

Mining Large-scale TV Group Viewing Patterns for Group Recommendation

November 4, 2013
Technical Report
MSR-TR-2013-114

Microsoft Research
Microsoft Corporation

Mining Large-scale TV Group Viewing Patterns for Group Recommendation

Allison J.B. Chaney*[†]
Department of Computer Science
Princeton University
achaney@cs.princeton.edu

Mike Gartrell*[†]
Department of Computer Science
University of Colorado Boulder
mike.gartrell@colorado.edu

Jake M. Hofman*
Microsoft Research
jmh@microsoft.com

John Guiver
Microsoft Research
joguiver@microsoft.com

Noam Koenigstein
Microsoft Research
noamko@microsoft.com

Pushmeet Kohli
Microsoft Research
pkohli@microsoft.com

Ulrich Paquet
Microsoft Research
ulripa@microsoft.com

ABSTRACT

We present a large-scale study of television viewing habits, focusing on how individuals adapt their preferences when consuming content in group settings. While there has been a great deal of recent work on modeling individual preferences, there has been considerably less work studying the behavior and preferences of groups, due mostly to the difficulty of data collection in these settings. In contrast to past work that has relied either on small-scale surveys or prototypes, we explore more than 4 million logged views paired with individual-level demographic and co-viewing information to uncover variation in the viewing patterns of individuals and groups. Our analysis reveals which genres are popular among specific demographic groups when viewed individually, how often individuals from different demographic categories participate in group viewing, and how viewing patterns change in various group contexts. Furthermore, we leverage this large-scale dataset to directly estimate how individual preferences are combined in group settings, finding subtle deviations from traditional preference aggregation functions. We present a simple model which captures these effects and discuss the impact of these findings on the design of group recommendation systems.

1. INTRODUCTION

We are in the midst of a technological revolution in which the primary means of home entertainment is shifting from traditional television sets to online and Web services (e.g., Netflix, Hulu, and Xbox) which contain a rapidly expanding catalogue of content. As a result, there is a growing body of research that investigates user behavior in online systems, largely for the purpose of aiding individuals in discovering novel content. At the same time, the Web is becoming an increasingly social space in which individuals interact and impact each other’s experiences, underscoring the importance of understanding how user preferences change in group contexts.

While there is a substantial body of literature on context-aware recommendations [2], there is a relatively small amount of empirical work on group recommendations, mostly owing to the difficulty of collecting co-viewing data. As we review in Section 2, older studies typically rely on small-scale, self-reported viewing data to draw qualitative conclusions

about group viewing, whereas large-scale log data typically lack the granularity to identify individuals within groups and track their activity both alone and together. As such, most work on group recommendations considers a limited, fixed set of strategies (e.g., average satisfaction, least misery, and maximum satisfaction) to aggregate the preferences of individuals within a group. While these approaches are somewhat successful, they obscure potentially more subtle group dynamics and interactions that affect group decision making.

In contrast to previous work, we use a large-scale and previously unpublished dataset of individual and group viewing patterns that was automatically recorded using the Nielsen ratings system¹ from a representative panel of U.S. viewers. Hence, our study presents a more accurate picture of not only program type viewing patterns but also of viewing duration. Furthermore, to the best of our knowledge, this paper is one of the first attempts at directly understanding the relationship between viewing patterns of groups and their constituent individuals from direct logged data. Our findings indicate that group context substantially impacts viewer activity and knowledge of the group’s composition can be informative with regard to determining group interests.

Our study makes several key contributions: First, we provide a novel large-scale analysis of viewing patterns with an emphasis on differences between groups and individuals. We present a detailed breakdown of what users watch alone, how often they engage in group viewing, and how their preferences change in these contexts. Second, we utilize the fact that our dataset is unique in its abundance of group viewing events to directly estimate how individual preferences are combined in group settings. In doing so, we observe that traditional approaches to preference aggregation substantially underestimate subtle, non-linear interactions between group members. Finally, we propose a simple yet effective approach to group recommendations based on the demographic information of the group’s constituent individuals. By capturing non-linearities between the constituents’ preferences, our approach is able to predict group preferences more accurately than existing group recommendation methods. This calls for more sophisticated non-linear aggregation functions that can better estimate the different dynamics between individuals within a group.

