2015 IEEE International Conference on Big Data (Big Data)

Combining Activity-evaluation Information with NMF for Trust-link Prediction in Social Media

Kanji Matsutani*, Masahito Kumano*, Masahiro Kimura*, Kazumi Saito†, Kouzou Ohara‡ and Hiroshi Motoda§¶
*Department of Electronics and Informatics, Ryukoku University, Japan
†School of Administration and Informatics, University of Shizuoka, Japan
‡Department of Integrated Information Technology, Aoyama Gakuin University, Japan
§Institute of Scientific and Industrial Research, Osaka University, Japan
¶School of Computing and Information Systems, University of Tasmania, Australia

***Abstract*—Acquiring a network of trust relations among users in social media sites, e.g., item-review sites, is important for analyzing users' behavior and efficiently finding reliable information on the Web. We address the problem of predicting trust-links among users for an item-review site. Non-negative matrix factorization (NMF) methods have recently been shown useful for trust-link prediction in such a site where both link and activity information is available. Here, a user activity in an item-review site means posting a review and giving a rating for an item. In this paper, for better trust-link prediction, we propose a new NMF method that incorporates people's evaluation of users' activities as well as trust-links and users' activities themselves. We further apply it to an analysis of users' behavior. Using two real world item-review sites, we experimentally demonstrate the effectiveness of the proposed method.**

***Keywords*—social media mining; trust-link prediction; behavioral analysis; non-negative matrix factorization**

I. Introduction

Recently, Social Media such as Digg, eBay, Epinions, Facebook, etc. has become popular, and allowed us to construct a large-scale network of trust relations in an online world. A trust network among social media users is a kind of social network, and helps efficiently find reliable information on the Web. News and opinions that are posted on social media sites can rapidly and widely spread through such a social network, and can be shared by a large number of people. This way, a trust network plays an important role for people's daily life. Thus, for mining a trust network (social network), researchers have made a variety of studies including information diffusion analysis (see e.g., [1], [2] [3], [4], [5], [6], [7]), and social-link prediction (see e.g., [8], [9], [10], [11], [12], [13], [14]).

For analyzing users' behavior in a social media site, it is indispensable to acquire the network of trust relations among users in the site. However, it is generally hard to obtain a complete trust network structure at a specified time-point for analysis, since a trust network continually evolves as time passes and there arise also privacy issues. In an online world, people's behavioral patterns and preferences may largely change over a long period of time. In this paper, we aim to develop an effective solution method for the problem of predicting trust-links to be created among recent active users in the very near future. Here, we note that the developed method is also applicable to the missing-link prediction problem.

Many of social media sites offer a rich set of activities as well as the opportunity for connecting trust-links, where users can select and perform one from a given set of activities. Examples include item-review sites, where as an activity, a user can post a review and give a rating for an item in a given set of items. For the trust-link prediction problem in a social media site offering trust-links and activities, Tang et al [14] presented a *non-negative matrix factorization (NMF)* method employing both link and activity information, where they exploited a social theory called *homophily*. Here, the homophily effect suggests that similar users (i.e., users performing similar activities) have a higher likelihood to establish trust relations [10], [15]. Their method called *hTrust* is based on NMF, where NMF [16] has been shown to be useful for many applications including collaborative filtering, document clustering and link prediction. Using real data of product-review sites, they experimentally demonstrated that *hTrust* outperforms conventional methods [14].

As is the case with Tang et al [14], we also focus on item-review sites. Recently, in many of item-review sites, people can post their appreciation messages to a review of a user for an item if they think it useful. In order to improve NMF methods for trust-link prediction, we consider combining such information with NMF. In this paper, for the trust-link prediction problem, we propose a new NMF method that incorporates people's evaluations of users' activities as well as trust-links and users' activities themselves, and also apply it to analysis of users' behavior in a social media site. We extensively evaluate the proposed method using real data of two item-review sites. First, we statistically analyze the datasets, and in particular confirm that the number of appreciation messages received correlates with the number of trust-links received, suggesting that incorporating the activity-evaluation information can be a promising approach. Next, we demonstrate that the proposed method outperforms *hTrust* and its variants for solving the trust-link prediction problem. Moreover, for the datasets, we present the analysis results for users' behavior in terms of creating trust-links.

The rest of the paper is organized as follows: In Section II, we formulate the trust-link prediction problem, and revisit the NMF methods that exploit both the link and activity information for trust-link prediction. We present the proposed NMF method in Section III, and report the results for the evaluation experiments in Section IV. We conclude the paper by summarizing the main results in Section V.

978-1-4799-9926-2/15/$31.00 ©2015 IEEE

II. NMF for Trust-link Prediction

Non-negative matrix factorization (NMF) has been shown to be useful for trust-link prediction in social media by Tang et al [14]. They proposed *hTrust*, an NMF method employing both link and activity information for trust-link prediction in social media. A similar NMF method was presented by Zhu et al [17]. They proposed *Joint Link-Content Matrix Factorization (JLCMF)*, an NMF method employing both link and content information for web page classification. In this section, we first formulate our problem of trust-link prediction in social media, and recall *hTrust* and *JLCMF*.

A. Problem Formulation

For a social media site offering trust-links and activities, we address a problem of predicting trust-links to be created among recent active users in the very near future. Here, we assume that an activity of a user in this site is to post a review and give a rating for an item in a given set of items. Moreover, we assume that people can post their appreciation message to a review of a user for an item in the site if they like it, that is, information of people's evaluations of users' activities is observable in the site.

We focus on a set $\mathcal{V}$ of recent active users: A recent active user is defined as a user who created at least n_1 trust-links and received at least n_2 trust-links during the most recent Δt_0 months (referred to as the *observation period* I_0), where it is assumed that the period I_0 is not long (for example, Δt_0 is about five or six). We set

$$\mathcal{V} \;=\; \{v_1, \ldots, v_N\}.$$

We denote by

$$G \;=\; (G_{i,j})_{i,j=1,\ldots,N}$$

the $N \times N$ matrix representing trust-links to date in $\mathcal{V}$, that is, $G_{i,j} = 1$ if there exists a trust-link from user v_i to user v_j until now, and $G_{i,j} = 0$ otherwise. In this paper, we deal with the problem of predicting trust-links that are to be created in $\mathcal{V}$ during the next Δt_1 months (referred to as the *prediction period* I_1), where it is assumed that the period I_1 is short (for example, Δt_1 is about two or three). Here, note that the values of $\Delta t_0 = |I_0|$ and $\Delta t_1 = |I_1|$ are set by taking into account the fact that users' behavioral patterns do not largely change in a short period of time (e.g., several months). In particular, we consider solving this prediction problem by the NMF approach that factorizes matrix G.

There is more information that is available for our trust-link prediction problem. Let

$$\mathcal{A} \;=\; \{a_1, \ldots, a_M\}$$

be the set of all the items for which users belonging to $\mathcal{V}$ have posted reviews and given ratings until now. Let $X = (X_{i,\alpha})$ denote the $N \times M$ matrix representing users' activities to date in $\mathcal{V}$, where $X_{i,\alpha}$ is the rating score which user v_i gave for item a_α. Here, we set $X_{i,\alpha} = 0$ if user v_i did not post a review for item a_α. Let $Y = (Y_{i,\alpha})$ denote the $N \times M$ matrix representing information of people's evaluations of users' activities to date in $\mathcal{V}$, where $Y_{i,\alpha}$ is the number of appreciation messages which user v_i received for the review of item a_α. In this paper, we exploit both matrices X and Y as well as matrix G while *hTrust* and *JLCMF* only exploit matrices G and X.

B. hTrust

The NMF approach seeks high-quality feature representations using latent spaces. *hTrust* exploits only one latent space. Let K be the dimension of the latent space. hTrust introduces a non-negative $N \times K$ matrix $U = (U_{i,k})$ and a non-negative $K \times K$ matrix $H = (H_{k,\ell})$, where $U_{i,k}$ represents the strength of user v_i for latent factor k, and $H_{k,\ell}$ represents the relationship strength from latent factor k to latent factor ℓ for creating trust-links. We consider minimizing the following function $\mathcal{F}_0(U, H)$ of U and H:

$$\begin{aligned}\mathcal{F}_0(U, H) \;=\; & \left\| G - UHU^T \right\|^2 + \lambda_U \|U\|^2 + \lambda_H \|H\|^2 \\ & + \lambda_X \operatorname{Tr}\left(U^T S_X U\right),\end{aligned} \qquad (1)$$

where λ_U, λ_H and λ_X are positive constants (hyper-parameters). For any matrix B, B^T, $\|B\|$ and $\operatorname{Tr}(B)$ stand for the transposed matrix of B, the Frobenius norm of B and the trace of B, respectively. $S_X = ((S_X)_{i,j})$ is the $N \times N$ symmetric matrix defined by

$$(S_X)_{i,j} \;=\; \bar{\xi}_i \, \delta_{i,j} \;-\; \xi_{i,j}, \qquad (2)$$

where $\delta_{i,j}$ is the Kronecker delta,

$$\xi_{i,j} \;=\; \frac{\sum_{\alpha=1}^{M} X_{i,\alpha} X_{j,\alpha}}{\sqrt{\sum_{\alpha=1}^{M} X_{i,\alpha}{}^2}\,\sqrt{\sum_{\alpha=1}^{M} X_{j,\alpha}{}^2}}, \qquad (3)$$

and

$$\bar{\xi}_i \;=\; \sum_{j=1}^{N} \xi_{i,j}. \qquad (4)$$

Note that

$$\operatorname{Tr}\left(U^T S_X U\right) \;=\; \frac{1}{2} \sum_{i,j=1}^{N} \xi_{i,j} \sum_{k=1}^{K} (U_{i,k} - U_{j,k})^2, \qquad (5)$$

and the term $\operatorname{Tr}\left(U^T S_X U\right)$ is called *homophily regularization* [14]. This term is introduced for incorporating the property that users with higher similarity are more likely to establish trust relations. Here, $\xi_{i,j}$ is a similarity measure between users v_i and v_j in terms of activity. Note that other choices for $\xi_{i,j}$ are possible, and we can use various similarity measures including Jaccard coefficient and Pearson correlation coefficient.

