Group Recommender Systems: Beyond Preference Aggregation

Judith Masthoff and Amra Delić

1 Introduction

Most work on recommender systems to date focuses on recommending items to individual users. For instance, they may select a book for a particular user to read based on a model of that user's preferences in the past. Here, preferences are considered to be either implicit (e.g., clicks, purchase, viewing time, etc.), or explicit (e.g., likes, ratings, rankings, etc.) [1]. The challenge recommender system designers traditionally faced is how to decide what would be optimal for an individual user. A lot of progress has been made on this, as evidenced by other chapters in this handbook (e.g., [2–4]).

In this chapter, we go one-step further. There are many situations when it would be good if we could recommend to a group of users rather than to an individual. For instance, a recommender system may select television programmes for a group to view or a sequence of songs to listen to, based on individual preferences of all group members. Recommending to groups is more complicated than recommending to individuals. Assuming that we know perfectly what is good for individual users, the issue arises how to define and find what is also good for a group composed by the same individuals. The usual approach to solve this issue usually revolves around combining individual user preferences into a group preference model, or recommendations tailored for individuals into group recommendations. Methods or algorithms that combine individual user preferences or recommendations are called preference aggregation strategies. In this chapter, we will discuss how group

J. Masthoff (✉)
Utrecht University, Utrecht, Netherlands
e-mail: j.f.m.masthoff@uu.nl

A. Delić
University of Sarajevo, Sarajevo, Bosnia and Herzegovina
e-mail: adelic@etf.unsa.ba

F. Ricci et al. (eds.), *Recommender Systems Handbook*,
https://doi.org/10.1007/978-1-0716-2197-4_10

recommendation works, what its problems are, and what advances have been made. Interestingly, we will show that group recommendation techniques have uses similarly to what has been found for individual recommendations. So, even if you are developing recommender systems aimed at individual users you may still want to read on (perhaps reading Sect. 8 first will convince you).

In recent years, there has been increased attention to so called "multistakeholder" recommendation [5] also in detail discussed in Chapter "Multistakeholder Recommender Systems", and it is worth noting its relationship to group recommendations. Multistakeholder recommendation aims to include the perspectives and utilities of multiple stakeholders, i.e., "any group or individual that can affect, or is affected by, the delivery of recommendations to users" [5]. In that sense, group recommendation is a special case where each individual group member is a stakeholder, and the recommendations are designed to satisfy distinct stakeholders in the group.

Moreover, there is increasing attention in machine learning, and hence also in recommender systems, on fairness, which is mainly concerned with ensuring all subgroups of users (e.g., different genders) are treated equally, and are given the same opportunities with the provided recommendations (e.g., job recommendations). This topic is especially interesting in the context of group recommendation (and other multistakeholder recommendation) where fairness-related criteria become relevant on multiple sides, perhaps also from multiple perspectives [6]. In group recommenders fairness is usually considered as a characteristic of an aggregation strategy which ensures that none of the group members is impaired by the group recommendations [7, 8].

There are other issues to consider when building a group recommender system which are outside the scope of this chapter. In particular:

- *How to acquire information about individual users' preferences.* The usual recommender techniques can be used (such as explicit ratings and implicit feedback collection, see other handbook chapters). There is a complication in that it is difficult to infer an individual's preferences when a group uses the system, but inferences can be made during individual use combined with a probabilistic model when using it in company. An additional complication is that an individual's ratings may depend on the group they are in. For instance, a teenager may be very happy to watch a programme with his younger siblings, but may not want to see it when with his friends.
- *How will the system know who is present?* Different solutions exist, such as users explicitly logging in, probabilistic mechanisms using the time of day to predict who is present, the use of tokens and tags, etc. [9]. More sophisticated approaches have been used in recent years. For example, the GAIN system divides the group into a known subgroup (users which it knows are there) and an unknown subgroup (users that cannot be recognized but should be there statistically) [10]. A group recommender in a public display system recognizes the gender, emotions and group structures of people present (which are alone and which with others) [11].

- *How to present and explain group recommendations?* As seen in this handbook's chapter on explanations, there are already many considerations when presenting and explaining *individual* recommendations. The case of group recommendations is even more difficult. More discussion on explaining group recommendations is provided in [12] and under Challenges in our final section.
- *How to help users to settle on a final decision?* In some group recommenders, users are given group recommendations, and based on these recommendations negotiate what to do. In other group recommenders this is not an issue (see Sect. 2.3 on the difference between passive and active groups). An overview of how users' decisions can be aided is provided in [12].

The next section highlights usage scenarios of group recommenders, and provides a classification of group recommenders inspired by differences between the scenarios. Section 3 discusses strategies for aggregating models of individual users to allow for group recommendation, what strategies have been used in existing systems, and what was learned from experiments in this area. Section 4 deals with the issue of order when we want to recommend a sequence of items. Section 5 provides an introduction into the modeling of affective state, including how an individual's affective state can be influenced by the affective states of other group members. Section 6 explores how such a model of affective state can be used to build more sophisticated aggregation strategies. Section 7 discusses other group attributes (such as personality of users) that can be used in aggregation strategies. Section 8 shows how group modeling and group recommendation techniques can be used when recommending to an individual user. Section 9 concludes this chapter and discusses future challenges.

2 Usage Scenarios and Classification of Group Recommenders

There are many circumstances in which recommendation to a group is needed rather than to an individual. Below, we present two scenarios that inspired our own work in this area, discuss the scenarios underlying related work, and provide a classification of group recommenders inspired by differences between the scenarios.

2.1 Usage Scenario 1: Interactive Television

Interactive television offers the possibility of personalized viewing experiences. For instance, instead of everybody watching the same news program, it could be personalized to the viewer. For Judith, this could mean adding more stories about the Netherlands (where she comes from), China (a country that fascinates her after having spent some holidays there) and football, but removing stories about cricket (a sport she hardly understands) and local crime. Similarly, music programs could be adapted to show music clips that a user actually likes.

There are two main differences between traditional recommendation as it applies to say PC-based software and the interactive TV scenarios sketched above. Firstly, in contrast to the use of PCs, television viewing is largely a family or social activity. So, instead of adapting the news to an individual viewer, the television would have to adapt it to the group of people sitting in front of it at that time. Secondly, traditional work on recommendation was often concerned recommending one particular thing to the user, so for instance, which movie the user should watch. In the scenarios sketched above, the television needs to adapt a sequence of items (news items, music clips) to the viewer. The combination of recommending to a group and recommending a sequence is very interesting, as it may allow you to keep all individuals in the group satisfied by compensating for items a particular user dislikes with other items in the sequence which they do like.

2.2 Usage Scenario 2: Ambient Intelligence

Ambient intelligence deals with designing physical environments that are sensitive and responsive to the presence of people. For instance, consider the case of a bookstore where sensors detect the presence of customers identified by some portable device (e.g., a Bluetooth-enabled mobile phone, or a fidelity card equipped with an active RFID tag). In this scenario, there are various sensors distributed among the shelves and sections of the bookstore which are able to detect the presence of individual customers. The bookstore can associate the identification of customers with their profiling information, such as preferences, buying patterns and so on.

With this infrastructure in place, the bookstore can provide customers with a responsive environment that would adapt to maximize their well-being with a view to increasing sales. For instance, the device playing the background music should take into account the preferences of the group of customers within hearing distance. Similarly, LCD displays scattered in the store show recommended books based on the customers nearby, the lights on the shop's display window (showing new titles) can be rearranged to reflect the preferences and interests of the group of customers watching it, and so on. Clearly, group adaptation is needed, as most physical environments will be used by multiple people at the same time.

2.3 Usage Scenarios Underlying Related Work

In this section we discuss the scenarios underlying some of the best known group recommender systems:

- MusicFX [13] chooses a radio station for background music in a fitness center, to suit a group of people working out at a given time. This is similar to the Ambient Intelligence scenario discussed above.
- PolyLens [14] is an extension of MovieLens. MovieLens recommends movies based on an individual's taste as inferred from ratings and social filtering. PolyLens allows users to create groups and ask for group recommendations.
- Intrigue [15] recommends places to visit for tourist groups taking into account characteristics of subgroups within that group (such as children and the disabled).
- The Travel Decision Forum [16] helps a group to agree on the desired attributes of a planned joint holiday. Users indicate their preferences on a set of features (like sport and room facilities). For each feature, the system aggregates the individual preferences, and users interact with embodied conversational agents representing other group members to reach an accepted group preference.
- The Collaborative Advisory Travel System (CATS) [17] also recommends a joint holiday. Users consider holiday packages, and critique their features (e.g., 'like the one shown but with a swimming pool'). Based on these critiques, which are combined into a group preference model, the system recommends other holidays. Users also select holidays they like for other members to see, and these are annotated with how well they match the preferences of each group member.
- Yu's TV Recommender [18] recommends a television programme for a group to watch. It bases its recommendation on the individuals' preferences for programme features (such as genre, actors, keywords).
- The Group Adaptive Information and News system (GAIN) [10] adapts the display of news and advertisements to the group of people near it.
- The Reminiscence Therapy Enhanced Material Profiling in Alzheimers and other Dementias system (REMPAD) [19] recommends multimedia material to be used by a facilitator in a group reminiscence therapy session, based on the suitability of material for individual participants as inferred from their date of birth, locations lived in, and interest vectors.
- HappyMovie [20] recommends movies to groups, by enriching the group preference model with the individuals' personality (assertiveness and cooperativeness) and the relationship strengths (they call this social trust) between individuals.
- IntelliReq [21] supports groups in deciding which software requirements to implement. Users can view and discuss recommendations for group decisions based on already defined user preferences.
- Where2Eat [22] is a mobile group recommender for restaurants. It implements *"interactive multi-party critiquing"*, which allows group members to generate proposals and counter-proposals until an agreement is reached.
- Choicla [23] is a group decision support environment that allows users to configure decision tasks and decision-making process in a domain-independent setting.

- HOOTLE [24] is a hotel group recommender that supports negotiation about the features of a desired hotel. After aggregating individual preferences into a group model, content-based filtering is used to generate group recommendations.
- STSGROUP (South Tyrol Suggests for Group) [25–27] is a chat-based, context-aware mobile application for recommending Points of Interest (POIs). An innovative feature of this system is that it combines users' long-term preferences (based on users' previous individual interactions with the system), with the dynamic preferences elicited during the group decision-making process.
- TOURREC [28] recommends routes to individuals and groups. A route is generated by finding POIs which best fit the user profile and the contextual information, and are located between the user-defined start and end point. Then, the system connects POIs in recommendations tailored for individuals. To generate a group recommendation, various social choice strategies were implemented to aggregate users' individual travel preferences. The system also allows the group to split up during the trip, so that every member is still able to visit their own favorite POIs.

