A Generative Bayesian Model for Item and User Recommendation in Social Rating Networks with Trust Relationships

Gianni Costa, Giuseppe Manco, Riccardo Ortale

ICAR-CNR
Via Bucci 41c
87036 Rende (CS)
{costa, manco, ortale}@icar.cnr.it

Abstract. A Bayesian generative model is presented for recommending interesting items and trustworthy users to the targeted users in social rating networks with asymmetric and directed trust relationships. The proposed model is the first unified approach to the combination of the two recommendation tasks. Within the devised model, each user is associated with two latent-factor vectors, i.e., her susceptibility and expertise. Items are also associated with corresponding latent-factor vector representations. The probabilistic factorization of the rating data and trust relationships is exploited to infer user susceptibility and expertise. Statistical social-network modeling is instead used to constrain the trust relationships from a user to another to be governed by their respective susceptibility and expertise. The inherently ambiguous meaning of unobserved trust relationships between users is suitably disambiguated. An intensive comparative experimentation on real-world social rating networks with trust relationships demonstrates the superior predictive performance of the presented model in terms of RMSE and AUC.

1 Introduction

The growing popularity gained by various online services for social networking has led to the increasing availability of online social rating networks [14], i.e., environments in which users rate items and establish connections to real-world acquaintances within their social networks. In particular, the presence of explicit trust relationships between users makes such environments an appealing setting for the development of realistic recommendation processes, in which the targeted users turn to their social networks for decision making and are more strongly influenced by (directly or indirectly) trusted real-world acquaintances.

Two fundamental tasks in social rating networks with trust relationships are item recommendation and user recommendation The former consists in taking advantage of the trust relationships to suggest unrated items, that are expected to be of interest to the targeted users. The latter instead consists in taking advantage of the trust relationships to suggest users having no relationships with the targeted users and still expected to be trusted by them.

Each individual task has been extensively studied in the literature in isolation. Previous research on rating prediction for item recommendation in social rating networks can be divided into two major areas of focus reflecting the nature of relationships in the underlying social networks, i.e., *unilateral* relationships (such as, e.g., trust) or *cooperative and mutual* relationships (such as, e.g., in the case of friends, classmates, colleagues, relatives and so forth) [19]. Rating prediction in trust networks has been the subject of several studies such as, e.g., [14, 17, 18]. A variety of other research efforts including [10, 19, 23, 27, 31] has instead focused on social rating networks with mutual relationships. Instead, the existing approaches to link prediction can be classified into two distinct classes, i.e., *temporal* and *structural* approaches. The temporal approaches predict links between the nodes of a graph, whose evolution involves new links and new nodes with respective ties. The structural approaches assume graphs with fixed sets of nodes and, thus, they are concerned only with the prediction of new links between already observed nodes. Temporal and structural approaches can be further divided into *unsupervised* or *supervised*. Unsupervised approaches [15] do not involve a learning phase. Rather, they compute predefined scores based on graph topology alone. On the contrary, link prediction is treated as a binary classification task in supervised approaches [1, 2, 6, 8, 7, 20, 21, 30, 32], which essentially learn some suitable model with which to predict scores for pairs of nodes [20]. Certain supervised approaches also allow the optional exploitation of side information on the nodes, e.g., [21, 20]. Rating and link prediction are both instances of *dyadic prediction*, which is the more general problem of predicting a label for unobserved interactions between pairs of entities [13, 20]. Nonetheless, modeling and studying them jointly has been so far unexplored.

In this paper, to the best of our knowledge, we propose the first unified approach to trust-aware recommendation of both items and users in social rating networks. The devised approach consists in a Bayesian nonparametric hierarchical model, in which the interactions from users to users as well as between users and items are assumed to be explained by some suitable number of latent factors. More precisely, each user is associated with two real-valued latent-factor vectors, namely, her *susceptibility* and *expertise*, similarly to [3]. The entries of the susceptibility vector are the degree to which the user is sensible to the corresponding latent factors. The entries of the expertise vector are the extent to which the user can meet the susceptibility requirements of other users on the corresponding latent factors. Additionally, each item is associated with a real-valued latent-factor vector, whose entries indicate the degree to which the item is characterized by the corresponding latent factors.

The proposed model combines ideas from Bayesian probabilistic matrix factorization [25] and statistical social-network modeling to infer and exploit the foresaid latent-factor vector representations of users and items. Specifically, the seminal Bayesian approach in [25] is extended to infer the susceptibility and expertise of each user as well as the latent-factor vector representation of every item through the probabilistic factorization of the user-rating and trust-relationship matrices. Statistical social-network modeling is instead employed for a twofold purpose. On one hand, it is used to model trust relationships governed by the susceptibility and expertise of the trusting and trusted users, respectively. On the other hand, it is leveraged to properly deal with the inherent ambiguity of the unobserved trust relationships. Therein, a missing trust relationship between two users may mean either actual lack of trust or lack of awareness. Such possibilities are mixed together across the unobserved trust relationships of the social rating network at hand and, in general, cannot be distinguished beforehand. An especially interesting and novel aspect of the devised model is that each unobserved trust relationship is associated with a respective binary latent variable, whose inferred value allows to suitably account for its actual meaning.

Unlike previous approaches to item recommendation, the devised model infers the susceptibility and expertise of the individual users by accounting for both the available ratings as well as the trust relationships. Such representations are shared across rating and link prediction, which enables performing both tasks jointly. Moreover, differently from existing approaches to link prediction, the establishment of a link from a user to another is ruled only by their respective susceptibility and expertise. Yet, unobserved trust relationships are treated by drawing from research in one-class collaborative filtering (e.g., [22, 28]).

The presented model is comparatively investigated over real-world social rating networks. The empirical evidence demonstrates the superiority of its predictive performance in terms of both RMSE and AUC.

The contents of this paper are organized as follows. Section 2 introduces notation and some preliminary concepts. Section 3 covers the proposed model. Section 4 develops approximate posterior inference within the proposed model. Section 5 presents the empirical results of an intensive comparative evaluation of the proposed model against state-of-the-art competitors on real-world social rating networks. Finally, Section 6 concludes and highlights future research.

2 Preliminaries and problem statement

A social rating network [14] can be formalized as a tuple $\mathcal{N} = \langle N, A, \mathcal{I}, R\rangle$ where N is a set of n users and $A \subseteq N \times N \times \{0, 1\}$ is a set of directed links between users. The underlying graph $\mathcal{G} = \langle N, A\rangle$ represents trust relationships between users. In particular, a positive link $u \overset{1}{\to} v$ means that u trusts v and, dually, a negative link $u \overset{0}{\to} v$ denotes lack of u's trust in v. We will generically use matrix notation $A_{u,v}$ to succinctly denote $u \to v$. Clearly, $A_{u,v}$ is either 0 or 1, according to whether the link $u \to v$ is negative or positive. In the following, we assume to be aware only of positive links and, thus, a missing link from u to v can denote either lack of trust, or lack of awareness (i.e., u is not aware of v).

