Modeling Uncertainty in Group Recommendations

Dimitris Sacharidis
TU Wien
E-Commerce Research Unit
Austria
dimitris@ec.tuwien.ac.at

ABSTRACT

In many settings, it is required that items are recommended to a group of users instead of a single user. Most often, when the decision criteria and preferences of the group as a whole are not known, the gold standard is to aggregate individual member preferences or recommendations. Such techniques typically presuppose some process under which group members reach consensus, e.g., least misery, maximum satisfaction, disregarding any uncertainty on whether this presumption is accurate. We propose a different approach that explicitly models the system's uncertainty in the way members might agree on a group ranking. The basic idea is to quantify the likelihood of hypothetical group rankings based on the observed member's individual rankings. Then, the systems recommends a ranking that has the highest expected reward with respect to the hypothetical rankings. Experiments with real and synthetic groups demonstrate the superiority of this approach compared to previous work based on aggregation strategies and to recent fairness-aware techniques.

CCS CONCEPTS

• **Information systems** → **Recommender systems**.

KEYWORDS

Group Recommender Systems; Aggregation Strategies; Uncertainty

ACM Reference Format:
Dimitris Sacharidis. 2019. Modeling Uncertainty in Group Recommendations. In *27th Conference on User Modeling, Adaptation and Personalization Adjunct (UMAP'19 Adjunct), June 9–12, 2019, Larnaca, Cyprus.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3314183.3324987

1 INTRODUCTION

In group recommender systems, the goal is to provide a recommendation to a group of people, e.g., to friends planning their summer vacation destination [1], or to a family deciding on a TV program to watch [24]. In the case of persistent, long-standing groups, the system can derive a profile of the group, e.g., from a history of group-item interactions, and essentially treat the group as a virtual user. The most interesting setting is when the system must recommend to an ad-hoc, ephemeral group, for which no historical information is known. This work focuses on group recommendations for ad-hoc groups, and proposes a model of the system's uncertainty regarding how group members might decide on items.

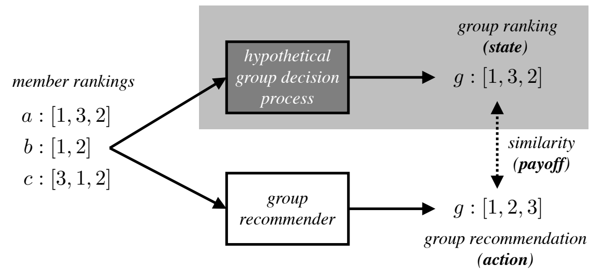


Figure 1: Group recommendation task

We formulate the group recommendation task as follows. For each user in the group, we assume the system knows her *member ranking* on a subset of items. For example, member rankings can be extracted via preference elicitation and feedback, or estimated with collaborative filtering or some other technique. We define the *group ranking* as the "optimal" ranking of a subset of items for the group, i.e., the ranking members would agree on if they would convene, had perfect knowledge, and were able to reach a consensus. Of course, such a scenario is only hypothetical, and thus the group ranking is *unknown* to the system. The goal of a group recommender is to propose a *group recommendation list* that, to the best of the system's knowledge, matches the unknown group ranking.

Figure 1 illustrates the task of a group recommender. Assume a group of three users, $\{a, b, c\}$, and three items $1, 2, 3$. Each member has individual preferences on these items, that are expressed as the three depicted group member rankings; for instance, user a ranks 1 the highest, followed by 3 and 2. The shaded part of the figure illustrates the unknown/hypothetical group decision process. The three users reach a consensus on their ideal group ranking $[1, 3, 2]$. The unshaded part of the figure presents the group recommender as a black box that accepts as input the member rankings, and outputs a group recommendation list. In our depiction, the system employs some group recommendation algorithm and produces the list $[1, 2, 3]$.

There has been a plethora of group recommendation algorithms for a thorough overview of this research area refer to [6, 9, 14]. Briefly, the existing approaches for ad-hoc groups can be classified as follows. Profile aggregation methods, e.g., [3, 13, 16, 22, 24], merge the profiles of members to create a group profile upon which to make recommendations. Score aggregation methods, e.g., [1, 14, 17], are inspired by social choice theory and aggregate predicted member rankings; in some cases external factors are also used taken

UMAP'19, June 2019, Larnaca, Cyprus

ACM ISBN 978-1-4503-6711-0/19/06…$15.00
https://doi.org/10.1145/3314183.3324987

Table 1: States, Actions, and payoffs

State	Prob.	Action [1, 2, 3]	[1, 3, 2]	[2, 1, 3]	[2, 3, 1]	[3, 1, 2]	[3, 2, 1]
[1, 2, 3]	2/9	1	1/3	1/3	-1/3	-1/3	-1
[1, 3, 2]	2/3	1/3	1	-1/3	1/3	-1	-1/3
[2, 1, 3]	0	1/3	-1	1	-1	1/3	-1/3
[2, 3, 1]	0	-1/3	1/3	-1	1	-1/3	1/3
[3, 1, 2]	2/3	-1/3	-1	1/3	-1/3	1	1/3
[3, 2, 1]	0	-1	-1/3	-1/3	1/3	1/3	1
Expected payoffs		2/9	2/27	2/27	-2/27	-2/27	-2/9

into account [7, 12, 15, 20, 21]. Rank aggregation techniques, e.g., [2], fuse the predicted member rankings considering only the relative order of items. More recently, the focus has been on fairness-aware methods, e.g., [11, 18, 23], that seek to recommend items so that to minimize dissatisfaction among members.