For the problem of minimizing the function $\mathcal{F}_0(U, H)$ under the condition of non-negative matrices $U \geq 0$ and $H \geq 0$, *hTrust* provides an iterative update algorithm of U and H (see [14] for more details). Let $U^* = (U^*_{i,k})$ and $H^* = (H^*_{k,\ell})$ denote the optimal U and H derived by the algorithm, respectively. We refer to a pair (i, j) with $G_{i,j} = 0$ as a *link-candidate*. *hTrust* predicts trust-links that are to be created in the period I_1 by ranking link-candidate (i, j) according to the value

$$G^*_{i,j} \;=\; \sum_{k,\ell=1}^{K} U^*_{i,k} \, H^*_{k,\ell} \, U^*_{j,\ell}.$$

C. Joint Link-Content Matrix Factorization

JLCMF also exploits only one latent space. Let K be the dimension of the latent space. *JLCMF* introduces not only a non-negative $N \times K$ matrix $U = (U_{i,k})$ and a non-negative $K \times K$ matrix $H = (H_{k,\ell})$, but also a non-negative $M \times K$ matrix $\Phi = (\Phi_{\alpha,k})$, where $\Phi_{\alpha,k}$ represents the relationship strength between item a_α and latent factor k from the point of view of posting reviews for items. We consider minimizing the following function $\mathcal{F}_1(U, H, \Phi)$ of U, H and Φ:

$$\begin{aligned}\mathcal{F}_1(U, H, \Phi) \;&=\; \left\|G - UHU^T\right\|^2 \,+\, \lambda_H \,\|H\|^2 \,+\, \lambda_\Phi \,\|\Phi\|^2 \\ &\quad + \lambda_X \left\|X - U\,\Phi^T\right\|^2 \end{aligned} \tag{6}$$

where λ_H, λ_Φ and λ_X are positive constants (hyper-parameters).

For the problem of minimizing the function $\mathcal{F}_1(U, H, \Phi)$ under the condition $U \geq 0$, $H \geq 0$ and $\Phi \geq 0$, we can easily derive an iterative update algorithm of U, H and Φ by applying *hTrust* [14] and an ordinary NMF [16]. Although *JLCMF* aims at classifying web pages, it can also be applied to the prediction of trust-links. Thus, in this paper, we regard *JLCMF*as an alternative NMF method exploiting both link and activity information for trust-link prediction.

III. Proposed Method

We propose a new NMF method that incorporates information of trust-links, users' activities and people's evaluations of users' activities, that is, an NMF method employing matrices G, X and Y. First, we propose a novel NMF model for trust-link prediction. Next, we derive an optimization algorithm for the proposed NMF model. Finally, we present a method of applying the proposed NMF model to an analysis of users' behavior.

A. NMF Model for Trust-link Prediction

We consider distinguishing a concept of fields which users prefer and a concept of fields for which users gain trust. The former fields are referred to as *P-fields*, and the latter fields are referred to as *T-fields*. Unlike *hTrust* and *JLCMF*, the proposed NMF model employs two latent spaces. One corresponds to the space of P-fields (called the *PF-space*), and the other corresponds to the space of T-fields (called the *TF-space*). Thus, P-field and T-field are also referred to as *latent P-factor* and *latent T-factor*, respectively.

Let K be the dimension of the PF-space and L the dimension of the TF-space. The proposed NMF model introduces a non-negative $N \times K$ matrix $U = (U_{i,k})$, a non-negative $N \times L$ matrix $W = (W_{i,\ell})$, and a non-negative $K \times L$ matrix $H = (H_{k,\ell})$, where $U_{i,k}$ represents the strength of user v_i for latent P-factor k, $W_{i,k}$ represents the strength of user v_i for latent T-factor k, and $H_{k,\ell}$ represents the relationship strength from latent P-factor k to latent T-factor ℓ for creating trust-links. We consider minimizing the following function $\mathcal{F}(U, W, H)$ of U, W and H:

$$\begin{aligned}&\mathcal{F}(U, W, H) = \\ &\qquad \left\|G - UHW^T\right\|^2 \,+\, \lambda_U \,\|U\|^2 \,+\, \lambda_W \,\|W\|^2 \,+\, \lambda_H \,\|H\|^2 \\ &\qquad + \lambda_X \,\mathrm{Tr}\left(U^T S_X U\right) \,+\, \lambda_Y \,\mathrm{Tr}\left(W^T S_Y W\right),\end{aligned} \tag{7}$$

where λ_U, λ_W, λ_H, λ_X and λ_Y are positive constants (hyper-parameters). $S_X = ((S_X)_{i,j})$ is the $N \times N$ symmetric matrix defined by Equations (2), (3) and (4). $S_Y = ((S_Y)_{i,j})$ is the $N \times N$ symmetric matrix defined by

$$(S_Y)_{i,j} \;=\; \bar{\eta}_i\,\delta_{i,j} \;-\; \eta_{i,j},$$

where

$$\eta_{i,j} \;=\; \frac{\sum_{\alpha=1}^{M} Y_{i,\alpha}\, Y_{j,\alpha}}{\sqrt{\sum_{\alpha=1}^{M} {Y_{i,\alpha}}^2}\,\sqrt{\sum_{\alpha=1}^{M} {Y_{j,\alpha}}^2}}$$

and

$$\bar{\eta}_i \;=\; \sum_{j=1}^{N} \eta_{i,j}.$$

Note that

$$\mathrm{Tr}\left(W^T S_Y W\right) \;=\; \frac{1}{2}\sum_{i,j=1}^{N} \eta_{i,j} \sum_{\ell=1}^{L} (W_{i,\ell} - W_{j,\ell})^2. \tag{8}$$

The proposed method incorporates information of both users' activities and people's evaluations of users' activities through the regularization terms $\mathrm{Tr}\left(U^T S_X U\right)$ and $\mathrm{Tr}\left(W^T S_Y W\right)$. Here, ξ_{ij} is a similarity measure between users v_i and v_j in terms of activity, and η_{ij} is a similarity measure between users v_i and v_j in terms of activity-evaluation. Note again that other choices for $\xi_{i,j}$ and $\eta_{i,j}$ are possible. Equations (5) and (8) imply that users of similar activities have similar representations in terms of latent P-factors, and users of similar activity-evaluations have similar representations in terms of latent T-factors. Moreover, the terms $\|U\|^2$, $\|W\|^2$ and $\|H\|^2$ are added to avoid overfitting as smoothness regularizations.

We present an iterative update algorithm of U, W and H (see Section III-B) to minimize the function $\mathcal{F}(U, W, H)$ under the condition $U \geq 0$, $W \geq 0$ and $H \geq 0$. Let $U^* = \left(U^*_{i,k}\right)$, $W^* = \left(W^*_{i,\ell}\right)$ and $H^* = \left(H^*_{k,\ell}\right)$ denote the optimal U, W and H found by the algorithm, respectively. The proposed method predicts trust-links to be created in the period I_1 by ranking link-candidate (i, j) according to the value

$$G^*_{i,j} \;=\; \sum_{k=1}^{K}\sum_{\ell=1}^{L} U^*_{i,k}\, H^*_{k,\ell}\, W^*_{j,\ell}.$$

B. Optimization Algorithm

For non-negative $N \times K$ matrix $U = (U_{i,k})$, non-negative $N \times L$ matrix $W = (W_{i,\ell})$ and non-negative $K \times L$ matrix $H = (H_{k,\ell})$, we consider the problem of minimizing the function $\mathcal{F}(U, W, H)$ defined by Equation (7). In what follows, we present an iterative update algorithm of U, W and H for solving this optimization problem. Hereafter, for any matrix B, $B_{p,q}$ denotes the (p, q) entry of B.

