

# Link Prediction via Generalized Coupled Tensor Factorisation

Beyza Ermis<sup>1</sup>, Evrim Acar<sup>2</sup>, and A. Taylan Cemgil<sup>1</sup>

<sup>1</sup> Boğaziçi University  
34342, Bebek, Istanbul, Turkey,  
beyza.ermis@boun.edu.tr, taylan.cemgil@boun.edu.tr,  
<sup>2</sup> University of Copenhagen  
DK-1958 Frederiksberg C, Denmark  
evrim@life.ku.dk

**Abstract.** This study deals with the *missing link prediction* problem: the problem of predicting the existence of missing connections between entities of interest. We address link prediction using coupled analysis of relational datasets represented as heterogeneous data, i.e., datasets in the form of matrices and higher-order tensors. We propose to use an approach based on probabilistic interpretation of tensor factorisation models, i.e., Generalised Coupled Tensor Factorisation, which can simultaneously fit *a large class of tensor models* to higher-order tensors/matrices with common latent factors using *different loss functions*. Numerical experiments demonstrate that joint analysis of data from multiple sources via coupled factorisation improves the link prediction performance and the selection of right loss function and tensor model is crucial for accurately predicting missing links.

**Keywords:** Coupled tensor factorisation, link prediction, missing data

## 1 Introduction

Recent technological advances, such as the Internet, multi-media devices or social networks provide abundance of relational data. For instance, in retail recommender systems, in addition to retail data showing *who has bought which items*, we may also have access to customers' social networks, i.e., *who is friends with whom*. In such complex problems, jointly analyzing data from multiple sources has great potential to increase our ability for capturing the underlying structure in data. Data fusion, therefore, is a viable candidate for addressing the challenging link prediction problem. Applications in many areas including recommender systems and social network analysis deal with link prediction, i.e., the problem of inferring whether there is a relation between the entities of interest. For instance, if a customer buys an item, the customer and the item can be considered to be linked. The task of recommending other items the customer may be interested in can be cast as a missing link prediction problem. However, the results are likely to be poor if the prediction is done in isolation on a single view of data. Such

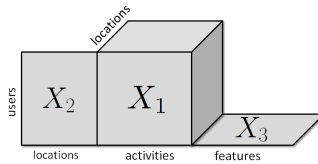


Fig. 1: A third-order tensor coupled with two matrices in two different modes.

datasets, whilst large in dimension, are already very sparse [1] and potentially represent only a very incomplete picture of the reality [2]. Therefore, relational data from other sources is often incorporated into link prediction models [3].

Matrix factorisations have proved to be very useful in recommender systems [4]. An effective way of including side information via additional relational data in a link prediction model is to represent different relations as a collection of matrices. Subsequently, this collection of matrices are jointly analyzed using collective matrix factorisation [5,6]. In many applications, however, matrices are not sufficient for a faithful representation of multiple attributes, and higher-order tensor and matrix factorisation methods are needed. An influential study in this direction is by Banerjee et al. [7], where a general clustering method for joint analysis of heterogeneous data has been studied. The goal here is clustering entities based on multiple relations, where each relation is represented as a matrix (e.g., movies by review words matrix showing movie reviews) or a higher-order tensor (e.g., movies by viewers by actors tensor showing viewers' ratings).

In this paper, we address link prediction problem using coupled analysis of datasets in the form of matrices and higher-order tensors. As an example application, we study a real-world GPS (Global Positioning System) dataset [8] for location-activity recommendation such that given an incomplete dataset showing which users perform which activities at various locations, we would like to fill in the missing links between (user, activity, location) triplets ( $X_1$ ). We also make use of additional sources of information showing the locations visited by users based on GPS trajectories ( $X_2$ ) and the features of locations in terms of number of different points of interest at each location ( $X_3$ ) (Figure 1).

Various algorithms have been proposed in the literature for coupled analysis of heterogeneous data. Lin et al. [9] addresses the community extraction problem on multi-relational data using a coupled factorisation approach modeling higher-order tensors using a specific tensor model, i.e., CANDECOMP/PARAFAC (CP) [10,11], and a Kullback-Leibler (KL) divergence-based cost function. Also, a recent study by Narita et al. [12] has addressed the tensor completion problem using additional data using a Euclidean distance-based loss function. Unlike previous studies, we use an approach, i.e., Generalized Coupled Tensor Factorisations (GCTF) [13], which enables us to investigate alternative tensor models and cost functions for coupled analysis of heterogeneous data. The main contributions of this paper can be summarized as follows:

- We consider different tensor models, i.e., CP and Tucker [14], and loss functions, i.e., KL-divergence and Euclidean distance, for joint analysis of heterogeneous data for link prediction using the GCTF framework.