The common theme among previous work is that it makes a *presumption*, explicit or not, on how a group would decide, and produces the group recommendation list accordingly. For instance, a recommender following the least-misery principle assumes that the group is only satisfied as much as its least satisfied member, and thus ranks items decreasingly by their minimum utility among members. In practice, user studies [5, 13] have demonstrated that groups may reach consensus on group rankings that differ substantially to those produced by the aggregation strategies typically employed by group recommenders. Despite these observations, existing group recommenders are rigid and do not account for uncertainty in the decision process.

In this work, we model the group recommendation task as a decision theoretic problem under structured uncertainty. Specifically, we see the optimal group ranking as the uncertain *state* of the world. In the example of Figure 1, group ranking $[1, 3, 2]$ is a state representing the world in which the group agrees to this particular ranking. Moreover, we assume that uncertainty is probabilistic, meaning that each state corresponds to a possible world that occurs with some probability. For example, state $[1, 3, 2]$ may occur with probability $3/7$. Such state probabilities are computed using the external information available to the system — in our case, the member rankings — in a manner we discuss in Section 2.

Furthermore, we see the task of producing a group recommendation list as the *action* the recommender must take. In the example of Figure 1, the recommendation list $[1, 2, 3]$ is the action of the system. Note that, in our model, both a state and an action correspond to rankings; the difference is that the former is the unknown/hypothetical ground truth of the state, while the latter is the system's output. For a specific state, an action is associated with a *payoff*. In our model, the payoff of a group recommendation list captures its similarity with the state's ground truth, i.e., the group ranking. For example, in Figure 1, action $[1, 2, 3]$ for the state $[1, 3, 2]$ leads to a payoff of $1/3$, derived as the Kendall's tau similarity between the two rankings.

Under this probabilistic formulation of the uncertainty in the group recommendation task, the system's optimal action is the one that maximizes the *expected payoff*, i.e., the probability-weighted average of the payoffs an action receives across all states. This translates into the recommender outputting the list that has the highest similarity to the group ranking *in expectation*.

Table 1 presents our model for the recommendation task of Figure 1. As there are only three items, there are six possible rankings and thus as many states and actions. Each state is accompanied by its probability. While we defer details for Section 2, note that the three states that place item 2 before 1 have zero probability; this is because no member prefers 1 over 2 in their rankings. Also note that states $[1, 3, 2]$ and $[3, 1, 2]$ are the most probable states as they exactly match one member's ranking (a's or c's) and extend another members's ranking (b's).

The depicted payoffs for each action-state pair are computed using Kendall's tau similarity, ranging from -1 to 1. The expected payoffs for each action are depicted in the last row, where action $[1, 2, 3]$ is found to have the highest value and is thus the output of our group recommender. It is worth noting that although rankings $[1, 3, 2]$ and $[3, 1, 2]$ have the highest probability of being group rankings (as states), they have less expected payoff (as actions) when all possible states of the world are considered. This is also the strength of our model: it does not recommend the most probable ranking, but hedges its bets accounting for the uncertainty in the group decision process.

The remaining of this paper is structured as follows. Section 2 describes our model of the group recommendation task, and explains how it used to make recommendations. Section 3 evaluates our group recommender against related work, and Section 4 presents our conclusions.

2 PAYOFF-BASED GROUP RECOMMENDER

Section 2.1 introduces the model employed by our recommender, and Section 2.2 presents how group recommendations are derived.

2.1 A Model of Uncertainty for the Group Recommendation Task

Our group recommender is based on a model capturing the uncertainty of the group recommendation task. The key element of this model is the notion of state probabilities. Once they are determined, the expected payoffs can be straightforwardly computed and the appropriate action found. In what follows, we explain how state probabilities are defined. Briefly, the pairwise preferences for each group member are extracted from the sole input of the recommender, the members rankings. Then, for each pairwise preference we assign a probability, which in turn defines the probability of each possible ranking. We start the discussion by defining pairwise preferences and rankings in general, and then introduce their probabilities of occurrence when considered in the context of a group.