$\mathcal{G}$ can be viewed as a graph with attributes by also accounting for additional node information. We focus on the degrees of preference (or ratings) assigned by the individual users from N to the elements of a set $\mathcal{I}$ of m items. Such preference degrees are summarized into the ratings $R \subseteq N \times \mathcal{I} \times \mathcal{V}$, whose generic entry $\langle u, i, r\rangle$ denotes the rating $r \in \mathcal{V} = \{1, \ldots, V\}$ assigned by user $u \in N$ to item $i \in \mathcal{I}$. Hereafter, we denote the rating r relative to the entry $\langle u, i, r\rangle$ as $R_{u,i}$.

We assume that trust relationships between users as well as their ratings to items can be explained in terms of a number of latent (i.e., unobserved and unknown) factors, that also contribute to characterize the individual items. More precisely, each user is associated with an extent of *susceptibility* and *expertise* with respect to the individual latent factors. A rating is governed by the combination of the susceptibility and expertise of a user with the extent to which the targeted item is characterized by each latent factor. A trust relationship from a user to another is determined by their respective susceptibility and expertise. Given a generic social rating network $\mathcal{N}$, we aim to infer a probabilistic model from the trust relationships observed in $\mathcal{G}$, that allows to recommend both interesting items and further trustworthy users to the targeted users within the network. The recommendation of interesting items is essentially a rating prediction task. Given a user $u \in N$ and an item $i \in \mathcal{I}$ such that $R_{u,i}$ is unknown, the degree of u's preference for i is predicted using $\mathcal{G}$ and R. In particular, if the trusted neighbors of u in $\mathcal{G}$ enjoyed i, then $R_{u,i}$ should be predicted accordingly. Analogously, the recommendation of trustworthy users is a trust prediction task. Given a pair of users $u, v \in N$ such that $u \rightarrow v \notin A$, the trust of u in v is again predicted using $\mathcal{G}$ and R. Specifically, if the trusted neighbors of u in $\mathcal{G}$ trust v because of her ratings, then u should trust v as well and, hence, a trust relationship should be established in $\mathcal{G}$ from u to v, i.e., the positive link $u \xrightarrow{1} v$ should be added to A. Instead, if trusted neighbors of u do not trust v, or if v's ratings significantly differ from u's known ratings, then a negative link $u \xrightarrow{0} v$ should be established. Trust relationships A and ratings R are the only observed data in $\mathcal{N}$. All other aforementioned aspects of interest cannot be measured directly.

3 The devised Bayesian generative model

We propose a Bayesian hierarchical model, that combines probabilistic matrix factorization and network modeling for the recommendation of items and users in a social rating network $\mathcal{N}$. Specifically, matrix factorization is exploited to explicitly capture the latent factors governing both trust relationships and item ratings. Network modeling contributes to determine user susceptibility and expertise. Probabilistic matrix factorization and statistical network modeling are seamlessly integrated for performing collaborative filtering, in order to suggest interesting items and establish missing relationships with trustable users.

In the following K is the overall number of latent factors behind the observed trust relationships in $\mathcal{G}$. Each user $u \in N$ is associated with two column vectors $\mathbf{P}_u, \mathbf{F}_u \in \mathbb{R}^K$. The generic k-th entry of $\mathbf{P}_u$ indicates the susceptibility of u to the latent topic k. Analogously, the k-th entry of $\mathbf{F}_u$ denotes the degree of expertise exhibited by u with regard to k. The susceptibility and expertise of all users in $\mathcal{G}$ are collectively denoted by means of matrices $\mathbf{P}$ and $\mathbf{F}$, respectively, where $\mathbf{P}, \mathbf{F} \in \mathbb{R}^{K \times M}$. A representation based on the latent factors is also adopted for the items in the set $\mathcal{I}$. The generic item $i \in \mathcal{I}$ is associated with one column vector $\mathbf{Q}_i \in \mathbb{R}^K$, whose k-th entry is the extent at which the latent factor k characterizes the item i. The latent factor representations of all items are collectively represented by the matrix $\mathbf{Q}$, where $\mathbf{Q} \in \mathbb{R}^{K \times N}$. Ratings $R_{u,i}$ for all $u \in N$ and $i \in \mathcal{I}$ are considered as random variables ranging in the set $\mathcal{V}$ of admissible values. Thus, in the proposed model the data likelihood, i.e., the conditional distribution over the observed data in R and A is given by

$$\Pr(R|\mathbf{P}, \mathbf{Q}, \mathbf{F}, \alpha) = \prod_{u \in N} \prod_{i \in \mathcal{I}} \mathcal{N}(R_{u,i}; \theta_{u,i}, \alpha^{-1})^{\delta_{u,i}} \tag{3.1}$$

$$\Pr\left(A|\mathbf{P}, \mathbf{Q}, \mathbf{F}, \mathbf{Z}, \beta\right) = \prod_{u \to v \in A} \mathcal{N}(A_{u,v}; \vartheta_{u,v}, \beta^{-1}) \tag{3.2}$$

where

$$\theta_{u,i} = (\mathbf{P}_u + \mathbf{F}_u)' \, \mathbf{Q}_j \text{ and } \vartheta_{u,v} = \mathbf{P}'_u \mathbf{F}_v$$

and $\mathcal{N}(x|\mu, \alpha^{-1})$ is the Gaussian distribution with mean μ and precision α. In particular, the observed links are centered around the dot product between the susceptibility of the start user and the expertise of the end user, which can be interpreted as the capability of the latter of satisfying the requirements of the former. Ratings involve the dot product of the sum of user susceptibility and expertise with the latent-factor representation of items, which entirely captures the interaction between users and items. Function $\delta_{u,i}$ is instead a binary indicator, which equals 1 if $R_{u,i} > 0$ (i.e., if u actually rated i) and 0 otherwise.

The representations in terms of latent-factors associated with users (i.e., their susceptibility and expertise) as well as items are drawn from prior distributions, which are assumed to be Gaussian with parameters $\Theta_{\mathbf{P}} = \{\mu_{\mathbf{P}}, \Lambda_{\mathbf{P}}\}$, $\Theta_{\mathbf{Q}} = \{\mu_{\mathbf{Q}}, \Lambda_{\mathbf{Q}}\}$ and $\Theta_{\mathbf{F}} = \{\mu_{\mathbf{F}}, \Lambda_{\mathbf{F}}\}$, respectively. In addition, Gaussian-Wishart prior distributions (denoted $\mathcal{NW}$ in the following) are placed on such parameters. For a generic parameter set $\Theta = \{\mu, \Lambda\}$, we have

$$\Pr(\Theta|\Theta_0) = \mathcal{N}\left(\mu; \mu_0, [\beta_0 \Lambda]^{-1}\right) \cdot \mathcal{W}\left(\Lambda; \nu_0, \mathbf{W}_0\right)$$

where $\Theta_0 = \{\mu_0, \beta_0, \nu_0, \mathbf{W}_0\}$ is the set of hyperparameters for the prior distribution placed on $\Theta = \{\mu, \Lambda\}$ and $\mathcal{W}\left(\Lambda; \nu_0, \mathbf{W}_0\right)$ is the Wishart distribution.