- Using a real GPS data set, we demonstrate that coupled tensor factorisations outperform low-rank approximations of a single tensor in terms of missing link prediction and the selection of the tensor model as well as the loss function is significant in terms of link prediction performance.
- We also demonstrate that it is possible to address the cold-start problem in link prediction using the proposed coupled models.

The rest of the paper is organized as follows. In §2, we survey the related work on link prediction as well as joint factorisation of data. §3 introduces our algorithmic framework, i.e., GCTF, while §4 discusses its adaptation for the link prediction problem. Experimental results on a real dataset are presented in §5. Finally, we conclude with future work in §6.

## 2 Related Work

In order to deal with the challenging task of link prediction, many studies have proposed to exploit multi-relational nature of the data and showed improved link prediction performance by incorporating related sources of information in their modeling framework. For instance, Taskar et al [15] use relational Markov networks that model links between entities as well as their attributes. Popescul and Ungan [16] extract relational features to learn the existence of links (see [3] for a comprehensive list of similar studies).

For analysis of multi-relational data, Singh and Gordon [6] as well as Long et al. [5] introduce collective matrix factorisation. Matrix factorisation-based techniques have proved useful in terms of capturing the underlying patterns in data, e.g., in recommender systems [4], and joint analysis of matrices has been widely applied in numerous disciplines including signal processing [17] and bioinformatics [18]. Recent studies extend collective matrix factorisation to coupled analysis of multi-relational data in the form of matrices and higher-order tensors [19,7] since in many disciplines, relations can be defined among more than two entities, e.g., when a user engages in an activity at a certain location, a relation can be defined over user, activity and location entities. Banerjee et al. [7] introduced a multi-way clustering approach for relational and multi-relational data where coupled analysis of heterogeneous data was studied using minimum Bregman information. Lin et al. [9] also discussed coupled matrix and tensor factorisations using KL-divergence modeling higher-order tensors by fitting a CP model. While these studies use alternating algorithms, Acar et al. [20] proposed an all-at-once optimization approach for coupled analysis.

Missing link prediction is also closely related to matrix and tensor completion studies. By using a low-rank structure of a data set, it is possible to recover missing entries for matrices [21] and higher-order tensors [22].

Note that we focus on missing link prediction in this paper and do not address the temporal link prediction problem, where snapshots of the set of links up to time  $t$  are given and the goal is to predict the links at time  $t + 1$ . Tensor factorisations have previously been used for temporal link prediction [23] but using only a single source of information.

### 3 Methodology

In this study, we discuss Generalized Coupled Tensor Factorisation framework [13] for coupled factorisation of several tensors and matrices to fill in the missing links in observed data. A generalized tensor factorisation problem is specified by an observed tensor  $X$  (with possibly missing entries) and a collection of latent tensors to be estimated,  $Z_{1:|\alpha|} = \{Z_\alpha\}$  for  $\alpha = 1 \dots |\alpha|$ .

GCTF framework is a generalisation of the Probabilistic Latent Tensor Factorisation (PLTF) [24] to coupled factorisation. In this framework, the goal is to compute an approximate factorisation of  $X$  in terms of a product of individual factors  $Z_\alpha$ . Here, we define  $V$  as the set of all indices in a model,  $V_0$  as the set of visible indices,  $V_\alpha$  as the set of indices in  $Z_\alpha$ , and  $\bar{V}_\alpha = V - V_\alpha$  as the set of all indices not in  $Z_\alpha$ . We use small letters as  $v_\alpha$  to refer to a particular setting of indices in  $V_\alpha$ .

PLTF tries to solve the following approximation problem

$$X(v_0) \approx \hat{X}(v_0) = \sum_{\bar{v}_0} \prod_{\alpha} Z_\alpha(v_\alpha), \quad (1)$$

Since the product  $\prod_{\alpha} Z_\alpha(v_\alpha)$  is collapsed over a set of indices, the factorisation is latent. The approximation problem is cast as an optimisation problem minimizing the divergence  $d(X, \hat{X})$ , where  $d$  is a divergence (a quasi-squared-distance) between the observed data  $X$  and model prediction  $\hat{X}$ . In applications,  $d$  is typically Euclidean (EUC), Kullback-Leibler (KL) or Itakura-Saito (IS) [13].