Pairwise Preferences and Rankings. We assume a set I of items. A *pairwise preference* of item i over j is denoted as $i \prec j$. Items i, j are *equi-preferred* if $i \prec j$ and $j \prec i$. On the other hand, items i, j are *incomparable* if neither $i \prec j$ nor $j \prec i$ hold, in which case we write $i \nsim j$. A *preference set* is a set of pairwise preferences. For example, preference set $\{1 \prec 2, 1 \prec 3\}$ says that among the items, 1 is the most preferred, while 2 and 3 are incomparable.

A *ranking* is a preference set that is a total order on a subset of items, i.e., the pairwise preferences define a transitive, antisymmetric, and complete (there exists a preference between each pair

Table 2: Member Rankings and Preference Probabilities

(a) Member Rankings

Member	Pairwise Preferences
a	$1 \prec 2, 1 \prec 3, 3 \prec 2$
b	$1 \prec 2$
c	$3 \prec 1, 1 \prec 2, 3 \prec 2$

(b) Probabilities for Group Pairwise Preferences

Pref.	Prob.	Pref.	Prob.	Pref.	Prob.
$1 \prec 2$	1	$1 \prec 3$	1/3	$2 \prec 3$	0
$2 \prec 1$	0	$3 \prec 1$	1/3	$3 \prec 2$	2/3
$1 \nsim 2$	0	$1 \nsim 3$	1/3	$2 \nsim 3$	1/3

(c) Probabilities for Group Preference Sets

Preference Set	Prob.	Preference Set	Prob.
$\{1 \prec 2\}$	1/9	$\{1 \prec 2, 3 \prec 2\}$	2/9
$\{1 \prec 2, 1 \prec 3\}$	1/9	$\{1 \prec 2, 1 \prec 3, 3 \prec 2\}$	2/9
$\{1 \prec 2, 3 \prec 1\}$	1/9	$\{1 \prec 2, 3 \prec 1, 3 \prec 2\}$	2/9

of items) relation. For example, preference set $\{1 \prec 2, 1 \prec 3, 2 \prec 3\}$ is a ranking, denoted as [1, 2, 3]. A ranking is *compatible* with a preference set, if the latter is a subset of the former. For example, ranking [1, 3, 2] is compatible with preference set $\{1 \prec 2, 1 \prec 3, 3 \prec 2\}$, as is with $\{1 \prec 2\}$.

Table 2a shows the pairwise preferences contained in the rankings of the three group members of Figure 1.

Group Pairwise Preferences, Group Rankings, and Probabilities. We assume a group of users, where for each group member we know her pairwise preferences. As explained, member pairwise preferences are straightforwardly derived from a ranking; note that there are even systems that specifically elicit such types of preferences [4].

A pairwise preference $i \prec j$ for the group (unlike for a user) is uncertain, and can be expressed with some probability $Pr(i \prec j)$. Here, we assume that the probability is simply the fraction of members that express the corresponding pairwise preference. Of course, more elaborate mechanisms for computing probabilities are possible; e.g., when a group has provided some feedback, or when group roles, social influence strength, etc. among members are known. Table 2b depicts the probabilities for each among the six pairwise preferences. In addition, the table also depicts the probability $Pr(i \nsim j)$ of incomparability for two items computed as $1 - Pr(i \prec j) - Pr(j \prec i)$.

Using the probabilities of pairwise preferences, we can define the probability of a preference set as the product of the probabilities of pairwise preferences it contains as well as all incomparability relations it implies.[1] For instance, the probability of preference set $\{1 \prec 2, 1 \prec 3\}$ is computed as $Pr(1 \prec 2) \cdot Pr(1 \prec 3) \cdot Pr(2 \nsim 3) = 1/9$. Table 2c shows the probabilities of preference sets; all sets not depicted have zero probability. Note that the probabilities naturally sum to one.

Finally, we can define the probability of a group ranking in the following manner. Recall that a ranking is compatible with multiple preference sets. Equivalently, a preference set may be completed in different ways to define a ranking. Therefore, the probability with which a group ranking occurs is the probability that one among its compatible preference sets occurs. This probability is thus computed as the sum of probabilities for all compatible preference sets. Intuitively, a ranking can have multiple interpretations, its compatible preference sets, and each one contributes separately to the ranking's probability.

Table 3 depicts the probabilities for all possible rankings of the three items. For instance, ranking [1, 2, 3] has only two compatible preference sets with nonzero probability, shown in the first row. As each occurs with probability 1/9 (see Table 2c), the probability of [1, 2, 3] is 2/9. Note that because rankings are not mutually exclusive events (e.g., [1, 2, 3] and [1, 3, 2] share preference set $\{1 \prec 2\}$), their probabilities may sum to over one.

[1]The underlying assumption here is that preferences are independent, which is the same as what pairwise learning methods (e.g., [19]) make.

