Preference-oriented Social Networks: Group Recommendation and Inference

Amirali Salehi-Abari
Department of Computer Science
University of Toronto
abari@cs.toronto.edu

Craig Boutilier*
Department of Computer Science
University of Toronto
cebly@cs.toronto.edu

ABSTRACT

Social networks facilitate a variety of social, economic, and political interactions. Homophily and social influence suggest that preferences (e.g., over products, services, political parties) are likely to be correlated among people whom directly interact in a social network. We develop a model, *preference-oriented social networks*, that captures such correlations of individual preferences, where preferences take the form of rankings over a set of options. We develop probabilistic inference methods for predicting individual preferences given observed social connections and partial observations of the preferences of others in the network. We exploit these predictions in a *social choice* context to make group decisions or recommendations even when the preferences of some group members are unobserved. Experiments demonstrate the effectiveness of our algorithms and the improvements made possible by accounting for social ties.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*Information filtering, Retrieval models, Selection process*

General Terms

Algorithms, Experimentation, Measurement

Keywords

Group recommendation; social networks; preferences; probabilistic models; probabilistic inference

1. INTRODUCTION

Social networks play a crucial role in many social and economic interactions [17], including discovery of job opportunities [15], the products we consume [13], or even weight gain [10]. Due to such factors, it is widely recognized that individuals' behaviors and preferences are correlated with those of their friends or connections (e.g., music tastes [21]).

Because of this, and increasing availability of user preference and behavioral data, it is essential to study the interplay of social network structure and individual behaviour, attitudes, and preferences. This has lead to research focused on inferring individual attributes and behaviour using social connections, e.g., inference of ratings over items [23, 24, 18], latent group membership [19], or latent "social positions" [16]. Yet surprisingly, using social networks to infer individual preferences—in the form of rankings of alternatives—has received little attention. Methods for inference and learning of preference rankings are studied in econometrics, psychometrics, statistics, and machine learning and data mining; in the latter case, they find application to recommender systems, information retrieval and group decision/recommendation problems (i.e., social choice), especially when faced with partial information. In contrast to cardinal utilities, preference rankings (or ordinal preferences) are of special interest in social choice and group recommendation, since they help circumvent, to some extent, the problem of interpersonal comparisons of utilities [2, 36].

In this work, we address how to use *social network structure* to support more accurate inference of preference rankings and to make group decisions when some individual preferences are unknown. Specifically, we exploit the fact that *homophily* or *social selection*—association with similar individuals—and *social influence*—adoption of properties and attitudes of those to whom one is connected—can be used to infer individual preferences more efficiently and with less data. This can in turn support more accurate group decision making with partial preferences.

To capture correlations of preference rankings over social networks, we introduce *preference-oriented social networks (POSNs)*, a generative model in which the similarity of the preference rankings of two individuals determines the odds with which they are connected. We exploit this model to infer unobserved individual preferences given observed preferences of others in the network. Intuitively, if we know something about the preferences of an individual's friends, family or colleagues—or their friends, etc.—we should be able to more accurately predict their preference ranking if homophily or social influence shapes network dynamics. Moreover, we demonstrate how network structure, by allowing such predictions, can be used to support effective group recommendations/decisions with incomplete preferences.

*Currently on leave at Google, Inc. Mountain View.

RecSys'15, September 16–20, 2015, Vienna, Austria.

DOI: http://dx.doi.org/10.1145/2792838.2800190.

2. PREFERENCE-ORIENTED NETWORKS

We start by outlining our basic model (we contrast it with existing network generation models in Sec. 3). A *preference-oriented social network (POSN)* consists of: (i) a social network, where nodes represent individuals, and edges represent some social relationship; and (ii) a finite set of options over which individuals have preferences, where these preferences take the form of an ordering or ranking. The model also includes a probabilistic generative process used to generate individual preferences and connections that induce correlated preferences.

The network in a POSN is an undirected graph $G = (\mathcal{N}, E)$ over individuals $\mathcal{N} = \{1, \ldots, n\}$. We use a binary adjacency matrix $[e_{ij}]$ where $e_{ij} = 1$ iff $(i,j) \in E$. We assume a finite set of *alternatives (or options)* $\mathcal{A} = \{a_1, \ldots, a_m\}$, e.g., a set of products, political candidates, policies, genre of movies, etc. over which individuals have preferences. The preference of node i is a *ranking* (or strict total order) r_i over $\mathcal{A}$. Let $\Omega(\mathcal{A})$ denote the set of all $m!$ rankings over $\mathcal{A}$.

The generative process for POSNs has two stages: first, individual preferences are drawn from a ranking distribution; then individuals form connections with a probability increasing with the *similarity* of their preferences. Each node i's preference ranking r_i is drawn independently from some distribution $\rho(r|\boldsymbol{\eta})$ over $\Omega(\mathcal{A})$ with parameters $\boldsymbol{\eta}$. Many ranking distributions can be used, e.g., Plackett-Luce, Bradley-Terry, etc. [25]. Here, we focus on the *Mallows ϕ-model*, characterized by a "modal" *reference ranking* σ and a *dispersion parameter* $\phi \in [0, 1)$, with the probability of a ranking r decreasing exponentially with its *τ-distance* from σ:

$$\rho(r|\sigma, \phi) = \frac{1}{z(\phi)} \phi^{d_\tau(r,\sigma)} , \tag{1}$$

where $d_\tau(r, \sigma)$ is Kendall's τ distance between r and σ (see below) and $z(\phi)$ is a normalization constant.