In this paper, we use non-negative variants of the two most widely-used low-rank tensor factorisation models, i.e., Tucker model, and the more restricted CP model, as baseline methods in §5. These models can be defined in the PLTF notation as follows. Given a three-way tensor  $X$ , its CP model is defined as:

$$X(i, j, k) \approx \hat{X}(i, j, k) = \sum_r Z_1(i, r) Z_2(j, r) Z_3(k, r) \quad (2)$$

where index sets  $V = \{i, j, k, r\}$ ,  $V_0 = \{i, j, k\}$ ,  $V_1 = \{i, r\}$ ,  $V_2 = \{j, r\}$  and  $V_3 = \{k, r\}$ . A Tucker model of  $X$  is defined in the PLTF notation as follows:

$$X(i, j, k) \approx \hat{X}(i, j, k) = \sum_{p, q, r} Z_1(i, p) Z_2(j, q) Z_3(k, r) Z_4(p, q, r) \quad (3)$$

where index sets  $V = \{i, j, k, p, q, r\}$ ,  $V_0 = \{i, j, k\}$ ,  $V_1 = \{i, p\}$ ,  $V_2 = \{j, q\}$ ,  $V_3 = \{k, r\}$  and  $V_4 = \{p, q, r\}$ .

The update equation for non-negative generalized tensor factorisation can be used for both (2) and (3) and is expressed as:

$$Z_\alpha \leftarrow Z_\alpha \circ \frac{\Delta_\alpha(M \circ \hat{X}^{-p} \circ X)}{\Delta_\alpha(M \circ \hat{X}^{1-p})} \quad s.t. \quad Z_\alpha(v_\alpha) > 0. \quad (4)$$

where  $\circ$  is the Hadamard product (element-wise product),  $M$  is a 0 – 1 mask array with  $M(v_0) = 1$  ( $M(v_0) = 0$ ) if  $X(v_0)$  is observed (missing). Here  $p$  determines the cost function, i.e.,  $p = \{0, 1, 2\}$  correspond to the  $\beta$ -divergence

[25] that unifies EUC, KL, and IS cost functions, respectively. In this iteration, we define the tensor valued function  $\Delta_\alpha(A)$  as:

$$\Delta_\alpha(A) = \sum_{\bar{v}_\alpha} A(v_0) \prod_{\alpha' \neq \alpha} Z_{\alpha'}(v_{\alpha'}) \quad (5)$$

$\Delta_\alpha(A)$  is an object, the same size of  $Z_\alpha$ , obtained simply by multiplying all factors other than the one being updated with an object of the order of the data. Hence the key observation is that the  $\Delta_\alpha$  function is just computing a tensor product and collapses this product over indices not appearing in  $Z_\alpha$ , which is algebraically equivalent to computing a marginal sum.

As an example, for KL cost, we rewrite (4) more compactly as:

$$Z_\alpha \leftarrow Z_\alpha \circ \Delta_\alpha(M \circ X / \hat{X}) / \Delta_\alpha(M) \quad (6)$$

This update rule can be used iteratively for all non-negative  $Z_\alpha$  and converges to a local minimum provided we start from some non-negative initial values. For updating  $Z_\alpha$ , we need to compute the  $\Delta$  function twice for arguments  $A = M_\nu \circ \hat{X}_\nu^{-p} \circ X_\nu$  and  $A = M_\nu \circ \hat{X}_\nu^{1-p}$ .

### 3.1 Generalized Coupled Tensor Factorisation

The Generalised Coupled Tensor Factorisation model takes the PLTF model one step further where, we have multiple observed tensors  $X_\nu$  that are factorised simultaneously:

$$X_\nu(v_{0,\nu}) \approx \hat{X}_\nu(v_{0,\nu}) = \sum_{\bar{v}_{0,\nu}} \prod_{\alpha} Z_\alpha(v_\alpha)^{R^{\nu,\alpha}} \quad (7)$$

where  $\nu = 1, \dots, |\nu|$  and  $R$  is a *coupling matrix* defined as follows:

$$R^{\nu,\alpha} = \begin{cases} 1 & \text{if } X_\nu \text{ and } Z_\alpha \text{ connected} \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

Note that, distinct from PLTF model, there are multiple visible index sets  $(V_{0,\nu})$  in the GCTF model, each specifying the attributes of the observed tensor  $X_\nu$ .