The overall generative process is graphically represented in Fig. 1, and can be devised as in Fig. 2. Notice that $A_{u,v}$ is a binary random variable and that its value is sampled from a continuous distribution. This is essentially accomplished by choosing the value of $A_{u,v}$ that is nearest to the mean $\mathbf{P}'_u \mathbf{F}_v$ of the distribution. More precisely, the discretization procedure looks at the densities $\Pr(A_{u,v} = 1|\mathbf{P}'_u \mathbf{F}_v, \beta^{-1})$ and $\Pr(A_{u,v} = 0|\mathbf{P}'_u \mathbf{F}_v, \beta^{-1})$ (whose sum differs from 1). Then, $A_{u,v}$ is set to 1 if $\Pr(A_{u,v} = 1|\mathbf{P}'_u \mathbf{F}_v, \beta^{-1}) > \Pr(A_{u,v} = 0|\mathbf{P}'_u \mathbf{F}_v, \beta^{-1})$ or 0 if $\Pr(A_{u,v} = 0|\mathbf{P}'_u \mathbf{F}_v, \beta^{-1}) > \Pr(A_{u,v} = 1|\mathbf{P}'_u \mathbf{F}_v, \beta^{-1})$. To elucidate, $A_{u,v} = 1$ in the case of Fig. 3(a) being closest to $\mathbf{P}'_u \mathbf{F}_v$. Instead, $A_{u,v} = 0$ in the case of Fig. 3(b), since this value is closest to $\mathbf{P}'_u \mathbf{F}_v$.

Predicting u's interest $R^*_{u,i}$ in an unrated item i or a missing trust relationship $A^*_{u,v}$ from u to v in the context of the Bayesian hierarchical model described so far requires, respectively, the predictive distributions $\Pr(R^*_{uj}|R, A, \boldsymbol{\Xi})$ and

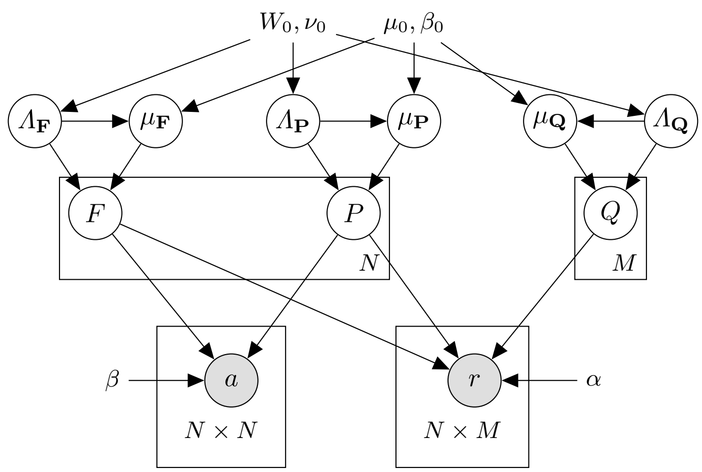


Fig. 1. Graphical representation of the proposed Bayesian hierarchical model.

1. Sample

$$\Theta_{\mathbf{P}} \sim \mathcal{NW}(\Theta_0)$$
$$\Theta_{\mathbf{Q}} \sim \mathcal{NW}(\Theta_0)$$
$$\Theta_{\mathbf{F}} \sim \mathcal{NW}(\Theta_0)$$

2. For each item $i \in \mathcal{I}$ sample

$$\mathbf{Q}_i \sim \mathcal{N}(\mu_{\mathbf{Q}}, \Lambda_{\mathbf{Q}}^{-1})$$

3. For each user $u \in N$ sample

$$\mathbf{P}_u \sim \mathcal{N}(\mu_{\mathbf{P}}, \Lambda_{\mathbf{P}}^{-1})$$
$$\mathbf{F}_u \sim \mathcal{N}(\mu_{\mathbf{F}}, \Lambda_{\mathbf{F}}^{-1})$$

4. For each pair $\langle u, v \rangle \in N \times N$ sample

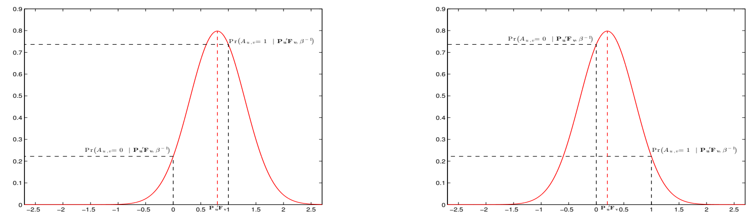
$$A_{u,v} \sim \mathcal{N}(\left(\mathbf{P}'_u \mathbf{F}_v\right), \beta^{-1})$$

5. For each pair $\langle u, i \rangle \in N \times \mathcal{I}$ sample

$$R_{u,i} \sim \mathcal{N}((\mathbf{P}_u + \mathbf{F}_u)\, \mathbf{Q}'_j, \alpha^{-1})$$

Fig. 2. Generative process for the proposed Bayesian hierarchical model.

$\Pr(A^*_{uv} | R, A, \boldsymbol{\Xi})$ relative to the prior $\boldsymbol{\Xi} = \{\Theta_0, \beta, \alpha\}$. Exact inference consists in computing these predictive distributions as reported at Eq. 3.3 and Eq. 3.4, where we set $\boldsymbol{\Theta} = \{\mathbf{P}, \Theta_{\mathbf{P}}, \mathbf{F}, \Theta_{\mathbf{F}}, \mathbf{Q}, \Theta_{\mathbf{Q}}\}$ for readability sake.



(a) $A_{u,v} = 1$ (b) $A_{u,v} = 0$

Fig. 3. The procedure to sample a binary $A_{u,v}$ value from a Gaussian with mean $\mathbf{P}_u'\mathbf{F}_v$

$$\Pr(R^*_{u,i}|R, A, \boldsymbol{\Xi}) = \int \Pr(R^*_{u,i}|\mathbf{P}_u, \mathbf{F}_u, \mathbf{Q}_i, \alpha) \Pr(\boldsymbol{\Theta}|A, R, \boldsymbol{\Xi}) \, \mathrm{d}\boldsymbol{\Theta} \tag{3.3}$$

$$\Pr(A^*_{u,v}|A, R, \boldsymbol{\Xi}) = \int \Pr(A^*_{u,v}|\mathbf{P}_u, \mathbf{F}_v, \beta) \Pr(\boldsymbol{\Theta}|A, R, \boldsymbol{\Xi}) \, \mathrm{d}\boldsymbol{\Theta} \tag{3.4}$$

However, the initial assumption that A only contains positive links introduces a severe bias in the model, as clearly no negative trust can be directly inferred through the posterior $\Pr(\boldsymbol{\Theta}|A, R, \boldsymbol{\Xi})$. If we consider A as an adjacency matrix, the latter generally tends to be extremely sparse. Therefore, only a very small percentage of its entries are labeled as positive, and ambiguity arises in the interpretation of all other entries, since in such cases actual lack of trust and lack of awareness cannot be distinguished. To handle this, we explicitly model awareness through a binary latent variable $Y_{u,v}$ relative to a pair (u, v) such that $u \to v \notin A$. The value $Y_{u,v} = 1$ denotes confidence in the lack of u's trust in v, whereas $Y_{u,v} = 0$ indicates confidence in the fact that u is not aware of v. The matrix of all variables is denoted by $\mathbf{Y}$ in the following.