States, Actions, and Payoffs. A *state* and an *action* is a ranking on a subset of N items. The *probability of a state* is the probability of the ranking associated with it. The *payoff* of an action given a state is the similarity between the action and state rankings. In this work, we quantify payoff using Kendall's tau similarity, which is a measure of how many item pairs are ranked concordantly; however, other metrics, including non-symmetric ones like NDCG, could be used.[2]

The *expected payoff* of an action is the state probability weighted average of its payoffs. The goal of the group recommender is to identify the action that has the *maximum expected payoff*.

2.2 Making Group Recommendations

The process of deriving the state probabilities can be expensive when we need to consider a large number of items; note that there exist $O(|I|^2)$ item pairs, $O(3^{|I|^2})$ preference sets, and $O(|I|!)$ rankings. In what follows, we describe an algorithm that avoids enumerating all item-pairs, and instead directly estimates the probabilities of rankings. The key observation is that the systems does not need to know the probabilities for all pairwise preferences, particularly when it will only recommend a list of $N \ll |I|$ items. Rather, what is important is the ability to identify those items that are likely to be among the top-N. The group recommendation algorithm entails three steps, (1) identify the items that are most likely to rank high in group rankings, (2) generate some of their permutations and compute their probabilities, and (3) deriving the maximum expected payoff ranking.

Identify items that most likely rank high. Our goal is to identify a basis of N items, which will be eventually recommended. Let $Pr(i; k)$ denote the probability that item i ranks among the top-k items in the (unknown) group ranking, and let I_k represent the subset of items with nonzero $Pr(i; k)$. Note that there may exist items with zero such probability; in our example, item 2 has zero

[2]Another option would be to measure similarity directly between preference sets (e.g., [10]) instead of rankings.

Table 3: Probabilities of Group Rankings

Ranking	Compatible Preference Sets	Prob.
[1, 2, 3]	$\{1 \prec 2\}$, $\{1 \prec 2, 1 \prec 3\}$	2/9
[1, 3, 2]	$\{1 \prec 2\}$, $\{1 \prec 2, 1 \prec 3\}$, $\{1 \prec 2, 3 \prec 2\}$, $\{1 \prec 2, 1 \prec 3, 3 \prec 2\}$	2/3
[2, 1, 3]	–	0
[2, 3, 1]	–	0
[3, 1, 2]	$\{1 \prec 2\}$, $\{1 \prec 2, 3 \prec 1\}$, $\{1 \prec 2, 3 \prec 2\}$, $\{1 \prec 2, 3 \prec 1, 3 \prec 2\}$	2/3
[3, 2, 1]	–	0

probability of being the top item ($k = 1$), as all three members prefer 1 over 2. Clearly, I_k monotonically increases with k: if an item has nonzero probability for ranking at k, it certainly has nonzero probability ranking at any $k' > k$. Therefore, for any item there exists a minimal value of k for which its probability is nonzero. In our example, item 2 has nonzero probability for ranking second, an increase over the zero probability for ranking first; to see this note that there exists a group ranking, [1, 2, 3], with nonzero probability (see Table 3).

Motivated by this observation, we select the basis of N items by finding the lowest k such that $|I_k| \geq N$, and then picking among I_k those items with the highest probabilities $Pr(i; k)$. The difficult task is how to estimate the $Pr(i; k)$ probabilities, without exhaustively considering all possible group rankings, which we address as follows. Each member ranking can be seen as assigning utility scores to items; the simplest way is via Borda counts. Therefore, an item can be represent by its utility vector, where each coordinate corresponds to a group member's utility. It is easy to verify that any vector-valued function that aggregates utilities and is monotonic along every coordinate (an increase in some utility score, implies a non-decrease in the function's output) induces a group ranking of items with nonzero probability. For our purposes, we simply use a linear vector-valued function with random weights to generate a large number (but much less than $N!$) of possible group rankings. Then, the probability $Pr(i; k)$ of an item i ranking within the top-k is estimated as the fraction of the generated rankings that place i in the top-k.

Therefore, the procedure for selecting a basis of N items goes as follows. Starting with $k = 1$, we progressively increase its value until we get at least N items with nonzero $Pr(\cdot; k)$ probability. For each k, we estimate the item probabilities using the aforementioned method of generating random rankings. Once the smallest k^* such that $|I_{k^*}| \geq N$ is found, we pick the N items with the highest $Pr(\cdot; k^*)$ probabilities.

Generate rankings and compute their probabilities. In the previous step, we have identified a basis of N items, and for each of them computed the probability of ranking within the top-k^*, for some value of k^* appropriately derived. Next, we generate some rankings and estimate their probability. Specifically, we construct a ranking of the basis items by iteratively placing items at descending position. For position n, we draw a single item from the remaining $N - n + 1$ basis items with *selection probabilities* that are proportional to the items' top-k^* probabilities $Pr(\cdot; k^*)$. Thus items with high $Pr(\cdot; k)$ are more likely to rank high in the generated ranking.

