnr + nr^2)$ and $O(pnr^2 + nr^2)$ respectively, where p is the non-zero entries, n is the number of entities and r is the number of factors. Adding a k-NN classifier¹ in the update process of A , will only change the complexity of updating A to $O(pnr + nr^2 + r \log(n))$.

4 Experiments

Dataset. We conduct experiments on a real-world dataset. [Fakhraei *et al.*, 2015] was collected from Tagged.com and deals with spammer detection. It consists of 5,617,345 users, who interact with each other in seven (7) different ways. The dataset presents high sparsity with the average density being lower than 0.002%. Due to memory issues, we perform random sampling in order to decrease the size of the dataset. We sample N users labeled as spammers and N users labeled as non-spammers, and for these users we find all interactions among them that are reported in the dataset. If there are no interactions for a user we discard them. Finally, we come up with a $M \times M \times 7$ tensor, where M is the number of users left after we remove the users without interactions. For the experiments we set $M=6.733$.

4.1 Analysis

As a performance measure we use the precision-recall curves. Precision is computed as the number of influential users identified at that point relative to the number of points examined. Recall is computed as the number of influentials identified at that point relative to the total number of influentials.

Number of factors. Before comparing our method against other approaches we want to measure the effect of the number of factors on its performance. We experiment with various $r = 2, 3, 5, 10, 15, 20, 30$ (see Figure 4). A great increase in precision-recall curve from $r = 3$ to $r = 5$. As the number of

¹with a kd-tree implementation

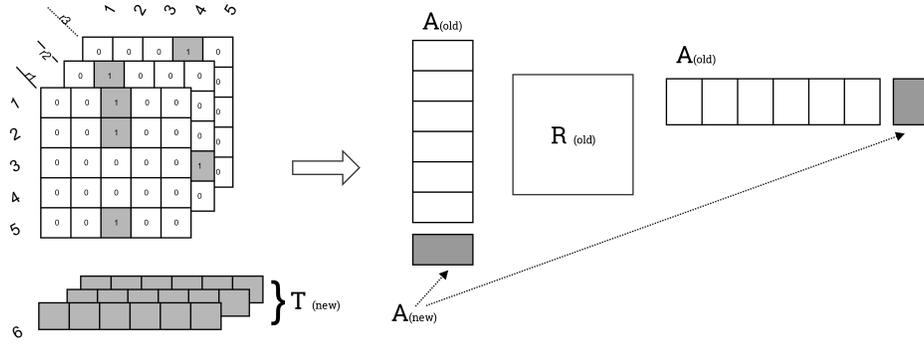


Figure 3: The *folding-in* process of CLASS-RESICAL. Note that R remains unchanged

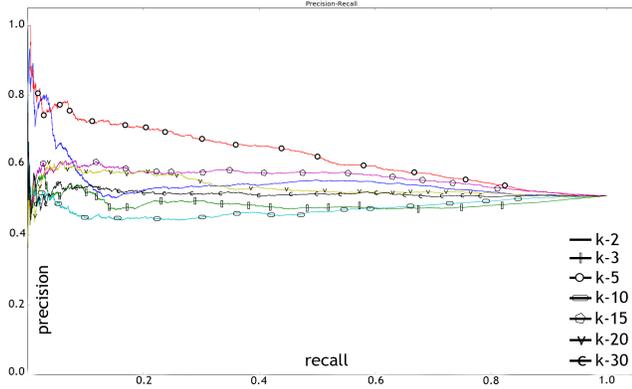


Figure 4: Behavior of CLASS-RESICAL while using different number of factors. For our experiments we use $r = 5$.