To compute connection probabilities, we define the similarity of two rankings using the τ metric, frequently used in psychometrics and social choice:

$$d_\tau(r_i, r_j) = \sum_{k \neq l} I[r_i(a_k) > r_i(a_l) \text{ and } r_j(a_k) < r_j(a_l)],$$

Intuitively, $d_\tau(r_i, r_j)$ measures the number of pairwise swaps needed to transform r_i to r_j.[1] A strictly decreasing *connection probability function* $c(d) : [0, \infty) \to [0, 1]$ specifies the probability that two nodes i, j are connected given the distance $d_\tau(r_i, r_j)$ between their corresponding rankings. We use the following connection function [38]:

$$c(d|\boldsymbol{\lambda}) = \gamma \left(1 + \frac{d}{\beta}\right)^{-\alpha} . \tag{2}$$

Here β controls average node degree and $\alpha > 1$ determines the extent of homophily (greater α implies more homophily). We use $\gamma \in (0, 1]$ to control the odds of connecting nodes with the same ranking (accounting for the discrete nature of ranking space). We sometimes write the connection probability as $c(r_i, r_j)$. Denote the parameters of c by $\boldsymbol{\lambda} = (\alpha, \beta, \gamma)$; the parameters of the ranking distribution by $\boldsymbol{\eta} = (\sigma, \phi)$; and all POSN parameters by $\boldsymbol{\theta} = (\boldsymbol{\lambda}, \boldsymbol{\eta})$. Fig.1 illustrates a small POSN, where individuals have preferences over three options; nodes with similar preferences are more

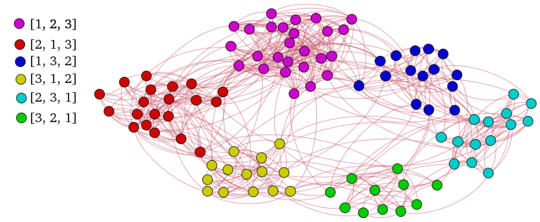

Figure 1: A POSN ($\alpha\!:\!2; \gamma\!:\!0.5; \beta\!:\!0.5; m\!:\!3; n\!:\!100; \phi\!:\!0.7$).

densely connected. Our POSN model is an instance of a more general notion of a *ranking network*, a latent space network model (see Sec. 3), in which latent attributes are generic rankings over options. We analyze general topological properties of this model in [33]; here we focus directly on inference and group recommendation.

[1]Other distance metrics for rankings can be used as well, e.g., Spearman's rho or footrule, or Hamming distance.

3. INFERENCE AND SOCIAL CHOICE

We now address two tightly connected problems, *preference inference* and *single-option group recommendation* (or consensus decision making). While preference inference is interesting in its own right, it plays a vital role in group recommendation when preferences of some group members are unobserved.

3.1 Preference Inference

We assume that individuals are partitioned into two sets: $O \subseteq \mathcal{N}$, whose complete preference rankings are observed (e.g., elicited or otherwise revealed); and $U = \mathcal{N} \setminus O$, whose preferences are unknown or "missing." Let $R^O = \{r_i | i \in O\}$ be the set of observed rankings and $R^U = \{r_i | i \in U\}$ be the set of random variables associated with unknown preferences. In this work, we assume that the network G and model parameters $\boldsymbol{\theta}$ are known. Learning model parameters given observed preferences is an important problem (and the subject of ongoing research); but learning can exploit our solution to the inference problem (e.g., when using EM).

Our goal in preference inference is to compute the posterior distribution over unobserved preferences $\Pr(R^U|G, R^O, \boldsymbol{\theta})$ given observed preferences R^O. We discuss sampling methods for approximating the posterior distribution in Sec. 5. Other inference problems include the *most probable explanation (MPE)*, i.e., finding the instantiation of R^U which maximizes the posterior:

$$R^{MPE} = \arg\max_{R^U} \Pr(R^U|G, R^O, \boldsymbol{\theta}).$$

We may also be interested in the posterior over the preferences of a single individual $i \in U$:

$$\Pr(r_i|G, R^O, \boldsymbol{\theta}) = \sum_{R^U \setminus \{r_i\}} \Pr(R^U|G, R^O, \boldsymbol{\theta})$$

and as well as the "individual MPE:"

$$r_i^{MPE} = \arg\max_{r \in \Omega(\mathcal{A})} \Pr(r_i = r|G, R^O, \boldsymbol{\theta}).$$

3.2 Group Recommendation

A key goal in this work is to exploit social network structure to make higher quality group decisions with incomplete preference information. Suppose we need to select an option from $\mathcal{A}$ for a group or "subpopulation" $\mathcal{S} \subseteq \mathcal{N}$ using some preference aggregation method (i.e., a *social choice function*, for example, a voting rule). We distinguish the subsegment

$\mathcal{S}$ (e.g., friends planning an activity, the electorate in a small district) from the larger society $\mathcal{N}$ (e.g., users of an online social network, eligible voters in a country): while many group decisions are local, they can be supported by knowledge of the preferences of individuals outside that group. We focus on the choice of a single option with an emphasis on "social welfare maximization" relative to a *scoring rule* $g : (\mathbb{N}, \mathbb{N}) \to \mathbb{R}^+$ where $g(k, m)$ is the positional score of an option ranked k^{th} relative to m options (the Borda and plurality score are common examples, we define Borda in Sec. 6). Define the *social welfare* of $a \in \mathcal{A}$:

$$sw(a, \mathcal{S}) = \sum_{i \in \mathcal{S}} g(r_i(a), m),$$

with the goal of selecting $a^* \in \mathcal{A}$ that maximizes $sw(., \mathcal{S})$.

In general, we will not know the preferences of all individuals in $\mathcal{S}$, requiring that we *infer* the social welfare of an option a given the observations at hand. Define $sw(a, \mathcal{S}|R^O, G, \boldsymbol{\theta})$ to be this *inferred social welfare*, which varies depending on the method of inference (we sometimes omit mention of G and $\boldsymbol{\theta}$). Assuming each individual's contribution to social welfare is independent, it can be decomposed into a *revealed component* $sw_{rev}(a, R^O_{\mathcal{S}})$ (corresponding to observed preferences) and an *inferred component* $sw_{inf}(a, R^U_{\mathcal{S}}|R^O)$ (for unobserved preferences):

$$sw(a, \mathcal{S}|R^O) = sw_{rev}(a, R^O_{\mathcal{S}}) + sw_{inf}(a, R^U_{\mathcal{S}}|R^O).$$

The revealed component is straightforward:

$$sw_{rev}(a, R^O_{\mathcal{S}}) = \sum_{r \in R^O_{\mathcal{S}}} g(r(a), m).$$

But there are various ways to define the inferred component.

Expected Score (ES). The most natural, principled way to define inferred component is the *expected score*:

$$sw^E_{inf}\left(a, R^U_{\mathcal{S}}|G, R^O\right) = \sum_{R^U_{\mathcal{S}}} \Pr(R^U_{\mathcal{S}}|G, R^O) \sum_{i \in U_{\mathcal{S}}} g(r_i(a), m),$$

which can be computed in $O(m!^{|U_{\mathcal{S}}|}|U_{\mathcal{S}}|)$ time (where $U_{\mathcal{S}}$ are those individuals with unobserved preferences). If $\Pr(r_i|G, R^O)$ is pre-computed for each $i \in U_{\mathcal{S}}$, we can write

$$sw^E_{inf}\left(a, R^U_{\mathcal{S}}|G, R^O\right) = \sum_{i \in U_{\mathcal{S}}} \sum_{r_i} \Pr(r_i|G, R^O) g(r_i(a), m),$$

which can be computed in $O(m!|U_{\mathcal{S}}|)$ time.