* Authors contributed equally.

[†]Research performed while interning at Microsoft Research.

¹www.nielsen.com

We begin by reviewing related work, including historic television studies and more recent group recommendations work, in Section 2. Next, we present details of the Nielsen data set in Section 3 and a simple analysis of individual viewing patterns in Section 4. We continue with a comprehensive description of group viewing activity in Section 5, including details of who tends to view content in groups, what content groups of different types tend to consume, and how this deviates from individual viewing. We conclude with an in-depth analysis of predicting group views, highlighting the shortcomings of traditional preference aggregation functions and exploiting subtle interactions among group members to improve the quality of group recommendations.

2. RELATED WORK

Analysis of viewing patterns is of interest to researchers working in a number of different fields including broadcasting, advertising, and more recently, machine learning and recommendation systems. Apart from historic studies of group viewing patterns, there has been a recent surge in work on group recommender systems. We review some of the relevant literature here.

Historic TV Viewing Studies. In the early eighties, Webster and Wakshlag [26] analyzed viewing patterns and program-type loyalty in group viewing. Their study analyzed how viewing behavior over two categories of programs—‘situational-comedies’ and ‘crime-action’—differed in individuals and groups. They found that groups that changed their composition over time exhibited a large variance in their viewing habit. On the other hand, groups that did not change over time showed more program-type loyalty, and mirrored the viewing trends of individual users. The analysis did not consider how the composition of the group affected their viewing patterns. To the best of our knowledge, this question has largely remained unstudied.

Most historic studies of users’ viewing behavior relied on surveys where respondents recorded program views in diaries [11, 26]. These studies were based on self-reported data that had a few hundred respondents. The small size made the results of these studies prone to subject selection biases. As later studies [17] show, television viewing behavior was affected by demographic characteristics such as age, gender, income and educational qualifications. Our work tries to overcome these problems by using a large, actively recorded dataset of viewing patterns that comes with detailed demographic information for a representative sample of viewers.

Recommendation Systems for Groups. The problem of group recommendation has been investigated in a number of works [3, 7, 8, 14, 20, 23, 25, 27]. Across this spectrum, various techniques target different types of items (e.g., movies, TV programs, music) and groups (e.g., family, friends, dynamic social groups).

Most group recommendation techniques consider the preferences of individual users and propose different strategies to either combine the individual user profiles into a single group (or pseudo user) profile, and make recommendations for the pseudo user, or generate recommendation lists for individual group members and merge the lists for group recommendation. Jameson and Smyth’s three main strategies for merging individual recommendations are *average satisfaction*, *least misery*, and *maximum satisfaction* [14]; these form the bedrock of group recommendations [3, 8, 16]. In

this paper, the three strategies are referred to as “preference aggregation functions”. Average satisfaction, which assumes equal importance across all group members, is used in several group recommendation systems [7, 27, 28]; there is evidence that both average satisfaction and least misery are plausible candidates for group decisions [16]. Different weights (like weights of family members) have also been used in aggregation models, rather than an average satisfaction strategy [5]. A more involved consensus function that utilizes the dissimilarity among group members on top of average satisfaction and least misery strategies, is also plausible [3]. This consensus function is open to extension, as it does not take other factors that may affect a group decision into consideration. Social connections and content interests can equally be utilized in heuristic group consensus functions [9]. The dynamic aspect of group recommendations can also be overlooked if the group is guaranteed to remain static. For instance, instead of combining the TV preferences of individual family members, a family-based TV program recommender can base recommendations on the view history of each household [25]. All of the aforementioned work involves relatively small-scale studies or prototypes, while other work on group recommendation relied in synthetically generated data from the MovieLens data set [1, 4, 15, 19]. In contrast, in this paper we analyze a large-scale dataset consisting of over a million TV program viewings, of which a quarter are group views.