Let $\hat{U}$, $\hat{W}$ and $\hat{H}$ be the current estimates of U, W and H, respectively. Below, we derive update formulas of U, W and H. First, we define three auxiliary functions $\mathcal{E}_U(U, \hat{U}; W, H)$, $\mathcal{E}_W(W, \hat{W}; U, H)$ and $\mathcal{E}_H(H, \hat{H}; U, W)$ as follows:

$$\begin{aligned}&\mathcal{E}_U(U, \hat{U}; W, H) = \\ &\qquad \|G\|^2 \,+\, \sum_{i=1}^{N}\sum_{k=1}^{K} \frac{\left(\hat{U}HW^TWH^T\right)_{i,k}}{\hat{U}_{i,k}}\, {U_{i,k}}^2 \\ &\qquad - 2\sum_{i=1}^{N}\sum_{k=1}^{K} \left(GWH^T\right)_{i,k} \hat{U}_{i,k} \left(1 \,+\, \log\frac{U_{i,k}}{\hat{U}_{i,k}}\right) \,+\, \lambda_U \|U\|^2 \\ &\qquad + \lambda_X \sum_{i,j=1}^{N} \xi_{i,j} \sum_{k=1}^{K} \left\{ {U_{i,k}}^2 \,-\, \hat{U}_{i,k}\hat{U}_{j,k} \left(1 \,+\, \log\frac{U_{i,k}U_{j,k}}{\hat{U}_{i,k}\hat{U}_{j,k}}\right)\right\} \\ &\qquad + \lambda_W \|W\|^2 \,+\, \lambda_H \|H\|^2 \,+\, \lambda_Y \mathrm{Tr}\left(W^T S_Y W\right),\end{aligned} \tag{9}$$

$$\begin{aligned}
&\mathcal{E}_W(W,\hat{W};U,H) = \\
&\quad \|G\|^2 + \sum_{i=1}^{N}\sum_{\ell=1}^{L} \frac{\left(\hat{W}H^TU^TUH\right)_{i,\ell}}{\hat{W}_{i,\ell}} {W_{i,\ell}}^2 \\
&\quad -2\sum_{i=1}^{N}\sum_{\ell=1}^{L}\left(G^TUH\right)_{i,\ell}\hat{W}_{i,\ell}\left(1+\log\frac{W_{i,\ell}}{\hat{W}_{i,\ell}}\right) + \lambda_W\|W\|^2 \\
&\quad +\lambda_Y\sum_{i,j=1}^{N}\eta_{i,j}\sum_{\ell=1}^{L}\left\{{W_{i,\ell}}^2 - \hat{W}_{i,\ell}\hat{W}_{j,\ell}\left(1+\log\frac{W_{i,\ell}W_{j,\ell}}{\hat{W}_{i,\ell}\hat{W}_{j,\ell}}\right)\right\} \\
&\quad +\lambda_U\|U\|^2 + \lambda_H\|H\|^2 + \lambda_X\mathrm{Tr}\left(U^TS_XU\right), \qquad (10)
\end{aligned}$$

$$\begin{aligned}
&\mathcal{E}_H(H,\hat{H};U,W) = \\
&\quad \|G\|^2 + \sum_{k=1}^{K}\sum_{\ell=1}^{L}\frac{\left(U^TU\hat{H}W^TW\right)_{k,\ell}}{\hat{H}_{k,\ell}}{H_{k,\ell}}^2 + \lambda_H\|H\|^2 \\
&\quad -2\sum_{k=1}^{K}\sum_{\ell=1}^{L}\left(U^TGW\right)_{k,\ell}\hat{H}_{k,\ell}\left(1+\log\frac{H_{k,\ell}}{\hat{H}_{k,\ell}}\right) + \lambda_U\|U\|^2 \\
&\quad +\lambda_W\|W\|^2 + \lambda_X\mathrm{Tr}\left(U^TS_XU\right) + \lambda_Y\mathrm{Tr}\left(W^TS_YW\right). \qquad (11)
\end{aligned}$$

Then, we can prove the following inequalities and equations (see Lemma 1 in Appendix):

$$\begin{aligned}
\mathcal{E}_U(U,\hat{U};W,H) &\geq \mathcal{F}(U,W,H), \\
\mathcal{E}_W(W,\hat{W};U,H) &\geq \mathcal{F}(U,W,H), \\
\mathcal{E}_H(H,\hat{H};U,W) &\geq \mathcal{F}(U,W,H), \\
\mathcal{F}(U,W,H) &= \mathcal{E}_U(U,U;W,H) = \mathcal{E}_W(W,W;U,H) \\
&= \mathcal{E}_H(H,H;U,W). \qquad (12)
\end{aligned}$$

Thus, from Equations (9) and (12), we can derive an update formula for $U = (U_{ik})$ by minimizing $\mathcal{E}_U(U,\hat{U};W,H)$ with respect to U as follows (see Lemma 2 in Appendix for more details):

$$U_{i,k} = \hat{U}_{i,k}\sqrt{\frac{(GWH^T)_{i,k} + \lambda_X\sum_{j=1}^{N}\xi_{i,j}\hat{U}_{j,k}}{\left(\hat{U}HW^TWH^T\right)_{i,k} + \lambda_U\hat{U}_{i,k} + \lambda_X\bar{\xi}_i\hat{U}_{i,k}}} \qquad (13)$$

Also, from Equations (10) and (12), we can derive an update formula for $W = (W_{i\ell})$ by minimizing $\mathcal{E}_W(W,\hat{W};U,H)$ with respect to W as follows (see Lemma 2 in Appendix for more details):

$$W_{i,\ell} = \hat{W}_{i,\ell}\sqrt{\frac{(G^TUH)_{i,\ell} + \lambda_Y\sum_{j=1}^{N}\eta_{i,j}\hat{W}_{j,\ell}}{\left(\hat{W}H^TU^TUH\right)_{i,\ell} + \lambda_W\hat{W}_{i,\ell} + \lambda_Y\bar{\eta}_i\hat{W}_{i,\ell}}} \qquad (14)$$

Likewise, from Equations (11) and (12), we can derive an update formula for $H = (H_{k,\ell})$ by minimizing $\mathcal{E}_H(H,\hat{H};U,W)$ with respect to H as follows (see Lemma 2 in Appendix for more details):

$$H_{k,\ell} = \hat{H}_{k,\ell}\sqrt{\frac{(U^TGW)_{k,\ell}}{\left(U^TU\hat{H}W^TW\right)_{k,\ell} + \lambda_H\hat{H}_{k,\ell}}} \qquad (15)$$

Now, we explain the iterative update algorithm of U, W and H, which iteratively uses the update formulas (13), (14) and (15). For any non-negative integer t, let U_t, W_t and H_t denote the t-th update values of U, W and H, respectively. We define U_{t+1} by

$$U_{t+1} = \arg\min_U \mathcal{E}_U(U,U_t;W_t,H_t),$$

which is obtained from Equation (13). Also, we define W_{t+1} by

$$W_{t+1} = \arg\min_W \mathcal{E}_W(W,W_t;U_{t+1},H_t),$$

which is obtained from Equation (14). Likewise, we define H_{t+1} by

$$H_{t+1} = \arg\min_H \mathcal{E}_H(H,H_t;U_{t+1},W_{t+1}),$$

which is obtained from Equation (15). Then, from Equation (12), we have

$$\begin{aligned}
\mathcal{F}(U_t,W_t,H_t) &= \mathcal{E}_U(U_t,U_t;W_t,H_t) \geq \mathcal{E}_U(U_{t+1},U_t;W_t,H_t) \\
&\geq \mathcal{F}(U_{t+1},W_t,H_t) = \mathcal{E}_W(W_t,W_t;U_{t+1},H_t) \\
&\geq \mathcal{E}_W(W_{t+1},W_t;U_{t+1},H_t) \geq \mathcal{F}(U_{t+1},W_{t+1},H_t) \\
&= \mathcal{E}_H(H_t,H_t;U_{t+1},W_{t+1}) \geq \mathcal{E}_H(H_{t+1},H_t;U_{t+1},W_{t+1}) \\
&\geq \mathcal{F}(U_{t+1},W_{t+1},H_{t+1}).
\end{aligned}$$

Therefore, the objective function $\mathcal{F}(U,W,H)$ is monotonically decreasing under the iterative update algorithm, and so the algorithm converges.

C. Application to Behavioral Analysis

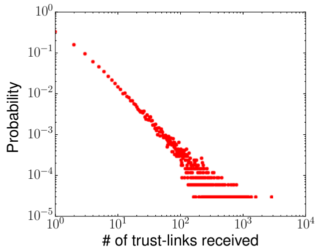
Using the optimal U, W and H found by the proposed algorithm, i.e., $U^* = \left(U^*_{i,k}\right)$, $W^* = \left(W^*_{i,\ell}\right)$ and $H^* = \left(H^*_{k,\ell}\right)$, we analyze users' behavior in the social media site under consideration. In order to investigate the latent P-factors (i.e., P-fields) and the latent T-factors (i.e., T-fields), we introduce a non-negative $M \times K$ matrix $\Phi = (\Phi_{\alpha,k})$ and a non-negative $M \times L$ matrix $\Psi = (\Psi_{\alpha,\ell})$, where $\Phi_{\alpha,k}$ represents the relationship strength between item a_α and latent P-factor k from the point of view of posting reviews for items, and $\Psi_{\alpha,\ell}$ represents the relationship strength between item a_α and latent T-factor ℓ from the point of view of receiving appreciation messages for reviews of items. Here, we consider minimizing the following functions $\mathcal{G}_X(\Phi)$ and $\mathcal{G}_Y(\Psi)$:

$$\begin{aligned}
\mathcal{G}_X(\Phi) &= \left\|X - U^*\Phi^T\right\|^2 + \lambda_\Phi\|\Phi\|^2, \\
\mathcal{G}_Y(\Psi) &= \left\|Y - W^*\Psi^T\right\|^2 + \lambda_\Psi\|\Psi\|^2,
\end{aligned}$$

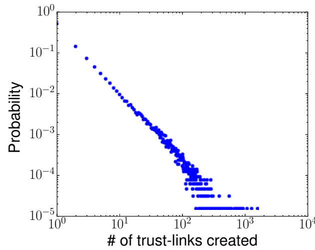
where λ_Φ and λ_Ψ are positive constants (hyper-parameters). For these minimization problems, we can derive iterative update algorithms of Φ and Ψ in the same way as Section III-B. Let $\Phi^* = \left(\Phi^*_{\alpha,k}\right)$ and $\Psi^* = \left(\Psi^*_{\alpha,\ell}\right)$ denote the optimal Φ and Ψ found by the algorithms, respectively. Using Φ^* and Ψ^*, we attempt to interpret the latent P-factors and the latent T-factors in terms of items. Based on those interpretations, we analyze the behavior of users in the site from the perspective of trust-link creation.