For each generated ranking, we assign a likelihood score that is the product of the aforementioned selection probabilities of the items. This likelihood serves as an estimate of the unnormalized probability of the state corresponding to this ranking.

Derive the maximum expected payoff ranking. From the previous step, we can now populate a partial payoff table involving the generated rankings acting as states and actions. To determine the best group ranking to return, we propose two approaches. The first, denoted as MEP, is to select the ranking that has the highest expected payoff with respect to the generated rankings and derived state probabilities. In other words, MEP assumes the payoff table is complete and state probabilities accurate, and makes the optimal choice. The second approach, termed MEP*, admits that the payoff table is only partial and creates a ranking by fusing all generated rankings. Specifically, we implement a weighted version of Borda counting, where the score of each item is the expected payoff-weighted sum of its Borda counts.

3 EVALUATION

Methods. We compare our payoff-based approaches, denoted as MEP and MEP*, with score aggregation, rank aggregation, and fairness-based methods. AVG, LM, and MUL implement the additive, least-misery, and multiplicative score aggregation strategies, respectively [14]; we exclude the maximum-pleasure strategy due to its low performance. BORDA and MEDIAN are two rank aggregation strategies [2], that assign to an item its average or median rank, respectively; MEDIAN is preferred over the Spearman's footrule-based method of [2] as an approximation of the Kemeny optimal rank, due to its lowest time complexity. GRF corresponds to the group rating fairness method of [18], SPG and EPG to the single proportionality and envy-freeness greedy algorithms of [23], and GVAR to the greedy variance algorithm of [11].

Data. The first dataset, denoted as TRAVEL [5], contains 60 groups of 2–4 members, where each user provides a rating, on a 5-point scale, for 11 travel destinations, and each group jointly agrees on their top-2 destinations. As member rankings, we use the users' individual rankings of the 11 items. As hypothetical group ranking, we take the group's ranked list of the two favorable items.

The second dataset, denoted as MovieLens, is based on the MovieLens 1M dataset [8]. Groups of size 3–20 are synthesized under two schemes: in *random*, users are chosen uniformly at random, while in *similar*, starting from a randomly selected user, the group is build incrementally by adding the most similar (in terms of mean-centered cosine similarity) user to the group. Similar to previous work [11, 18, 23], we use a simple matrix factorization technique to fill in the missing ratings in the dataset. As member rankings, we extract the top-N items. For the hypothetical group ranking, we generate a random ranking according to the framework described in Section 2.2.

Metrics. The main evaluation metric, and the optimization target of our methods, is the *Kendall's tau similarity* of the recommended group ranking with the hypothetical group ranking. Recall that our methods are not restricted to Kendall's tau, as the payoff can be defined in terms of any similarity metric between rankings.

As a secondary metric, we use the *normalized discounted cumulative gain* at rank k (NDCG@k). For this metric, each item in the hypothetical group ranking should have a relevance score. As in [11], we use Borda semantics and set the relevance of an item at position r in the top-N hypothetical group ranking to $N - r + 1$. Then, DCG@$k = \sum_{r=1}^{k} \frac{2^{s_r}-1}{\log(r+1)}$, where s_r is the relevance score of the item at position r in the recommended group ranking; IDCG@k is the maximum possible DCG@k; and NDCG@k = DCG@k/IDCG@k.

The reported values are the averages among the 60 groups in TRAVEL, and among 100 randomly generated groups in MovieLens.

Table 4: NDCG@2 for TRAVEL

AVG	LM	MUL	BORDA	MEDIAN	GRF	SPG	EFG	GVAR	MEP	MEP*
0.434	0.264	0.389	0.384	0.363	0.464	0.438	0.368	0.386	**0.488**	0.485

Results on TRAVEL. When evaluating against a ground truth ranking of only two items, Kendall's tau is not a particularly meaningful metric, as it can only take its two extreme values, -1 and 1. On the other hand, NDCG is meaningful and Table 4 presents NDCG@2 for all methods. Our proposals, MEP and MEP* achieve the best scores, followed by two fairness methods GRF and SPG that score slightly better than plain AVG.

Results on MovieLens. In the first experiment, we vary the size of the group from 3 to 20, and ask for the top-20 group recommendations. Figure 2 shows Kendall's tau similarity for random and similar groups. A general observation is that the case of similar groups is easier with all methods achieving higher similarity scores than for random groups. For all group sizes and types, MEP and MEP* are the two best methods by a large margin. Among the aggregation methods, AVG is the best in all cases, a result that is in accordance to previous work [2, 11]. Among the fairness-based methods, GVAR has the best performance, and the main reason is that it weighs fairness and average utility, producing thus rankings very similar to AVG. Specifically, for groups of 5, MEP and MEP* have Kendall's tau similarity of about 0.57 (resp. 0.59) compared to AVG's 0.44 (resp. 0.50) and GVAR's 0.45 (resp. 0.49) for random (resp. similar) groups.