The latent variables $Y_{u,v}$ for all the pairs $u \to v \notin A$ are drawn from a Bernoulli distribution with parameter $\epsilon_{u,v}$.

$$\Pr(Y_{u,v}) = \epsilon_{u,v}^{Y_{u,v}} \left(1 - \epsilon_{u,v}\right)^{1 - Y_{u,v}} \tag{3.5}$$

Again, we can provide a full Bayesian treatment by placing a Beta prior distribution with hyperparameter $\boldsymbol{\gamma} = \{\gamma_1, \gamma_2\}$ on the individual parameters $\epsilon_{u,v}$:

$$\Pr(\epsilon_{u,v}|\boldsymbol{\gamma}) = \frac{1}{B(\gamma_1, \gamma_2)} \epsilon_{u,v}^{\gamma_1 - 1} \left(1 - \epsilon_{u,v}\right)^{\gamma_2 - 1} \tag{3.6}$$

The adoption of the latent variables $\mathbf{Y}$ allows us to provide an unbiased estimate of the posterior $\Pr(\boldsymbol{\Theta}|A, R, \boldsymbol{\Xi})$ as

$$\Pr(\boldsymbol{\Theta}|A, R, \boldsymbol{\Xi}) = \int \sum_{\mathbf{Y}} \Pr(\boldsymbol{\Theta}, \mathbf{Y}, \boldsymbol{\epsilon}|A, R, \boldsymbol{\Xi}, \gamma) \, \mathrm{d}\boldsymbol{\epsilon}, \tag{3.7}$$

which can be plugged directly into equations 3.3 and 3.4. Also, we can further decompose the posterior as follows:

$$\begin{aligned}\Pr(\boldsymbol{\Theta}, \mathbf{Y}, \boldsymbol{\epsilon} | A, R, \boldsymbol{\Xi}, \gamma) \propto & \Pr(R|\boldsymbol{\Theta}, \alpha) \Pr(A|\boldsymbol{\Theta}, \mathbf{Y}, \beta) \\ & \cdot Pr(\boldsymbol{\Theta}|\Theta_0) \Pr(\mathbf{Y}|\boldsymbol{\epsilon}) \Pr(\boldsymbol{\epsilon}|\gamma)\end{aligned}$$

where finally the term $\Pr(A|\boldsymbol{\Theta}, \mathbf{Y}, \beta)$ can be devised as

$$\Pr(A|\mathbf{P}, \mathbf{F}, \mathbf{Y}, \beta) = \prod_{u \to v \in A} \mathcal{N}(1; \vartheta_{u,v}, \beta^{-1}) \cdot \prod_{u \to v \notin A} \mathcal{N}(0; \vartheta_{u,v}, \beta^{-1})^{Y_{u,v}} \tag{3.8}$$

4 Inference

The exact computation of both Eq. 3.3 and Eq. 3.4 is analytically intractable, because of the complexity of the posterior $\Pr(\boldsymbol{\Theta}, \mathbf{Y}, \boldsymbol{\epsilon}|A, R, \boldsymbol{\Xi}, \gamma)$. Therefore, we resort to *Monte-Carlo* approximation that allows to estimate the predictive distributions by averaging over samples of the model parameters:

$$\Pr(R^*_{u,i}|R, A, \boldsymbol{\Xi}, \boldsymbol{\gamma}) \approx \frac{1}{H} \sum_{h=1}^{H} \Pr(R^*_{u,i}|\mathbf{P}^{(h)}_u, \mathbf{F}^{(h)}_u, \mathbf{Q}^{(h)}_i, \alpha) \tag{4.1}$$

$$\Pr(A^*_{u,v}|A, R, \boldsymbol{\Xi}, \boldsymbol{\gamma}) \approx \frac{1}{H} \sum_{h=1}^{H} \Pr(A^*_{u,v}|\mathbf{P}^{(h)}_u, \mathbf{F}^{(h)}_v, \beta). \tag{4.2}$$

Here, the matrices $\mathbf{P}^{(h)}$, $\mathbf{F}^{(h)}$ and $\mathbf{Q}^{(h)}$ are sampled by running a Markov chain, whose stationary distribution approaches the posterior $\Pr(\boldsymbol{\Theta}, \mathbf{Y}, \boldsymbol{\epsilon}|A, R, \boldsymbol{\Xi}, \gamma)$. In particular, we exploit the Gibbs sampling technique, that provides simple inference algorithms even when the underlying model has a very large number of hidden variables. The Markov chain is built by sequentially considering a variable $\varphi \in \{\mathbf{P}_u, \mathbf{F}_u, \mathbf{Q}_i, Y_{u,v}, \epsilon_{u,v}\}_{u,v \in N, i \in \mathcal{I}}$ and sampling according to the probability $\Pr(\varphi|Rest)$, where *Rest* represents all remaining variables in $\{\mathbf{P}_u, \mathbf{F}_u, \mathbf{Q}_i, Y_{u,v}, \epsilon_{u,v}\}_{u,v \in N, i \in \mathcal{I}}$. Thus, inference in the context of our probabilistic model involves computing the full conditional distributions of the latent variables, which are discussed in the following.

Sampling **P**, **F** *and* **Q**. By exploiting conjugacy, the full conditional of each factor can be expressed as a multivariate gaussian. For example, for **P**, we can observe that

$$\begin{aligned}\Pr(\mathbf{P}_u|Rest) \propto & \Pr(\mathbf{P}_u|\Theta_{\mathbf{P}}) \prod_{i \in \mathcal{I}} \Pr(\mathbf{R}_{u,i}|\mathbf{P}_u, \mathbf{F}_u, \mathbf{Q}_i, \alpha)^{\delta_{u,i}} \\ & \cdot \prod_{v: u \to v \in A} \Pr(1|\mathbf{P}_u, \mathbf{F}_u, \beta) \prod_{v: u \to v \notin A} \Pr(0|\mathbf{P}_u, \mathbf{F}_u, \beta)^{Y_{u,v}},\end{aligned}$$

which results in

$$\mathbf{P}_u \sim \mathcal{N}\left(\mu_P^{*(u)}, \left[\Lambda_P^{*(u)}\right]^{-1}\right)$$

with

$$\Lambda_P^{*(u)} = \Lambda_{\mathbf{P}} + \alpha \sum_{i \in \mathcal{I}} \delta_{u,i} \mathbf{Q}_i \mathbf{Q}_i' + \beta \sum_{v \in N} \tilde{Y}_{u,v} \mathbf{F}_v \mathbf{F}_v'$$

and

$$\mu_P^{*(u)} = \left[\Lambda_P^{*(u)}\right]^{-1} \left[\alpha \sum_{i \in \mathcal{I}} \delta_{u,i} \mathbf{Q}_j R_{u,i} - \alpha \left(\sum_{i \in \mathcal{I}} \delta_{u,i} \mathbf{Q}_i \mathbf{Q}_i'\right) \mathbf{F}_u + \beta \sum_{v: u \to v \in A} \mathbf{F}_v + \Lambda_{\mathbf{P}} \mu_{\mathbf{P}}\right]$$