Joint Most Probable Explanation Score (JMPES). This uses the unobserved preferences, R^{MPE}, which maximizes the joint posterior $\Pr(R^U|G, R^O, \boldsymbol{\theta})$:

$$sw^{JM}_{inf}(a, R^{MPE}_{\mathcal{S}}) = \sum_{r \in R^{MPE}_{\mathcal{S}}} g(r(a), m).$$

It can be computed in $O(|U_{\mathcal{S}}|)$ time if R^{MPE} is given.

Individual Most Probable Explanation Score (IMPES). This uses the instantiation r_j^{MPE} for each $j \in U$ that maximizes the posterior $\Pr(r_j|G, R^O, \boldsymbol{\theta})$:

$$sw^{IM}_{inf}(a, \{r_j^{MPE}\}) = \sum_{j \in U_{\mathcal{S}}} g(r_j^{MPE}(a), m).$$

It is computable in $O(|U_{\mathcal{S}}|)$ time if the r_j^{MPE} are given.

3.3 Related Work and Models

We review the related work on group recommendation, network formation models, nodal attribute inference, preference ranking learning, collaborative filtering methods using social networks, and decision making on social networks.

Group Recommendation. Group recommendation can be broadly categorized as follows: (i) *Virtual/artificial profile methods* (see, e.g., [29]), where joint artificial user profiles for each group of users are created to keep track of their joint revealed/elicited preferences; (ii) *Profile-merging methods* (see, e.g., [42, 5]), which merge group member profiles to form a group profile, based on which recommendations are made; (iii) *Recommendation/scoring aggregation* methods (see, e.g., [27, 3, 1, 35, 12]), which aggregate the recommendations (or inferred preferences) for each group member into single group recommendation list (or recommended option). This aggregation is usually conducted by a *group consensus function* (or social choice function). Our method falls into this third category.

Network Formation Models. Our POSN model lies in the class of random, static network formation models [31]. It is also a *spatial (or latent space) networks* [16, 4], where nodes possess *latent attributes* and are connected with odds determined by these attributes. The *Waxman model* [39] distributes nodes uniformly at random on the plane with node connection probabilities decreasing exponentially with Euclidean distance. Hoff *et al.* [16] develop a similar model where nodes are points in a d-dimensional "social space". The *hidden variable model* [7] generalizes the Waxman model, giving nodes a hidden (real-valued or integer) random attribute drawn independently from a specified distribution.

Nodal Attribute Inference. Inference of nodal attributes, given social network structure, has also received attention. Hoff *et al.* [16] develop inference and learning methods for spatial models. Kim and Leskovec [19] propose a *variational EM* method for learning model parameters given network structure and node attributes, and for inferring latent attributes. Other related work includes *collective classification* [37] and active learning over networks [6].

Learning Preference Rankings. Distributional models of rankings are widely studied in statistics, psychometrics and machine learning, though accounting for social network structure has been unaddressed. EM has been used to learn model parameters of mixtures of distance-based ranking models given completely or partially observed individual rankings [30, 9, 22]. Our model is distinct from those above as it models preference correlations induced by social ties, requiring new sampling and inference methods.

Collaborative Filtering and Social Networks. Collaborative Filtering (CF) methods which exploit social networks for rating prediction have recently become popular (see for example, [23, 18, 24, 41, 14, 26, 20] and [40] for a recent survey). These methods—along with traditional CF methods—fall into two broad categories, *memory-based* [14, 26, 20] and *model-based* [23, 18, 24, 41] approaches. In memory-based approaches, the social network structure is usually taken into account when computing the pairwise similarity scores (or trust values) between users [14, 26]. These scores are then used for prediction of missing ratings. Model-based approaches focus largely on latent space probabilistic models in which users and items are embedded in a low-dimensional latent feature space, and ratings are generated by combining these feature vectors while accounting for social network structure. Our model differs in that it considers preference ranking correlations rather than ratings correlations over social networks, and in its focus on group rather than individual recommendation.

Group Recommendation using Social Factors. Group recommendations based on social factors or interaction patterns have recently drawn a fair amount of attention. Masthoff and Gatt [28] analyse the effect of group member relationship types on their emotional conformity and contagion in a group recommendation task. Social relationship strength has been considered in a group collaborative filtering context [32]. Salehi-Abari and Boutilier [34] study *empathetic social choice* in social networks, in which individuals derive benefit based on both their own intrinsic preferences and empathetic preferences, the latter determined by the satisfaction of their neighbors.

4. TARGET DISTRIBUTIONS

We describe the form and structure of the joint distribution induced by POSNs. Assuming that the preference distribution parameters $\boldsymbol{\eta}$ are given, the joint over R^U is:

$$\Pr(R^U|\boldsymbol{\eta}) = \prod_{r \in R^U} \rho(r|\boldsymbol{\eta}), \tag{3}$$

where $\rho(r|\boldsymbol{\eta})$ is the preference distribution. To specify the distribution over G, given $\boldsymbol{\theta}$, we first focus on the probability $\Pr(e_{ij} = 1)$ with which an edge occurs between two nodes i and j in G. We define it under three conditions: (1) the preferences of both i and j are unobserved (and drawn independently from $\rho(r|\boldsymbol{\eta})$); (2) one is observed and the other unobserved; and (3) both are observed.

Unobserved preferences for both nodes. In this case, $\Pr(e_{ij} = 1|\boldsymbol{\theta})$ is the chance of an edge between two nodes whose preferences are drawn independently from $\rho(r|\boldsymbol{\eta})$:

$$\mathcal{E}(\boldsymbol{\theta}) = \sum_{r \in \Omega(\mathcal{A})} \sum_{r' \in \Omega(\mathcal{A})} \rho(r|\boldsymbol{\eta})\rho(r'|\boldsymbol{\eta})c\left(d_\tau(r', r)|\boldsymbol{\lambda}\right), \tag{4}$$

(The expected number of edges in a POSN is $\binom{n}{2}\mathcal{E}(\boldsymbol{\theta})$.)