The inference, i.e., estimation of the shared latent factors  $Z_\alpha$ , can be achieved via iterative optimisation (see [13]). For non-negative data and factors, one can obtain the following compact fixed point equation where each  $Z_\alpha$  is updated in an alternating fashion fixing the other factors  $Z_{\alpha'}$ , for  $\alpha' \neq \alpha$ :

$$Z_\alpha \leftarrow Z_\alpha \circ \frac{\sum_{\nu} R^{\nu,\alpha} \Delta_{\alpha,\nu}(M_\nu \circ \hat{X}_\nu^{-p} \circ X_\nu)}{\sum_{\nu} R^{\nu,\alpha} \Delta_{\alpha,\nu}(M_\nu \circ \hat{X}_\nu^{1-p})} \quad (9)$$

where  $M_\nu$  is a 0 – 1 mask array with  $M_\nu(v_{0,\nu}) = 1$  ( $M_\nu(v_{0,\nu}) = 0$ ) if  $X_\nu(v_{0,\nu})$  is observed (missing). Here  $p$ , as in (4), determines the cost function, i.e.,  $p = \{0, 1\}$  corresponds to EUC and KL cost functions, respectively (see Table 1). In this iteration, the key quantity is the  $\Delta_{\alpha,\nu}$  function defined as follows:

$$\Delta_{\alpha,\nu}(A) = \left[ \sum_{v_{0,\nu} \cap \bar{v}_\alpha} A(v_{0,\nu}) \sum_{\bar{v}_{0,\nu} \cap \bar{v}_\alpha} \prod_{\alpha' \neq \alpha} Z_{\alpha'}(v_{\alpha'})^{R^{\nu,\alpha'}} \right] \quad (10)$$

Table 1: Update rules for different  $p$  values

$p$	Cost Function	Multiplicative Update Rule
0	Euclidean	$Z_\alpha \leftarrow Z_\alpha \circ \frac{\sum_\nu R^{\nu,\alpha} \Delta_{\alpha,\nu}(M_\nu \circ \hat{X}_\nu)}{\sum_\nu R^{\nu,\alpha} \Delta_{\alpha,\nu}(M_\nu \circ \hat{X}_\nu)}$
1	Kullback-Leibler	$Z_\alpha \leftarrow Z_\alpha \circ \frac{\sum_\nu R^{\nu,\alpha} \Delta_{\alpha,\nu}(M_\nu \circ \hat{X}_\nu^{-1} \circ X_\nu)}{\sum_\nu R^{\nu,\alpha} \Delta_{\alpha,\nu}(M_\nu)}$

Assuming that all datasets have equal number of dimensions, i.e., a tensor is an  $N \times N \times N$  array while the coupled matrix is of size  $N \times N$ , then the leading term in the computational complexity of the coupled model will be due to the updates for the tensor model. For an  $F$ -component CP model, for instance, that would be  $O(N^3 F)$ . The updates can be implemented by taking into account the sparsity pattern of the data.

## 4 Link Prediction with Coupled Tensor Factorisation

In this section, by using the GCTF framework, we propose a solution for link prediction task with different coupled models and loss functions. We have a three-way observation tensor  $X_1$  with elements 0 and 1, where 0 denotes a known absent link and 1 denotes a known present link, and two auxiliary matrices  $X_2$  and  $X_3$  that provide side information. Our aim is to restore the missing links in  $X_1$ . This is a difficult link prediction problem since  $X_1$  contains less than 1% of all possible links or an entire slice of  $X_1$  may be missing. Using low-rank factorisation of a tensor to estimate missing entries will be ineffective, in particular, in the case of structured missing data such as missing slices. In terms of coupled models, we are not restricted to a specific model topology, i.e., since we use the GCTF framework, we can design application-specific models. The choice of a particular factorisation model is strongly guided by the application; therefore, we first give a brief description of the data set.

UCLAF dataset<sup>3</sup> [8] is extracted from the GPS data that include information of three types of entities: user, location and activity. The relations between user-location-activity triplets are used to construct a three-way tensor  $X_1$ . In tensor  $X_1$ , an entry  $X_1(i, j, k)$  indicates the frequency of a user  $i$  visiting location  $j$  and doing activity  $k$  there; otherwise, it is 0. Since we address the link prediction problem in this study, we define the user-location-activity tensor  $X_1$  as:

$$X_1(i, j, k) = \begin{cases} 1 & \text{if user } i \text{ visits location } j \text{ and performs activity } k \text{ there} \\ 0 & \text{otherwise} \end{cases}$$

The dataset has been constructed by clustering raw GPS points into 168 meaningful locations and manually parsing the user comments attached to the GPS data into activity annotations for the 168 locations. Consequently, the data consists of 164 users, 168 locations and 5 different types of activities. (See [8] for details).