2.4 A Classification of Group Recommenders

The scenarios provided above differ on several dimensions, which provide a way to classify group recommender systems:

- *Individual preferences are known versus developed over time.* In most scenarios, the group recommender starts with individual preferences. In contrast, in CATS, individual preferences develop over time, using a critiquing style approach. Others have also adopted this critiquing approach (e.g., [22]). In INTELLIREQ, individual preferences can be influenced by the group discussion and the group recommendation based on preferences defined so far. In STSGROUP individual preferences are known, but group-related preferences are developed over time, and the profile of an individual user is also updated based on the group interaction.
- *Recommended items are experienced by the group versus presented as options.* In the Interactive TV scenario, the group experiences the news items. In the Ambient Intelligence, GAIN, and MUSICFX scenarios, they experience the music and advertisements. In contrast, in the other scenarios, they are presented with a list of recommendations, e.g., POLYLENS.
- *The group is passive versus active.* In most scenarios, the group does not interact with the way individual preferences are aggregated. However, in the TRAVEL DECISION FORUM and CATS the group negotiates the group model. In INTELLIREQ, the group does not influence the aggregation, but may influence the ratings provided.
- *Negotiation versus the single-shot recommendations.* Here we distinguish whether the system allows the group to interact with the suggested

recommendations, thus, allowing them to go into several rounds of negotiation [29]. In MUSICFX, recommendations are delivered to a group without an option to interact with the system (e.g., skip a recommended item). In a similar fashion, POLYLENS delivers recommendations in a "single-shot". In contrast, TRAVEL DECISION FORUM, CATS, INTELLIREQ, STSGROUP allow users to discuss proposals made by the system.
- *Recommending a single item versus a sequence.* In the scenarios of MUSICFX, POLYLENS, and YU'S TV RECOMMENDER it is sufficient to recommend individual items: people normally only see one movie per evening, radio stations can play forever, and YU'S TV RECOMMENDER chooses one TV program only. In contrast, in our Interactive TV scenario, a sequence of items is recommended, for example making up a complete news broadcast. Similarly, in INTRIGUE and TOURREC, it is quite likely that a tourist group would visit multiple attractions during their trip, so would be interested in a sequence of attractions to visit. Also, in the Ambient Intelligence scenario it is likely that a user will hear multiple songs, or see multiple items on in-store displays. In GAIN, the display shows multiple items simultaneously; additionally, the display is updated every 7 min, so people are likely to see a sequence as well.

DeCampos et al.'s classification of group recommenders also distinguishes between passive and active groups [30]. In addition, it uses two other dimensions:

- *How individual preferences are estimated.* They distinguish between content-based and collaborative filtering. Of the systems mentioned above, POLYLENS and HAPPYMOVIE use collaborative filtering; the others use content-based filtering (e.g., REMPAD).
- *Whether profiles or recommendations are aggregated.* In the first case, profiles, containing group members' individual preferences usually expressed numerically for each item, are aggregated into a group model. In the second case, recommendations are produced for individuals and then aggregated into a group recommendation. The two approaches are slightly different, since preferences are usually numerical values, while recommendations are items. However, recommendations can be generated (1) as ranked lists, and then individual lists are aggregated into the group list, or (2) as numerical estimations assigned to individual recommendations which then enable their aggregation in a similar fashion as aggregating profiles. They mention INTRIGUE and POLYLENS as aggregating recommendations, while the others tend to aggregate profiles. Aggregating profiles can happen in multiple ways. In this chapter, we will look at the aggregation of preference ratings. It is also possible to aggregate content: for example, GroupReM aggregates individuals' tag cloud profiles to produce a group tag cloud profile [31]. It is also possible to use a combination of aggregating profiles and aggregating recommendations: [32] proposes a hybrid switching approach that uses aggregated recommendations when user data is sparse and aggregated profiles otherwise. Following their example, [33] also uses a combination.

These two dimensions are related to how the group recommender is implemented rather than being inherent to the usage scenario. In this chapter, we focus on aggregating profiles, but the same aggregation strategies apply when aggregating recommendations. The material presented in this chapter is independent of how the individual preferences are obtained.

3 Aggregation Strategies

The main problem group recommendation needs to solve is how to adapt to the group as a whole based on information about individual users' likes and dislikes. For instance, suppose the group contains three people: Peter, Jane and Mary. Suppose a system is aware that these three individuals are present and knows their interest in each of a set of items (e.g., music clips or advertisements). Table 1 gives example ratings on a scale of 1 (really hate) to 10 (really like). Which items should the system recommend, given time for four items?

3.1 Overview of Aggregation Strategies

Many strategies exist for aggregating individual ratings into a group rating (e.g., used in elections and when selecting a party leader). For example, the Least Misery Strategy uses the minimum of ratings to avoid misery for group members (Table 2). These strategies can then be used to rank items according to their relevance for the group at hand, hence to estimate what should be recommended to the group.

Eleven aggregation strategies inspired by Social Choice Theory are summarized in Table 3 (see [9] for more details). In [34], aggregation strategies are classified into (1) *majority-based strategies* that use the most popular items (e.g., Plurality Voting), (2) *consensus-based strategies* that consider the preferences of all group members (e.g., Average, Average without Misery, Fairness), and (3) *borderline strategies* that only consider a subset (e.g., Dictatorship, Least Misery, Most Pleasure).

Table 1 Example of individual ratings for ten items (A to J)

	A	B	C	D	E	F	G	H	I	J
Peter	10	4	3	6	10	9	6	8	10	8
Jane	1	9	8	9	7	9	6	9	3	8
Mary	10	5	2	7	9	8	5	6	7	6

Table 2 Example of the least misery strategy

	A	B	C	D	E	F	G	H	I	J
Peter	10	4	3	6	10	9	6	8	10	8
Jane	1	9	8	9	7	9	6	9	3	8
Mary	10	5	2	7	9	8	5	6	7	6
Group rating	1	4	2	6	7	8	5	6	3	6

Table 3 Overview of aggregation strategies

Strategy	How it works	Example
Plurality/majority voting	Uses 'first past the post': repetitively, the item with the most votes is chosen.	A is chosen first, as it has the highest rating for the majority of the group, followed by E (which has the highest rating for the majority when excluding A)
Average	Averages individual ratings	B's group rating is 6, namely (4+9+5)/3
Multiplicative	Multiplies individual ratings	B's group rating is 180, namely 4*9*5
Borda count	Counts points from items' rankings in the individuals' preference lists, with bottom item getting 0 points, next one up getting one point, etc.	A's group rating is 17, namely 0 (last for Jane) + 9 (first for Mary) + 8 (shared top 3 for Peter)
Copeland rule	Counts how often an item beats other items (using majority vote[a]) minus how often it looses	F's group rating is 5, as F beats 7 items (B,C,D,G,H,I,J) and looses from 2 (A,E)
Approval voting	Counts the individuals with ratings for the item above approval threshold (e.g. 6)	B's group rating is 1 and F's is 3
Least Misery	Takes the minimum of individual ratings	B's group rating is 4, namely the smallest of 4,9,5
Most pleasure	Takes the maximum of individual ratings	B's group rating is 9, namely the largest of 4,9,5
Average without Misery	Averages individual ratings, after excluding items with individual ratings below a certain threshold (say 4).	J's group rating is 7.3 (the average of 8,8,6), while A is excluded because Jane hates it
Fairness	Items are ranked as if individuals are choosing them in turn.	Item E may be chosen first (highest for Peter), followed by F (highest for Jane) and A (highest for Mary)
Most respected person (or Dictatorship)	Uses the rating of the most respected individual.	If Jane is the most respected person, then A's group rating is 1. If Mary is most respected, then it is 10

[a] If the majority of group members have a higher rating for an item X than for an item Y, then item X beats item Y

3.2 *Aggregation Strategies Used in Related Work*

Most of the related work uses one of the aggregation strategies in Table 3 (sometimes with a small variation), and they differ in the one used:

- INTRIGUE uses a weighted form of the Average Strategy. It bases its group recommendations on the preferences of subgroups, such as children or disabled.

It takes the average, with weights depending on the number of people in the subgroup and the subgroup's relevance (children and disabled were given a higher relevance).

- POLYLENS uses the Least Misery Strategy, assuming groups of people going to watch a movie together tend to be small and that a small group tends to be as happy as its least happy member.
- MUSICFX uses a variant of the Average Without Misery Strategy. Users rate all radio stations, from +2 (really love this music) to −2 (really hate this music). These ratings are converted to positive numbers (by adding 2) and then squared to widen the gap between popular and less popular stations. An Average Without Misery Strategy is used to generate a group list: the average of ratings is taken but only for those items with individual ratings all above a threshold. To avoid starvation and always picking the same station, a weighted random selection is made from the top stations of the list.
- YU'S TV RECOMMENDER uses a variant of the Average Strategy. It bases its group recommendation on individuals' ratings of program features: −1 (dislikes the feature), +1 (likes the feature) and 0 (neutral). The feature vector for the group minimizes its distance compared to individual members' feature vectors (see [18] for detail). This is similar to taking the average rating per feature.
- The TRAVEL DECISION FORUM has implemented multiple strategies, including the Average Strategy and the Median Strategy. The Median Strategy (not in Table 3) uses the middle value of the ratings. So, in our example, this results in group ratings of 10 for A, and 9 for F. The Median Strategy was chosen because it is non-manipulable: users cannot steer the outcome to their advantage by deliberately giving extreme ratings that do not truly reflect their opinions. In contrast, for example, with the Least Misery strategy devious users can avoid getting items they dislike slightly, by giving extremely negative ratings. The issue of manipulability is most relevant when users provide explicit ratings, used for group recommendation only, and are aware of others' ratings, all of which is the case in the TRAVEL DECISION FORUM. It is less relevant when ratings are inferred from user behavior, also used for individual recommendations, and users are unaware of the ratings of others (or even of the aggregation strategy used).
- In CATS, users indicate through critiquing which features a holiday needs to have. For certain features, users indicate whether they are required (e.g., ice skating required). For others, they indicate quantities (e.g., at least 3 ski lifts required). The group model contains the requirements of all users, and the item which fulfills most requirements is recommended. Users can also completely discard holidays, so, the strategy has a Without Misery aspect.
- GAIN uses a variant of the Average Strategy, with different weights for users that the system knows are near the system and for unrecognized users who should be there statistically.
- The REMPAD system, having such a specific and especially sensitive task, has to reduce any negative effects that the suggested media material could cause for individuals in the therapy group, and to this end uses the Least Misery Strategy.

- HappyMovie uses the Average Strategy, but prior to the aggregation, the system updates individuals' ratings as a result of influence in the group which is estimated according to group members' personality (assertiveness and cooperativeness) and their inter-personal trust levels.
- IntelliReq allows participants to select their preferred configuration of software requirements, defined upon multiple dimensions, and applies Plurality Voting.
- Where2Eat does not aggregate individual preferences of group members. Instead, it allows group members to adjust (critique) the features of a restaurant proposed by a fellow group member in the previous step, and to make a counter proposal. The counter proposal is automatically explained to the fellow group member in terms of differences to her original proposal.
- Choicla supports a configurable, domain-independent group decision-making process, where the initiator defines the decision task as well as some basic settings of the decision process. The settings also include the selection of the aggregation strategy, and some of the available options are Majority Vote, Average Vote, Least Misery, Most Pleasure.
- Hootle aggregates group members' individual preferences on the set of desirable attributes of a hotel. Hence, each group member gives an importance score for her most important attributes, or she can put a veto on an attribute. However, group members can express their preferences only on a number of attributes where the number depends on how restrictive the attribute is for the output of the recommender system. To aggregate individual preferences the system uses a variant of the Borda Count, so that it accounts not only for the individuals' importance scores, but also for vetoed options and restrictions [35].
- STSGroup is a content-based recommender system which considers users' individual long-term preferences, as well as session-based preferences which the group members expressed as reactions on each-others' proposals (i.e., best choice, like, dislike). At the beginning of the discussion process, group members' individual utility vectors are aggregated to a group utility vector with the Average Strategy. However, when the interaction begins, the weighted average strategy is used, and the weights are calculated based on the proportion of the user's actions (POI proposals, POI evaluations and POI comments) over the total number of actions acquired in the group [29].
- TourRec introduces two extensions to the standard aggregation strategies that allow groups to separate during their POI sequence route. In the first one, *Split*, after aggregating individual "profits" to visit a POI, and generating group recommendations, individual recommendations are also generated, which the system uses to check whether there are individuals for whom it would pay off to exchange a certain POI on the group route with their individually more preferred POI, but making sure that the distance to and from that POI is reasonable. In the second one, *Connect Segments (CS)*, individual recommendations are aggregated in a way that each group member has two of their favourite POIs in the group route.