Unobserved preference for one node. When only one node's preference is observed (say i) $\Pr(e_{ij} = 1|r_i, \boldsymbol{\theta})$ is:

$$\mathcal{D}(r, \boldsymbol{\theta}) = \sum_{r' \in \Omega(\mathcal{A})} \rho(r'|\boldsymbol{\eta})c\left(d_\tau(r', r)|\boldsymbol{\lambda}\right). \tag{5}$$

$\mathcal{D}(r, \boldsymbol{\theta})$ also determines the expected degree of a node with ranking r, which is simply $(n-1)\mathcal{D}(r, \boldsymbol{\theta})$.

Observed preferences for both nodes. The edge probability between i and j when both r_i and r_j are observed is $\Pr(e_{ij} = 1|r_i, r_j, \boldsymbol{\theta}) = c(d_\tau(r_i, r_j)|\boldsymbol{\lambda})$.

Using these edge probabilities, the probability $\Pr(G|R^O, \boldsymbol{\theta})$ of graph structure G given observed preferences R^O is:

$$\begin{aligned}\Pr(G|R^O, \boldsymbol{\theta}) = &\prod_{i,j \in U;\, i<j} \mathcal{E}(\boldsymbol{\theta})^{e_{ij}} \left(1 - \mathcal{E}(\boldsymbol{\theta})\right)^{1-e_{ij}} \times \\ &\prod_{i \in O, j \in U} \mathcal{D}(r_i, \boldsymbol{\theta})^{e_{ij}} \left(1 - \mathcal{D}(r_i, \boldsymbol{\theta})\right)^{1-e_{ij}} \times \\ &\prod_{i,j \in O;\, i<j} c(r_i, r_j|\boldsymbol{\lambda})^{e_{ij}} \left(1 - c(r_i, r_j|\boldsymbol{\lambda})\right)^{1-e_{ij}}.\end{aligned} \tag{6}$$

We formulate $\Pr(G|R^U, R^O)$ by focusing on the probability $\Pr(e_{ij}|r_i, r_j)$ of an edge between i and j. Since $\Pr(e_{ij}|r_i, r_j) = c(r_i, r_j)^{e_{ij}}(1 - c(r_i, r_j))^{1-e_{ij}}$, we have:

$$\Pr(G|R^O, R^U) = \prod_{\substack{i,j \in \mathcal{N} \\ i<j}} c(r_i, r_j)^{e_{ij}} \left(1 - c\left(r_i, r_j\right)\right)^{1-e_{ij}} \tag{7}$$

Using Bayes rule, the posterior over R^U given observed preferences R^O and network G is given by:

$$\Pr(R^U|G, R^O, \boldsymbol{\theta}) = \frac{\Pr(G|R^U, R^O)\Pr(R^U|\boldsymbol{\eta})}{\Pr(G|R^O, \boldsymbol{\theta})}, \tag{8}$$

where one can use Eq. 3, Eq. 6, and Eq. 7 for computation of $\Pr(R^U|\boldsymbol{\eta})$, $\Pr(G|R^O, \boldsymbol{\theta})$, and $\Pr(G|R^U, R^O)$, respectively.

5. SAMPLING METHODS

For both preference inference and group recommendation, we must compute the joint posterior $\Pr(R^U|G, R^O, \boldsymbol{\theta})$. Exact computation is, not surprisingly, computationally expensive. So we develop sampling methods to approximate the posterior. At a high level, we sample L *preference profiles* $R^{(1)}, \ldots, R^{(L)}$ from the posterior, where each profile consists of a preference ranking for each individual $i \in U$. Let $R_i^{(t)}$ denote the sampled preference ranking of individual $i \in U$ in the t^{th} profile. We approximate the posterior preference of any individual $i \in U$, $\Pr(r_i = l|G, R^O, \boldsymbol{\theta})$, $\forall\ l \in \Omega(\mathcal{A})$, and the expected score of the inferred component of social welfare $sw^E_{inf}\left(a, G, R^O, O\right)$ as follows:

$$\Pr(r_i = l|G, R^O, \boldsymbol{\theta}) \approx \frac{1}{L}\sum_{t=1}^{L} I[R_i^{(t)} = l], \text{and}$$

$$sw^E_{inf}\left(a, G, R^O\right) \approx \frac{1}{L}\sum_{t=1}^{L}\sum_{i \in U} g(R_i^{(t)}(a), m)$$

Here $g(R_i^{(t)}(a), m)$ is the positional score of option a in i's preference ranking for the t^{th} sample. L must be sufficiently large to ensure a good approximation. More critically, we must be able to draw independent samples from the (unknown) posterior. To do this, we use an MCMC algorithm, specifically, *Gibbs sampling*, where individual variables are in turn sampled using *Metropolis sampling*.

We use iterative Gibbs sampling to sample unobserved preferences R^U. It begins with an initial preference profile $R^{(0)}$, completing the rankings for all unobserved preferences. At each iteration l, we sample $r_i^{(l)}$ for each $i \in U$ from the conditional distribution

$$\Pr(r_i|R_1^{(l)}, \ldots, R_{i-1}^{(l)}, R_{i+1}^{(l-1)}, \ldots, R_{|U|}^{(l-1)}, R^O).$$

The order in which preferences are sampled can impact the efficiency of the method. The order can be deterministic or stochastic, and may be based on node degree or the number of observed preferences of their neighbors. In our experiments, we use a fixed arbitrary ordering.

To sample r_i from the distribution $\Pr(r_i|R_{\setminus i})$ we use *Metropolis sampling*. By Eqs. 6–8 and 3, the probability of r_i given all other individual preferences is $\Pr(r_i|R_{\setminus i}) \propto \tilde{p}(r_i)$, where

$$\tilde{p}(r_i) = \prod_{j \in U} \Pr(e_{ij}|r_i, r_j) \prod_{j \in O} \Pr(e_{ij}|r_i, r_j)\phi^{d_\tau(\sigma, r_i)},$$

which can be computed in $O(n)$ time. To sample r_i at iteration l of Gibbs, we sample r^* from a *conditional proposal* *distribution*

$$q(r^*|R_i^{(l-1)}) = \frac{1}{z(\hat{\phi}_i)} \hat{\phi}_i^{\,d_\tau(r_i,\sigma)},$$

which is a Mallows distribution that uses the previous sample of i's preference $R_i^{(l-1)}$ as its reference ranking and a dispersion parameter $\hat{\phi}_i$ (below we fix $\hat{\phi}_i = \phi$, preference dispersion parameter). We accept proposal r^* as $R_i^{(l)}$ (i.e., set $R_i^{(l)} = r^*$) with probability

$$A(r^*, R_i^{(l-1)}) = \min\left(1, \frac{\tilde{p}(r^*)}{\tilde{p}(R_i^{(l-1)})}\right);$$

otherwise, we set $R_i^{(l)} = R_i^{(l-1)}$. To sample from the Mallows model $q(.)$, we use the *repeated insertion model* [11]. One can sample L preference profiles given $|R^U|$ unobserved preferences in $O(L|R^U|nm)$ time using our proposed method. Assuming $|R^U|$ is a constant fraction of n, our sampling methods runs in $O(Ln^2m)$ time, which may prove intractable for very large networks. Designing more scalable sampling methods is an important future direction.