<sup>3</sup> <http://www.cse.ust.hk/~vincentz/aaai10.uclaf.data.mat>

Additionally, the collected data includes side information: the location features from the POI (points of interest) database as well as the user-location preferences from the GPS trajectory data, represented by matrix  $X_2$  and  $X_3$  respectively. In our model, the user-location preferences matrix of size  $I \times J$  has entries  $X_2(i, m)$ , where  $I$  is the number of users and  $J$  is the number of locations and we use index  $m$  as the location index instead of  $j$ . The rationale behind this choice is to relax the model as the entries in  $X_1$  and  $X_2$  are measuring distinct quantities:  $X_2(i, m)$  represents the frequency of a user  $i$  visiting location  $m$  and stayed there over a time threshold while  $X_1$  only indicates an activity by a specific user  $i$  at location  $j$ . The relation between the location entries  $j$  and  $m$  in  $X_1$  and  $X_2$  are coupled via a common factor over the users. Finally, we represent the location-feature values with matrix  $X_3$  of size  $J \times N$ , where  $J$  is the number of locations, that has the same location type in  $X_1$ , and  $N$  is the number of features. In particular, an entry  $X_3(j, n)$  represents the number of different POIs at location  $j$ . Using the location features, we could gain information about location similarities based on their feature values.

In this data set, 18 users have no location and activity information. Therefore, we have used the remaining 146 users. In order to decrease the effect of outliers, location-feature matrix is preprocessed as follows:  $X_3(j, n) = 1 + \log(X_3(j, n))$  if  $X_3(j, n) > 0$ ; otherwise,  $X_3(j, n) = 0$ . In our experiments, number of users is  $I = 146$ , number of locations  $J = 168$ , number of activities  $K = 5$  and number of location features  $N = 14$ .

We form two coupled models to fill in the missing links in tensor  $X_1$ . For both models, we use KL divergence and Euclidean as the cost functions in our non-negative decomposition problems. In the first model, we applied the coupled approach to a CP-style tensor model by analysing the tensor  $X_1$  jointly with the additional matrices  $X_2$  and  $X_3$ . This gives us the following model:

$$\hat{X}_1(i, j, k) = \sum_r A(i, r)B(j, r)C(k, r) \quad (11)$$

$$\hat{X}_2(i, m) = \sum_r A(i, r)D(m, r) \quad (12)$$

$$\hat{X}_3(j, n) = \sum_r B(j, r)E(n, r) \quad (13)$$

Here, we have three observed tensors, that share common factors; therefore, we have a coupled tensor factorisation problem. The coupling matrix  $R$  with  $|\alpha| = 5$ ,  $|\nu| = 3$  for this model is defined as follows:

$$R = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \end{bmatrix} \quad \begin{aligned} \hat{X}_1 &= \sum A^1 B^1 C^1 D^0 E^0 \\ \text{with } \hat{X}_2 &= \sum A^1 B^0 C^0 D^1 E^0 \\ \hat{X}_3 &= \sum A^0 B^1 C^0 D^0 E^1 \end{aligned} \quad (14)$$

Note that,  $X_1$  and  $X_2$  share the common factor matrix  $A$  with entries  $A(i, r)$ ; we can interpret each row of  $A(i, \cdot)$  as user  $i$ 's latent position in a  $|r|$  dimensional 'preferences' space. The factor matrix  $B$  with entries  $B(j, r)$  represents the latent position of the location  $j$  in the same preferences space. The user  $i$  at location  $j$  tends to perform activity  $k$  where the weight  $A(i, r)B(j, r)$  is large for at least one

$r$ , i.e., there is a match between the users preference and what the location ‘has to offer’. The location specific factor  $B$  is also influenced by the location-feature matrix  $X_3$ .

Following the same line of thought, we apply the coupled approach using a Tucker decomposition to form our second model, which is as follows:

$$\hat{X}_1(i, j, k) = \sum_{p, q, r} A(i, p)B(j, q)C(k, r)D(p, q, r) \quad (15)$$

$$\hat{X}_2(i, m) = \sum_p A(i, p)E(m, p) \quad (16)$$

$$\hat{X}_3(j, n) = \sum_r B(j, q)F(n, q) \quad (17)$$

In this model, once again, the factor  $A$  is shared by  $X_1$  and  $X_2$ , while the factor  $B$  is shared by  $X_1$  and  $X_3$ . In contrast to the coupled CP model in (11), this model assumes that user  $i$  at location  $j$  tends to perform activity  $k$  with the weight  $\sum_{p, q} A(i, p)B(j, q)D(p, q, r)$ . Here, a latent preference space interpretation is less intuitive but the model has more freedom to represent the dependence.