It should be noted that Yu's TV Recommender, the Travel Decision Forum and IntelliReq aggregate preferences for each feature without using the idea of fairness: loosing out on one feature is not compensated by getting your way on another. In addition to the strategies in Table 3, more complex strategies and recommendation algorithms have been used:[1]

Heuristics-Based Methods In this group of strategies, we combine methods that make certain assumptions about group decision-making process, individual and group preferences, and accordingly define the aggregation strategy. The *graph-based ranking* algorithm [36] uses (1) a graph with users and items as nodes, with positive links between users and items rated above the user's average item rating, and negative links for items rated below (with weights of how much above/below), (2) a user neighborhood graph linking users with similar rating patterns, and (3) an item neighborhood graph linking items that have been rated similarly. Group recommendations are based on two random walks over the graphs, with the assumption that highly visited items over positive links would tend to be liked by the group, and items highly visited by a random walk over negative links would tend to be disliked by the group. The *Spearman footrule rank* aggregation [37] defines that the aggregated list for the group is a list with minimum distance to the individual lists. The Spearman footrule distance between two lists is the summation of absolute differences between the ranks of the items in the lists. The *Nash equilibrium* [38] models group members as players in a non-cooperative game, and players' actions as item recommendations (choosing from their top 3 items). Group satisfaction is achieved by finding the Nash equilibrium in the game. Finally, the *purity* and *completeness* strategies [39] are statistics-based strategies. The purity is a statistical dispersion strategy that aims to satisfy as many group members' preferences as possible (considering the deviation in preferences). The completeness models group recommendation as a negotiation, favoring high scores whilst penalizing large differences between members.

Influence-Based Methods In this group of strategies, we combine approaches which define preference-based influence and use it in the aggregation method. The *Pre-GROD* and *GROD* [40] first calculate a pairwise similarity matrix, considered as a "trust" matrix, where elements represent how much importance the group members assign to the opinion of others in the group. As it is assumed that members update their opinions according to the influence in the group, the approach updates their individual ratings according to the similarity matrix until a "stability point" (consensus) is reached. Group scores for individual items are the average of the updated opinions. When a group consensus cannot be reached, the GROD approach updates the similarity matrix by adding elements that ensure consensus reaching. The *Preference network (PrefNet)* and *Non-linear preferences (Non-lin)* methods [41] modify importance and preferences of group members (respectively) prior to applying the standard aggregation strategies, e.g., average. PrefNet represents user

[1] These strategies are too complicated to fully explain here, see the original papers for details.

preferences in a network structure based on group members' pairwise similarities, and weights group members according to their centrality in that network - giving more importance to those who share a greater deal of preferences with others in the group. Non-lin transforms individual preferences to a non-linear scale, thus allowing stronger positive influence of highly ranked items and stronger negative influence of items ranked at the bottom of individual lists. Finally, *MAGReS* (Multi-Agent Group Recommender System) [42] implements user agents that follow the *Monotonic Concession Protocol* (MPC). Agents suggest items according to the utility value which that item has for their corresponding user, and assesses items proposed by other agents in the group. An agreement is reached when an agent suggests an item which is at least as good, for any other agent, as their own current suggestion. If an agreement is not reached an agent makes a concession. The procedure is repeated until an agreement is found or until no agent can concede anymore.

Matrix Factorization and Deep Learning Methods The *After Factorization (AF)*, *Before Factorization (BF)* and *Weighted Before Factorization (WBF)* methods [43] all employ Matrix Factorization (MF), with the only difference whether the individual preferences are aggregated (with standard aggregation strategies) prior or after the MF. The interesting approach here is WBF which aggregates users' individual ratings, but by assigning more weights to items that have been rated by the majority of the group, and have similar ratings by the users in the group. The *AGREE* approach [44] uses a neural network to learn the aggregation strategy and item-dependant weights of each group member. The network extends the structure of the well known Neural Collaborative Filtering (NCF) approach [45] by learning not only the user-item interactions, but as well, group-item interactions. *MoSAN* [46] is another neural approach that models member user-user interactions with the help of attention networks, for ad-hoc groups which did not have previous interactions. For each group, a set of sub-attention networks is created (one for each group member). The network aims to capture decisions of each group member, given the decisions of others in that group, hence the influence. Finally, the *GroupIM* approach [47] trains a neural network to compute probabilities that a group would interact with an item by minimizing (1) the group recommendation loss between history of group-item interactions and the predicted item-probabilities, (2) the contextually weighted user-item loss, and (3) maximizing the mutual information between the group and the member user.

Fairness-Maximization Methods The *SPGreedy* and *EFGreedy* strategies [8] use a greedy algorithm to add items to a package recommendation that maximize fairness proportionality or envy-freeness. Fairness proportionality ensures that each group member gets at least *m* items that she likes, compared to items not in the package, whole envy-freeness looks into the other group members and makes sure that for each member there is a sufficient number of items that she likes more than other group members. The *Greedy-LM* and *Greedy-Var* strategies [7], in a similar fashion, use a greedy algorithm and scalarization method to find a Pareto Efficient solution to a multi-objective optimization problem which maximizes the social welfare of a group and fairness. The social welfare is defined as the average of group members' individual utilities, while fairness is defined with the least misery approach (Greedy-LM), or with the variance of the individual utilities (Greedy-Var).

3.3 Which Strategy Performs the Best

Though some exploratory evaluation of MUSICFX, POLYLENS and CATS has taken place, for none of these systems it has been investigated how effective their strategy really is, and what the effect would be of using a different strategy. The experiments presented in this section shed some light on this question. Moreover, systems like WHERE2EAT, CHOICLA, STSGROUP, and HOOTLE each adopt a quite unique approach of delivering recommendations, and hence do not evaluate the performance of different aggregation strategies, but they do evaluate in their own context the usability of the system and performance of the proposed approach. In contrast, some evaluation of YU'S TV RECOMMENDER has taken place [18]. They found that their aggregation worked well when the group was quite homogeneous, but that results were disliked when the group was quite heterogeneous. This is as we would expect, given the Average Strategy will make individuals quite happy if they are quite similar, but will cause misery when tastes differ widely.

In [9], a series of experiments were conducted to investigate which strategy from Table 3 is the best in terms of (perceived) group satisfaction. The Experiment 1 (see Fig. 1) investigated how people would solve the group recommendation problem, using the User as Wizard evaluation method [48]. Participants were given individual ratings identical to those in Table 1. These ratings were chosen to be able to distinguish between strategies. Participants were asked which items the group should watch, if there was time for one, two,…, seven items. Participants' decisions were compared and rationale with those of the aggregation strategies. It was found that participants cared about fairness, and about preventing misery and starvation ("this one is for Mary, as she has had nothing she liked so far"). Participants' behavior reflected that of several of the strategies (e.g. the Average, Least Misery, and Average Without Misery were used), while other strategies (e.g. Borda count, Copeland rule) were clearly not used.[2]

In Experiment 2 (see Fig. 2), participants were given item sequences chosen by the aggregation strategies as well as the individual ratings in Table 1. They rated how satisfied they thought the group members would be with those sequences, and explained their ratings. We found that the Multiplicative Strategy performed the best, in the sense that it was the only strategy for which *all* participants thought its sequence would keep all members of the group satisfied. Borda count, Average, Average without Misery and Most Pleasure also performed quite well. Several

[2] This does not necessarily mean that these strategies are bad, as complexity can also play a role. In fact, in Experiment 2 Borda count was amongst the best performing strategies.

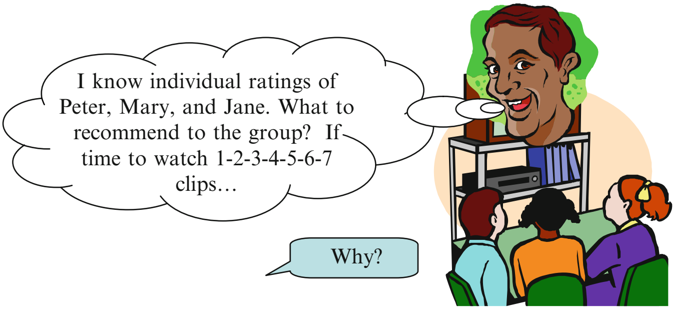

Fig. 1 Experiment 1: which sequence of items do people select if given the system's task

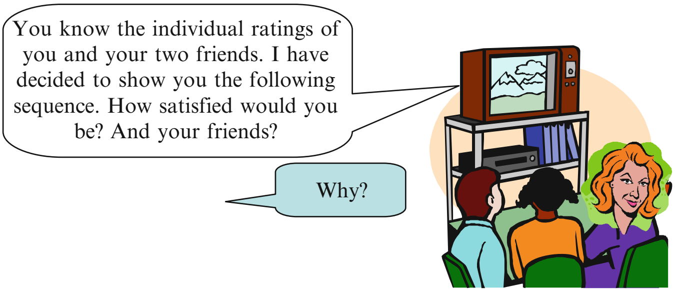

Fig. 2 Experiment 2: What do people like?

strategies (such as Copeland rule, Plurality voting, Least misery) could be discarded as they clearly were judged to result in misery for group members.

The participants' judgments were also compared with predictions of simple satisfaction modeling functions. Amongst other, it was found that more accurate predictions[3] resulted from using:

- quadratic ratings,[4] which e.g., makes the difference between a rating of 9 and 10 bigger than that between a rating of 5 and 6.
- normalization,[5] which takes into account that people rate in different ways, e.g., some always use the extremes, while others only use the middle of the scale.

[3] In terms of satisfaction functions predicting the same relative satisfaction scores for group members as predicted by participants, see [9] for details.

[4] A rating r was transformed into $(\text{r-scale\_midpoint})^2$ if r$\geq$ scale_midpoint, and $-(\text{r-scale\_midpoint})^2$ if r<scale_midpoint.

[5] A rating r was transformed by a user u into r $\times$ (TotalRatingsAverage $\div$ TotalRatings(u)), where TotalRatingsAverage is the sum for all items of the average ratings by all users, and TotalRatings(u) is the sum for all items of u's rating.

In [49], a further study using simulated users based on models of affective state (see next Section) was conducted. It was found that the Multiplicative Strategy performed the best.

There are other studies investigating the effect of different aggregation strategies from Table 3, and the more advanced methods. Table 4 provides an overview of evaluations of aggregation strategies, but before looking into the results of these studies, we will talk a bit more about evaluating group recommender systems.

3.4 Evaluating Group Recommender Systems

To evaluate and measure the quality of group recommender systems user studies as well as off-line experiments have been employed [50]. However, how to properly evaluate a group recommender system is still a major research topic.

3.4.1 User Studies

A user study is carried out when the criteria used to measure the system performance are related to system usability and user experiences (e.g., perceived user's satisfaction or recommendation quality), such as those in WHERE2EAT [22], CHOICLA [23], STSGROUP [25], or TOURREC [28]. This type of study can be conducted by directly interviewing participants or through crowd sourcing sites such as Amazon Mechanical Turk [51]. This approach, however, cannot be the sole method for evaluating the efficacy of a system as the approach does not scale. Nevertheless, user studies are crucial in determining the success of a proposed approach, to evaluate whether or not the system is even accepted by the users [1].

3.4.2 Off-Line Evaluations

As when evaluating single-user recommendations, off-line evaluations are also used in group recommenders research. However, these approaches are hindered since there is a lack of publicly available data sets that capture the preferences of users in actual group settings.