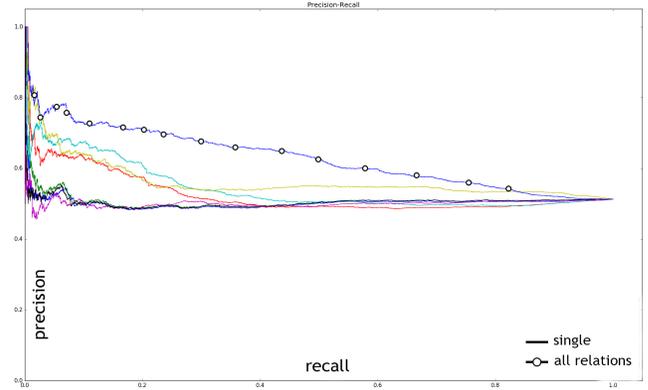


Figure 5: Performance of CLASS-RESICAL for various values of the number of relations.

factors increases, the computation time of the factorizations also increases. As a result for the rest of the experiments we use $r = 5$ which seems to be a good compromise between performance and computational cost.

Number of relations. We also examine the role of the number of relations in the performance of the factorization. In Figure 5 we show that the performance improves when using all available information (all seven relations).

4.2 Identifying Key Nodes in Social Networks

Evaluation Methodology. In this experiment, we compare the performance of our method against other related approaches in the Tagged.com dataset. As CLASS-RESICAL is a semi-supervised method, we evaluate it over various training sizes $s \in [10, 40]$ percent, averaging the results over 10 random runs. In each run, we sample a random number of instances from each label to use as the training set, and consider the remaining as the test set. With the use of the training set we obtain the semi-supervised tensor decomposition. Then, for each instance in the test set we compute its latent representation using the *folding-in* technique presented in Section 3. Thereupon, we compute a ranking with respect to the ranking scores of users, $\mathbf{a}_i^{\text{score}}$, as described in Section 3. The results reported correspond to the performance of the methods on the test set only.

We employ a widely used metric to evaluate the performance of different methods, namely the area under the precision-recall curve (AUPR).

Methods. To compare our model, we choose methods that (a) can be applied to multi-relational data and (b) deal with node identification :

- **SVM-degree:** This is a baseline SVM model built by using as features the degree of a node in each relation.
- **RESICAL:** This is the original RESICAL factorization.
- **MultiRank:** MultiRank is a co-ranking method based on random walks in multi-relational data. It constructs two transition probability tensors, one for objects and the other for relations. Based on these tensorial representation of multi-relational data it performs a pagerank like random walk and computes a separate multirank score for objects and another for relations.
- **HARrank:** This is a multi-relational analog of the HITS algorithm. It constructs three different transition probability tensors, one for hubs, one for authorities and the last one for relations. Similar to MultiRank, it performs a random walk in the created tensors and computes three different scores (hub, authority, relation).
- **TripleRank:** This is a CANDECOMP/PARAFAC based

Table 1: The AUPR curve scores for all methods.

AUPR	Tagged.com
SVM-degree	0.51
RESCAL	0.46
MultiRank	0.52
HARrank	0.49
TripleRank	0.51
CLASS-RESCAL	0.64

tensor factorization model for ranking entities in linked data. It uses the factor scores produced from the tensor decomposition to rank entities.

In the experiments, we set the necessary parameters for each method. For our method, we set the number of factors $r = 5$ and hyperparameters $\lambda_g = 0.1$ and $\lambda_A, \lambda_R = 0.5$, after performing grid search. We also use $r = 5$ for TripleRank. For the rest methods we use the values proposed by the authors.

Experimental Results. Table 1 shows the performance of all methods on the Tagged.com dataset. CLASS-RESCAL outperforms the other methods, while RESCAL yields the lowest performance. SVM-degree performs better than RESCAL, which indicates the importance of supervision. MultiRank seems to perform slightly better than the other methods. In conclusion, the user of class-label information in the decomposition of CLASS-RESCAL, seems to lead to a more effective ranking, confirming our original intuition.

Running Times. For the runtime comparison we performed various tests experimenting with the size of the tensors. An empirical runtime comparison for predicting a ranked list of users is depicted in Figure 6. As we can see, the runtimes of all methods are dependent on the size of the tensor and hence on the size of the dataset. As the size increases, the runtime of all methods increases. However, the runtime of CLASS-RESCAL differentiates from the rest. The runtimes of MultiRank and HARRank are reasonable since they perform random walks on existing edges which results to lower running times when the sparsity of the dataset is high. The runtime of TripleRank is problematic even for small datasets and becomes unfeasible as the size of the tensor increases.