## 5 Experimental Results

In this section, we assess the performance of the coupled models proposed in the previous section in terms of missing link prediction. First, we demonstrate that coupled tensor factorisations outperform low-rank approximations of a single tensor in terms of missing link prediction. Then we compare different tensor models and loss functions and show that selection of the tensor model and loss function is significant in terms of link prediction performance, especially when the fraction of unobserved elements is high. Furthermore, we study the case with completely missing slices, which corresponds to the cold-start problem in our link prediction setting and demonstrate that it is still possible to predict missing links using the proposed coupled models.

### 5.1 Experimental Setting

We design experiments to evaluate the performance of our models in terms of link prediction. By setting different amounts of data to missing in user-location-activity tensor  $X_1$ , we compare the following models using both KL-divergence and the Euclidean as cost functions:

- *Low-rank approximations of a single tensor:* (i) CP and (ii) Tucker factorisation of user-location-activity tensor  $X_1$ ,
- *Coupled tensor factorisations:* (i) CP factorisation of  $X_1$  coupled with factorisation of user-location matrix  $X_2$  and location-feature matrix  $X_3$  (ii) Tucker factorisation of  $X_1$  coupled with factorisation of  $X_2$  and  $X_3$ .

We use two missing data patterns: (i) randomly missing entries, (ii) randomly missing slices. In all experiments, number of components, i.e., number of columns in each factor matrix,  $Z_i$ , is set to 2. To measure the link prediction performance, we use AUC (Area Under the Receiver Operating Characteristic Curve).

## 5.2 Results

In order to demonstrate the power of coupled analysis, we compared the link prediction performance of standard CP and Tucker models with coupled ones using EUC and KL cost functions at different amounts, i.e.,  $\{40, 60, 80, 90, 95\}$ , of randomly unobserved elements. For all cases, coupled models outperform the standard models clearly. Figure 2 shows the comparison of CP and coupled CP models with different cost functions when 80% of the data is missing. As we can see, coupled models using additional information perform better than the standard models; in particular, when the percentage of missing data is high. When the fraction of missing data was more than 80%, the standard models could not find a solution.

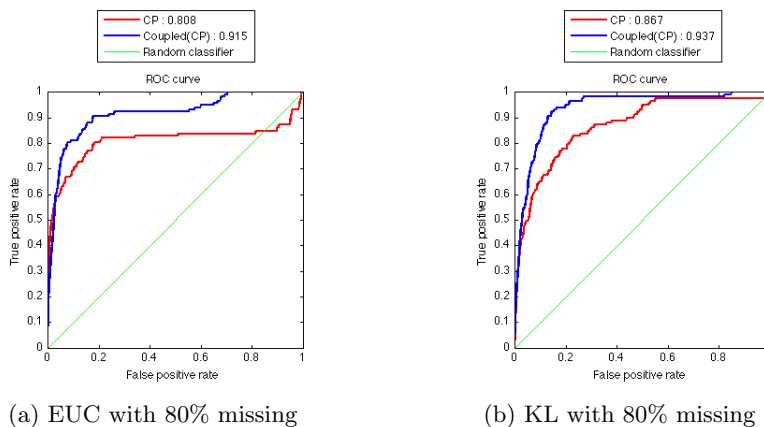


Fig. 2: Comparison of CP and Coupled(CP) models

In order to demonstrate the effect of the cost function modeling the data, we have also carried out experiments on both coupled CP and Tucker models at different missing data fractions. For all cases, the KL cost function seems to perform better than EUC, especially when the fraction of missing entries is high. Figure 3 illustrates the performance of Euclidean distance and Kullback-Leibler divergence for both coupled CP and Tucker models when 90% of the data is unobserved.

Finally, Figure 4 shows the comparison of coupled CP and Tucker models in order to illustrate the tensor model modeling the data best. We can see that Tucker model outperforms the CP model; because Tucker model is more flexible due to the full core tensor which is helpful for us to explore the structural information embedded in the data.

**Cold Start Problem:** We also study the missing slice problem, which is particularly important in link prediction because we may often have new users starting to use an application, e.g., a location-activity recommender system. Since they are new users, they will have no entry in  $X_1$ , i.e., a completely missing slice. It