As one of the solutions for this problem, researchers have used so called synthetic groups. Synthetic groups are artificial groups sampled from standard data sets such as MovieLens [33, 37]. The main challenge of this solution is how to properly evaluate effectiveness of group recommendations once they are delivered to such groups. One approach is to compare the group recommendations with the joint group assessment of the recommended items, however this is indeed problematic when in essence the actual group choice is not known. Therefore, it is assumed that the "true" group preferences can be derived from individual preferences with the Average Strategy, which is exactly the problem that group recommenders aim

Table 4 Evaluation of aggregation strategies

Who	Domain	Evaluation methodology	Groups	Strategies	Results
Masthoff [9]	TV	Experiment 1 above user as wizard	Size: 3 Friends Heterogeneous	All from Table 3.	USED: average, average without misery, least misery. NOT USED: borda, approval, plurality, copeland
Masthoff [9]	TV	Experiment 2 above user as wizard	Size: 3 friends heterogeneous	All from Table 3.	BEST: multiplicative. OK: borda, average, average without misery, most pleasure. WORST: copeland, plurality, least misery
Masthoff and Gatt [49]	TV	Simulated users metric: satisfaction functions.	Size: 3 friends heterogeneous	All from Table 3.	BEST: multiplicative. WORST: borda, plurality, most pleasure
Senot et al. [34]	TV	Historic TV use, including individual and group data	Size: 2–5 Family groups	Plurality voting, least misery, most pleasure, dictatorship, average	BEST: average for most groups, Dictatorship in 20% of groups. WORST: most pleasure, least misery
Bourke et al. [53]	TV movies	User study. Metric: average of individual satisfaction	Size: 3,5,10 Types: experts, high similarity, social relationships	Multiplicative, borda, approval voting, least misery, most pleasure, respect	BEST: multiplicative and respect. WORST: most pleasure. Least misery better in larger groups than most pleasure
De Pessemier et al. [33]	Movies	Synthetic data metric: average of individual satisfaction	Size: 2,5	Average, average without misery, dictatorship, least misery, most pleasure	BEST: average and average without misery. Dictatorship low accuracy when aggregating recommendations
Berkovsky and Freyne [32]	Recipes	User study	Size: 2,3,4 family types: homogeneous, heterogeneous	Average, weighted average (based on activity, roles, etc.)	BEST: Weighted average with weights based on activity

to solve, and we do not know which aggregation strategy truly represents the group opinion. Unsurprisingly, those studies tend to find that the Average Strategy performs well (as do strategies that resemble it). Another approach is to compare group recommendations with the individual preferences. It assumes that group members' individual satisfaction with the group recommendation depends only on their individual preferences, and the match/mismatch of these individual preferences with the group recommendations, therefore, that preferences of individuals are independent from the group, which is actually not the case in most scenarios. In fact, the opinions or judgments of users in a group are likely to be influenced by the other group members (i.e., emotional contagion and conformity effects [49]). As a result, specific approaches come into play, as the two experiments that we previously elaborated.

There are notable exceptions to using synthetic groups, where researchers collected data about individual preferences as well as about what those individuals select when in groups with others. For instance, in [34], data with a history of TV use by individuals and groups was collected. In [44], two data sets were used, (1) Mafengwo is collected from a website of the same name, where users can record their traveled venues, create or join a group travel, and at the moment it is not publicly available; (2) CAMRa2011 contains movie ratings of individuals and household and is publicly available. In [46], another data set containing information about real groups was used, i.e., the Plancast data, collected from the event-based social network, contains information about events and participants of those events, hence each event is considered a group and event participants as group members. Finally, [52] provides a detailed description of a study collecting individual and group preferences about travel destinations, and simulates real face-to-face group interactions where the goal is to select a destination to visit jointly. In addition the data contains individual and pairwise characteristics such as personality and social relationships. This provides a more accurate view on what actually happens in groups. The only drawback of that approach is that what happens in real groups does not necessarily lead to optimal group satisfaction. For example, in [34], when a Dictatorship Strategy is used (as seems to have happened in 20% of their groups), this may have left others in the group unsatisfied, and it is possible that the group as a whole would have been more satisfied if a different approach had been used (though sometimes due to for example participant personality, individuals may well be satisfied when Dictatorship is used). This raises the question whether group recommenders should mimic what happens in real groups or should try to do better.

In the context of interactive group recommenders, Nguyen and Ricci [27] proposed a group discussion simulation model where the impact of (1) alternative combinations of long-term and session-based preferences, and (2) different individual group members' behavioral types (conflict resolution types) on the recommendation performance in different group scenarios was studied.

Having this in mind, Table 4 also contains information about the used data sets and the domain. Moreover, most studies compare group sizes and often also compare between homogeneous groups (where users' preferences are similar) and more heterogeneous groups. Studies typically find that aggregation strategies perform better for more homogeneous and smaller groups.

More advanced methods, that we previously shortly introduced, were as well evaluated, and usually by comparing them to the standard aggregation strategies from Table 3, if applicable. When such a comparison was not applicable, for instance for the neural network approaches, then the comparisons were done within the group of methods (e.g., AGREE was compared against NCF combined with standard aggregation strategies, Greedy-LM and Greedy-Var were compared against SPGreedy and EFGreedy, among others).

4 Impact of Sequence Order

As mentioned in Sect. 2, we are also interested in recommending a *sequence* of items. In Sect. 3, we have tackled the issue on what items to select if there is time for a certain number of items. For example, for a personalized news program on TV, a recommender may select seven news items to be shown to the group. To select the items, it can use an aggregation strategy (such as the Multiplicative Strategy) to combine individual preferences, and then select the seven items with the highest group ratings.

In this section, we are interested in the *order* of items in the sequence. For example, once seven news items have been selected, the question arises in what order to show them in the news program. Many options exist: for instance, the news program could show the items in descending order of group rating, starting with the highest rated item and ending with the lowest rated one. Or, it could mix up the items, showing them in a random order. An example of this approach, for package recommendations, was implemented in [8], where the authors use the greedy algorithm that in each iteration adds an item with a maximum possible utility for the group, where utility could be the highest average rating, the highest fairness score, etc.

However, the problem is actually far more complicated than that. Firstly, in responsive environments, the group membership changes continuously, so deciding on the next seven items to show based on the current members seems not a sensible strategy, as in the worse case, none of these members may be present anymore when the seventh item is shown.

Secondly, overall satisfaction with a sequence may depend more on the order of the items than one would expect. For example, for optimal satisfaction, we may need to ensure that our news program has:

- *A good narrative flow.* It may be best to show topically related items together. For example, if we have two news items about Michael Jackson (say about his funeral and about a tribute tour) then it seems best if these items are presented together. Similarly, it would make sense to present all sports' items together.

- *Mood consistency*. It may be best to show items with similar moods together. For example, viewers may not like seeing a sad item (such as a soldier's death) in the middle of two happy items (such as a decrease in unemployment).
- *A strong ending*. It may be best to end with a well-liked item, as viewers may remember the end of the sequence most.

Similar ordering issues arise in other recommendation domains, e.g., a music programme may want to consider rhythm when sequencing items. The recommender may need additional information (such as items' mood, topics, rhythm) to optimize ordering. It is beyond the topic of this chapter to discuss how this can be done (and is very domain specific). We just want to highlight that the items already shown may well influence what the best next item is. For instance, suppose the top four songs in a music recommender were all Blues. It may well be that another Blues song ranked sixth may be a better next selection than a Classical Opera song ranked fifth. Similarly, the group may prefer something from a different music genre after a sequence of songs from one genre, even if the song ranked the best next is of the same genre.

In Experiment 3 (see Fig. 3 and more detail in [9]), it was investigated, in the news domain, how a previous item may influence the impact of the next item. Participants rated a set of news items. They were then shown one news item[6] and rated how interested they were in it and how it made them feel, and re-rated the original items to see if their ratings would have changed. Amongst others, we found

Fig. 3 Experiment 3: Investigating the effect of mood and topic

[6] In a between-subject design, two different topics were used evoking different moods.

that mood (resulting from the previous item) and topical relatedness can influence ratings for subsequent news items.

This means that aggregating individual profiles into a group profile should be done repeatedly, every time a decision needs to be made about the next item to display. So, instead of first selecting say seven items to show and then deciding on the order, only one item is selected, and then it needs to be decided which item from *all* remaining ones is the best to show next, given that the first item may have an impact on the ratings of the remaining ones.

5 Modeling Affective State

When recommending to a group of people, you cannot give everybody what they like all the time. However, you do not want anybody to get too dissatisfied. For instance, in a shop it would be bad if a customer were to leave and never come back, because they really cannot stand the background music. Many shops currently opt to play music that nobody really hates, but most people not love either. This may prevent loosing customers, but would not result in increasing sales. An ideal shop would adapt the music to the customers in hearing range in such a way that they get songs they really like most of the time (increasing the likelihood of sales and returns to the shop). To achieve this, it is unavoidable that customers will occasionally get songs they hate, but this should happen at a moment when they can cope with it (e.g., when being in a good mood because they loved previous songs). Therefore, it is important to monitor continuously how satisfied each group member is. Of course, it would put an unacceptable burden on the customers if they had to rate their satisfaction (on music, advertisements, etc.) all the time. Similarly, measuring this satisfaction via sensors (such as heart rate monitors or facial expression classifiers) is not yet an option, as they tend to be too intrusive, inaccurate or expensive. So, it was proposed to model group members' satisfaction; predicting it based on what we know about their likes and dislikes.

Modeling individual's satisfaction is a topic on its own, even without the group dimension considered. For instance, in [49] various satisfaction functions to model an individual's satisfaction with a sequence of items were evaluated. The satisfaction function that performed the best defines the satisfaction of a user with new item after having seen a sequence of items, as a weighted sum of the satisfaction with the previously seen items, and the estimated impact that the new item will have on the user's satisfaction. The function implements two crucial effects (1) the effect of satisfaction decay over time, and (2) the influence of the user's satisfaction after experiencing previous items on the impact (perception) of a new item.

Motivated by these findings, works [54] and [55] implemented a sequence recommendation approach for a group of people by accounting for the decaying effect (i.e., items that were experienced more recently have a greater impact on the user's overall satisfaction with the recommended sequence). In [54], a sequence of artworks on a certain path within a museum is recommended. Here, in addition to modeling satisfaction and the decay effect, the physical order of the artworks has to be considered. To this end, the museum (artworks and paths) has been modelled as a Directed Acyclic Graph, where the goal is to find a maximum satisfaction path bounded by the time available for the museum visit. User satisfaction depends on the previously seen artworks on the path, as well as on the next artwork added to that path. The authors showed that sequence recommendations with a decay effect performed evidently better than a sequence of TOP-k recommendations. In [55], two methods were proposed, i.e., “Balancing without Decay” and ”Balancing with Decay”. The first step for both methods is to generate candidate items, that is, those with highest aggregated ratings for a group. In the second step, satisfaction of each user in a group is modelled as a cumulative satisfaction without (with) decay, for previously experienced items and for the next item that is to be added to the sequence. To select the next item for the sequence, pairwise differences between group members’ cumulative satisfactions is summed, and the item with the lowest sum of differences is added next. Evaluated with a user study, the authors showed that the ”Balancing with Decay” method outperformed “Balancing without Decay”.

5.1 Effects of the Group on an Individual’s Satisfaction

In the previously described approaches the satisfaction of other users in the group was not taken into account, which may well influence a user’s satisfaction. As argued in [49] based on social psychology, two main processes can take place.

Emotional Contagion Firstly, the satisfaction of other users can lead to so-called emotional contagion: other users being satisfied may increase a user’s satisfaction (e.g., if somebody smiles at you, you may automatically smile back and feel better as a result). The opposite may also happen: other users being dissatisfied may decrease a user’s satisfaction (e.g., if you are watching a film with a group of friends then the fact that your friends are clearly not enjoying it may negatively impact your own satisfaction). Emotional contagion may depend on your personality (some people are more easily contaged than others), and your relationship with the other person. Anthropologists and social psychologists have found substantial evidence for the existence of four basic types of relationships, see Fig. 4. In Experiment 5 (see Fig. 5), participants were given a description of a hypothetical person they were watching TV with (using the relationship types in Fig. 4) and asked how their own emotion would be impacted (on a scale from ‘decrease a lot’ to ‘increase a lot’) by that person’s strong positive or negative emotions (see detail in [49]). Results confirmed that emotional contagion indeed depends on the relationship you have: you are more likely to be contaged by somebody you love (such as your best friend) or respect (such as your mother or boss) than by somebody you are on equal footing with or are in competition with.