6. EMPIRICAL ANALYSIS

We conduct experiments to assess the effectiveness of our inference and group recommendation algorithms. We measure the accuracy of preference inference, and more importantly, assess the quality of the group decisions reached when exploiting network structure to better deal with missing preferences of certain group members.

Experimental Setup. We experiment on three types of data sets: *two-sided synthetic data* in which both preferences and networks are randomly generated; *one-sided real-world data* in which preferences are derived from Irish electoral data, but networks are synthetically generated; and *two-sided real-world data* in which both preferences and network structure are derived from Flixster. We assume the model parameters $\boldsymbol{\theta}$ are known (e.g., learned from a similar population). Unless otherwise noted, we set $(\alpha, \gamma, \beta) = (2, 0.7, 1)$ and $n = 200$. We use Borda as our scoring rule where $g(r(a), m) = m - r(a)$. While other rules can be used, Borda is a useful surrogate for random utility models [8] and serves to illustrate the value of the POSN model.

We vary the degree to which preferences are observed with parameter $\psi \in [0, 1]$, the probability that any node's ranking is observed. By varying ψ, we can assess the impact of preference observability on the efficiency of our methods. We select the *decision making group* $\mathcal{S} \subseteq \mathcal{N}$ (with n_s members), for whom a group recommendation is to be made, using one of three methods. *RSA (Random Selection from All)* selects n_s individuals uniformly at random from $\mathcal{N}$. *RSU (Random Selection from Unobserved)* select n_s individuals uniformly at random from U (e.g., reflecting a company with access to a social network and the preferences of existing customers, and wanting to market to new prospects without knowing their preferences). *RSC (Random Selection from Community)* selects a connected community: it first selects a "seed" individual at random, then extends the group by selecting $n_s - 1$ friends of this seed at random; if this set is smaller than $n_s - 1$, friends of these friends are selected at random to complete the group.

Performance Metrics. To measure prediction accuracy, we determine how close inferred preferences are to the *true unobserved preferences* from a held out test set. We measure closeness using *mean scaled expected Kendall-τ (MSEK)*:

$$MSEK = \frac{1}{\binom{m}{2}|U|} \sum_{i \in U} \sum_{r \in \Omega(\mathcal{A})} \Pr(r|G, R^O, \boldsymbol{\theta}) d_\tau(r, \hat{r}_i),$$

where $\hat{r}_i$ is the true preference of i and $\binom{m}{2}$ is maximum τ-distance between two rankings over m options. MSEK lies in $[0, 1]$: $MSEK = 0$ if all preferences are inferred correctly, while $MSEK = 1$ implies maximum "inaccuracy."

To examine the decision/recommendation quality using inferred preferences, we compare its social welfare with that of *the decision that would be made had actual preferences been observed*. Let $sw(\cdot)$ denote social welfare with true preferences, and a^* and a^*_{inf} be the optimal options under given actual and inferred preferences. Rather than directly comparing social welfare, we define *relative social welfare loss (RSWL)* to be $[sw(a^*) - sw(a^*_{inf})]/sw(a^*)$ (we report it as a percentage).

Benchmarks. We consider several other ways of dealing with missing preferences in decision making, and use these as benchmarks. In *ϕ-mallows inference (PM)*, we assume that all unobserved preferences are independent and are drawn from a ϕ-mallows model (with parameters identical to those in the POSN model). We calculate the same inferred social welfare functions as in our model, namely, ES, JMPES, IMPES. Note that ES will be the same for all unobserved preferences and can be computed once. Moreover, JMPES and IMPES must be the same as the reference ranking σ. Another approach to missing preferences, dubbed *Discard Unobserved (DU)*, is to ignore them and make a decision using only observed preferences.

For each fixed setting, we generate 10 partially observed POSNs. In each, we burn-in 1000 samples, then collect 1000 samples using our Gibbs-Metropolis method. We report MSEK averaged over the 10 instances. For each instance, we also randomly select 40 decision making groups of fixed sizes $\{3, 5, 10, 15, 20\}$ using RSA, RSU, or RSC, giving 400 social choice instances per an experimental setting. RSWL is reported as the average over these 400 instances.

m / ψ	$\psi = 0.5$	$\psi = 0.6$	$\psi = 0.7$	$\psi = 0.8$
$m = 3$	0.000	0.000	0.000	0.000
$m = 4$	0.009	0.008	0.007	0.006
$m = 5$	0.168	0.158	0.152	0.148
$m = 6$	0.378	0.367	0.349	0.335

Table 1: Avg. MSEK (10 instances), various m, ψ.

Two-sided synthetic. We set $\phi = 0.85$, $\sigma = (1, \ldots, m)$, and $\boldsymbol{\lambda}$ as stated above. Table 1 shows average MSEK for various ψ and m. Unsurprisingly, MSEK increases with m and decreases with ψ. As m increases, the number of rankings increases factorially, as does the the support of the ranking distribution. In such cases, lower MSEK requires more information for accurate prediction. When $m = 3$, $n = 200$ is sufficient to push MSEK to almost 0. With $m = 4$, it remains very low. To examine the effect of n on MSEK, we fix $m = 6$ and $\psi = 0.8$ but vary n: Table 2 shows that MSEK decreases with n as expected. Decision quality of our methods in this setting is qualitatively similar to those discussed below. Our ES and IMPES methods outperform the other benchmark methods in most settings, including over all group sizes, group selection methods, and various m (even for $m = 6$ with relatively high MSEK).