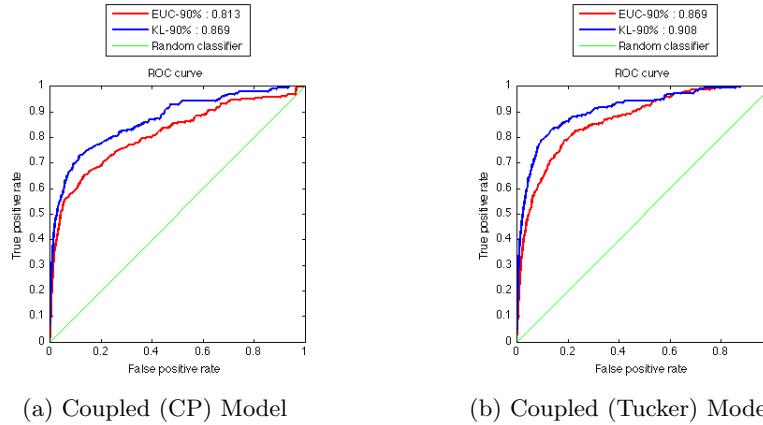


Fig. 3: Comparison of EUC distance and KL divergence with 90% missing data

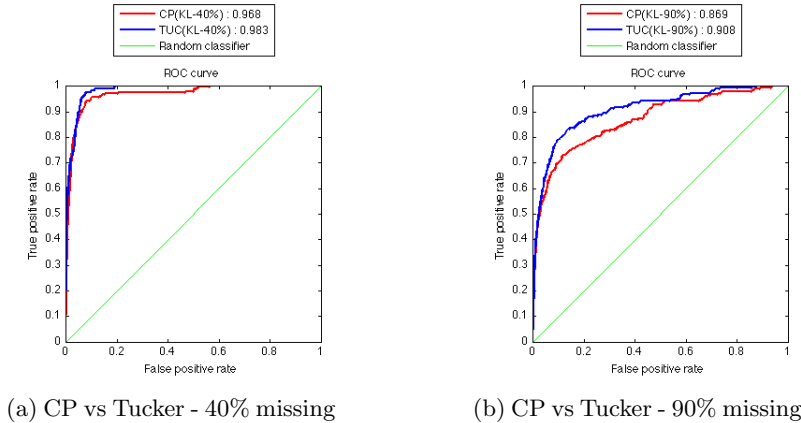


Fig. 4: Comparison of Coupled CP and Tucker models with KL

is not possible to reconstruct a missing slice of a tensor using its low-rank approximation. A similar argument is valid in the case of matrices for completely missing rows/columns [21]. In such cases, additional sources of information will be useful to make recommendations to new users. We observe that our coupled models could predict the links when there is no information about a user in tensor  $X_1$ , by utilizing the additional sources of information. We test this case by setting slices to missing randomly in  $X_1$ . Figure 5 demonstrates the performance of coupled models with KL divergence when 10 users' data and 50 users' data are missing. Note that Tucker is superior to CP as the amount of missing data increases.

## 6 Conclusions

In this study, we have studied link prediction problem using coupled analysis of relational data represented as datasets in the form of matrices and higher-order

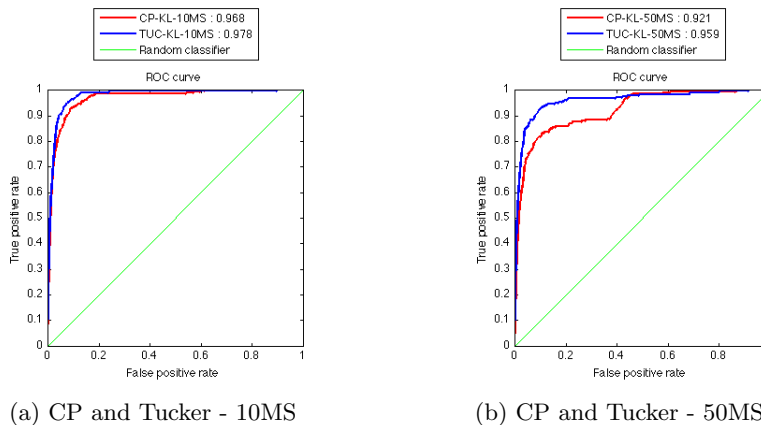


Fig. 5: Link prediction result with missing slices and KL cost

tensors. The problem is formulated as simultaneous factorisation of higher-order tensors/matrices extracting common latent factors from the shared modes. While most existing studies on coupled analysis have been developed to fit a specific type of a tensor model using a particular loss function, we have used Generalized Coupled Tensor Factorisation framework, which enables us to develop coupled models for joint analysis of multiple data sets using various tensor models and cost functions. In our coupled analysis for link prediction, we have studied both KL-divergence and Euclidean distance-based cost functions as well as different tensor models. Our numerical results on a real GPS data demonstrate that selection of the tensor model and the loss function is important in terms of link prediction performance. While our experiments have been limited to a dataset, which is not large-scale, the updates used in GCTF respect the sparsity pattern in the data; therefore, the proposed approach scales to large data. We plan to extend our study in that direction and show its applicability on large-scale data.