5 Conclusion - Discussion

In this paper, we presented an extension of the tensor factorization method RESCAL, for identifying the social position of users in a multi-relational network by employing supervision. The factorization we proposed incorporates a classification error function in the optimization process achieving class-driven modeling of the tensorial data. Our method models the tensorial data while minimizing classification error. Therefore we obtain a semi-supervised tensor factorization which assimilates class-label information. In our evaluation, we conducted experiments with ground truth data that demonstrate the usefulness of labeled data for tensor factorization and illustrate the effectiveness of our approach.

Among our immediate plans to extend this work, we are studying other tasks of Social Network Analysis, like link

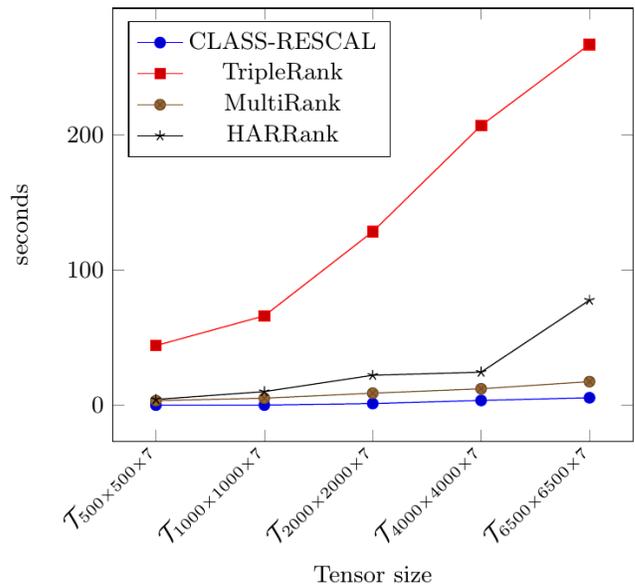


Figure 6: Running times of all methods for different tensor sizes.

prediction and recommendation. At the same time we are exploring different tensor factorization and classification methods. Additionally, we plan to evaluate our approach extensively with more datasets that contain more types of relations.

Acknowledgments

This paper is supported by the project "HOBBIT- Holistic Benchmarking of Big Linked Data", which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 688227.

References

- [Acar *et al.*, 2013] Evrim Acar, Morten Arendt Rasmussen, Francesco Savorani, Tormod Næs, and Rasmus Bro. Understanding data fusion within the framework of coupled matrix and tensor factorizations. *Chemometrics and Intelligent Laboratory Systems*, 129:53–63, 2013.
- [Bader *et al.*, 2007] Brett W Bader, Richard A Harshman, and Tamara G Kolda. Temporal analysis of semantic graphs using alsan. In *Data Mining, 2007. ICDM 2007. Seventh IEEE International Conference on*, pages 33–42. IEEE, 2007.
- [Erdos and Miettinen, 2013] Dóra Erdos and Pauli Miettinen. Discovering facts with boolean tensor tucker decomposition. In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*, pages 1569–1572. ACM, 2013.
- [Fakhraei *et al.*, 2015] Shobeir Fakhraei, James Foulds, Madhusudana Shashanka, and Lise Getoor. Collective spammer detection in evolving multi-relational social networks. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1769–1778. ACM, 2015.