Conformity Secondly, the opinion of other users may influence your own expressed opinion, based on the so-called process of conformity. Figure 6 shows the

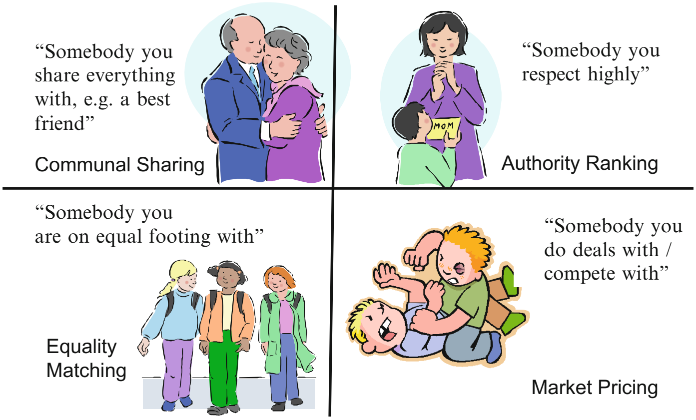

Fig. 4 Types of relationship

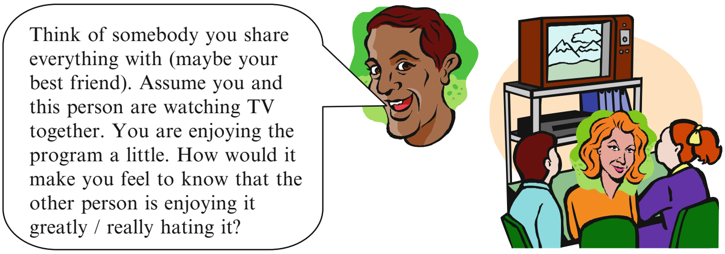

Fig. 5 Experiment 5: Impact of relationship type on emotional contagion

famous conformity experiment by Asch [56]. Participants were given a very easy task to do, such as decide which of the four lines has the same orientation as the line in Card A. They thought they were surrounded by other participants, but in fact the others where part of the experiment team. The others all answered the question before them, picking the same wrong answer. It was shown that most participants then pick that same wrong answer as well.

Two types of conformity exist: (1) normative influence, in which you want to be part of the group and express an opinion like the rest of the group even though inside you still belief differently, and (2) informational influence, in which your own opinion changes because you believe the group must be right. Informational influence would change your own satisfaction, while normative influence can change the satisfaction of others through emotional contagion because of the (insincere) emotions you are portraying.

More complicated satisfaction functions are presented in [49] to model emotional contagion and both types of conformity. These functions also serve as a basis for work in [57].

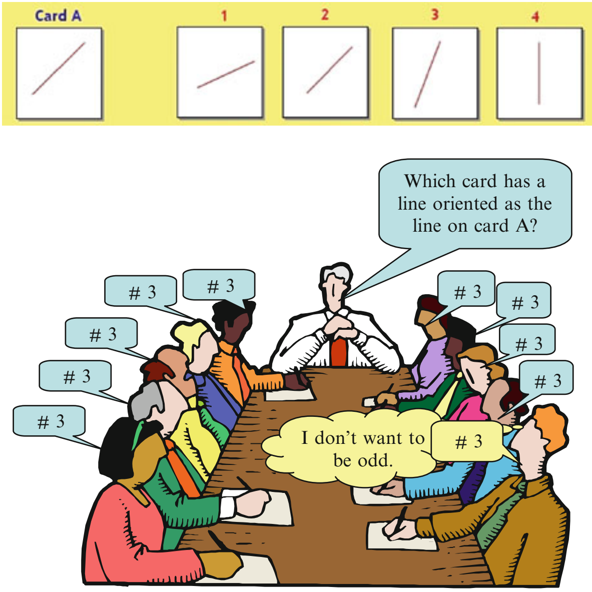

Fig. 6 Conformity experiment by Asch

6 Using Satisfaction Inside Aggregation Strategies

Once you have an accurate model of the individual users' satisfaction, which predicts how satisfied each group member is after a sequence of items, it would be nice to use this model to improve the group aggregation strategies. For instance, the aggregation strategy could set out to please the member of the group who is least satisfied with the sequence of items chosen so far. This can be done in many different ways, and we have only started to explore this issue. For example:

- *Strongly Support Grumpiest strategy.* This strategy picks the item which is *most liked* by the least satisfied member. If multiple of these items exist, it uses one of the standard aggregation strategies, for instance the Multiplicative Strategy, to distinguish between them.
- *Weakly Support Grumpiest strategy.* This strategy selects the items that are *quite liked* by the least satisfied member, for instance items with a rating of 8 or above. It uses one of the standard aggregation strategies, such as the Multiplicative Strategy, to choose between these items.
- *Weighted strategy.* This strategy assign weights to users depending on their satisfaction, and then use a weighted form of a standard aggregation strategy. For instance, Table 5 shows the effect of assigning double the weight to Jane when using the Average Strategy. Note that weights are impossible to apply to a strategy such as the Least Misery Strategy.

Table 5 Results of average strategy with equal weights and with twice the weight for Jane

	A	B	C	D	E	F	G	H	I	J
Peter	10	4	3	6	10	9	6	8	10	8
Jane	1	9	8	9	7	9	6	9	3	8
Mary	10	5	2	7	9	8	5	6	7	6
Average (equal weights)	7	6	4.3	7.3	8.7	8.7	5.7	7.7	6.7	7.3
Average (Jane twice)	5.5	6.8	5.3	8.3	8.3	8.8	5.8	8	5.8	7.5

In [58], this was discussed in more detail, an agent-based architecture was proposed for applying these ideas to the Ambient Intelligence scenario, and an implemented prototype was described. Preliminary work in [59], also uses a strategy which balances user satisfaction. Clearly, empirical research is needed to investigate the best way of using affective state inside an aggregation strategy.

7 Incorporating Group Attributes

Above, we discussed how an individual's satisfaction can be influenced by others in the group due to emotional contagion and normative behavior. Individuals' personality (e.g., propensity to emotional contagion) and social relationships between individuals played a role in this. For instance, in [60, 61], the authors showed that group members' individual satisfaction with group decisions is not only related to the match/mismatch between their own individual preferences and the group choice, but also to their personality types, and strength of their social relationships to others in the group. Moreover, in [62], it was demonstrated that the group score on *Conscientiousnes* (i.e., the extent to which one is precise, careful and reliable, or rather sloppy, careless, and undependable), and diversity of group members' preferences are significant predictors of the decision-reaching approach the groups adopted in their travel-related discussions (more details about various personality models most often used in recommender systems can be found in Chapter "Personality and Recommender Systems"). Individuals' personality and social relationships were incorporated into the models of satisfaction [49], which were then used in aggregation strategies [58]. Instead of using group attributes indirectly, via satisfaction models, it is also possible to incorporate them more directly into aggregation strategies.

Firstly, attributes can be used of individual group members, typically giving more weight to certain group members than others in the preference aggregation step:

- *Demographics and Roles*. As mentioned above, INTRIGUE [15] distinguishes different user types (children, adults with and without disability), and uses higher weights for more vulnerable user types. The recipe group recommender in [32] distinguishes between user roles (applicant, partner, child) and varies the weights based on their presumed level of engagement with the system (lowest for child, highest for applicant). An analysis of groups whose behavior corresponded to a Dictatorship strategy in [34] showed many cases in which teenagers acted as dictators in the company of children and where adults acted as dictators in the company of teenagers or children. So, different roles may influence what happens in groups, though [34] does not present the group composition when Dictatorship is not used.

- *Personality: Assertiveness and Cooperativeness*. HappyMovie [20] uses how *assertive* (extent to which person attempts to satisfy own concerns) and how *cooperative* (extent to which person attempts to satisfy others concerns) group members are, and gives a higher weight to assertive members and a lower weight to cooperative members. Moreover, findings from a simulation study [63] indicated that (1) groups of similar preferences whose members have non-cooperative conflict resolution styles achieved a lower utility compared to groups whose members have cooperative styles; (2) groups of diverse preferences and non-cooperative members often select recommendations that ensure equal utility losses for everybody, which may result in a greater loss on average, while the opposite happened for cooperative groups; 3) groups of mixed conflict resolution styles manage to make the group decision with the smallest average individual's utility loss, even though the difference in their utility is the largest.
- *Expertise*. As reported in [64], according to Social Psychology expertise may provide influence, so in normal group processes experts may have more influence on the group's decision.[7] Gatrell et al. [65] apply higher weights in the generation of recommendations to people with more expertise.[8] They infer expertise from activity, namely the number of movies rated, but only considering a pre-selected set of movies (i.e., 100 popular items). The recipe group recommender in [32] also uses higher weights for family members who have engaged more. In a similar fashion, in [66] group members' expertise and their corresponding weights were evaluated from the group member's activity in the system, i.e., the more item-ratings a group member provided the greater the weight would be. It is noteworthy that none of these approaches observed the group behavior or the actual activity of a group member in the group.
- *Personal impact*.[9] Liu et al. [67] incorporate the concept of *personal impact* into their group recommender algorithm, to model that different members will have different impacts on group decisions. They consider decisions made in the past to decide on personal impact. Similarly, in [68], personal impact is based on the match/mismatch between users' individual choice and the group choice in which

[7] Additionally, it seems plausible (but requires investigation) that users would be more dissatisfied with disliked selected items when their expertise is higher than that of other group members.

[8] This may work well when using an Additive or Multiplicative strategy, but does not really work for the Least Misery and Most Pleasure strategies used in [65], and hence unfortunately some of the formulas in [65] which incorporate expertise lack validity.

[9] Personal impact is not completely distinct from Role: somebody's role may influence their personal impact. However, it is still possible for people with the same official roles to have a different cognitive centrality.

the user has participated in the past. For example, if a user was a member of six different groups and her preferred option was selected as the group choice in three out of six cases which will be her weight in the impact-based model. Herr et al. [64] advocate using the social psychology concept of *cognitive centrality*: the degree to which a group member's cognitive information is shared within the group. They propose that the degree of centrality can be used to infer a person's importance and that more important members should be given higher weights. In fact, the underlying idea is that group members who have more in common with others in the group are given more weights, thus as well, items that are liked by these people should be items that are liked by others in the group, and items that are disliked by these members should be items that are disliked by others as well. This approach, applied to group members' shared preferences, was evaluated in [41], indicating that this kind of weighting approach can provide valuable information to aggregation strategies. This approach was even employed for individual recommendations [69], where a group formation method based on the pairwise similarity of users' ratings was proposed.

Secondly, people may behave differently in different groups, for instance depending on their closeness to the group, hence they might want to satisfy others in close relationships, while in not so close groups, they simply might want to make sure no body is too dissatisfied. To this end, attributes can be used of the group as a whole, typically using a different aggregation strategy based on the type of group:

- *Relationship strength*. The authors of [65, 70] advocate using different aggregation strategies depending on the group's relationship strength. They propose to use a Maximum Pleasure strategy for groups with a strong relationship (such as couples and close friends' groups), a Least Misery strategy for groups with a weak relationship (such as first-acquaintance groups), and an Average strategy for groups with an intermediate relationship (such as acquaintance groups).
- *Relationship type*. Wang et al. [57] distinguish between *positionally homogeneous*[10] and *positionally heterogeneous* groups. In positionally homogeneous groups, such as friend and tourist groups, the position of members is equal. In heterogeneous groups, such as family groups, the position of members is unequal. They also distinguish between *tightly-coupled* (strong relationship: members are close and intercommunication is important) and *loosely-coupled* (weak relationship: members are relatively estranged, and intercommunication is less frequent and less important) groups. Based on these two dimensions, Wang et al. define four different group types: tightly-coupled homogeneous (e.g., a friends' group), loosely coupled homogeneous (e.g., a tourist group), tightly-coupled heterogeneous (e.g., a family group), and loosely coupled heterogeneous (e.g., a staff group including managers).