IV. Evaluation Experiments

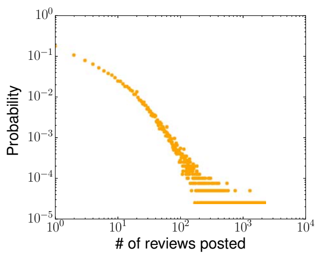
Using real data from two item-review sites, Epinions and @cosme. we evaluate the proposed method. We begin by statistically analyzing the datasets, and evaluate the proposed method for solving the trust-link prediction problem stated in Section II-A. Next, by applying the proposed method, we try to analyze the properties of users' behavior of each site from the perspective of trust-link creation.



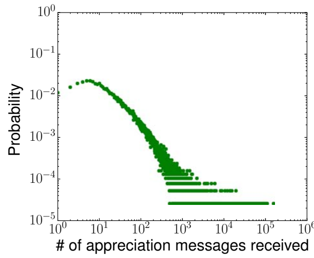
(a) Indegree distribution



(b) Outdegree distribution

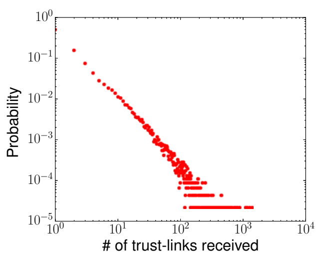


(c) Actvity distribution

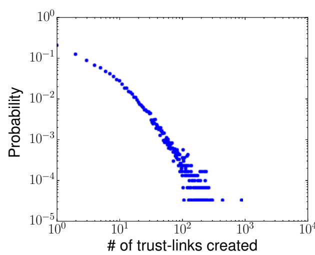


(d) Activity-evaluation distribution

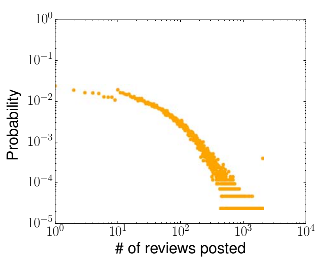
Fig. 1: Fundamental statistical analysis of the Epinions data.



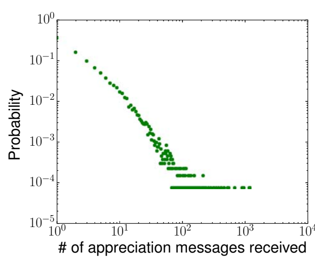
(a) Indegree distribution



(b) Outdegree distribution



(c) Activity distribution

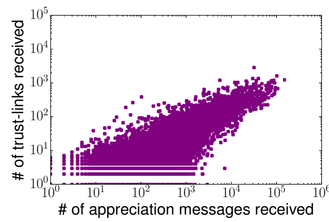


(d) Activity-evaluation distribution

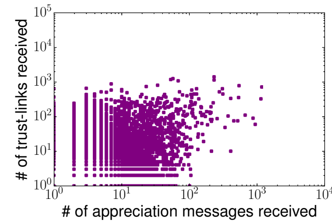
Fig. 2: Fundamental statistical analysis of the @cosme data.

A. Social Media Data

We collected real data from two item-review sites, Epinions[1] and @cosme[2], where Epinions is a social media site of product reviews and consumer reports, and @cosme is a Japanese word-of-mouth communication site for cosmetics. In both sites, a user can not only create a trust-link (or fan-link) to another user, but also post a review and give a rating for an item in a given set of items. Moreover, people can post their appreciation messages to a review of a user for an item if they found it useful. As for Epinions, we traced the



(a) Results for the Epinions data



(b) Results for the @cosme data

Fig. 3: Correlation between the number of activity-evaluations and indegree.

trust-links by the breadth-first search from a user who was featured as the most popular user in October 2012 until no new users appeared, and collected a set of trust-links, a set of reviews and ratings, and a set of appreciation messages. In a similar way, we also collected such data for @cosme in June 2010. The collected data includes $64,268$ users, $509,293$ trust-links, $809,517$ reviews and ratings for $268,891$ items, and $18,960,792$ appreciation messages for Epinions, and $30,369$ users, $359,817$ trust-links (fan-links), $3,815,622$ reviews and ratings for $122,927$ items, and $92,807$ appreciation messages for @cosme.

For the Epinions data and the @cosme data, we first investigated the indegree distribution (i.e., the fraction of the number of trust-links a user received), the outdegree distribution (i.e., the fraction of the number of trust-links a user created), the activity distribution (i.e., the fraction of the number of reviews a user posted) and the activity-evaluation distribution (i.e., the fraction of the number of appreciation messages a user received) in order to analyze their fundamental statistical properties. Figures 1 and 2 display the results for the Epinions data and the @cosme data, respectively. We observe that all the distributions exhibit power-law tails. These results imply that both the Epinions data and the @cosme data satisfy the typical properties of social data in an online world. Next, we examined a correlation between the number of activity-evaluations and indegree. Figure 3 indicates the results for the Epinions data and the @cosme data. We see that the number of appreciation messages a user received positively correlates with the number of trust-links the user received, This suggests that incorporating the activity-evaluation information can be a promising approach for improving an NMF method of trust-link prediction.

B. Experimental Settings

Using the Epinions data and the @cosme data, we evaluated the performance of the proposed method for solving the trust-link prediction problem (see Section II-A). For each data, we constructed four datasets D_1, D_2, D_3 and D_4 in the following way. First, we set $n_1 = n_2 = 1$ $\Delta t_0 = 6$, and $\Delta t_1 = 3$ (see Section II-A). Also, we let the prediction period I_1 be January to March for D_1, April to June for D_2, July to September for D_3, and October to December for D_4, respectively. By taking into account the stability for the number of trust-links created, we try to predict the trust-links created in 2006 for the Epinions data and 2009 for the @cosme data, respectively. Here, for example, for the dataset D_1 of the Epinions data, I_1 is January to March in 2006 and the

[1]http://www.epinions.com/
[2]http://www.cosme.net/

observation period I_0 is July to December in 2005. Tables I and II indicate the fundamental statistics of the datasets.

TABLE I: Fundamental statistics of the Epinions datasets

	D_1	D_2	D_3	D_4
# of users, N	771	722	734	727
# of items, M	56,642	57,886	59,522	61,327
# of observed trust-links	27,154	25,382	26,096	26,581
# of activities	83,786	83,933	86,396	88,980
# of activity-evaluations	1,074,042	1,039,505	1,127,326	1,170,950
# of trust-links created in I_1	1,670	1,308	1,441	1,684

TABLE II: Fundamental statistics of the @cosme datasets

	D_1	D_2	D_3	D_4
# of users, N	2,805	2,945	2,715	2,321
# of items, M	63,304	66,628	65,625	63,388
# of observed trust-links	22,733	27,129	24,497	20,274
# of activities	481,062	518,429	480,477	414,661
# of activity-evaluations	5,708	7,228	9,433	11,129
# of trust-links created in I_1	2,618	1,975	2,056	1,748

For the trust-link prediction problem, we compare the proposed method with *hTrust* (see Equation (1)) and *JLCMF* (see Equation (6)). Also, we define a variant of the proposed method by extending *JLCFM*, and compare the proposed method with it. Here, for a non-negative $N{\times}K$ matrix U, a non-negative $N{\times}L$ matrix W, a non-negative $K{\times}L$ matrix H, a non-negative $M{\times}K$ matrix Φ and a non-negative $M{\times}L$ matrix Ψ, we consider minimizing the following function $\mathcal{F}_2(U, W, H, \Phi, \Psi)$:

$$\begin{aligned}\mathcal{F}_2(U, W, H, \Phi, \Psi) \; = \\ & \left\| G - U H W^T \right\|^2 \; + \; \lambda_U \, \|U\|^2 \; + \; \lambda_W \, \|W\|^2 \; + \; \lambda_H \, \|H\|^2 \\ & + \lambda_X \left\| X - U\, \Phi^T \right\|^2 + \lambda_Y \left\| Y - W\, \Psi^T \right\|^2 \\ & + \lambda_\Phi \, \|\Phi\|^2 \; + \; \lambda_\Psi \, \|\Psi\|^2\end{aligned}$$

where λ_U, λ_W, λ_H, λ_X, λ_Y, λ_Φ and λ_Ψ are positive constants (hyper-parameters). For this minimization problem, we can easily derive an iterative update algorithm of U, W, H, Φ and Ψ in the same way as the proposed method (see Section III-B). We refer to this method as *JLCMF2*.

The hyper-parameters in all the methods can be determined through cross validation. In this paper, we simply set their values according to [14]. Specifically, we set $\lambda_U = \lambda_W = \lambda_H = \lambda_\Phi = \lambda_\Psi = 0.01$, and also set $K = L = 10$, $\lambda_X = \lambda_Y = 10$ for the Epinions data, and $K = L = 20$, $\lambda_X = \lambda_Y = 2.5$ for the @cosme data[3].