Smaller practical systems include PolyLens, a group-based movie recommender that targets small, private, and persistent groups [20]. PolyLens includes facets like the nature of groups, rights of group members, social value functions, and interfaces for displaying group recommendations. PartyVote provides a simple democratic mechanism for selecting and playing music at social events, such that each group member is guaranteed to have at least one of her preferred songs played [23].

Recently, the first available large scale group preference datasets have begun to emerge. The 2011 Challenge on Context-Aware Movie Recommendation (CAMRa 2011), held in conjunction with the ACM Conference on Recommender Systems, utilized a large scale group preference dataset from the Moviepilot Web site consisting of over 170,000 users, over 24,000 movies, and nearly 4.4 million ratings [22]. This dataset also provides information on the household membership for most users. The “group” component is substantially smaller: there are only 290 households in which the household membership accompanies a user’s rating, and “group ratings” are lacking. This dataset is not publicly available. A number of group recommendation approaches have been proposed and evaluated using this dataset, including [6, 10, 12, 13, 18]. Our work differs in that we use a large dataset with hundreds of thousands of implicit group preferences available in the data in the form of program views and the time that a user spent watching a program, along with substantial metadata for individuals, households, and programs.

3. DATASET

The Nielsen Company maintains a panel of U.S. households and collects TV viewing data through both electronic metering and paper diaries. In the month of June 2012, Nielsen recorded 4,331,851 program views by 75,329 users via their electronic People Meter system, which records both

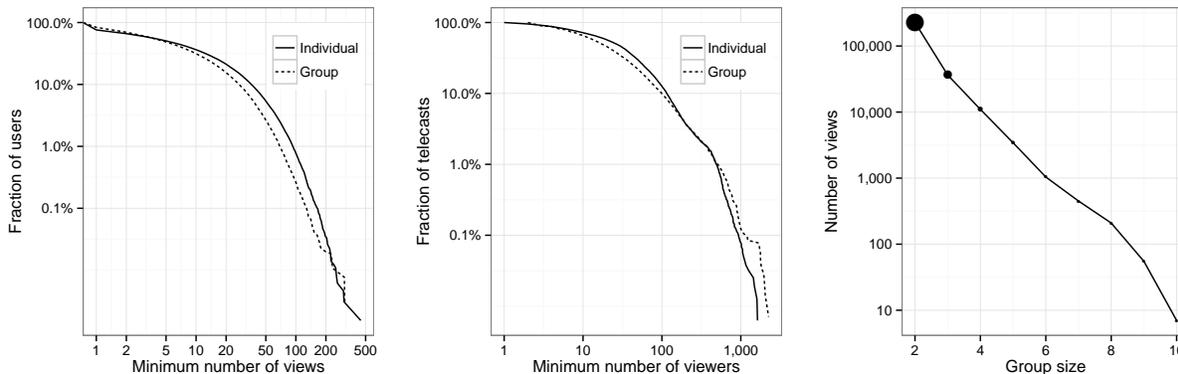


Figure 1: (a) Cumulative distribution of user activity split by individual and group views. (b) Cumulative distribution of telecast popularity by number of viewers. (c) Number of views by group size.

what content is being broadcast and who is consuming that content. We restrict this dataset to events where at least half of the program was viewed², resulting in a collection of 1,093,161 program views by 50,200 users. These views are comprised of 2,417 shows with 16,546 unique telecasts (e.g., individual series episodes, sports events, and movie broadcasts). Each program is associated with one of 34 genres and other metadata, including the distributor and optional sub-genre.

Users also have associated metadata, including age and gender, and are assigned to households, allowing a simple heuristic for identifying group viewing activity. We define a group view as one where at least two members of a household each watch at least half of the same telecast on the same day. There are 279,546 such group views in our dataset. When a user watches the majority of a telecast alone, we define this an individual view; 813,615 individual views are present our dataset.