Here, $\tilde{Y}_{u,v} = 1$ if either $u \to v \in A$ or $Y_{uv} = 1$ (that is to say, $\tilde{Y}_{u,v}$ models awareness of u for v). Similarly, we have

$$\begin{aligned} \mathbf{F}_u \sim & \mathcal{N}\left(\mu_F^{*(u)}, \left[\Lambda_F^{*(u)}\right]^{-1}\right) \\ \mathbf{Q}_i \sim & \mathcal{N}\left(\mu_Q^{*(i)}, \left[\Lambda_Q^{*(i)}\right]^{-1}\right) \end{aligned}$$

where

$$\begin{aligned} \Lambda_F^{*(u)} = & \Lambda_F + \alpha \sum_{i \in \mathcal{I}} \delta_{u,i} \mathbf{Q}_i \mathbf{Q}_i' + \beta \sum_{v \in N} \tilde{Y}_{u,v} \mathbf{P}_v \mathbf{P}_v' \\ \Lambda_Q^{*(i)} = & \Lambda_{\mathbf{Q}} + \alpha \sum_{u \in N} \delta_{u,i} \left(\mathbf{P}_u + \mathbf{F}_u\right) \left(\mathbf{P}_u + \mathbf{F}_u\right)' \end{aligned}$$

and

$$\begin{aligned} \mu_F^{*(u)} &= \left[\Lambda_F^{*(u)}\right]^{-1} \left[\alpha \sum_{i \in \mathcal{I}} \delta_{u,i} \mathbf{Q}_i R_{u,i} - \alpha \left(\sum_{i \in \mathcal{I}} \delta_{u,i} \mathbf{Q}_i \mathbf{Q}_i'\right) \mathbf{P}_u + \beta \sum_{v: u \to v \in A} \mathbf{P}_v + \Lambda_{\mathbf{F}} \mu_{\mathbf{F}}\right] \\ \mu_Q^{*(i)} &= \left[\Lambda_Q]^{*(i)}\right]^{-1} \left[\alpha \sum_{u \in N} \left(\mathbf{P}_u + \mathbf{F}_u\right) \delta_{u,i} R_{u,i} + \Lambda_{\mathbf{Q}} \mu_{\mathbf{Q}}\right] \end{aligned}$$

Sampling $\mathbf{Y}$ *and* $\boldsymbol{\epsilon}$. For each pair (u, v) such that $u \to v \notin A$, we can express the full conditional likelihood as

$$\Pr(Y_{u,v} | \epsilon_{u,v}, A, \mathbf{P}_u, \mathbf{F}_v, \beta) \propto \Pr(0 | \mathbf{P}_u, \mathbf{F}_v, \beta)^{Y_{u,v}} \cdot \Pr(Y_{u,v} | \epsilon_{u,v}).$$

which yields the equation

$$\Pr(Y_{u,v} | \epsilon_{u,v}, A, \mathbf{P}_u, \mathbf{F}_v, \beta) = \frac{\exp\left\{-\beta/2 \left(\mathbf{P}_u' \mathbf{F}_v\right)^2 + \eta_{uv}\right\}}{\exp\left\{-\beta/2 \left(\mathbf{P}_u' \mathbf{F}_v\right)^2 + \eta_u\right\} + 1} \tag{4.3}$$

with $\eta_{uv} = \log \epsilon_{u,v} / (1 - \epsilon_{u,v})$.

The distribution over the individual $\epsilon_{u,v}$ (for each (u, v) such that $u \to v \notin A$) can be obtained by conditioning on their respective Markov blanket. By exploiting conjugacy, we obtain

$$\Pr\left(\epsilon_{u,v} | Y_{u,v}, \gamma\right) = \frac{\gamma_1 + Y_{u,v}}{\gamma_1 + \gamma_2 + 1} \tag{4.4}$$

Sampling $\Theta_{\mathbf{P}}$, $\Theta_{\mathbf{Q}}$ and $\Theta_{\mathbf{F}}$. Again, the conjugacy of the Gaussian-Wishart to the multivariate normal distribution provides a simplification of the full conditional into a Gaussian-Wishart [9, pp. 178]. In general, for a multivariate normal sample $\mathbf{X} \equiv \mathbf{x}_1, \ldots, \mathbf{x}_n$, the posterior $\Pr(\Theta|\mathbf{X}, \Theta_0)$ results into a $\mathcal{NW}(\Theta; \Theta_n)$ where $\Theta_n = \{\mu_n, \beta_n, \nu_n, \mathbf{W}_n\}$ and

$$\mu_n = \frac{\beta_0 \mu_0 + n}{\beta_0 + n}, \qquad \beta_n = \beta_0 + n, \qquad \nu_n = \nu_0 + n$$
$$[\mathbf{W}_n]^{-1} = \mathbf{W}_0^{-1} + \mathbf{S}_{\mathbf{X}} + \frac{\beta_0 n}{\beta_0 + n}(\mu_0 - \overline{\mathbf{x}})(\mu_0 - \overline{\mathbf{x}})'$$

with $\overline{\mathbf{x}} = 1/n \sum_i \mathbf{x}_i$ and $\mathbf{S}_{\mathbf{X}} = \sum_i (\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})'$.

Thus, the posteriors for $\Theta_{\mathbf{P}}$, $\Theta_{\mathbf{F}}$ and $\Theta_{\mathbf{Q}}$ are obtained by updating the respective statistics from which the corresponding hyperparameters depend.

The Gibbs sampling algorithm for approximate inference. Fig. 4 illustrates the Gibbs sampler used to perform approximate inference within the devised model. An execution of the sampler essentially consists in the repetition of a certain number of iterations (lines 3-20). The generic iteration h divides into two stages. The first stage is devoted to sampling hyperparameters $\Theta_{\mathbf{P}}^{(h)}$, $\Theta_{\mathbf{F}}^{(h)}$ and $\Theta_{\mathbf{Q}}^{(h)}$ and $\boldsymbol{\epsilon}_{u,v}$ (lines 4-9). Model parameters $\mathbf{Y}$, $\mathbf{P}_u$, $\mathbf{F}_u$ and $\mathbf{Q}_j$ are then sampled at the second stage (lines 10-19).

Notice that running the Markov chain to its equilibrium through a maximum number of iteration is a widely-adopted convergence-criterion [16]. The overall number of iterations must be carefully set, so that to the probability of transitions of the sampler between latent states converges to a stationary distribution after a preliminary burn-in period. This permits to gather samples drawn after convergence for prediction (as discussed in Sec.4), while discarding burn-in samples which are sensible to the initialization of the sampler.