C. Trust-link Prediction

For the trust-link prediction problem (see Section II-A), we compared the proposed method with *hTrust*, *JLCMF* and *JLCMF2*. In the experiments, we measured the prediction performance in terms of the area under the ROC curve (AUC). Figures 4 and 5 show the results for the Epinions data and the @cosme data, respectively. Furthermore, Figure 6 indicates the ROC curves for the Epinions datasets, and Figure 7 indicates the ROC curves for the @cosme datasets. We observe that the proposed method performed the best and *hTrust* followed the next. The performance difference depends on datasets,

[3]Our immediate future work includes investigating the effects of hyper-parameters in detail.

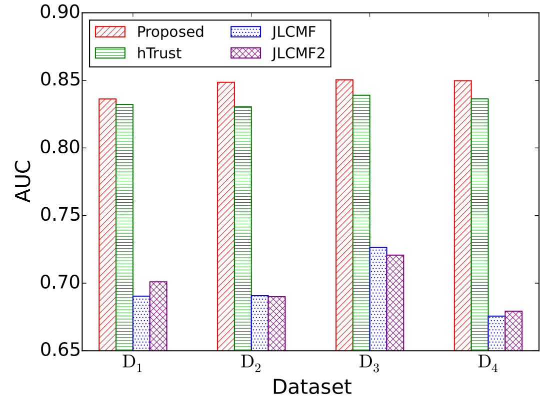


Fig. 4: Performance comparison of trust-link prediction for the Epinions data.

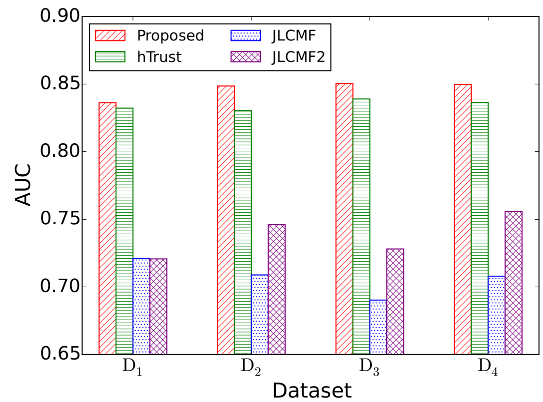


Fig. 5: Performance comparison of trust-link prediction for the @cosme data.

but *JLCMF* and *JLCMF2* were always worse than these two methods. These results show the importance of incorporating the activity-evaluation information, and demonstrate the effectiveness of the proposed method that appropriately combines two kinds of latent factors (i.e., two latent spaces): latent P-factors and latent T-factors.

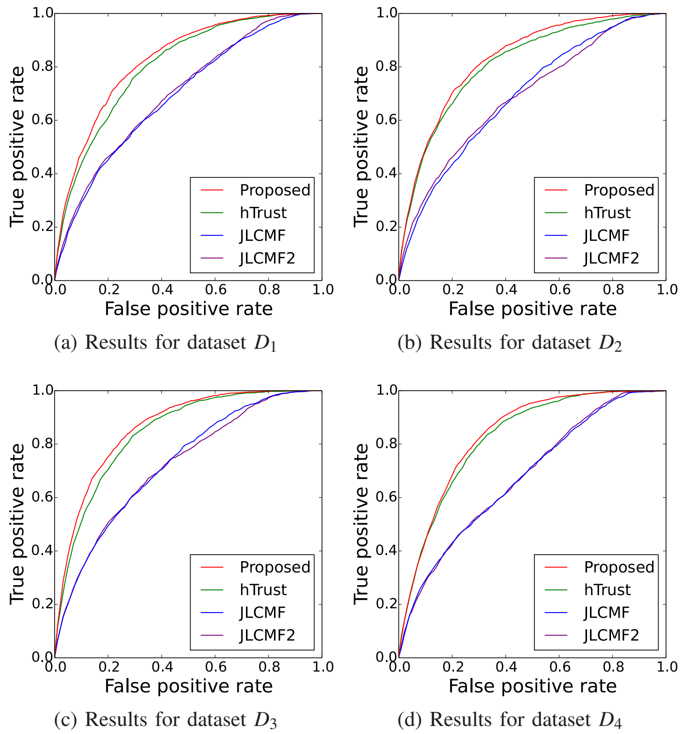
D. Behavioral Analysis

By applying the proposed method, we try to analyze users' behavior for trust-link creation in Epinions and @cosme. Here, we only report the analysis results for the dataset D_3.

Figure 8 shows the visualization results of $H^* = \left(H^*_{k,\ell}\right)$ for Epinions and @cosme, where the brightness of the entry in the k-th row and the ℓ-th column indicates the value of

TABLE III: Example of latent P-factor for the Epinions D_3 dataset ($k = 6$).

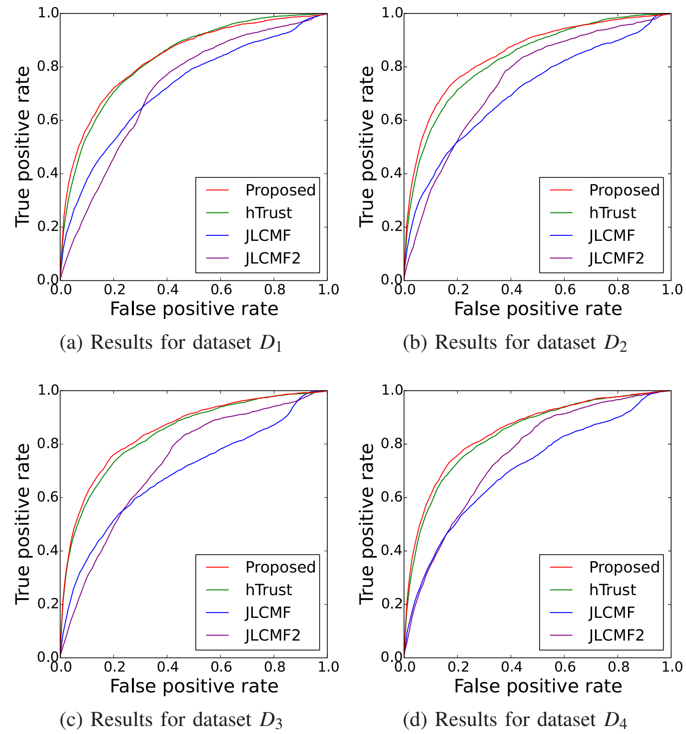
Item	Category
Nintendo Game Cube White Console	Video Game Consoles
Sony PlayStation 2 Slimline Console	Video Game Consoles
Star Wars Episode III: Revenge of the Sith	Movies
Sega Dreamcast Grey Console	Video Game Consoles



(a) Results for dataset D_1 (b) Results for dataset D_2

(c) Results for dataset D_3 (d) Results for dataset D_4

Fig. 6: ROC curves for the Epinions datasets.



(a) Results for dataset D_1 (b) Results for dataset D_2

(c) Results for dataset D_3 (d) Results for dataset D_4

Fig. 7: ROC curves for the @cosme datasets.

TABLE IV: Example of latent T-factor for the Epinions D_3 dataset ($\ell = 6$).

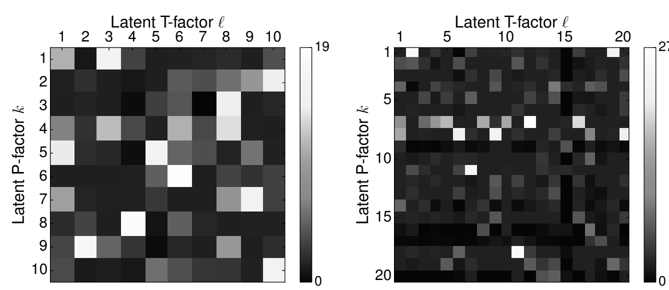
Item	Category
Lord of the Rings: The Return of the King	Movies
Star Wars Episode III: Revenge of the Sith	Movies
The Incredibles by Pixar & Disney	Movies
Lord of the Rings: The Fellowship of the Ring	Movies

relationship strength $H^*_{k,\ell}$ from latent P-factor k to latent T-factor ℓ for trust-link creation. We observe that the maximum value of $H^*_{k,\ell}$ is attained at $(k, \ell) = (6, 6)$ for the Epinions data and $(k, \ell) = (7, 12)$ for the @cosme data. Using $\Phi^* = \left(\Phi^*_{\alpha,k}\right)$ and $\Psi^* = \left(\Psi^*_{\alpha,\ell}\right)$, we also investigated the set of latent P-factors $\{k\}$ and the set of latent T-factors $\{\ell\}$. Here, we note that there were clear differences between the latent P-factors and the latent T-factors.

Tables III and IV show the top four items $\{a_\alpha\}$ of the Epinions data with respect to $\Phi^*_{\alpha,6}$ and $\Psi^*_{\alpha,6}$, respectively. Note that these results illustrate latent P-factor $k = 6$ and latent T-factor $\ell = 6$ for the Epinions data. Thus, we see that trust-links tend to be created from users who prefer video games

TABLE V: Example of latent P-factor for the @cosme D_3 dataset ($k = 7$).

Item (Brand)	Category
Stylo Yeux Waterproof (CHANEL / France)	Pencil Eyeliner
Ombres Couleurs (Cle de Peau-Beaute / Japan)	Powder Eyeshadow
Les Quatre Ombres (CHANEL / France)	Powder Eyeshadow
Decorte Liposome Treatment Liquid (Cosme Decorte / Japan)	Beauty Essence



(a) Results for the Epinions data (b) Results for the @cosme data

Fig. 8: Visualization of $H^* = \left(H^*_{k\ell}\right)$ for dataset D_3.

and science fiction movies to users who gain trust for science fiction & fantasy movies for the Epinions D_3 dataset.