## 7 Acknowledgments

This work is funded by the TUBITAK grant number 110E292, Bayesian matrix and tensor factorisations (BAYTEN) and Boğaziçi University research fund BAP5723. It is also funded in part by the Danish Council for Independent Research — Technology and Production Sciences and Sapere Aude Program under the projects 11-116328 and 11-120947.

## References

1. Getoor, L., Diehl, C.P.: Link mining: a survey. *ACM SIGKDD Explorations Newsletter* **7**(2) (2005) 3–12
2. Clauset, A., Moore, C., Newman, M.: Hierarchical structure and the prediction of missing links in networks. *Nature* **453** (2008)
3. Hasan, M.A., Zaki, M.J.: A survey of link prediction in social networks. In Aggarwal, C.C., ed.: *Social Network Data Analytics*. Springer US (2011) 243–275

4. Koren, Y., Bell, R., Volinsky, C.: Matrix factorization techniques for recommender systems. *Computer* **42**(8) (2009) 30–37
5. Long, B., Zhang, Z.M., Wu, X., Yu, P.S.: Spectral clustering for multi-type relational data. In: *ICML'06*. (2006) 585–592
6. Singh, A.P., Gordon, G.J.: Relational learning via collective matrix factorization. In: *KDD'08*. (2008) 650–658
7. Banerjee, A., Basu, S., Merugu, S.: Multi-way clustering on relation graphs. In: *SDM'07*. (2007) 145–156
8. Zheng, V.W., Cao, B., Zheng, Y., Xie, X., Yang, Q.: Collaborative filtering meets mobile recommendation: A user-centered approach. In: *AAAI'10*. (2010)
9. Lin, Y.R., Sun, J., Castro, P., Konuru, R., Sundaram, H., Kelliher, A.: Metafac: community discovery via relational hypergraph factorization. In: *KDD '09*. (2009) 527–536
10. Harshman, R.A.: Foundations of the PARAFAC procedure: Models and conditions for an “explanatory” multi-modal factor analysis. *UCLA working papers in phonetics* **16** (1970) 1–84
11. Carroll, J.D., Chang, J.J.: Analysis of individual differences in multidimensional scaling via an N-way generalization of “Eckart-Young” decomposition. *Psychometrika* **35** (1970) 283–319
12. Narita, A., Hayashi, K., Tomioka, R., Kashima, H.: Tensor factorization using auxiliary information. In: *ECML PKDD '11*. (2011) 501–516
13. Yilmaz, Y.K., Cemgil, A.T., Simsekli, U.: Generalised coupled tensor factorisation. In: *NIPS'11*. (2011)
14. Tucker, L.R.: Some mathematical notes on three-mode factor analysis. *Psychometrika* **31** (1966) 279–311
15. Taskar, B., Wong, M.F., Abbeel, P., Koller, D.: Link prediction in relational data. In: *NIPS03*. (2003)
16. Popescul, A., Ungar, L.H.: Statistical relational learning for link prediction. In: *IJCAI'03*. (2003)
17. Yoo, J., Kim, M., Kang, K., Choi, S.: Nonnegative matrix partial co-factorization for drum source separation. In: *ICASSP'10*. (2010) 1942–1945
18. Alter, O., Brown, P.O., Botstein, D.: Generalized singular value decomposition for comparative analysis of genome-scale expression data sets of two different organisms. *PNAS* **100** (2003) 3351–3356
19. Smilde, A., Westerhuis, J.A., Boque, R.: Multiway multiblock component and covariates regression models. *Journal of Chemometrics* **14** (2000) 301–331
20. Acar, E., Kolda, T.G., Dunlavy, D.M.: All-at-once optimization for coupled matrix and tensor factorizations. In: *KDD'11 Workshop Proceedings*. (2011)
21. Candès, E.J., Plan, Y.: Matrix completion with noise. *Proceedings of the IEEE* **98** (2010) 925–936
22. Acar, E., Dunlavy, D., Kolda, T., Morup, M.: Scalable tensor factorizations for incomplete data. *Chemometr. Intell. Lab* **106** (2011) 41–56
23. Dunlavy, D.M., Kolda, T.G., Acar, E.: Temporal link prediction using matrix and tensor factorizations. *ACM TKDD* **5**(2) (2011) Article 10
24. Yilmaz, Y.K., Cemgil, A.T.: Probabilistic latent tensor factorization. In: *LVA/ICA*. (2010) 346–353
25. Cichoki, A., Zdunek, R., Phan, A., Amari, S.: *Nonnegative Matrix and Tensor Factorization*. Wiley (2009)