Figure 3 presents the NDCG@5 metric for the same experiment. MEP and MEP* are still the best two methods, but by a smaller margin. Generally, MEP is slightly better than MEP*, except in the case of large groups with similar members. Methods AVG and GVAR are again strong, this time joined by MUL.

Figure 4 fixes the group size at 5 and shows Kendall's tau similarity as we vary the number N of requested recommendations from 5 up to 50. In all cases, our payoff-aware methods outperform all competitors, with the margin increasing with N.

Finally, Figure 5 depicts NDCG at various ranks starting from 2 up to 20 when making top-20 recommendations. The finding are consistent in that MEP and MEP* being the best methods, while AVG and GVAR closely follow. In terms of NDCG, apparently all score aggregation strategies perform equally well for similar groups.

Discussion. Our payoff-aware methods outperform the state of the art in terms of Kendall's tau similarity, as well as in NDCG, meaning that the important top ranks of the recommendation lists are quite accurate. MEP* improves on MEP in the case of large similar groups. We observe that the simplest possible aggregation strategy, AVG, is a very strong competitor, especially in small and similar groups. Fairness methods, with the exception of GVAR, sacrifice accuracy and perform consistently worse than aggregation strategies.

4 CONCLUSION

This work introduced a group recommendation method that explicitly models the uncertainty in the way members might agree on an appropriate group ranking of items. The basic idea is to compute probabilities to hypothetical group rankings by observing member's individual rankings. Then, the systems recommends a ranking that has the highest expected payoff with respect to the hypothetical rankings. Experiments with real and synthetic groups demonstrate the superiority of this approach compared to previous work based on aggregation strategies and to recent fairness-aware techniques.

REFERENCES

[1] Liliana Ardissono, Anna Goy, Giovanna Petrone, Marino Segnan, and Pietro Torasso. 2003. Intrigue: Personalized Recommendation of Tourist Attractions for Desktop and Hand Held Devices. *Applied Artificial Intelligence* 17, 8-9 (2003), 687–714. https://doi.org/10.1080/713827254

[2] Linas Baltrunas, Tadas Makcinskas, and Francesco Ricci. 2010. Group recommendations with rank aggregation and collaborative filtering. In *RecSys*. 119–126. https://doi.org/10.1145/1864708.1864733

[3] Shlomo Berkovsky and Jill Freyne. 2010. Group-based recipe recommendations: analysis of data aggregation strategies. In *RecSys*. 111–118. https://doi.org/10.1145/1864708.1864732

[4] Laura Blédaité and Francesco Ricci. 2015. Pairwise Preferences Elicitation and Exploitation for Conversational Collaborative Filtering. In *Proceedings of the 26th ACM Conference on Hypertext & Social Media, HT 2015, Guzelyurt, TRNC, Cyprus, September 1-4, 2015*. ACM, 231–236. https://doi.org/10.1145/2700171.2791049

[5] Amra Delic, Julia Neidhardt, Thuy Ngoc Nguyen, Francesco Ricci, Laurens Rook, Hannes Werthner, and Markus Zanker. 2016. Observing Group Decision Making Processes. In *RecSys*. 147–150. https://doi.org/10.1145/2959100.2959168

[6] Alexander Felfernig, Martin Stettinger, Ludovico Boratto, and Marko Tkalcic. 2018. *Group Recommender Systems: An Introduction*. Springer US.

[7] Mike Gartrell, Xinyu Xing, Qin Lv, Aaron Beach, Richard Han, Shivakant Mishra, and Karim Seada. 2010. Enhancing group recommendation by incorporating social relationship interactions. In *Proceedings of the 2010 International ACM SIGGROUP Conference on Supporting Group Work, GROUP 2010, Sanibel Island, Florida, USA, November 6-10, 2010*. ACM, 97–106. https://doi.org/10.1145/1880071.1880087

[8] F. Maxwell Harper and Joseph A. Konstan. 2016. The MovieLens Datasets: History and Context. *TiiS* 5, 4 (2016), 19:1–19:19. https://doi.org/10.1145/2827872

[9] Anthony Jameson and Barry Smyth. 2007. Recommendation to Groups. In *The Adaptive Web, Methods and Strategies of Web Personalization*. Springer, 596–627.