Also, concerning $\mathbf{Y}$, we do not sample the whole set of pairs (u, v) such that $u \to v \notin A$. This is a crucial efficiency issue. In practice, we are assuming $\mathbf{Y}$ contains several unknown values, and hence only a limited amount of unconnected pairs in a corresponding set U has to be considered. The underlying assumption is that the number $|U|$ of pairs to sample is the result of a prior Poisson process, fixed in the beginning and not reported here for lack of space.

5 Experimental Evaluation

The joint modeling of users' trust networks and ratings provides a powerful framework to detect and understand different patterns within the input social rating network. In this section we analyze the application of the proposed model to real-world social rating networks. More specifically, we are interested in evaluating the effectiveness of our approach in three respects.

- Firstly, we measure its accuracy in rating prediction.
- Secondly, we evaluate the accuracy in predicting trust between pairs of users by measuring the AUC of the proposed model.

Gibbs sampling($\mathcal{N}$, $\Theta_0 = \{\mu_0, \beta_0, \nu_0, \mathbf{W}_0\}$, $\boldsymbol{\gamma}$, α, β)

Sample a subset $U \subseteq N \times N$ such that $u \to v \notin A$;
Initialize $\mathbf{P}^{(0)}$, $\mathbf{F}^{(0)}$, $\mathbf{Q}^{(0)}$, $\mathbf{Y}^{(0)}$;
for $h = 1$ **to** H **do**
  Sample $\Theta_{\mathbf{P}}^{(h)} \sim \mathcal{NW}(\Theta_n)$ where Θ_n is computed by updating Θ_0 with $\overline{\mathbf{P}}$, $\mathbf{S_P}$;
  Sample $\Theta_{\mathbf{F}}^{(h)} \sim \mathcal{NW}(\Theta_n)$ where Θ_n is computed by updating Θ_0 with $\overline{\mathbf{F}}$, $\mathbf{S_F}$;
  Sample $\Theta_{\mathbf{F}}^{(h)} \sim \mathcal{NW}(\Theta_n)$ where Θ_n is computed by updating Θ_0 with $\overline{\mathbf{Q}}$, $\mathbf{S_Q}$
  for each $(u, v) \in U$ **do**
    Sample $\epsilon_{u,v}^{(h)}$ according to Eq. 4.4;
  end for
  for each $(u, v) \in U$ **do**
    Sample $Y_{uv}^{(h)}$ according to Eq. 4.3;
  end for
  for each $u \in N$ **do**
    Sample $\mathbf{P}_u \sim \mathcal{N}\left(\mu_P^{*(u)}, \left[\Lambda_P^{*(u)}\right]^{-1}\right)$;
    Sample $\mathbf{F}_u \sim \mathcal{N}\left(\mu_F^{*(u)}, \left[\Lambda_F^{*(u)}\right]^{-1}\right)$;
  end for
  for each $i \in \mathcal{I}$ **do**
    Sample $\mathbf{Q}_i \sim \mathcal{N}\left(\mu_Q^{*(i)}, \left[\Lambda_Q^{*(i)}\right]^{-1}\right)$;
  end for
end for

Fig. 4. The scheme of Gibbs sampling algorithm in pseudo code

	Ciao	*Epinions*
Users	7,375	49,289
Trust Relationships	111,781	487,181
Items	106,797	139,738
Ratings	282,618	664,823
InDegree (Avg/Median/Min/Max)	15.16/6/1/100	9.8/2/1/2589
OutDegree (Avg/Median/Min/Max)	16.46/4/1/804	14.35/3/1/1760
Ratings on items (Avg/Median/Min/Max)	2.68/1/1/915	4.75/1/1/2026
Ratings by Users (Avg/Median/Min/Max)	38.32/18/4/1543	16.55/6/1/1023

Table 1. Summary of the chosen social rating networks.

– Thirdly, we analyze the structure of the model and investigate the properties that can be derived, such as relationships among factors and propensities of users within given factors.

Datasets. We conducted experiments on two datasets representing social rating networks from the popular product review sites *Epinions* and *Ciao*, described in [29]. Users in these sites can share their reviews about products. Also they can establish their trust networks from which they may seek advice to make decisions. Both sites employ a 5-star rating system. Some statistics of the datasets are shown in Table 1 and in Fig. 5. We can notice that both the trust relationships and the rating distributions are heavy-tailed. *Epinions* exhibits a larger number of users, as well as a larger sparsity coefficient on A.

Evaluation setting. We chose some state-of-the-art baselines from the current literature. For rating prediction, we compared our approach against *SocialMF* [14]. The metric used here is the standard RMSE. We exploited the implementation of *SocialMF* made available at http://mymedialite.net. For trust prediction, we

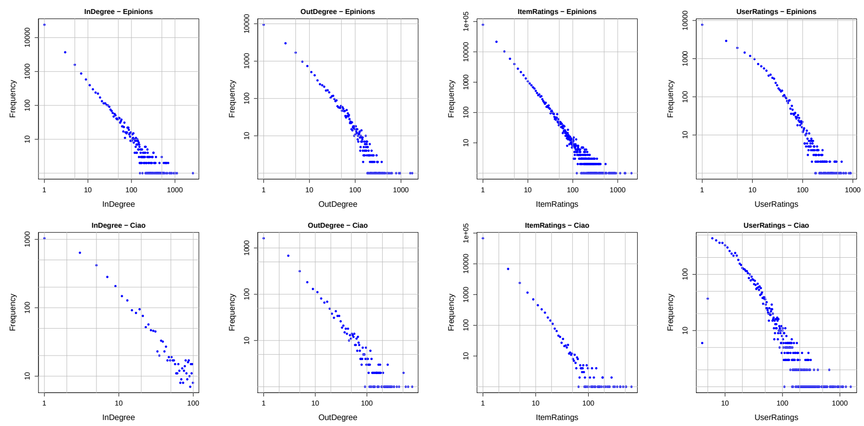


Fig. 5. Distributions of trust relationships and ratings in *Epinions* and *Ciao*.

adapted the framework described in [20]. For each user, we considered the ratings as user features and we trained the factorization model which minimizes the AUC loss. We exploited the implementation made available by the authors at http://cseweb.ucsd.edu/ akmenon/code. We refer to this method as *AUC-MF* in the following. In addition, we considered a further comparison in terms of both RMSE and AUC against a basic matrix factorization approach based on SVD named *Joint SVD (JSVD)* [11]. We computed a low-rank factorization of the joint adjacency/feature matrix $\mathbf{X} = [A\ R]$ as $\mathbf{X} \approx \mathbf{U} \cdot diag(\sigma_1, \ldots, \sigma_K) \cdot \mathbf{V}^T$, where K is the rank of the decomposition and $\sigma_1, \ldots, \sigma_K$ are the square roots of the K greatest eigenvalues of $\mathbf{X}^T\mathbf{X}$. The matrices $\mathbf{U}$ and $\mathbf{V}$ resemble the roles of $\mathbf{P}$, $\mathbf{F}$ and $\mathbf{Q}$: The term $U_{u,k}$ can be interpreted as the tendency of u to trust users, relative to factor k. Analogously, $V_{u,k}$ represents the tendency of u to be trusted, and $V_{i,k}$ represents the rating tendency of item i in k. The score can be hence computed as [26] $score(u,x) = \sum_{k=1}^{K} U_{u,k}\sigma_k V_{x,k}$, where x denotes either a user v or an item i.