- [Ferrante, 2012] Joan Ferrante. *Sociology: A global perspective*. Nelson Education, 2012.
- [Franz *et al.*, 2009] Thomas Franz, Antje Schultz, Sergej Sizov, and Steffen Staab. Triplerank: Ranking semantic web data by tensor decomposition. In *International semantic web conference*, pages 213–228. Springer, 2009.
- [Gleich *et al.*, 2015] David F Gleich, Lek-Heng Lim, and Yongyang Yu. Multilinear pagerank. *SIAM Journal on Matrix Analysis and Applications*, 36(4):1507–1541, 2015.
- [Harshman and Lundy, 1994] Richard A Harshman and Margaret E Lundy. Parafac: Parallel factor analysis. *Computational Statistics & Data Analysis*, 18(1):39–72, 1994.
- [Jingjing *et al.*, 2013] Wei Jingjing, Tang Changhong, Liao Xiangwen, and Chen Guolong. Mining social influence in microblogging via tensor factorization approach. In *Cloud Computing and Big Data (CloudCom-Asia), 2013 International Conference on*, pages 583–591. IEEE, 2013.
- [Kolda and Bader, 2009] Tamara G. Kolda and Brett W. Bader. Tensor decompositions and applications. *SIAM Rev.*, 51(3):455–500, August 2009.
- [Li and Ngom, 2010] Yifeng Li and Alioune Ngom. Non-negative matrix and tensor factorization based classification of clinical microarray gene expression data. *2010 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 438–443, 2010.
- [Li *et al.*, 2012] Xutao Li, Michael K Ng, and Yunming Ye. Har: hub, authority and relevance scores in multi-relational data for query search. In *Proceedings of the 2012 SIAM International Conference on Data Mining*, pages 141–152. SIAM, 2012.
- [Liu *et al.*, 2010] Lu Liu, Jie Tang, Jiawei Han, Meng Jiang, and Shiqiang Yang. Mining topic-level influence in heterogeneous networks. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management, CIKM '10*, pages 199–208, New York, NY, USA, 2010. ACM.
- [McPherson *et al.*, 2001] Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27(1):415–444, 2001.
- [Narita *et al.*, 2011] Atsuhiko Narita, Kohei Hayashi, Ryota Tomioka, and Hisashi Kashima. Tensor factorization using auxiliary information. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 501–516. Springer, 2011.
- [Ng *et al.*, 2011] Michael Kwok-Po Ng, Xutao Li, and Yunming Ye. Multirank: co-ranking for objects and relations in multi-relational data. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1217–1225. ACM, 2011.
- [Nickel *et al.*, 2011] Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. A three-way model for collective learning on multi-relational data. In *Proceedings of the 28th international conference on machine learning (ICML-11)*, pages 809–816, 2011.
- [Nickel *et al.*, 2012] Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. Factorizing yago: Scalable machine learning for linked data. In *Proceedings of the 21st International Conference on World Wide Web, WWW '12*, pages 271–280, New York, NY, USA, 2012. ACM.
- [Nimishakavi *et al.*, 2016] Madhav Nimishakavi, Uday Singh Saini, and Partha Talukdar. Relation schema induction using tensor factorization with side information. *arXiv preprint arXiv:1605.04227*, 2016.
- [Rendle *et al.*, 2009] Steffen Rendle, Leandro Balby Marinho, Alexandros Nanopoulos, and Lars Schmidt-Thieme. Learning optimal ranking with tensor factorization for tag recommendation. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 727–736. ACM, 2009.
- [Tian Dai *et al.*, 2012] Bing Tian Dai, Freddy Chong Tat Chua, and Ee-Peng Lim. Structural analysis in multi-relational social networks. In *Proceedings of the 2012 SIAM International Conference on Data Mining*, pages 451–462. SIAM, 2012.
- [Tucker, 1966] Ledyard R Tucker. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31(3):279–311, 1966.
- [Wu *et al.*, 2013] Fei Wu, Xu Tan, Yi Yang, Dacheng Tao, Siliang Tang, and Yueting Zhuang. Supervised nonnegative tensor factorization with maximum-margin constraint. In *AAAI*, 2013.
- [Zhaoyun *et al.*, 2013] Ding Zhaoyun, Jia Yan, Zhou Bin, and Han Yi. Mining topical influencers based on the multi-relational network in micro-blogging sites. *China Communications*, 10(1):93–104, 2013.