[10] Saikishore Kalloori, Francesco Ricci, and Marko Tkalcic. 2016. Pairwise Preferences Based Matrix Factorization and Nearest Neighbor Recommendation Techniques. In *Proceedings of the 10th ACM Conference on Recommender Systems, Boston, MA, USA, September 15-19, 2016*. ACM, 143–146. https://doi.org/10.1145/2959100.2959142

[11] Xiao Lin, Min Zhang, Yongfeng Zhang, Zhaoquan Gu, Yiqun Liu, and Shaoping Ma. 2017. Fairness-Aware Group Recommendation with Pareto-Efficiency. In *Proceedings of the Eleventh ACM Conference on Recommender Systems, RecSys 2017, Como, Italy, August 27-31, 2017*. ACM, 107–115. https://doi.org/10.1145/3109859.3109887

[12] Xingjie Liu, Yuan Tian, Mao Ye, and Wang-Chien Lee. 2012. Exploring personal impact for group recommendation. In *21st ACM International Conference on Information and Knowledge Management, CIKM'12, Maui, HI, USA, October 29 - November 02, 2012*. ACM, 674–683. https://doi.org/10.1145/2396761.2396848

[13] Judith Masthoff. 2004. Group Modeling: Selecting a Sequence of Television Items to Suit a Group of Viewers. *User Model. User-Adapt. Interact.* 14, 1 (2004), 37–85. https://doi.org/10.1023/B:USER.0000010138.79319.fd

[14] Judith Masthoff. 2015. Group Recommender Systems: Aggregation, Satisfaction and Group Attributes. In *Recommender Systems Handbook*. 743–776. https://doi.org/10.1007/978-1-4899-7637-6_22

[15] Judith Masthoff and Albert Gatt. 2006. In pursuit of satisfaction and the prevention of embarrassment: affective state in group recommender systems. *User Model. User-Adapt. Interact.* 16, 3-4 (2006), 281–319. https://doi.org/10.1007/s11257-006-9008-3

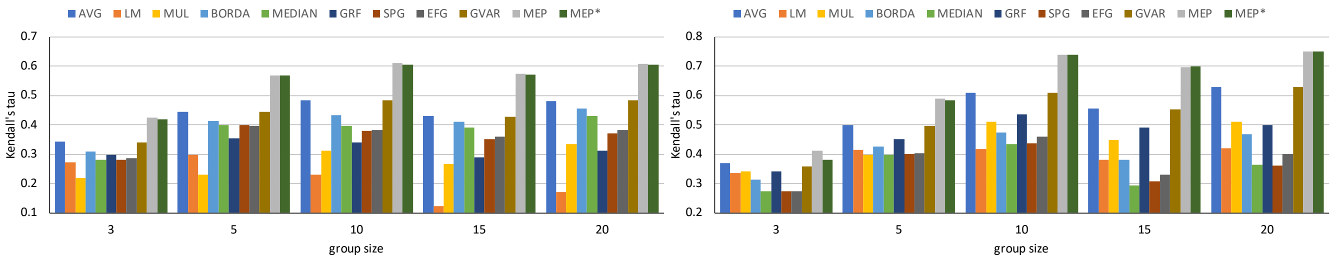
[16] Joseph F. McCarthy and Theodore D. Anagnost. 1998. MusicFX: An Arbiter of Group Preferences for Computer Supported Collaborative Workouts. In *Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work (CSCW '98)*. ACM, New York, NY, USA, 363–372. https://doi.org/10.1145/289444.289511

[17] Mark O'Connor, Dan Cosley, Joseph A. Konstan, and John Riedl. 2001. PolyLens: A recommender system for groups of user. In *ECSCW*. 199–218.

[18] Shuyao Qi, Nikos Mamoulis, Evaggelia Pitoura, and Panayiotis Tsaparas. 2018. Recommending packages with validity constraints to groups of users. *Knowl. Inf. Syst.* 54, 2 (2018), 345–374. https://doi.org/10.1007/s10115-017-1082-9

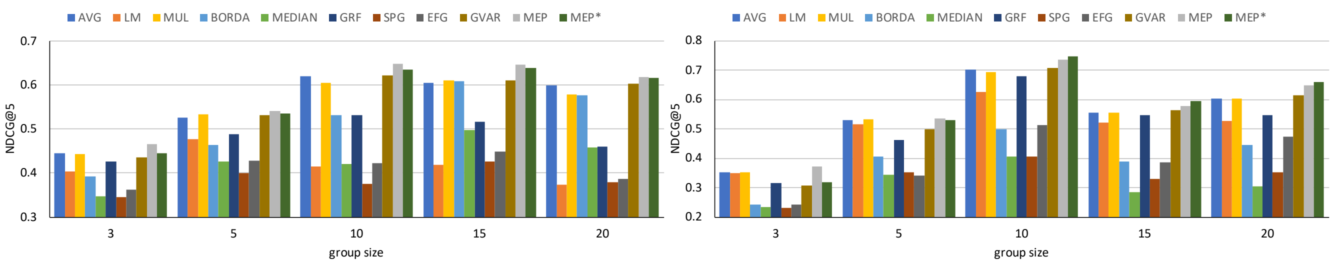
[19] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian Personalized Ranking from Implicit Feedback. In *UAI 2009, Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence, Montreal, QC, Canada, June 18-21, 2009*. AUAI Press, 452–461. https://dslpitt.org/uai/displayArticleDetails.jsp?mmnu=1&smnu=2&article_id=1630&proceeding_id=25