In all the experiments, we performed a Monte-Carlo Cross Validation, by performing 5 training/test splits. Within the partitions, 70% of the data were retained as training, and the remaining 30% as test. The splitting was accomplished for the sole data upon which to measure the performance (i.e., ratings for the RMSE and links for the AUC).

Concerning the AUC, it is worth noticing that *Epinions* and *Ciao* only contain positive trust relationships, and the computation of the AUC relies on the presence of negative values. Negative values are indeed crucial in the approach [20], since the latter relies on a loss function which penalizes situations where the score of negative links is higher than the score of positive links. In principle, we can consider all links in the test-set as positive examples, and all non-existing links as negative example. However, the sparsity of the networks poses a major tractability issue, as it would make the computation of the AUC infeasible. A better estimation strategy in [2, 26] consists in narrowing the negative examples to all the 2-hops non-existing links, i.e., all triplets (u, v, w) where both (u, v) and (v, w) exhibit a trust relationship in A, but (u, w) does not.

Results. Fig. 6 reports the averaged results of the evaluation. We ran the experiments on a variable number of latent factors, ranging from 4 to 128. We can notice that the proposed hierarchical model, denoted as *HBPMF*, achieves the minimum RMSE on both datasets. There is a tendency of the RMSE to progressively decrease. However, this tendency is more evident on *SocialMF*, while the other two methods exhibit negligible differences.

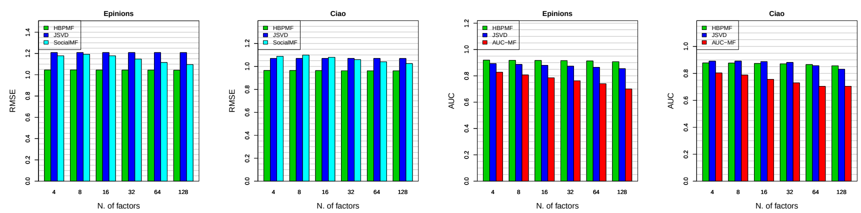


Fig. 6. Prediction results.

The opposite trend is observed in trust prediction. Here, all methods tend to prefer a low number of factors, as the best results are achieved with $K = 4$. The devised *HBPMF* model achieves the maximum AUC on the *Epinions* dataset, and results comparable to *JSVD* on *Ciao*. The detailed results are shown in Fig. 7, where the ROC curves are reported. In general, the predictive accuracy of the Bayesian hierarchical model is stable with regards to the number of factors. This is a direct result of the Bayesian modeling, which makes the model robust to the growth of the model complexity. Fig. 8 also shows how the accuracy varies according to the distributions which characterize the data. We can notice a correlation between accuracy and node degrees, as well as the number of ratings provided by a user or received by an item.

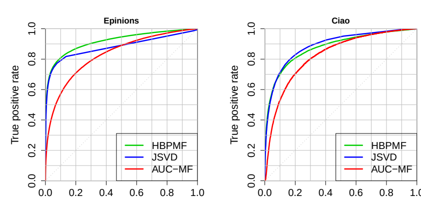


Fig. 7. ROC curves on trust prediction for $K = 4$.

To evaluate the effects of the joint modeling of both the trust relationships and the ratings, we conducted some further experiments with $K = 4$. In a first experiment, we performed the sampling without considering the trust relationships.

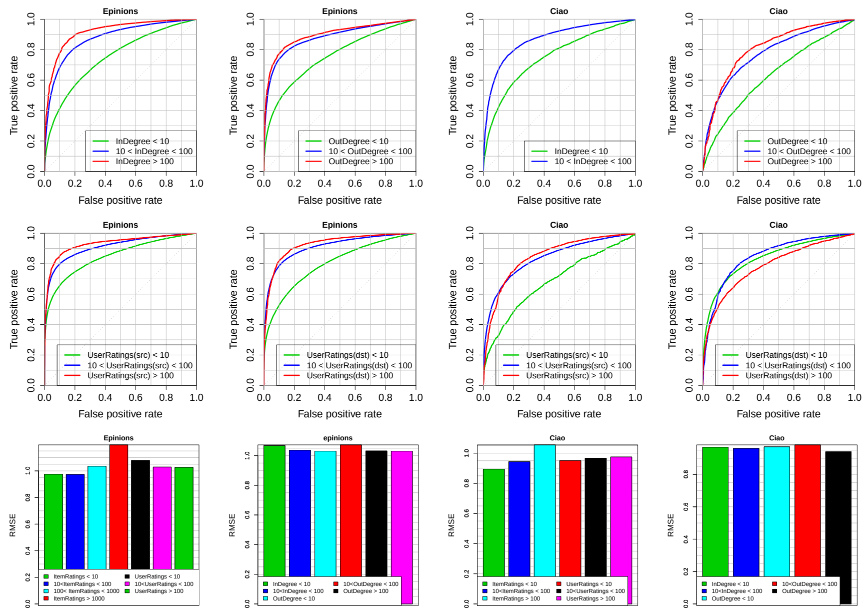


Fig. 8. Data distribution vs. AUC and rating prediction.

More precisely, we performed a simple *BPMF* (as described in [25]). Dually, we discarded the rating matrix and performed the sampling by only considering the trust relationships. The first graph of Fig. 9 shows the comparison between the results of these partial models against those achieved through the full *HBPMF* model. The effects of the joint modeling can be appreciated on the RMSE: in practice, the additional information provided by the trust relationships refines the modeling of the data, thus lowering the RMSE. By contrast, the effects of the joint modeling on the AUC do not highlight substantial improvements.

Finally, the last two graphs of Fig. 9 report the running times relative to the methods. For the *HBPMF*, we achieved stable results for the RMSE after 100 iterations, whereas the AUC result was stable after 20 iterations. Both *SocialMF* and *AUC-MF* exhibited stable results with 20 iterations. The computational overhead of the Gibbs Sampling procedure plays a crucial role here. Therein, it would be interesting to investigate alternative inference strategies based on variational approximation, which are known to guarantee fast convergence.