Tables V and VI show the top four items $\{a_\alpha\}$ of the @cosme data with respect to $\Phi^*_{\alpha,7}$ and $\Psi^*_{\alpha,12}$, respectively. Note also that these results illustrate latent P-factor $k = 7$ and latent T-factor $\ell = 12$ for the @cosme data. Thus, we can observe

TABLE VI: Example of latent T-factor for the @cosme D_3 dataset ($\ell = 12$).

Item (Brand)	Category
Parure Pearly White Brightening Fluid Foundation (GUERLAIN / France)	Liquid-type Foundation
Double Wear Light Stay-in-Place Makeup SPF10 / PA++ (Estee Lauder / America)	Liquid-type Foundation
Dior Addict Ultra Gloss Reflect (Christian Dior / France)	Lip Gloss
Sublimage Eye (CHANEL / France)	Eye Care

that trust-links tend to be created from users who prefer a group of luxury makeup product to users who gain trust for another group of luxury makeup product for the @cosme D_3 dataset.

V. Conclusion

In this paper, we addressed the problem of predicting trust-links that are to be created among recent active users of an item-review site (a social media site) in the very near future. We adopted an NMF approach for this problem since the recently reported *hTrust* that uses an NMF method has been shown to be effective. *hTrust* exploits information of trust-links and users' activities, where a user activity means posting a review and giving a rating for an item. In many of item-review sites, people can post their appreciation messages to a review of a user for an item if they like it, that is, information of people's evaluations of users' activities is observable. Aiming to improve NMF methods for trust-link prediction, we proposed a new NMF method that incorporates the activity-evaluation information as well as both trust-link and activity information. We evaluated the proposed method using real data of two item-review sites: @cosme and Epinions.

First, we confirmed that the number of appreciation messages received correlates with the number of trust-links received, suggesting that incorporating the activity-evaluation information is promising. Next, we demonstrated that the proposed method outperforms *hTrust* and its variants *JLCMF* and *JLCMF2* in a trust-link prediction problem. These results show the importance of employing the activity-evaluation information for trust-link prediction, and demonstrate the effectiveness of the proposed method that appropriately combines two kinds of latent factors (i.e., two latent spaces): latent P-factors and latent T-factors. Further, we applied the proposed method to an analysis of users' behavior in an item-review site, and found several characteristic properties for Epinions and @cosme from the perspective of trust-link creation.

Acknowledgements

This work was partly supported by Asian Office of Aerospace Research and Development, Air Force Office of Scientific Research - United States, under Grant No. AOARD-13-4042, and JSPS Grant-in-Aid for Scientific Research (C) (No. 26330352), Japan.

References

[1] D. Kempe, J. Kleinberg, and E. Tardos, "Maximizing the spread of influence through a social network," in *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'03)*, 2003, pp. 137–146.

[2] M. Kimura, K. Saito, and H. Motoda, "Blocking links to minimize contamination spread in a social network," *ACM Transactions on Knowledge Discovery from Data*, vol. 3, pp. 9:1–9:23, 2009.

[3] M. Kimura, K. Saito, R. Nakano, and H. Motoda, "Extracting influential nodes on a social network for information diffusion," *Data Mining and Knowledge Discovery*, vol. 20, pp. 70–97, 2010.

[4] K. Saito, M. Kimura, K. Ohara, and H. Motoda, "Selecting information diffusion models over social networks for behavioral analysis," in *Proceedings of 2010 European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD'10)*, ser. LNAI, vol. 6323. Springer, 2010, pp. 180–195.

[5] W. Chen, L. Lakshmanan, and C. Castillo, "Information and influence propagation in social networks," *Synthesis Lectures on Data Management*, vol. 5, pp. 1–177, 2013.

[6] M. Kimura, K. Saito, K. Ohara, and H. Motoda, "Learning to predict opinion share and detect anti-majority opinionists in social networks," *Journal of Intelligent Information Systems*, vol. 41, pp. 5–37, 2013.

[7] K. Saito, M. Kimura, K. Ohara, and H. Motoda, "Detecting changes in information diffusion patterns over social networks," *ACM Transactions on Intelligent Systems and Technology*, vol. 4, pp. 55:1–55:23, 2013.

[8] R. Guha, R. Kumar, P. Raghavan, and A. Tomkins, "Propagation of trust and distrust," in *Proceedings of the 13th International Conference on World Wide Web (WWW'04)*, 2004, pp. 403–412.

[9] D. Liben-Nowell and J. Kleinberg, "The link-prediction problem for social networks," *Journal of the American Society for Information Science and Technology*, vol. 58, pp. 1019–1031, 2007.

[10] H. Liu, E. Lim, H. Lauw, M. Le, A. Sun, J. Srivastava, and Y. Kim, "Predicting trusts among users of online communities: an epinion case study," in *Proceedings of the 9th ACM Conference on Electronic Commerce (EC'08)*, 2008, pp. 310–319.

[11] H. Mannila and E. Terzi, "Finding links and initiators: a graph-reconstruction problem," in *Proceedings of 2009 SIAM International Conference on Data Mining (SDM'09)*, 2009, pp. 1207–1217.

[12] V. Nguyen, E. Lim, J. Jiang, and A. Sun, "To trust or not to trust? predicting online trusts using trust antecedent framework," in *Proceedings of the 9th IEEE International Conference on Data Mining (ICDM'09)*, 2009, pp. 896–901.

[13] M. Gomez-Rodriguez, J. Leskovec, and A. Krause, "Inferring networks of diffusion and influence," in *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'10)*, 2010, pp. 1019–1028.

[14] J. Tang, H. Gao, X. Hu, and H. Liu, "Exploiting homophyily effect for trust prediction," in *Proceedings of the 6th ACM International Conference on Web Search and Data Mining (WSDM'13)*, 2013, pp. 53–62.

[15] D. Crandall, D. Cosley, D. Huttenlocher, J. Kleinberg, and S. Suri, "Feedback effects between similarity and social influence in online communities," in *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'08)*, 2008, pp. 160–168.

[16] D. Lee and H. Seung, "Algorithms for non-negative matrix factorization," in *Advances in Neural Information Processing Systems 13 (NIPS-2000)*, 2001, pp. 556–562.

[17] S. Zhu, K. Yu, Y. Chi, and Y. Gong, "Combining content and link for classification using matrix factorization," in *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'07)*, 2007, pp. 487–494.

Appendix
Properties of Auxiliary Functions

Lemma 1: The following inequalities hold:

$$\begin{aligned}\mathcal{E}_U(U,\hat{U};W,H) &\geq \mathcal{F}(U,W,H)\\ \mathcal{E}_W(W,\hat{W};U,H) &\geq \mathcal{F}(U,W,H)\\ \mathcal{E}_H(H,\hat{H};U,W) &\geq \mathcal{F}(U,W,H)\end{aligned}$$

Proof: We first prove the inequality

$$\mathcal{E}_U(U,\hat{U};W,H) \geq \mathcal{F}(U,W,H).$$

From the definition of Frobenius norm and Equation (5), the following equation holds:

$$\begin{aligned}&\left\|G - UHW^T\right\|^2 + \lambda_X \operatorname{Tr}\left(U^T S_X U\right)\\ &= \sum_{i,j}\left(G_{i,j} - \sum_{k,\ell} U_{i,k} W_{j,\ell} H_{k,\ell}\right)^2 + \frac{\lambda_X}{2}\sum_{i,j}\xi_{i,j}\sum_k (U_{i,k} - U_{j,k})^2\end{aligned}$$

$$= \|G\|^2 + \sum_{i,j}\left(\sum_{k,\ell} U_{i,k}W_{j,\ell}H_{k,\ell}\right)^2 - 2\sum_{i,j}\sum_{k,\ell} G_{i,j}U_{i,k}W_{j,\ell}H_{k,\ell} + \lambda_X \sum_{i,j}\xi_{i,j}\sum_k \left({U_{i,k}}^2 - U_{i,k}U_{j,k}\right) \tag{16}$$

Then, we obtain the following inequality:

$$\begin{aligned}
&\sum_{i,j}\left(\sum_{k,\ell} U_{i,k}W_{j,\ell}H_{k,\ell}\right)^2 \\
&\quad = \sum_i \sum_{k,k'} \left(HW^TWH^T\right)_{k,k'} U_{i,k}U_{i,k'} \\
&\quad = \sum_i \sum_{k,k'} \left(HW^TWH^T\right)_{k,k'} \hat{U}_{i,k}\hat{U}_{i,k'} \frac{U_{i,k}}{\hat{U}_{i,k}} \frac{U_{i,k'}}{\hat{U}_{i,k'}} \\
&\quad \le \sum_i \sum_{k,k'} \left(HW^TWH^T\right)_{k,k'} \hat{U}_{i,k}\hat{U}_{i,k'} \frac{1}{2}\left(\frac{{U_{i,k}}^2}{{\hat{U}_{i,k}}^2} + \frac{{U_{i,k'}}^2}{{\hat{U}_{i,k'}}^2}\right) \\
&\quad = \sum_i \sum_{k,k'} \left(HW^TWH^T\right)_{k,k'} \hat{U}_{i,k}\hat{U}_{i,k'} \frac{{U_{i,k}}^2}{{\hat{U}_{i,k}}^2} \\
&\quad = \sum_i \sum_{k} \left(\hat{U}HW^TWH^T\right)_{k,k'} \frac{{U_{i,k}}^2}{\hat{U}_{i,k}}
\end{aligned} \tag{17}$$