The number of programs watched by users exhibit a heavy-tailed distribution, with many users viewing only a handful of telecasts while a few heavy users consume substantially more content. Figure 1a shows that roughly half of all users have viewed at least 5 telecasts individually and another (probably overlapping) 50% of users have viewed at least 5 telecasts in a group. Similarly, most programs are watched relatively infrequently, with a few being very popular. For example, Figure 1b shows that less than 10% of telecasts have been viewed by at least 100 users. We note that telecast popularity is slightly higher in group settings because each individual in a group view is counted separately here, so that a show watched by a pair of individuals is counted as two views for that broadcast. Finally, as shown in Figure 1c, we see that upwards of 80% of co-viewing occurs in groups of size 2, with larger groups occurring substantially less frequently.

4. INDIVIDUAL VIEWING PATTERNS

In this section, we analyze how individual viewing behavior varies with age and gender. For this purpose, we compute the genre-specific view counts in the context of demographic information. Figure 2 depicts how users of varying

²This 50% threshold simplifies our analysis so that at most one telecast can be viewed by each user in a given time slot.

age and gender distribute their attention across genres at the aggregate level. Panels are ordered by decreasing overall popularity, and point size shows the relative fraction of overall views accounted for by each demographic group in each genre. Table 1 provides an alternative view of these data, showing the top genres by view count for individuals of different age and gender. We discuss several clear age and gender patterns below. Note that these viewing patterns are limited to individual views only.

4.1 Age Effects

We observe that age has a strong influence on the viewing of certain genres like general drama, child multi-weekly, evening animation, news, popular music, general variety and news documentary. For instance, we observe that older viewers spend a large fraction (about 20-30%) of their time watching news, while teenage viewers devote more of their attention to popular music shows. General documentaries are more popular with adults and seniors than with children. Child multi-weekly programs are popular for children and much less popular with adults and seniors, as one would expect. General dramas are quite popular for every age and gender demographic we examined.

While the relative popularity of genres such as general drama, news, and news documentary increases with age (regardless of gender) viewing of genres such as Child multi-weekly, the popularity of genres such as evening animation and feature film decrease with age. Viewing of genres such as situational comedies and popular music are also affected by age but in this case only certain age ranges are affected and there is no general increase/decrease with age.

4.2 Gender Effects

We also see gender differences, with females spending more of their time watching talk shows, drama, and music relative to males, who spend more time on animation, documentaries, and sports. Point size indicates the number of views within each demographic. Sports events tend to be more popular with males than with females, across all ages.

5. GROUP VIEWING PATTERNS

Having briefly investigated individual viewing activity, we turn to the main analysis of the paper and analyze group