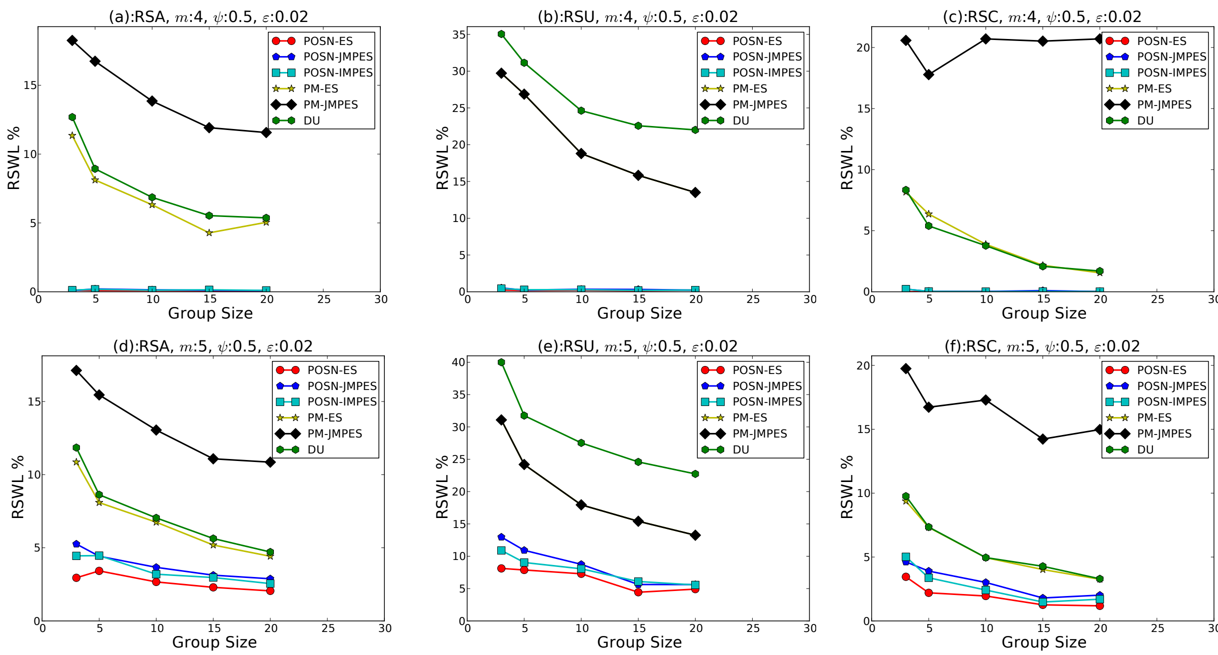

Figure 2: Avg. RSWL (over 400 instances) for various group sizes n_s, group selection methods, and m but fixed $\psi = 0.5$, $\varepsilon = 0.2$.

n	200	400	600	800	1000
MSEK	0.335	0.2474	0.185	0.141	0.106

Table 2: Avg. MSEK (10 instances), $\psi = 0.8$, $m = 6$.

m / ε	$\varepsilon = 0.0$	$\varepsilon = 0.01$	$\varepsilon = 0.02$	$\varepsilon = 0.05$
$m = 4$	0.007	0.009	0.009	0.022
$m = 5$	0.172	0.198	0.214	0.274

Table 3: Avg. MSEK, $\psi = 0.5$, $n = 200$, Irish data set.

Irish data. We test our methods using real-world preferences from the 2002 Irish Election, Dublin West Constituency, with 9 candidates and 29,989 ballots of the top-t form, of which 3800 are complete rankings. We created preference data sets with various values m from these complete preferences, by choosing m candidates with highest aggregate Borda score, and limiting each individual's preferences to these m options.

For each m, we learn ϕ and σ from its corresponding filtered data set and used those parameters in our methods (hence we have a loose prior over preferences, but not a precise prior for specific group , see below). For each experimental setting, we generate 10 partially observed POSNs with $\psi = 0.5$ and 200 individuals with preferences drawn from the filtered Irish data set. We then generate the POSN using our model, but with additional noise: we randomly change the parity of each e_{ij} (i.e., delete or add an edge) with probability ε. Though we create a synthetic social network using our POSN model, adding noise in this fashion reflects scenarios in which the social network is not generated using our specific model, or when learned model parameters provide a less-than-ideal fit to the underlying data.

Table 3 reports average MSEK when ε varies ($m = 4, 5$). Unsurprisingly, MSEK increases with both m and ε when n and ψ are fixed. MSEK is very low when $m = 4$, even with high $\varepsilon = 0.05$ (10 edge flips per node in expectation). Tables 1 and 3 show comparable MSEK values for $m = 4, 5$, suggesting that that even in scenarios where the preference distribution $\rho(.)$ is not known *a priori*, but is a learned ϕ-mallows model, POSNs support effective inference.

Fig. 2 shows average RSWL with $\psi = 0.5$ and $\varepsilon = 0.02$ (400 expected edge flips in the network). We vary m, the group selection method and the inference method. Our POSN-ES and POSN-IMPES approaches outperform the other benchmarks in most settings, including: all situations in which no group preferences are observed (see Fig. 2(b) and 2(e)); and even with $m = 5$ (see Fig. 2(d)-2(f)) despite its relatively high MSEK (see Table 3). RSWL in all benchmark methods (PM-ES, PM-JMPES, DU) is very sensitive to group size, increasing dramatically as group size decreases (see Fig. 2(a)-(f)). However, POSN-ES and POSN-IMPES are more robust to group size (see Fig. 2(a)-(f)). POSN-IMPES approximates POSN-ES reasonably well, while POSN-JMPES also performs well.