Here, we used the inequality $\sqrt{xy} \le (x+y)/2$ for $x, y \ge 0$ and the symmetric property $\left(HW^TWH^T\right)_{k,k'} = \left(HW^TWH^T\right)_{k',k}$ for $1 \le k, k' \le K$. We also have the following inequality:

$$\begin{aligned}
&-2\sum_{i,j}\sum_{k,\ell} G_{i,j}U_{i,k}W_{j,\ell}H_{k,\ell} + \lambda_X \sum_{i,j}\xi_{i,j}\sum_k \left({U_{i,k}}^2 - U_{i,k}U_{j,k}\right) \\
&\quad = -2\sum_i\sum_k \left(GWH^T\right)_{i,k}\hat{U}_{i,k}\frac{U_{i,k}}{\hat{U}_{i,k}} \\
&\qquad + \lambda_X \sum_{i,j}\xi_{i,j}\sum_k\left({U_{i,k}}^2 - \hat{U}_{i,k}\hat{U}_{j,k}\frac{U_{i,k}U_{j,k}}{\hat{U}_{i,k}\hat{U}_{j,k}}\right) \\
&\quad \le -2\sum_i\sum_k \left(GWH^T\right)_{i,k}\hat{U}_{i,k}\left(1 + \log\frac{U_{i,k}}{\hat{U}_{i,k}}\right) \\
&\qquad + \lambda_X \sum_{i,j}\xi_{i,j}\sum_k\left\{{U_{i,k}}^2 - \hat{U}_{i,k}\hat{U}_{j,k}\left(1 + \log\frac{U_{i,k}U_{j,k}}{\hat{U}_{i,k}\hat{U}_{j,k}}\right)\right\}
\end{aligned} \tag{18}$$

Here, we used the inequality $x \ge 1 + \log x$ for $x > 0$. From Equations (7), (9), (16), (17) and (18), we can easily derive the inequality $\mathcal{E}_U(U, \hat{U}; W, H) \ge \mathcal{F}(U, W, H)$.

Next, we prove the inequality

$$\mathcal{E}_W(W, \hat{W}; U, H) \;\ge\; \mathcal{F}(U, W, H).$$

In the same way as Equation (16), we obtain

$$\begin{aligned}
&\left\|G - UHW^T\right\|^2 + \lambda_Y \operatorname{Tr}\left(W^TS_YW\right) \\
&= \|G\|^2 + \sum_{i,j}\left(\sum_{k,\ell} U_{i,k}W_{j,\ell}H_{k,\ell}\right)^2 - 2\sum_{i,j}\sum_{k,\ell} G_{i,j}U_{i,k}W_{j,\ell}H_{k,\ell} \\
&\quad + \lambda_Y \sum_{i,j}\eta_{i,j}\sum_k\left({W_{i,\ell}}^2 - W_{i,\ell}W_{j,\ell}\right).
\end{aligned} \tag{19}$$

Then, we have the following inequality:

$$\begin{aligned}
&\sum_{i,j}\left(\sum_{k,\ell} U_{i,k}W_{j,\ell}H_{k,\ell}\right)^2 \\
&\quad = \sum_i \sum_{\ell,\ell'} \left(H^TU^TUH\right)_{\ell,\ell'} W_{i,\ell}W_{i,\ell'} \\
&\quad = \sum_i \sum_{\ell,\ell'} \left(H^TU^TUH\right)_{\ell,\ell'} \hat{W}_{i,\ell}\hat{W}_{i,\ell'} \frac{W_{i,\ell}}{\hat{W}_{i,\ell}} \frac{W_{i,\ell'}}{\hat{W}_{i,\ell'}} \\
&\quad \le \sum_i \sum_{\ell,\ell'} \left(H^TU^TUH\right)_{\ell,\ell'} \hat{W}_{i,\ell}\hat{W}_{i,\ell'} \frac{1}{2}\left(\frac{{W_{i,\ell}}^2}{{\hat{W}_{i,\ell}}^2} + \frac{{W_{i,\ell'}}^2}{{\hat{W}_{i,\ell'}}^2}\right) \\
&\quad = \sum_i \sum_{\ell,\ell'} \left(H^TU^TUH\right)_{\ell,\ell'} \hat{W}_{i,\ell}\hat{W}_{i,\ell'} \frac{{W_{i,\ell}}^2}{{\hat{W}_{i,\ell}}^2} \\
&\quad = \sum_i \sum_{\ell} \left(\hat{W}H^TU^TUH\right)_{i,\ell} \frac{{W_{i,\ell}}^2}{\hat{W}_{i,\ell}}
\end{aligned} \tag{20}$$

Here, we used the the symmetric property $\left(H^TU^TUH\right)_{\ell,\ell'} = \left(H^TU^TUH\right)_{\ell',\ell}$ for $1 \le \ell, \ell' \le L$. We also get the following inequality:

$$\begin{aligned}
&-2\sum_{i,j}\sum_{k,\ell} G_{i,j}U_{i,k}W_{j,\ell}H_{k,\ell} + \lambda_Y \sum_{i,j}\eta_{i,j}\sum_\ell \left({W_{i,\ell}}^2 - W_{i,\ell}W_{j,\ell}\right) \\
&\quad = -2\sum_i\sum_\ell \left(G^TUH\right)_{i,\ell}\hat{W}_{i,\ell}\frac{W_{i,\ell}}{\hat{W}_{i,\ell}} \\
&\qquad + \lambda_Y \sum_{i,j}\eta_{i,j}\sum_\ell\left({W_{i,\ell}}^2 - \hat{W}_{i,\ell}\hat{W}_{j,\ell}\frac{W_{i,\ell}W_{j,\ell}}{\hat{W}_{i,\ell}\hat{W}_{j,\ell}}\right) \\
&\quad \le -2\sum_i\sum_\ell \left(G^TUH\right)_{i,\ell}\hat{W}_{i,\ell}\left(1 + \log\frac{W_{i,\ell}}{\hat{W}_{i,\ell}}\right) \\
&\qquad + \lambda_Y \sum_{i,j}\eta_{i,j}\sum_\ell\left\{{W_{i,\ell}}^2 - \hat{W}_{i,\ell}\hat{W}_{j,\ell}\left(1 + \log\frac{W_{i,\ell}W_{j,\ell}}{\hat{W}_{i,\ell}\hat{W}_{j,\ell}}\right)\right\}
\end{aligned} \tag{21}$$

Thus, from Equations (7), (10), (19), (20) and (21), we can easily show the inequality $\mathcal{E}_W(W, \hat{W}; U, H) \ge \mathcal{F}(U, W, H)$.

Finally, we prove the inequality

$$\mathcal{E}_H(H, \hat{H}; U, W) \;\ge\; \mathcal{F}(U, W, H).$$

First, it follows that

$$\begin{aligned}
&\left\|G - UHW^T\right\|^2 \\
&\quad = \sum_{i,j}\left(\sum_{k,\ell} U_{i,k}W_{j,\ell}H_{k,\ell}\right)^2 - 2\sum_{i,j}\sum_{k,\ell} G_{i,j}U_{i,k}W_{j,\ell}H_{k,\ell}.
\end{aligned} \tag{22}$$

Then, we obtain the following inequality:

$$\begin{aligned}
&\sum_{i,j}\left(\sum_{k,\ell} U_{i,k}W_{j,\ell}H_{k,\ell}\right)^2 \\
&\quad = \sum_{k,k'}\sum_{\ell,\ell'} \left(U^TU\right)_{k,k'}\left(W^TW\right)_{\ell,\ell'} H_{k,\ell}H_{k',\ell'} \\
&\quad = \sum_{k,k'}\sum_{\ell,\ell'} \left(U^TU\right)_{k,k'}\left(W^TW\right)_{\ell,\ell'} \hat{H}_{k,\ell}\hat{H}_{k',\ell'} \frac{H_{k,\ell}}{\hat{H}_{k,\ell}} \frac{H_{k',\ell'}}{\hat{H}_{k',\ell'}} \\
&\quad \le \sum_{k,k'}\sum_{\ell,\ell'} \left(U^TU\right)_{k,k'}\left(W^TW\right)_{\ell,\ell'} \hat{H}_{k,\ell}\hat{H}_{k',\ell'} \frac{1}{2}\left(\frac{{H_{k,\ell}}^2}{{\hat{H}_{k,\ell}}^2} + \frac{{H_{k',\ell'}}^2}{{\hat{H}_{k',\ell'}}^2}\right)
\end{aligned}$$

$$= \sum_{k,k'} \sum_{\ell,\ell'} \left(U^T U\right)_{k,k'} \left(W^T W\right)_{\ell,\ell'} \hat{H}_{k,\ell} \hat{H}_{k',\ell'} \frac{{H_{k,\ell}}^2}{{\hat{H}_{k,\ell}}^2}$$

$$= \sum_{k,\ell} \left(U^T U \hat{H} W^T W\right)_{k,\ell} \frac{{H_{k,\ell}}^2}{\hat{H}_{k,\ell}} \tag{23}$$