6 Conclusions and Future Research

We presented the first unified approach to the recommendation of interesting items and trustworthy users in social rating networks with trust relationships. The key intuition is that the interactions from users to users as well as between users and items are explained by the same latent factors, which ultimately allows to combine user and item recommendation into a simple and intuitive Bayesian

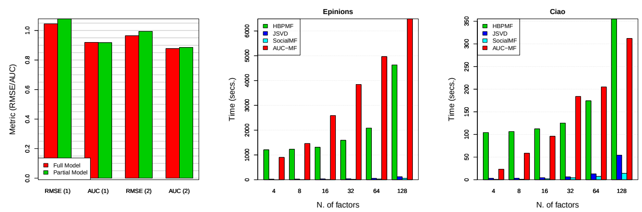


Fig. 9. (a) Effects of the joint modeling. (1 denotes *Epinions*, and 2 denotes *Ciao*). (b) Average running time for iteration (JSVD reports the total time).

generative model. A comparative experimentation over real-world social rating networks confirmed such an intuition: the devised model was shown to deliver a superior predictive performance in terms of both RMSE and AUC.

Future research will focus on two major directions. We planned to study an extension of our model in which the Indian Buffet Process [12] is exploited to automatically infer the most appropriate number of latent factors from the input social rating network. In addition, variational approximate inference and related learning algorithms will be studied to improve the computational efficiency. Finally, a further line of research is relative to how the proposed models can be adapted to support recommendation tasks behind rating prediction [24, 5, 4].

References

1. E.M. Airoldi, D.M. Blei, S.E. Fienberg, and E.P. Xing. Mixed membership stochastic blockmodels. *The Journal of Machine Learning Research*, 9:1981 – 2014, 2008.
2. L. Backstrom and J. Leskovec. Supervised random walks: Predicting and recommending links in social networks. In *Proc. ACM WSDM Conf.*, pages 635 – 644, 2011.
3. N. Barbieri, F. Bonchi, and G. Manco. Cascade-based community detection. In *Proc. of ACM WSDM Conf.*, pages 33–42, 2013.
4. N. Barbieri and G. Manco. An analysis of probabilistic methods for top-n recommendation in collaborative filtering. In *Proc. ECML/PKDD Conf.*, pages 172–187, 2011.
5. N. Barbieri, G. Manco, R. Ortale, and E. Ritacco. Balancing prediction and recommendation accuracy: Hierarchical latent factors for preference data. In *Proc. of SIAM Int. Conf. on Data Mining*, pages 1035 – 1046, 2012.
6. G. Costa and R. Ortale. A bayesian hierarchical approach for exploratory analysis of communities and roles in social networks. In *Proc. of the IEEE/ACM ASONAM Conf.*, pages 194 – 201, 2012.
7. G. Costa and R. Ortale. A Unified Generative Bayesian Model for Community Discovery and Role Assignment based upon Latent Interaction Factors. In *Proc. of the IEEE/ACM ASONAM Conf.*, To appear, 2014.
8. G. Costa and R. Ortale. Probabilistic analysis of communities and inner roles in networks: Bayesian generative models and approximate inference. *Social Network Analysis and Mining*, 3(4):1015 – 1038, 2013.
9. M. DeGroot. *Optimal Statistical Decisions.* McGraw-Hill, 1970.

10. J. Delporte, A. Karatzoglou, T. Matuszczyk, and S. Canu. Socially enabled preference learning from implicit feedback data. In *Proc. of Machine Learning and Knowledge Discovery in Databases*, pages 145 – 160, 2013.
11. N. Z. Gong et al. Joint link prediction and attribute inference using a social-attribute network. *ACM TIST*, 5(2), 2014.
12. T.L. Griffiths and Z. Ghahramani. The indian buffet process: An introduction and review. *The Journal of Machine Learning Research*, 12:1185 – 1224, 2011.
13. T. Hofman, J. Puzicha, and M.I. Jordan. Learning from dyadic data. In *Proc. NIPS Conf.*, pages 466 – 472, 1999.
14. M. Jamali and M. Ester. A matrix factorization technique with trust propagation for recommendation in social networks. In *Proc. of ACM RECSYS Conf.*, pages 135 – 142, 2010.
15. D. Liben-Nowell and J. Kleinberg. The link-prediction problem for social networks. *Journal of the American Society for Information Science and Technology*, 58(7):1019–1031, 2007.
16. J.S. Liu. *Monte Carlo Strategies in Scientific Computing.* Springer, 2001.
17. H. Ma, I. King, and M.R. Lyu. Learning to recommend with social trust ensemble. In *Proc. of Int. ACM SIGIR Conf.*, pages 203 – 210, 2009.
18. H. Ma, H. Yang, M.R. Lyu, and I. King. Sorec: Social recommendation using probabilistic matrix factorization. In *Proc. of ACM CIKM Conf.*, pages 931 – 940, 2008.
19. H. Ma, D. Zhou, C. Liu, M.R. Lyu, and I. King. Recommender systems with social regularization. In *Proc. of ACM WSDM Conf.*, pages 287 – 296, 2011.
20. A.K. Menon and C. Elkan. Link prediction via matrix factorization. In *Proc. ECML/PKDD Conf.*, pages 437 – 452, 2011.
21. K.T. Miller, T.L. Griffiths, and M.I. Jordan. Nonparametric latent feature models for link prediction. In *Proc. NIPS Conf.*, pages 1276 – 1284, 2009.
22. R. Pan, Y. Zhou, B. Cao, N.N. Liu, R.M. Lukose, M. Scholz, and Q. Yang. One-class collaborative filtering. In *Proc. IEEE ICDM Conf.*, pages 502 – 511, 2008.
23. S. Purushotham, Y. Liu, and C. c.J. Kuo. Collaborative topic regression with social matrix factorization for recommendation systems. In *Proc. ICML Conf.*, pages 759 – 766, 2012.
24. S. Rendle, Christoph F., Zeno G., and Lars S. Bpr: Bayesian personalized ranking from implicit feedback. In *Proc. UAI Conf.*, 2009.
25. R. Salakhutdinov and A. Mnih. Bayesian probabilistic matrix factorization using markov chain monte carlo. In *Proc. ICML Conf.*, pages 880 – 887, 2008.
26. Badrul M. Sarwar et al. Application of dimensionality reduction in recommender system – a case study. In *In ACM WEBKDD Workshop*, 2000.
27. Y. Shen and R. Jin. Learning personal+social latent factor model for social recommendation. In *Proc. of ACM SIGKDD Conf.*, pages 1303 – 1311, 2012.
28. V. Sindhwani, S.S. Bucak, J. Hu, and A. Mojsilovic. One-class matrix completion with low-density factorizations. In *Proc. of IEEE ICDM Conf.* , pages 1055 – 1060, 2010.
29. J. Tang, H. Gao, and H. Liu. mtrust: Discerning multi-faceted trust in a connected world. In *Proc. ACM WSDM Conf.*, pages 93–102, 2012.
30. S.-H. Yang, and others. Like like alike: Joint friendship and interest propagation in social networks. In *Proc. WWW Conf.* , pages 537 – 546, 2011.
31. X. Yang, H. Steck, and Y. Liu. Circle-based recommendation in online social networks. In *Proc. ACM SIGKDD Conf.* , pages 1267 – 1275, 2012.
32. J. Zhu. Max-margin nonparametric latent feature models for link prediction. In *Proc. NIPS Conf.*, pages 719 – 726, 2012.