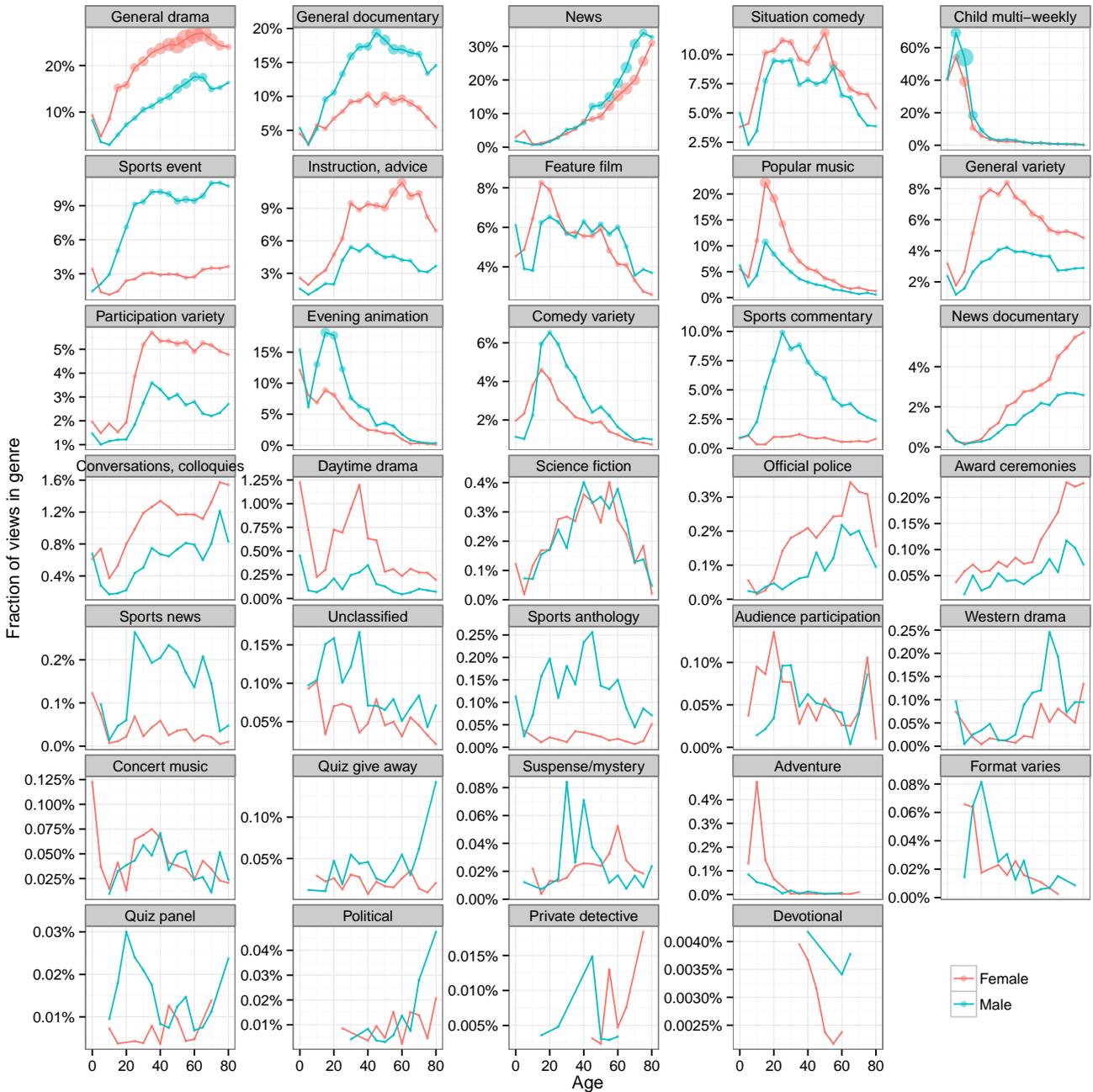


Figure 2: Distribution of views across genres by age and gender.

viewing patterns. We examine how often individuals engage in group viewing, which content is popular amongst groups, and how groups of various types distribute their viewing time. We conclude this section with a look at how individuals modify their viewing habits in group contexts.

5.1 Group Engagement

As noted in Section 4, roughly a quarter of all views in our dataset are by groups of size 2 or larger, comprising a sizable fraction of total activity. To gain further insight into the composition of groups, Figure 3 shows the relative

amount of group viewing by users of different ages and gender. The solid lines indicate the median fraction of group views for the specified demographic, with the top and bottom of the surrounding ribbon showing the upper and lower quartiles, respectively. We see that younger users spend the majority (~75%) of their time viewing in groups compared to older viewers. Viewers in their 20s and 30s spend roughly equal amounts of time viewing alone and in groups, whereas older viewers generally spend slightly more time watching individually. We see small gender effects for younger individuals and larger gender effects for older individuals, with

Table 1: Ranked list of genres for individuals with varying demographics.

Male child	Female teen	Male adult	Female adult	Male senior	Female senior
Child multi-weekly	Popular music	General documentary	General drama	News	General drama
General drama	General drama	General drama	General documentary	General drama	News
Feature film	Child multi-weekly	Sports event	Situation comedy	General documentary	General documentary
General documentary	Situation comedy	Situation comedy	Instruction, advice	Sports event	Instruction, advice
Evening animation	Feature film	Feature film	News	Situation comedy	Situation comedy
Sports event	Evening animation	News	General variety	Instruction, advice	Participation variety
Situation comedy	General documentary	Instruction, advice	Participation variety	Feature film	General variety
Popular music	General variety	Evening animation	Feature film	Participation variety	Sports event
Participation variety	Comedy variety	Participation variety	Popular music	News documentary	News documentary
Instruction, advice	Instruction, advice	Sports commentary	Sports event	General variety	Feature film