Here, we used the symmetric property $\left(U^T U\right)_{k,k'} \left(W^T W\right)_{\ell,\ell'} = \left(U^T U\right)_{k',k} \left(W^T W\right)_{\ell',\ell}$ for $1 \le k, k' \le K$, $1 \le \ell, \ell' \le L$. We also have the following inequality:

$$\begin{aligned} -2 \sum_{i,j} \sum_{k,\ell} & G_{i,j} U_{i,k} W_{j,\ell} H_{k,\ell} \\ &= -2 \sum_{k,\ell} \left(U^T G W\right)_{k,\ell} \hat{H}_{k,\ell} \frac{H_{k,\ell}}{\hat{H}_{k,\ell}} \\ &\le -2 \sum_{k,\ell} \left(U^T G W\right)_{k,\ell} \hat{H}_{k,\ell} \left(1 + \log \frac{H_{k,\ell}}{\hat{H}_{k,\ell}}\right) \end{aligned} \tag{24}$$

The inequality $\mathcal{E}_H(H, \hat{H}; U, W) \ge \mathcal{F}(U, W, H)$ follows immediately from Equations (7), (11), (22), (23) and (24). Hence, Lemma 1 is proved. □

Lemma 2: The functions $\mathcal{E}_U(U, \hat{U}; W, H)$, $\mathcal{E}_W(W, \hat{W}; U, H)$ and $\mathcal{E}_H(H, \hat{H}; U, W)$ have global minimum points at $U^\# = \left(U^\#_{i,k}\right)$, $W^\# = \left(W^\#_{i,\ell}\right)$ and $H^\# = \left(H^\#_{k,\ell}\right)$, resepectively:

$$U^\#_{i,k} = \hat{U}_{i,k} \sqrt{\frac{(GWH^T)_{i,k} + \lambda_X \sum_{j=1}^N \xi_{i,j} \hat{U}_{j,k}}{\left(\hat{U} H W^T W H^T\right)_{i,k} + \lambda_U \hat{U}_{i,k} + \lambda_X \bar{\xi}_i \hat{U}_{i,k}}}$$

$$W^\#_{i,\ell} = \hat{W}_{i,\ell} \sqrt{\frac{(G^T U H)_{i,\ell} + \lambda_Y \sum_{j=1}^N \eta_{i,j} \hat{W}_{j,\ell}}{\left(\hat{W} H^T U^T U H\right)_{i,\ell} + \lambda_W \hat{W}_{i,\ell} + \lambda_Y \bar{\eta}_i \hat{W}_{i,\ell}}}$$

$$H^\#_{k,\ell} = \hat{H}_{k,\ell} \sqrt{\frac{(U^T G W)_{k,\ell}}{\left(U^T U \hat{H} W^T W\right)_{k,\ell} + \lambda_H \hat{H}_{k,\ell}}}$$

Proof: By definition, we first note that the matrices $U^\#$, $W^\#$ and $H^\#$ are guaranteed to be non-negative. From Equations (9), (10) and (11), we obtain the following equations:

$$\begin{aligned} \frac{\partial \mathcal{E}_U}{\partial U_{i,k}} = & \frac{2\left(\hat{U} H W^T W H^T\right)_{i,k} U_{i,k}}{\hat{U}_{i,k}} - \frac{2\left(G W H^T\right)_{i,k} \hat{U}_{i,k}}{U_{i,k}} \\ & + 2\lambda_U U_{i,k} + 2\lambda_X \bar{\xi}_i U_{i,k} - \frac{2\lambda_X \hat{U}_{i,k} \sum_j \xi_{i,j} \hat{U}_{j,k}}{U_{i,k}} \end{aligned} \tag{25}$$

$$\begin{aligned} \frac{\partial \mathcal{E}_W}{\partial W_{i,\ell}} = & \frac{2\left(\hat{W} H^T U^T U H\right)_{i,\ell} W_{i,\ell}}{\hat{W}_{i,\ell}} - \frac{2\left(G^T U H\right)_{i,\ell} \hat{W}_{i,\ell}}{W_{i,\ell}} \\ & + 2\lambda_W W_{i,\ell} + 2\lambda_Y \bar{\eta}_i W_{i,\ell} - \frac{2\lambda_Y \hat{W}_{i,\ell} \sum_j \eta_{i,j} \hat{W}_{j,\ell}}{W_{i,\ell}} \end{aligned} \tag{26}$$

$$\begin{aligned} \frac{\partial \mathcal{E}_H}{\partial H_{k,\ell}} = & \frac{2\left(U^T U \hat{H} W^T W\right)_{k,\ell} H_{k,\ell}}{\hat{H}_{k,\ell}} - \frac{2\left(U^T G W\right)_{k,\ell} \hat{H}_{k,\ell}}{H_{k,\ell}} \\ & + 2\lambda_H H_{k,\ell} \end{aligned} \tag{27}$$

Then, it is easily shown that

$$\left.\frac{\partial \mathcal{E}_U}{\partial U_{i,k}}\right|_{U=U^\#} = 0 \qquad (i = 1, \ldots, N,\ k = 1, \ldots, K), \tag{28}$$

$$\left.\frac{\partial \mathcal{E}_W}{\partial W_{i,\ell}}\right|_{W=W^\#} = 0 \qquad (i = 1, \ldots, N,\ \ell = 1, \ldots, L), \tag{29}$$

$$\left.\frac{\partial \mathcal{E}_H}{\partial H_{k,\ell}}\right|_{H=H^\#} = 0 \qquad (k = 1, \ldots, K,\ \ell = 1, \ldots, L). \tag{30}$$

From Equations (25), (26) and (27), we also have the following equations:

$$\begin{aligned} \frac{\partial^2 \mathcal{E}_U}{\partial {U_{i,k}}^2} = & \frac{2\left(\hat{U} H W^T W H^T\right)_{i,k}}{\hat{U}_{i,k}} + \frac{2\left(G W H^T\right)_{i,k} \hat{U}_{i,k}}{{U_{i,k}}^2} \\ & + 2\lambda_U + 2\lambda_X \bar{\xi}_i + \frac{2\lambda_X \hat{U}_{i,k} \sum_j \xi_{i,j} \hat{U}_{j,k}}{{U_{i,k}}^2} \end{aligned} \tag{31}$$

$$\begin{aligned} \frac{\partial^2 \mathcal{E}_W}{\partial {W_{i,\ell}}^2} = & \frac{2\left(\hat{W} H^T U^T U H\right)_{i,\ell}}{\hat{W}_{i,\ell}} + \frac{2\left(G^T U H\right)_{i,\ell} \hat{W}_{i,\ell}}{{W_{i,\ell}}^2} \\ & + 2\lambda_W + 2\lambda_Y \bar{\eta}_i + \frac{2\lambda_Y \hat{W}_{i,\ell} \sum_j \eta_{i,j} \hat{W}_{j,\ell}}{{W_{i,\ell}}^2} \end{aligned} \tag{32}$$

$$\begin{aligned} \frac{\partial^2 \mathcal{E}_H}{\partial {H_{k,\ell}}^2} = & \frac{2\left(U^T U \hat{H} W^T W\right)_{k,\ell}}{\hat{H}_{k,\ell}} + \frac{2\left(U^T G W\right)_{k,\ell} \hat{H}_{k,\ell}}{{H_{k,\ell}}^2} \\ & + 2\lambda_H \end{aligned} \tag{33}$$

$$\frac{\partial^2 \mathcal{E}_U}{\partial U_{i,k}\, \partial U_{i',k'}} = \delta_{i,i'}\, \delta_{k,k'}\, \frac{\partial^2 \mathcal{E}_U}{\partial {U_{i,k}}^2} \tag{34}$$

$$\frac{\partial^2 \mathcal{E}_W}{\partial W_{i,\ell}\, \partial W_{i',\ell'}} = \delta_{i,i'}\, \delta_{\ell,\ell'}\, \frac{\partial^2 \mathcal{E}_W}{\partial {W_{i,\ell}}^2} \tag{35}$$

$$\frac{\partial^2 \mathcal{E}_H}{\partial H_{k,\ell}\, \partial H_{k',\ell'}} = \delta_{k,k'}\, \delta_{\ell,\ell'}\, \frac{\partial^2 \mathcal{E}_H}{\partial {H_{k,\ell}}^2} \tag{36}$$

By Equations (31), (32) and (33), we obtain

$$\frac{\partial^2 \mathcal{E}_U}{\partial {U_{i,k}}^2} > 0, \quad \frac{\partial^2 \mathcal{E}_W}{\partial {W_{i,\ell}}^2} > 0, \quad \frac{\partial^2 \mathcal{E}_H}{\partial {H_{k,\ell}}^2} > 0 \tag{37}$$

since the matrices to be considered are non-negative and all the hyper-parameters are positive. Thus, by Equations (34), (35), (36) and (37), three Hessian matrices

$$\left(\frac{\partial^2 \mathcal{E}_U}{\partial U_{i,k}\, \partial U_{i',k'}}\right), \quad \left(\frac{\partial^2 \mathcal{E}_W}{\partial W_{i,\ell}\, \partial W_{i',\ell'}}\right), \quad \left(\frac{\partial^2 \mathcal{E}_H}{\partial H_{k,\ell}\, \partial H_{k',\ell'}}\right)$$

are positive definite. Hence, it follows from Equations (28), (29) and (30) that functions $\mathcal{E}_U(U, \hat{U}; W, H)$, $\mathcal{E}_W(W, \hat{W}; U, H)$ and $\mathcal{E}_H(H, \hat{H}; U, W)$ attain minimum values at $U = U^\#$, $W = W^\#$ and $H = H^\#$, resepectively. This completes the proof of Lemma 2. □