Flixster data. The Flixster dataset [18] consists of a social network of movie watchers and their ratings of movies, and allows a test of our methods using both real-world network and preference data. Because movie ratings are sparse, we aggregate them into preferences over movie genres (genres were determined automatically using the Rotten Tomatoes and IMDB web sites). Let $\tilde{r}_{um}$ be the rating of user u for movie m where $\tilde{r}_{um} \in \{0.5, 1, \cdots, 5\}$ if u has rated m otherwise 0 (for missing ratings). For each user u and genre g, we define a *user-genre* score

$$S^C_{ug} = \frac{1}{\hat{I}_u} \sum_m \text{sign}(\tilde{r}_{um}) A_{mg},$$

where $\hat{I}_u = \sum \text{sign}(\tilde{r}_{um})$ is the number of movies rated by u, and $A_{mg} = 1$ if movie m has genre g (and $A_{mg} = 0$ otherwise). This score reflects the relative number of movies of each genre watched by a specific user. This is converted into a ranking of genres for each user u by ordering genres according to their scores S^C_{ug}. We limit our focus to four diverse genres—Comedy, Drama, Kids/Family

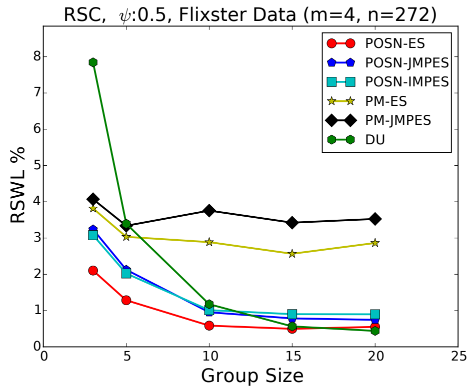

Figure 3: Avg. RSWL (400 instances), Flixster.

	n_s=3	n_s=5	n_s=10	n_s=15	n_s=20
POSN-ES	**7.34**	**5.04**	**2.86**	**2.57**	2.66
POSN-JMPES	9.89	7.67	4.33	3.69	3.29
POSN-IMPES	9.59	7.09	4.34	3.89	3.77
PM-ES	9.36	8.07	7.56	6.76	6.73
PM-JMPES	9.72	8.54	8.63	7.84	7.48
DU	20.61	12.07	5.31	2.92	**2.12**

Table 4: Std. RSWL in percentage (Flixster, ψ=0.5)

and Mystery/Suspense.[2] We run our methods on a 272-node subgraph of the Flixster data set, with 924 edges. We estimate a ϕ-Mallows model and POSN model parameters using maximum likelihood methods on this sub-network; the learned parameters are $(\alpha, \beta, \gamma, \phi) = (2.05, 1.06, 0.07, 0.33)$. For each run, we test our methods on 10 instances of a partially observed network, censoring each individual's genre preference with probability $\psi = 0.5$ or $\psi = 0.3$.

Average MSEK is 0.242 and 0.256 for $\psi = 0.5$ and $\psi = 0.3$ (resp.). This suggests that genre preferences are reasonably predictable using the POSN model. Fig. 3 shows decision making performance, i.e., average RSWL, for the various methods described above using RSC to select groups. Each of our POSN-sensitive methods—ES, IMPES, and JMPS—outperform the ϕ-Mallows benchmark for all group sizes, and outperform DU significantly for small groups. DU performs comparably to methods that account for network structure when groups are larger (15 or 20 individuals) since, in expectation, the preferences of 7–10 group members are observed: this is sufficient to make a good decision without estimating missing preferences explicitly due to normal sampling bounds from the underlying Mallows model. This, in addition to the fact that homophily across a large group makes it likely that the missing preferences are similar to those observed, means that making a group decision based only on observed preferences usually results in near-optimal decisions. Table 6 reports the std. dev. for these results when $\psi = 0.5$. ES has the smallest variance in RSWL in general, implying more robustness in the decisions made. Overall, ES is the most reliable method of those analyzed here. (Results for $\psi = 0.3$ are qualitatively similar).

[2] We focus on these four genres in part to increase data "density." Our choice of these genres may impact the results below; future investigation is needed to assess this impact.

7. CONCLUDING REMARKS

We introduced preference-oriented social networks (POSNs) to capture the correlation of preference rankings between individuals who interact in social networks. We developed effective inference methods to predict an individual's preferences by exploiting these correlations. We also developed methods for group recommendation when the preferences of some (or even all) group members are unobserved. Our experiments showed the value of accounting for social ties in inference and group recommendation when faced with missing preferences.

This work is a starting point for the deeper modeling of preferences in a social network context. Interesting future directions include: empirical investigation of preference correlations in real-world networks; scalable learning methods for estimating model parameters; more efficient sampling methods based on network topology; studying other aggregation functions (e.g., other social choice functions, voting rules, bargaining solution concepts, etc.), and extensions to other social choice problems (e.g., matchings, assignments).

Of practical importance is investigating the extent to which preference rankings are correlated and play a role in shaping connections in real-world social networks. Developing scalable methods for learning model parameters is essential; such learning techniques can exploit our inference methods as important building block (e.g., in EM-based algorithms). More efficient sampling methods can be designed by taking into account the presence or absence of subsets of possible edges. Our model can provide the basis for more effective preference elicitation. As decision making using MPE seems to provide a reasonable approximation to optimal decisions, studying how MPE can be computed or approximated without the use of sampling remains of interest. Similar to active learning methods [6], the tighter integration of inference and decision making methods would also be of value.

There are a number of potential extension to our POSN model. This includes accommodating partial information about the preferences of specific users (e.g., a small set of pairwise comparisons); and incorporating both the strength and types of relationships between individuals. Such generalizations may offer greater performance in certain preference inference and group recommendation settings.

Acknowledgements. This research was supported by NSERC.

8. REFERENCES

[1] S. Amer-Yahia, S. B. Roy, A. Chawlat, G. Das, and C. Yu. Group recommendation: Semantics and efficiency. *Proc. VLDB Endow.*, 2(1):754–765, 2009.

[2] K. Arrow. *Social Choice and Individual Values.* 1951.

[3] L. Baltrunas, T. Makcinskas, and F. Ricci. Group recommendations with rank aggregation and collaborative filtering. *Proc. 4th ACM Conf. on Recommender Systems (RecSys-10)*, pp.119–126, 2010.

[4] M. Barthélemy. Spatial networks. *Physics Reports*, 499(1):1–101, 2011.

[5] S. Berkovsky and J. Freyne. Group-based recipe recommendations: Analysis of data aggregation strategies. *Proc. 4th ACM Conf. on Recommender Systems (RecSys-10)*, pp.111–118, 2010.

[6] M. Bilgic, L. Mihalkova, and L. Getoor. Active learning for networked data. *Proc. 27th Intl. Conf. on Mach. Learn. (ICML-10)*, pp.79–86, 2010.

[7] M. Boguñá and R. Pastor-Satorras. Class of correlated random networks with hidden variables. *Physical Review E*, 68:036112, Sep 2003.

[8] C. Boutilier, I. Caragiannis, S. Haber, T. Lu, A. D. Procaccia, O. Sheffet. Optimal social choice functions: A utilitarian view. *Proc. 13th ACM Conf. on Electronic Commerce (EC-12)*, pp.197–214, 2012.