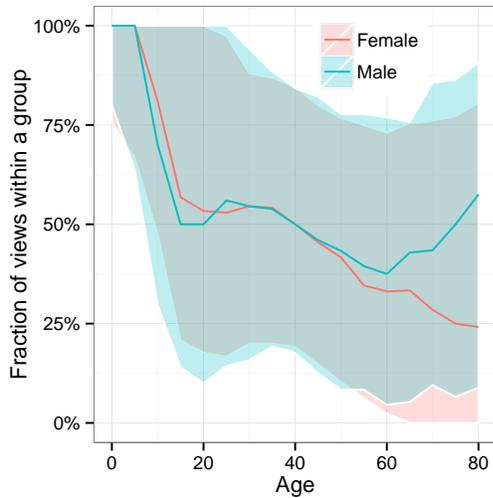


Figure 3: Fraction of views within a group by age and gender.

younger females and older males displaying a higher rate of group views relative to their counterparts.

Next we investigate the type of content consumed by these groups. As shown in Figure 4, the relative fraction of group viewing varies significantly by genre. While more than a third of views on quiz shows, drama, and sports events are within groups, only 20% of music, news, and politics views occur in groups settings. We note that many of the genres that are likely to be viewed by groups comprise a relatively small fraction of total activity, as indicated by point size. For example, while upwards of a third of all award ceremony views are in groups, there are relatively few such views overall.

5.2 Individual vs. Group Viewing

With this understanding of group engagement, we turn our attention to how individual viewing habits change in group settings. To do so, we compute viewing profiles for each user in the dataset under various group contexts and compare their individual and group profiles. Specifically, we characterize each user as either an adult or child (over/under 18, respectively) and male or female; likewise, we categorize each group view by its gender (all male/mixed gender/all female) and age (all adult/mixed gender/all child) breakdowns. For each user, we compute the fraction of time they

spend viewing each genre alone and in each of these nine possible group types. We then quantify the similarity between each user’s individual and group view profiles using Hellinger distance, a metric over probability distributions.³ Finally, we aggregate by user and group type and report the median similarity across users in each demographic when viewing in each group setting, as shown in Figure 5. From this plot we see that the similarity between individual and group viewing patterns varies substantially with the age composition of groups and less so with gender breakdown. For example, the bottom panel shows that activity by groups of all children looks most similar to views by individual children, compared to the mixed age groups in the top panel, which display the largest deviations from what members of those groups watch individually. Thus, the younger and more homogeneous the group, the higher the similarity between group and individual views.

For more details on how preferences shift in individual and group settings, Figures 6 and 7 show how attention is re-distributed across genres with different age and gender audience compositions, respectively. For example, Figure 6 reveals that feature films are more popular among mixed age groups than they are either for individuals or groups of the same age. Likewise, we see that children devote substantially more of their time to child multi-weekly shows when viewing in groups (~50%) compared to viewing alone (~30%). Adults watch more dramas, documentaries, and sports events in groups with other adults, and are more likely to watch news, sports commentary, and advice shows alone. We also see that adults and children both compromise on certain genres: one group watching more than usual and the other watching less. This occurs for many genres, including dramas and documentaries, where adults watch less than usual and children watch more, as well as popular music and evening animation, where children watch less and adult watch more together than they do separately. We see little compromise for adults on sports events and participation shows, possibly due to time sensitivity; in both of these cases, adults watch just as much as they do in groups with other adults, and children watch far more than they otherwise would.

We also see substantial shifts in preferences as gender composition varies in Figure 7. For instance, feature films are more popular with same gender groups than they are with either individuals or mixed gender groups, whereas the opposite effect is seen for news, which is more popular amongst individual males and females than in groups. We also see

³Hellinger distance is normalized to fall between 0 and 1; we measure similarity by the complement of Hellinger distance.