[10] We have added the word positional to the terms homogeneous and heterogeneous used in [57] to avoid confusion with the earlier use of these words to indicate how diverse group preferences are.

Thirdly, attributes of people pairs in the group can be used, typically to adjust the ratings of an individual in light of the rating of the other person in the pair:

- *Relationship strengths*. HappyMovie [20] uses relationship strengths (which they call *social trust*) into its aggregation strategy, adapting individuals' ratings based on the ratings of others depending on the relationship strength between individuals. In [71] social relationships within the group are also considered as an indicator of influence. However, in comparison to the HappyMovie model, the authors considered three social aspects: (1) Social trust as the affective relationship between pairs of group members, (2) Social similarity between pairs of group members, and (3) Social centrality of each group member in the social environment.
- *Personal impact*. The concept of personal impact from [67] mentioned above can also be used in this way. In [51], it was demonstrated that positive and negative "shifts" from group member's initial preferences can happen, depending whether the relationship to another person in the group is a friendly or conflicting one. Ioannidis et al. [72] argue that people will be influenced by some people in their group more than others. Their group recommender uses a cascading process, where the group can see the votes that have already been cast and by whom, and can comment on alternatives. They and Ye et al. [73] learn social influence values for use in the group aggregation strategy. Rossi et al. [74] also assume that some members are more dominant than the others when making group decisions. Hence they propose a model that exploits group members' interactions in online social networks to determine the impact of each member, and to weight their preferences in the aggregation strategies accordingly.

A study presented in [75] aims to consider these factors simultaneously, i.e., the combination of the effects of group members' personality, expertise, pairwise social relationships (intimacy) and preference similarity in a group preference model. A clear improvement in comparison to the standard aggregation strategies was demonstrated.

8 Applying Group Recommendation to Individual Users

So, what if you are developing an application that recommends to a single user? Group recommendation techniques can be useful in three ways: (1) to aggregate multiple criteria, (2) to solve the so-called cold-start problem, (3) to take into account opinions of others.

8.1 Multiple Criteria

Sometimes it is difficult to give recommendations because the problem is multi-dimensional: multiple criteria play a role. For instance, in a news recommender

Table 6 Ratings on criteria for 10 news items

	A	B	C	D	E	F	G	H	I	J
Topic	10	4	3	6	10	9	6	8	10	8
Location	1	9	8	9	7	9	6	9	3	8
Recency	10	5	2	7	9	8	5	6	7	6

Table 7 Average strategy ignoring unimportant criterion location

	A	B	C	D	E	F	G	H	I	J
Topic	10	4	3	6	10	9	6	8	10	8
Recency	10	5	2	7	9	8	5	6	7	6
Group	10	4.5	2.5	6.5	9.5	8.5	5.5	7	8.5	7

system, a user may have a preference for location (being more interested in stories close to home, or related to their favorite holiday place). The user may also prefer more recent news, and have topical preferences (e.g., preferring news about politics to news about sport). The recommender system may end up with a situation such as in Table 6, where different news story rate differently on the criteria. Which news stories should it now recommend?

Table 6 resembles the one we had for group recommendation above (Table 1), except that now instead of multiple users we have multiple criteria to satisfy. It is possible to apply our group recommendation techniques to this problem. However, there is an important difference between adapting to a group of people and adapting to a group of criteria. When adapting to a group of people, it seems sensible and morally correct to treat everybody equally. Of course, there may be some exceptions, for instance when the group contains adults as well as children, or when it is somebody's birthday. But in general, equality seems a good choice, and this was used in the group adaptation strategies discussed above. In contrast, when adapting to a group of criteria, there is no particular reason for assuming all criteria are as important. It is even quite likely that not all criteria are equally important to a particular person. Indeed, in an experiment we found that users treat criteria in different ways, giving more importance to some criteria (e.g., recency is seen as more important than location) [76]. So, how can we adapt the group recommendation strategies to deal with this? There are several ways in which this can be done:

- Apply the strategy to the most respected criteria only. The ratings of unimportant criteria are ignored completely. For instance, assume criterion Location is regarded unimportant, then its ratings are ignored. Table 7 shows the result of the Average Strategy when ignoring Location.
- Apply the strategy to all criteria but use weights. The ratings of unimportant criteria are given less weight. For instance, in the Average Strategy, the weight of a criterion is multiplied with its ratings to produce new ratings. For instance, suppose criteria Topic and Recency were three times as important as criterion Location. Table 8 shows the result of the Average Strategy using these weights.

Table 8 Average strategy with weights 3 for topic and recency and 1 for location

	A	B	C	D	E	F	G	H	I	J
Topic	30	12	9	18	30	27	18	24	30	24
Location	1	9	8	9	7	9	6	9	3	8
Recency	30	15	6	21	27	24	15	18	21	18
Group	8.7	5.1	3.3	6.8	9.1	8.6	5.6	7.3	7.7	7.1

Table 9 Unequal average without misery strategy with location unimportant and threshold 6

	A	B	C	D	E	F	G	H	I	J
Topic	10	4	3	6	10	9	6	8	10	8
Location	1	9	8	9	7	9	6	9	3	8
Recency	10	5	2	7	9	8	5	6	7	6
Group	7			7.3	8.6	8.6		7.6	6.6	7.3

In case of the Multiplicative Strategy, multiplying the ratings with weights does not have any effect. In that strategy, it is better to use the weights as exponents, so replace the ratings by the ratings to the power of the weight. Note that in both strategies, a weight of 0 results in ignoring the ratings completely, as above.

- Adapt a strategy to behave differently to important versus unimportant criteria: Unequal Average Without Misery. Misery is avoided for important criteria but not for unimportant ones. Assume criterion Location is again regarded as unimportant. Table 9 shows the results of the Unequal Average Without Misery strategy with threshold 6.

We have some evidence that people's behavior reflects the outcomes of these strategies [76], however, more research is clearly needed in this area to see which strategy is the best. Also, more research is needed to establish when to regard a criterion as "unimportant".

8.2 *Cold-Start Problem*

A major concern for recommender systems, as discussed in several other chapters in this book, is the so-called cold-start problem: to adapt to a user, the system needs to know what the user liked in the past. The group recommendation techniques presented in this chapter provide an alternative solution. When a user is new to the system, we simply provide recommendations to that new user that would keep the whole group of existing users happy (i.e., all the existing users in the system). We assume that our user will resemble one of our existing users, though we do not know which one, and that by recommending something that would keep everybody happy, the new user will be happy as well.[11]

Gradually, the system will learn about the new user's tastes, for instance, by them rating our recommended items or, more implicitly, by them spending time on the items or not. We provide recommendations to the new user that would keep the group of existing users happy including the new user (or more precisely, the person we now assume the new user to be). The weight attached to the new user will be low initially, as we do not know much about them yet, and will gradually increase. We also start to attach less weight to existing users whose taste now evidently differs from our new user.

A small-scale study using the MovieLens data set was conducted, in order to explore the effectiveness of this approach. Five movies and twelve users who had rated them were randomly selected: ten users as already known to the recommender, and two as new users. Using the Multiplicative Strategy on the group of known users, movies were ranked for the new users. Results were encouraging: the movie ranked highest was in fact the most preferred movie for the new users, and also the rest of the ranking was fine given the new users' profiles. Applying weights led to a further improvement of the ranking, and weights started to reflect the similarity of the new users with known users. More detail on the study and on applying group adaptation to solve the cold-start problem is given in [77]. A follow on study in [78] confirmed the usefulness of this method. The use of aggregate ratings to solve the cold-start problem is also discussed in [79].

8.3 Virtual Group Members

Finally, group adaptation can also be used when adapting to an individual by adding virtual members to the group. For instance, parents may want to influence what television their children watch. They may not mind their children watching certain entertainment programmes, but may prefer them watching educational programmes. When the child is alone, a profile representing the parent's opinions (about how suitable items are for their child) can be added to the group as a virtual group member, and the TV could try to satisfy both, establishing a balance between the opinions of the parent and child. Similarly, a virtual group member with a profile produced by a teacher could be added to a group of learners.

[11] This initially offers the user non-personalized recommendations, however not necessarily by purely using popularity (e.g. Average without Misery can be used and fairness principles can be applied towards the other group members when recommending a sequence).

9 Conclusions and Challenges

Group recommendation is a relatively new research area. This chapter is intended as an introduction in the area, and the various techniques to deliver group recommendations.

9.1 Main Issues Raised

The main issues raised in this chapter are:

- Adapting to groups is needed in many scenarios such as interactive TV, ambient intelligence, recommending to tourist groups, etc. Inspired by the differences between scenarios, group recommenders can be classified using multiple dimensions.
- Many strategies exist for aggregating individual preferences (see Table 3), and some perform better than others. Users seem to care about avoiding misery and fairness.
- Existing group recommenders differ on the classification dimensions and in the aggregation strategies used. See Table 10 for an overview.
- When recommending a sequence of items, aggregation of individual profiles has to occur at each step in the sequence, as earlier items may impact the ratings of later items.
- It is possible to construct satisfaction functions to predict how satisfied an individual will be at any time during a sequence. However, group interaction effects (such as emotional contagion and conformity) can make this complicated.
- It is possible to evaluate in experiments how good aggregation strategies and satisfaction functions are, though this is not an easy problem.
- Group aggregation strategies are not only important when recommending to groups of people, but can also be applied when recommending to individuals, e.g. to prevent the cold-start problem and deal with multiple criteria.

9.2 Caveat: Group Modeling

The term “group modeling” is also used for work that is quite different from that presented in this chapter. A lot of work has been on modeling common knowledge between group members (e.g. [80, 81], modeling how a group interacts (e.g. [82, 83]) and group formation based on individual models (e.g. [82, 84]).