[9] L. M. Busse, P. Orbanz, J. M. Buhmann. Cluster analysis of heterogeneous rank data. *Proc. 24th Intl. Conf. on Mach. Learn.*, pp.113–120, 2007.

[10] N. A. Christakis and J. H. Fowler. The spread of obesity in a large social network over 32 years. *New England J. Medicine*, 357(4):370–379, 2007.

[11] J.-P. Doignon, A. Pekec, and M. Regenwetter. The repeated insertion model for rankings: Missing link between two subset choice models. *Psychometrika*, 69(1):33–54, 2004.

[12] M. Gartrell, X. Xing, Q. Lv, A. Beach, R. Han, S. Mishra, K. Seada. Enhancing group recommendation by incorporating social relationship interactions. *Proc. 16th Intl. Conf. on Supp. Group Work*, pp.97–106, 2010.

[13] M. Gladwell. *The Tipping Point: How Little Things can Make a Big Difference*. Little Brown, 2000.

[14] J. Golbeck and J. Hendler. Filmtrust: Movie recommendations using trust in web-based social networks. *Proc. IEEE Consumer Communications and Networking Conf.*, Vol. 96, pp.282–286, 2006.

[15] M. S. Granovetter. The strength of weak ties. *American J. Sociology*, 78(6):1360–1380, 1973.

[16] P. D. Hoff, A. E. Raftery, and M. S. Handcock. Latent space approaches to social network analysis. *J. American Statistical Assoc.*, 97(460):1090–1098, 2002.

[17] M. Jackson. *Social and Economic Networks*. 2008.

[18] M. Jamali, M. Ester. A matrix factorization technique with trust propagation for recommendation in social networks. *Proc. 4th ACM Conf. on Recommender Systems (RecSys-10)*, pp.135–142, 2010.

[19] M. Kim and J. Leskovec. Latent multi-group membership graph model. *Proc. 29th Intl. Conf. on Machine Learning (ICML-12)*, pp.1719–1726, 2012.

[20] I. Konstas, V. Stathopoulos, J. Jose. On social networks and collaborative recommendation. *Proc. 32nd Intl. Conf. on R&D in Info. Retr. (SIGIR)*, pp.195–202, 2009.

[21] K. Lewis, M. Gonzalez, and J. Kaufman. Social selection and peer influence in an online social network. *PNAS*, 109(1):68–72, 2012.

[22] T. Lu and C. Boutilier. Learning mallows models with pairwise preferences. *Proc. 28th Intl. Conf. on Machine Learning (ICML-12)*, pp.145–152, 2011.

[23] H. Ma, I. King, M. Lyu. Learning to recommend with social trust ensemble. *Proc. 32nd Intl. Conf. on R&D in Info. Retr. (SIGIR)*, pp.203–210, 2009.

[24] H. Ma, H. Yang, M. R. Lyu, and I. King. SoRec: Social recommendation using probabilistic matrix factorization. *Proc. 17th ACM Conf. on Information and Knowledge Mgmt.*, pp.931–940, 2008.

[25] J. I. Marden. *Analyzing and Modeling Rank Data*. Chapman and Hall, London, 1995.

[26] P. Massa, P. Avesani. Controversial users demand local trust metrics: An experimental study on epinions.com community. *Proc. 20th National Conf. on Artificial Intelligence (AAAI-05)*, pp.121–126, 2005.

[27] J. Masthoff. Group modeling: Selecting a sequence of television items to suit a group of viewers. *User Model. & User-Adapted Interact.*, 14(1):37–85, 2004.

[28] J. Masthoff and A. Gatt. In pursuit of satisfaction and the prevention of embarrassment: Affective state in group recommender systems. *User Model. & User-Adapted Interact.*, 16(3-4):281–319, 2006.

[29] J. F. McCarthy, T. D. Anagnost. Musicfx: An arbiter of group preferences for computer supported collaborative workouts. *Proc. 1998 ACM Conf. on Computer-supported Coop. Work*, pp.363–372, 1998.

[30] T. Murphy, D. Martin. Mixtures of distance-based models for ranking data. *Comp. Stats. & Data Anal.*, 41(3):645–655, 2003.

[31] M. Newman. *Networks: An Intro.* Oxford, 2010.

[32] L. Quijano-Sanchez, J. Recio-Garcia, B. Diaz-Agudo, G. Jimenez-Diaz. Social factors in group recommender systems. *ACM Trans. Intel. Sys. & Tech.*, 4(1):8, 2013.

[33] A. Salehi-Abari and C. Boutilier. Ranking networks. *NIPS-13 Workshop on Frontiers of Network Analysis: Methods, Models, and Applications*, 2013.

[34] A. Salehi-Abari and C. Boutilier. Empathetic social choice on social networks. *Proc. 13th Intl. Conf. on Autonomous Agents and Multiagent Systems (AAMAS-14)*, pp.693–700, 2014.

[35] S. Seko, T. Yagi, M. Motegi, and S. Muto. Group recommendation using feature space representing behavioral tendency and power balance among members. *Proc. Fifth ACM Conf. on Recommender Systems (RecSys-11)*, pp.101–108, 2011.

[36] A. K. Sen. *Collective Choice and Social Welfare*. North-Holland, 1970.

[37] P. Sen, G. M. Namata, M. Bilgic, L. Getoor, B. Gallagher, and T. Eliassi-Rad. Collective classification in network data. *AI Magazine*, 29(3):93–106, 2008.

[38] M. A. Serrano, D. Krioukov, and M. Boguñá. Self-similarity of complex networks and hidden metric spaces. *Physical Review Letters*, 100:078701, Feb 2008.

[39] B. M. Waxman. Routing of multipoint connections. *IEEE J. Selected Areas in Communications*, 6(9):1617–1622, Dec. 1988.

[40] Y. Xiwang, Y. Guo, Y. Liu, and H. Steck. A survey of collaborative filtering based social recommender systems. *Computer Communications*, 41:1–10, 2014.

[41] X. Yang, H. Steck, Y. Liu. Circle-based recommendation in online social networks. *Proc. 18th Intl. Conf. on Knowl. Disc. and Data Mining*, pp.1267–1275, 2012.

[42] Z. Yu, X. Zhou, Y. Hao, and J. Gu. TV program recommendation for multiple viewers based on user profile merging. *User Model. & User-Adapted Interact.*, 16(1):63–82, 2006.