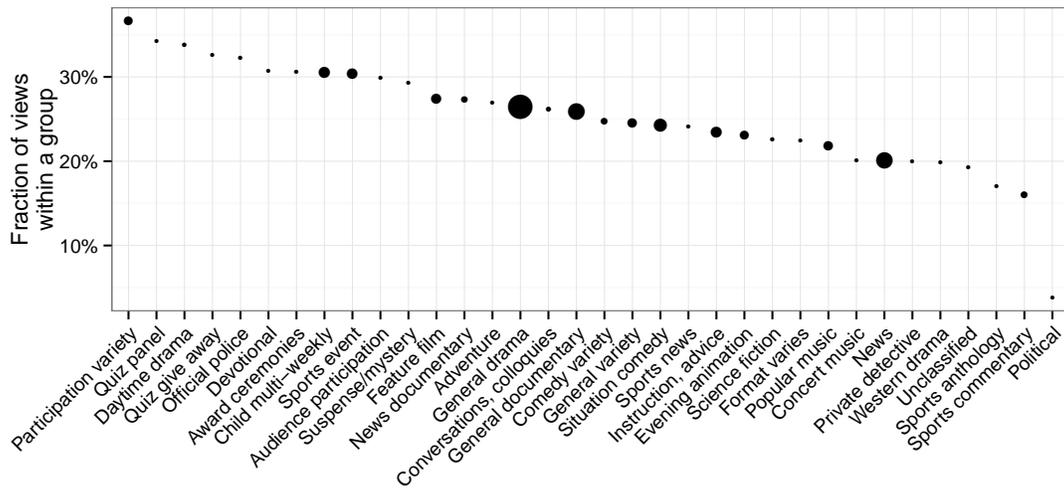


Figure 4: Fraction of views within a group by genre.

that news is more popular in mixed gender groups than in same-gender groups. We speculate that this effect is attributed to passive viewing patterns of couples in the same household, rather than an active desire to watch news as a group. While these changes are fairly similar between men and women, we note that other genres show gender-specific effects. For example, groups of men spend nearly double the

amount of their time watching sports compared to individual males, but no such change is seen for females. Likewise, all female groups spend substantially more of their time viewing popular music shows than do individual females. Finally, as with age effects, mixed gender groups appear to compromise on many categories. For dramas, advice, and sitcoms, men watch more and women watch less together than they do when the genders are separated. The reverse effect happens for documentaries, evening animation, and sports shows, with women watching more and men watching less.

Table 2, which shows a rank-ordered list of the most popular genres by audience type, provides a complementary perspective on this variation in preferences. For example, we see that while individual and groups of adults prefer to watch drama, news, and documentaries, children prefer multi-weekly shows, animation, and popular music; mixed age groups display a non-trivial blend of these preferences. Similarly, while drama, documentaries, and news remain prominent among groups of different gender composition, the popularity of animation, sports, and variety shows varies substantially between males and females.

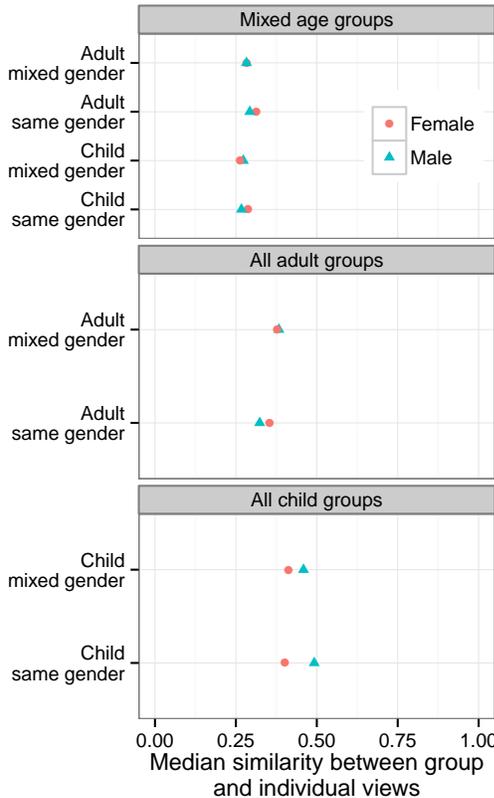


Figure 5: Similarity between group and individual viewing distributions for different group types.