Table 10 Group recommender systems

System	Usage scenario	Classification				Strategy used
		Preferences known	Direct Experience	Group Active	Recommends Sequence	
MusicFX [13]	Chooses radio station in fitness center based on people working out	Yes	Yes	No	No	Average Without Misery
PolyLens [14]	Proposes movies for a group to view	Yes	No	No	No	Least Misery
Intrigue [15]	Proposes tourist attractions to visit for a group based on characteristics of subgroups (such as children and the disabled)	Yes	No	No	Yes	Average
Travel Decision Forum [16]	Proposes a group model of desired attributes of a planned joint vacation and helps a group of users to agree on these	Yes	No	Yes	No	Median
Yu's TV Rec. [18]	Proposes a TV program for a group to watch based on individuals' ratings for multiple features	Yes	No	No	No	Average
CATS [17]	Helps users choose a joint holiday, based on individuals' critiques	No	No	Yes	No	Counts requirements met Uses Without Misery
Masthoff's [9, 49]	Chooses a sequence of music video clips for a group to watch	Yes	Yes	No	Yes	Multiplicative etc.
GAIN [10]	Displays information and advertisements adapted to the group present	Yes	Yes	No	Yes	Average
REMPAD [19]	Proposes multimedia material for a group reminiscence therapy session	Yes	No	No	No	Least Misery
HappyMovie [20]	Recommends movies to groups	Yes	No	No	No	Average
IntelliReq [21]	Supports groups in deciding which requirements to implement	No	No	Yes	Yes	Plurality voting

9.3 Challenges

Compared to work on individual recommendations, group recommendation is still quite a novel area. The work presented in this chapter is only a starting point. There are many challenging directions for further research, including:

- *Recommending item sequences to a group.* The work in [9, 49] and the preliminary work in [59] seem to be the only work to date on recommending balanced *sequences* that address the issue of fairness. Even though sequences are important for the usage scenario of INTRIGUE, their work has not investigated making sequences balanced nor has it looked at sequence order. Clearly, a lot more research is needed on recommending and ordering sequences, in particular on how already shown items should influence the ratings of other items. Some of this research will have to be recommender domain specific.
- *Modeling of affective state.* There is a lot more work needed to produce validated satisfaction functions. The work presented in this chapter and [49] is only the starting point. In particular, large scale evaluations are required, as are investigations on the affect of group size.
- *Incorporating satisfaction within an aggregation strategy* As noted in Sect. 6, there are many ways in which satisfaction can be used inside an aggregation strategy. We presented some initial ideas in this area, but extensive empirical research is required to investigate this further.
- *Explaining group recommendations: Transparency and Privacy* One might think that accurate predictions of individual satisfaction can also be used to improve the recommender's transparency: showing how satisfied other group members are could improve users' understanding of the recommendation process and perhaps make it easier to accept items they do not like. However, as also confirmed in a recent study by Najafian et al. [85], users' need for privacy is likely to conflict with their need for transparency. An important task of a group recommender system is to avoid embarrassment. Users often like to conform to the group to avoid being disliked (we discussed normative conformity as part of Sect. 5.1 on how others in the group can influence an individual's affective state). In [49], it was investigated how different group aggregation strategies may affect privacy. More work is needed on explanations of group recommendations, in particular on how to balance privacy with transparency and scrutability. Chapter "Beyond Explaining Single Item Recommendations" provides more detail on the different roles of explanations in recommender systems [86].
- *User interface design.* An individual's satisfaction with a group recommendation may be increased by good user interface design. For example, when showing an item, users could be shown what the next item will be (e.g., in a TV programme through a subtitle). This may inform users who do not like the current item that they will like the next one better. There is also a need for additional research on interfaces for supporting group decision making (for initial research see [87]).
- *Group aggregation strategies for cold-start problems.* In Sect. 8.2, we have sketched how group aggregation can be used to help solve the cold-start problem.

However, the study in this area was very limited, and a lot more work is required to validate and optimize this approach.

- *Dealing with uncertainty.* In this chapter, we have assumed that we have accurate profiles of individuals' preferences. For example, in Table 1, the recommender knows that Peter's rating of item B is 4. However, in reality we will often have probabilistic data. For example, we may know with 80% certainty that Peter's rating is 4. Adaptations of the aggregation strategies may be needed to deal with this. DeCampos et al try to deal with uncertainty by using Bayesian networks [30]. However, they have so far focused on the Average and Plurality Voting strategies, not yet tackling the avoidance of misery and fairness issues.
- *Dealing with group attributes*. In Sect. 7, we have discussed work on incorporating group attributes in group recommender systems. Additionally, as mentioned in [88], users may well have different preferences in the context of a particular group then when they are alone. Clearly more research is needed in this area.
- *Empirical Studies*. More empirical evaluations are vital to bring this field forwards. It is a challenge to design well-controlled, large scale empirical studies in a real-world setting, particularly when dealing with group recommendations and affective state. It is likely that different aggregation strategies may be effective for different kinds of groups and for different application domains (see [87] for initial work on group recommender application domains). Almost all research so far has either been on a small scale, in a contrived setting, using synthetic groups (with the problem of using an Average metric, see Sect. 3.4) or lacks control.

Acknowledgments Judith Masthoff's research has been partly supported by Nuffield Foundation Grant No. NAL/00258/G.

References

1. F. Ricci, L. Rokach, B. Shapira, Recommender systems: Introduction and challenges, in ed. by F. Ricci, L. Rokach, B. Shapira, *Recommender Systems Handbook*, 2nd edn. (Springer, Berlin, 2015), pp. 1–34
2. M.F. Dacrema, I. Cantador, I. Fernandez-Tobias, S. Berkovsky, P. Cremonesi, Design and evaluation of cross-domain recommender systems, in *Recommender Systems Handbook*, ed. by F. Ricci, L. Rokach, B. Shapira (Springer, New York, 2022)
3. C. Musto, M. de Gemmis, P. Lops, F. Narducci, G. Semeraro, Semantics and content-based recommendations, in *Recommender Systems Handbook*, ed. by F. Ricci, L. Rokach, B. Shapira (Springer, New York, 2022)
4. A.N. Nikolakopoulos, X. Ning, C. Desrosiers, G. Karypis, Trust your neighbors: a comprehensive survey of neighborhood-based methods for recommender systems, in *Recommender Systems Handbook*, ed. by F. Ricci, L. Rokach, B. Shapira (Springer, New York, 2022)
5. H. Abdollahpouri, G. Adomavicius, R. Burke, I. Guy, D. Jannach, T. Kamishima, J. Krasnodebski, L. Pizzato, Multistakeholder recommendation: Survey and research directions, in *User Modeling and User-Adapted Interaction* (Springer, Berlin, 2020), pp. 127–158
6. R. Burke, Multisided fairness for recommendation. Preprint arXiv:1707.00093l (2017)

7. L. Xiao, Z. Min, Z. Yongfeng, G. Zhaoquan, L. Yiqun, M. Shaoping, Fairness-aware group recommendation with pareto-efficiency, in *Proceedings of the 11th ACM Conference on Recommender Systems* (ACM, New York, 2017), pp. 107–115
8. D. Serbos, S. Qi, N. Mamoulis, E. Pitoura, P. Tsaparas, Fairness in package-to-group recommendations, in *Proceedings of the 26th International Conference on World Wide Web* (ACM, New York, 2017), pp. 371–379
9. J. Masthoff, Group modeling: selecting a sequence of television items to suit a group of viewers. User Model. User-Adap. Interac. **14**, 37–85 (2004)
10. B. De Carolis, Adapting news and advertisements to groups, in *Pervasive Advertising* (Springer, Berlin, 2011), pp. 227–246
11. E. Kurdyukova, S. Hammer, E. André, Personalization of content on public displays driven by the recognition of group context, in *Ambient Intelligence* (Springer, Berlin, 2012), pp. 272–287
12. A. Jameson, B. Smyth, Recommendation to groups, in ed. by P. Brusilovsky, A. Kobsa, W. Njedl, *The Adaptive Web Methods and Strategies of Web Personalization* (Springer, Berlin, 2007), pp. 596–627
13. J. McCarthy, T. Anagnost, MusicFX: An Arbiter of Group Preferences for Computer Supported Collaborative Workouts, in *Proceedings of the 1998 ACM conference on Computer supported cooperative work, CSCW*, Seattle, WA (1998), pp. 363–372
14. M. O' Conner, D. Cosley, J.A. Konstan, J. Riedl, PolyLens: A Recommender System for Groups of Users. ECSCW, Bonn (2001), pp. 199–218. As Accessed on http://www.cs.umn.edu/Research/GroupLens/poly-camera-final.pdf
15. L. Ardissono, A. Goy, G. Petrone, M. Segnan, P. Torasso, Tailoring the recommendation of tourist information to heterogeneous user groups, in ed. by S. Reich, M. Tzagarakis, P. De Bra, *Hypermedia: Openness, Structural Awareness, and Adaptivity, International Workshops OHS-7, SC-3, and AH-3*. Lecture Notes in Computer Science, vol. 2266 (Springer, Berlin, 2002), pp. 280–295
16. A. Jameson, More than the sum of its members: Challenges for group recommender systems, in *International Working Conference on Advanced Visual Interfaces, Gallipoli* (2004)
17. K. McCarthy, L. McGinty, B. Smyth, M. Salamo, The needs of the many: A case-based group recommender system, in *European Conference on Case-Based Reasoning* (Springer, Berlin, 2006), pp. 196–210
18. Z. Yu, X. Zhou, Y. Hao, J. Gu, TV program recommendation for multiple viewers based on user profile merging. User Model. User-Adap. Int. **16**, 63–82 (2006)
19. A. Bermingham, J. O'Rourke, C. Gurrin, R. Collins, K. Irving, A.F. Smeaton, Automatically recommending multimedia content for use in group reminiscence therapy, in *Proceedings of the 1st ACM International Workshop on Multimedia Indexing and Information Retrieval for Healthcare* (ACM, New York, 2013), pp. 49–58
20. L. Quijano-Sanchez, J.A. Recio-Garcia, B. Diaz-Agudo, G. Jimenez-Diaz, Social factors in group recommender systems. ACM Trans. Intell. Syst. Technol. **4**(1), 8 (2013)
21. A. Felfernig, C. Zehentner, G. Ninaus, H. Grabner, W. Maalej, D. Pagano, F. Reinfrank, Group decision support for requirements negotiation, in *Advances in User Modeling* (Springer, Berlin, 2012), pp. 105–116
22. F. Guzzi, F. Ricci, R. Burke, Interactive multi-party critiquing for group recommendation, in *Proceedings of the Fifth ACM Conference on Recommender Systems* (ACM, New York, 2011), pp. 265–268
23. M. Stettinger, Choicla: Towards domain-independent decision support for groups of users, in *Proceedings of the 8th ACM Conference on Recommender Systems* (ACM, New York, 2014), pp. 425–428
24. J.O.Á. Márquez, J. Ziegler, Negotiation and reconciliation of preferences in a group recommender system. J. Inf. Proc. **26**, 186–200 (2018). Information Processing Society of Japan
25. T.N. Nguyen, F. Ricci, Combining long-term and discussion-generated preferences in group recommendations, in *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization* (ACM, New York, 2017), pp. 377–378