[20] Amirali Salehi-Abari and Craig Boutilier. 2015. Preference-oriented Social Networks: Group Recommendation and Inference. In *Proceedings of the 9th ACM*



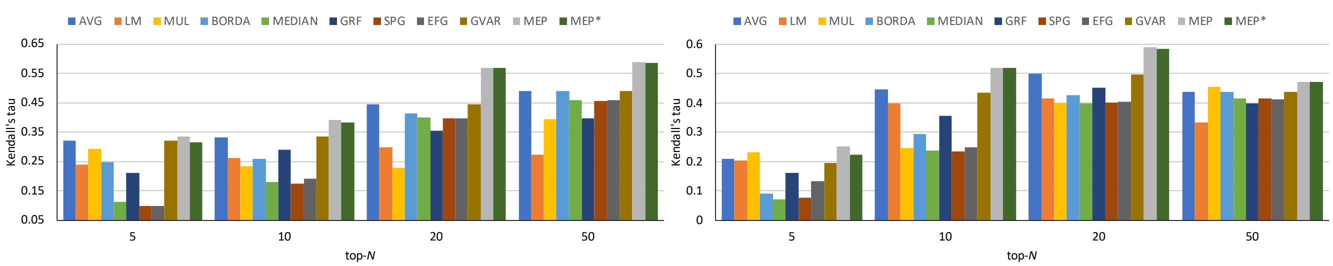
(a) Random groups (b) Similar groups

Figure 2: Kendall's tau similarity vs. group size



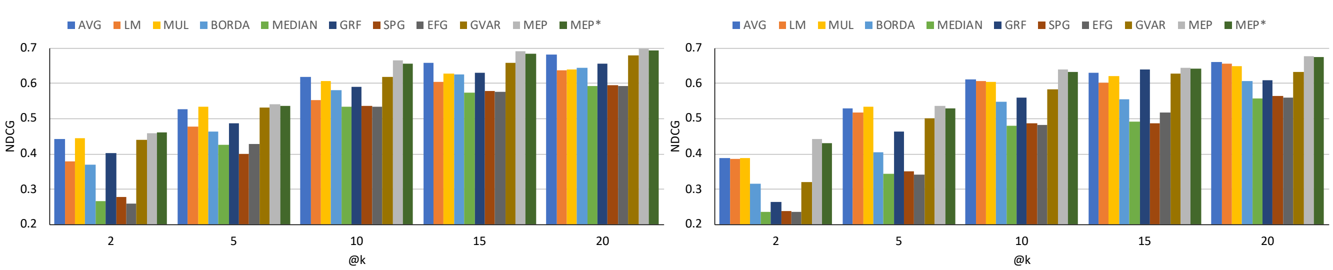
(a) Random groups (b) Similar groups

Figure 3: NDCG@5 vs. group size



(a) Random groups (b) Similar groups

Figure 4: Kendall's tau similarity vs. top-N



(a) Random groups (b) Similar groups

Figure 5: NDCG at various positions

Conference on Recommender Systems, RecSys 2015, Vienna, Austria, September 16-20, 2015. ACM, 35–42. https://doi.org/10.1145/2792838.2800190

[21] Lara Quijano Sánchez, Juan A. Recio-García, Belén Díaz-Agudo, and Guillermo Jiménez-Díaz. 2013. Social factors in group recommender systems. *ACM TIST* 4, 1 (2013), 8:1–8:30. https://doi.org/10.1145/2414425.2414433

[22] Christophe Senot, Dimitre Kostadinov, Makram Bouzid, Jérôme Picault, Armen Aghasaryan, and Cédric Bernier. 2010. Analysis of Strategies for Building Group Profiles. In *User Modeling, Adaptation, and Personalization, 18th International Conference, UMAP 2010, Big Island, HI, USA, June 20-24, 2010. Proceedings (Lecture Notes in Computer Science)*, Vol. 6075. Springer, 40–51. https://doi.org/10.1007/978-3-642-13470-8_6

[23] Dimitris Serbos, Shuyao Qi, Nikos Mamoulis, Evaggelia Pitoura, and Panayiotis Tsaparas. 2017. Fairness in Package-to-Group Recommendations. In *Proceedings of the 26th International Conference on World Wide Web, WWW 2017, Perth, Australia, April 3-7, 2017.* ACM, 371–379. https://doi.org/10.1145/3038912.3052612

[24] Zhiwen Yu, Xingshe Zhou, Yanbin Hao, and Jianhua Gu. 2006. TV Program Recommendation for Multiple Viewers Based on user Profile Merging. *User Model. User-Adapt. Interact.* 16, 1 (2006), 63–82. https://doi.org/10.1007/s11257-006-9005-6