26. T.N. Nguyen, F. Ricci, A chat-based group recommender system for tourism, in *Information and Communication Technologies in Tourism* (Springer, Berlin, 2017), pp. 17–30
27. T.N. Nguyen, F. Ricci, Situation-dependent combination of long-term and session-based preferences in group recommendations: An experimental analysis, in *Proceedings of the 33rd Annual ACM Symposium on Applied Computing* (ACM, New York, 2018), pp. 1366–1373
28. D. Herzog, W. Wörndl, User-centered evaluation of strategies for recommending sequences of points of interest to groups, in *Proceedings of the 13th ACM Conference on Recommender Systems* (ACM, New York, 2019), pp. 96–100
29. T.N. Nguyen, Supporting Group Discussions with Recommendation Techniques (2019). Free University of Bozen-Bolzano
30. L.M. de Campos, J.M. Fernandez-Luna, J.F. Huete, M.A. Rueda-Morales, Managing uncertainty in group recommending processes. User Model. User-Adap. Interac. **19**, 207–242 (2009)
31. M.S. Pera, Y.K. Ng, A group recommender for movies based on content similarity and popularity. Inf. Process. Manag. **49**(3), 673–687 (2013)
32. S. Berkovsky, J. Freyne, Group-based recipe recommendations: Analysis of data aggregation strategies, in *Proceedings of the Fourth ACM Conference on Recommender Systems* (ACM, New York, 2010), pp. 111–118
33. T. De Pessemier, S. Dooms, L. Martens, Comparison of group recommendation algorithms. Multimed. Tools Appl. **72**, 2497–2541 (2014)
34. C. Senot, D. Kostadinov, M. Bouzid, J. Picault, A. Aghasaryan, C. Bernier, Analysis of strategies for building group profiles, in *Proceedings of User Modeling, Adaptation, and Personalization* (Springer, Berlin, 2010), pp. 40–51
35. J.O.Á. Márquez, J. Ziegler, Preference elicitation and negotiation in a group recommender system, in *Proceedings of IFIP Conference on Human-Computer Interaction* (Springer, Berlin, 2015), pp. 20–37
36. H.N. Kim, M. Bloess, A. El Saddik, Folkommender: a group recommender system based on a graph-based ranking algorithm. Multimedia Syst. **19**, 1–17 (2013)
37. L. Baltrunas, T. Makcinskas, F. Ricci, Group recommendations with rank aggregation and collaborative filtering, in *Proceedings of the Fourth ACM Conference on Recommender Systems* (ACM, New York, 2010), pp. 119–126
38. L.A. Carvalho, H.T. Macedo, Users' satisfaction in recommendation systems for groups: An approach based on noncooperative games, in *Proceedings of the 22nd International Conference on World Wide Web Companion* (2013), pp. 951–958
39. M. Salamó, K. McCarthy, B. Smyth, Generating recommendations for consensus negotiation in group personalization services. Personal Ubiquitous Comput. **16**(5), 597–610 (2012)
40. J. Castro, J. Lu, G. Zhang, Y. Dong, L. Martínez, Opinion dynamics-based group recommender systems. IEEE Trans. Syst. Man Cybe. Syst. **48**, 2394–2406 (2017)
41. A. Delic, F. Ricci, J. Neidhardt, Preference networks and non-linear preferences in group recommendations, in *Proceedings of IEEE/WIC/ACM International Conference on Web Intelligence* (ACM, New York, 2019), pp. 403–407
42. S. Schiaffino, D. Godoy, J.A.D. Pace, Y. Demazeau, A MAS-based approach for POI group recommendation in LBSN, in *Proceedings of the International Conference on Practical Applications of Agents and Multi-Agent Systems* (Springer, Berlin, 2020), pp. 238–250
43. F. Ortega, A. Hernando, J. Bobadilla, J.H. Kang, Recommending items to group of users using matrix factorization based collaborative filtering. J. Inf. Sci. **345**, 313–324 (2016)
44. D. Cao, X. He, L. Miao, Y. An, C. Yang, R. Hong, Attentive group recommendation, in *Proceedings of The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval* (ACM, New York, 2018), pp. 645–654
45. X. He, L. Liao, H. Zhang, L. Nie, X. Hu, T.-S. Chua, Neural collaborative filtering, in *Proceedings of the 26th International Conference on World Wide Web* (2017), pp.173–182
46. L. Vinh Tran, T.-A. Nguyen Pham, Y. Tay, Y. Liu, G. Cong, X. Li, Interact and decide: Medley of sub-attention networks for effective group recommendation, in *The Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval* (ACM, New York, 2019), pp. 255–264

47. A. Sankar, Y. Wu, Y. Wu, W. Zhang, H. Yang, H. Sundaram, GroupIM: A mutual information maximization framework for neural group recommendation, in *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval* (ACM, New York, 2020), pp. 1279–1288
48. J. Masthoff, The user as wizard: A method for early involvement in the design and evaluation of adaptive systems, in *Fifth Workshop on User-Centred Design and Evaluation of Adaptive Systems* (2006)
49. J. Masthoff, A. Gatt, In pursuit of satisfaction and the prevention of embarrassment: affective state in group recommender systems. User Model. User-Adapt. Interac. **16**, 281–319 (2006)
50. C. Trattner, A. Said, L. Boratto, A. Felfernig, Evaluating group recommender systems, in ed. by A. Felfernig, L. Boratto, M. Stettinger, M. Tkalcic, *Group Recommender Systems: An Introduction*, 1st edn. (Springer, Berlin, 2018), pp. 59–71
51. F. Barile, J. Masthoff, S. Rossi, A detailed analysis of the impact of tie strength and conflicts on social influence, in *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization* (ACM, New York, 2017), pp. 227–230
52. A. Delic, J. Neidhardt, T.N. Nguyen, F. Ricci, An observational user study for group recommender systems in the tourism domain. J. Inf. Technol. Tourism **19**, 87–116 (2018). Springer
53. S. Bourke, K. McCarthy, B. Smyth, Using social ties in group recommendation, in *Proceedings of the 22nd Irish Conference on Artificial Intelligence and Cognitive Science*. University of Ulster-Magee. Intelligent Systems Research Centre (2011)
54. S. Rossi, F. Barile, C. Galdi, L. Russo, Artworks sequences recommendations for groups in museums, in *Proceedings of 12th International Conference on Signal-Image Technology & Internet-Based Systems* (IEEE, Piscataway, 2016), pp. 455–462
55. A. Piliponyte, F. Ricci, J. Koschwitz, Sequential music recommendations for groups by balancing user satisfaction, in *Proceedings of UMAP Workshops* (2013). CEUR
56. S.E. Asch, Studies of independence and conformity: a minority of one against a unanimous majority. Pschol. Monogr. **70**, 1–70 (1956)
57. Z. Wang, X. Zhou, Z. Yu, H. Wang, H. Ni, Quantitative evaluation of group user experience in smart spaces. Cybern. Syst. An Int. J. **41**(2), 105–122 (2010)
58. J. Masthoff, W.W. Vasconcelos, C. Aitken, F.S. Correa da Silva, Agent-based group modelling for ambient intelligence, in *AISB Symposium on Affective Smart Environments, Newcastle* (2007)
59. A. Piliponyte, F. Ricci, J. Koschwitz, Sequential music recommendations for groups by balancing user satisfaction, in *Proceedings of the Workshop on Group Recommender Systems: Concepts, Technology, Evaluation at UMAP13* (2013), pp. 6–11
60. A. Delic, J. Neidhardt, L. Rook, H. Werthner, M. Zanker, Researching individual satisfaction with group decisions in tourism: Experimental evidence, in *Information and Communication Technologies in Tourism* (Springer, Berlin, 2017), pp. 73–85
61. A. Delic, J. Masthoff, J. Neidhardt, H. Werthner, *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization* (ACM, New York, 2018), pp. 121–129
62. A. Delic, J. Neidhardt, H. Werthner, Group decision making and group recommendations, in *Proceedings of IEEE 20th Conference on Business Informatics (CBI)* (IEEE, Piscataway, 2018), pp. 79–88
63. T.N. Nguyen, F. Ricci, A. Delic, D. Bridge, Conflict resolution in group decision making: insights from a simulation study. User Model. User-Adap. Interac. **29**, 895–941 (2019).
64. S. Herr, A. Rösch, C. Beckmann, T. Gross, Informing the design of group recommender systems, in *CHI12 Extended Abstracts on Human Factors in Computing Systems* (ACM, New York, 2012), pp. 2507–2512
65. M. Gartrell, X. Xing, Q. Lv, A. Beach, R. Han, S. Mishra, K. Seada, Enhancing group recommendation by incorporating social relationship interactions, in *Proceedings of the 16th ACM International Conference on Supporting Group Work* (ACM, New York, 2010), pp. 97–106

66. I. Ali, S.-W. Kim, Group recommendations: approaches and evaluation, in *Proceedings of the 9th International Conference on Ubiquitous Information Management and Communication* (ACM, New York, 2015), pp. 1–6
67. X. Liu, Y. Tian, M. Ye, W.C. Lee, Exploring personal impact for group recommendation, in *Proceedings of the 21st ACM International Conference on Information and Knowledge Management* (ACM, New York, 2012), pp. 674–683
68. E. Quintarelli, E. Rabosio, L. Tanca, Recommending new items to ephemeral groups using contextual user influence, in *Proceedings of the 10th ACM Conference on Recommender Systems* (ACM, New York, 2016), pp. 285–292
69. H. Mahyar, K.E. Ghalebi, S.M. Morshedi, S. Khalili, R. Grosu, A. Movaghar, Centrality-based group formation in group recommender systems, in *Proceedings of the 26th International Conference on World Wide Web Companion* (2017), pp. 1187–1196. International World Wide Web Conferences Steering Committee.
70. J.S. Zhang, M. Gartrell, R. Han, Q. Lv, S. Mishra, GEVR: An event venue recommendation system for groups of mobile users, in *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* (ACM, New York, 2019), pp. 1–25
71. I. Alina Christensen, S. Schiaffino, Social influence in group recommender systems. J. Online Inf. Rev. **38**, 524–542 (2014). Emerald Group Publishing Limited.
72. S. Ioannidis, S. Muthukrishnan, J. Yan, A consensus-focused group recommender system. Preprint arXiv:1312.7076 (2013)
73. M. Ye, X. Liu, W.C. Lee, Exploring social influence for recommendation: A generative model approach, in *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval* (ACM, New York, 2012), pp. 671–680
74. S. Rossi, F. Barile, A. Caso, et al., Dominance weighted social choice functions for group recommendations. ADCAIJ: Adv. Distrib. Comput. Artif. Intell. J. **4**, 65–79 (2015). Ediciones Universidad de Salamanca (España)
75. J. Guo, Y. Zhu, A. Li, Q. Wang, W. Han, A social influence approach for group user modeling in group recommendation systems. IEEE Intell. Syst. **31**, 40–48 (2016). IEEE
76. J. Masthoff, Selecting news to suit a group of criteria: An exploration, in *4th Workshop on Personalization in Future TV - Methods, Technologies, Applications for Personalized TV, Eindhoven* (2004)
77. J. Masthoff, Modeling the multiple people that are me, in ed. by P. Brusilovsky, A.Corbett, F. de Rosis, *Proceedings of the 2003 User Modeling Conference, Johnstown, PA* (Springer, Berlin, 2003), pp. 258–262
78. W.L. de Mello Neto, A. Nowé, Insights on social recommender system, in *Proceedings of the Workshop on Recommendation Utility Evaluation: Beyond RMSE, at ACM RecSyS12* (2012), pp. 33–38
79. A. Umyarov, A. Tuzhilin, Using external aggregate ratings for improving individual recommendations. ACM Trans. Web **5**, 1–45 (2011)
80. J. Introne, R. Alterman, Using shared representations to improve coordination and intent inference. User model. User-Adap. Interac. **16**, 249–280 (2006)
81. S. Suebnukarn, P. Haddawy, Modeling individual and collaborative problem-solving in medical problem-based learning. User Model. User-Adap. Interac. **16**, 211–248 (2006)
82. T. Read, B. Barros, E. Bárcena, J. Pancorbo, Coalescing individual and collaborative learning to model user linguistic competences. User Model. User-Adap. Interac. **16**, 349–376 (2006)
83. A. Harrer, B.M. McLaren, E. Walker, L. Bollen, J. Sewall, Creating cognitive tutors for collaborative learning: steps toward realization. User Model User-Adap. Interac. **16**, 175–209 (2006)
84. E. Alfonseca, R.M. Carro, E. Martín, A. Ortigosa, P. Paredes, The impact of learning styles on student grouping for collaborative learning: a case study. User Model. User-Adap. Interac. **16**, 377–401 (2006)
85. S. Najafian, O. Inel, N. Tintarev, Someone really wanted that song but it was not me! Evaluating which information to disclose in explanations for group recommendations, in *Proceedings of the 25th International Conference on Intelligent User Interfaces Companion* (2020), pp. 85–86

86. N. Tintarev, J. Masthoff, Beyond explaining single item recommendations, in *Recommender Systems Handbook*, ed. by F. Ricci, L. Rokach, B. Shapira (Springer, New York, 2022)
87. M. Stettinger, G. Ninaus, M. Jeran, F. Reinfrank, S. Reiterer, WE-DECIDE: A decision support environment for groups of users, in *Recent Trends in Applied Artificial Intelligence* (Springer, Berlin, 2013), pp. 382–391
88. J. Gorla, N. Lathia, S. Robertson, J. Wang, Probabilistic group recommendation via information matching, in *Proceedings of the 22nd International Conference on World Wide Web* (2013), pp. 495–504