6. GROUP RECOMMENDATIONS

The previous section highlights the difference between a group’s preferences and those of its individual constituents. While these effects are large at the aggregate level, there is substantial variability in preferences within and across demographic groups, underscoring the potential difficulty in accurately modeling any particular group’s preferences. Here we investigate this problem in more detail—namely, assuming that we know what the members of a group like individually, how do we aggregate their preferences to predict what the group will view?

We approach this problem in two steps. First, we fit a standard matrix factorization model to approximate individual preferences, which demonstrates good empirical results in predicting individual viewing activity. Next, we evaluate popular baseline methods for aggregating each individual’s modeled preferences to predict group activity. Interestingly, we find that traditional aggregation methods fail to capture subtle non-linearities and interactions between indi-

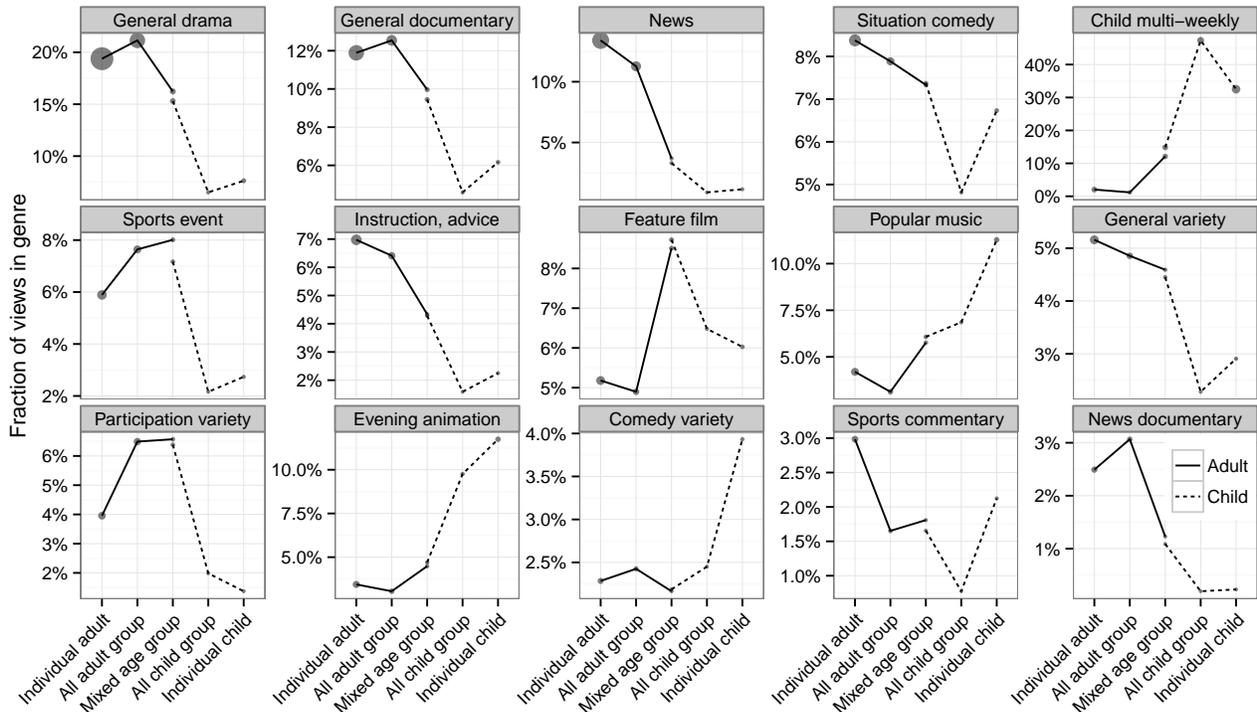


Figure 6: Distribution of views by genre for adults and children in different group contexts.

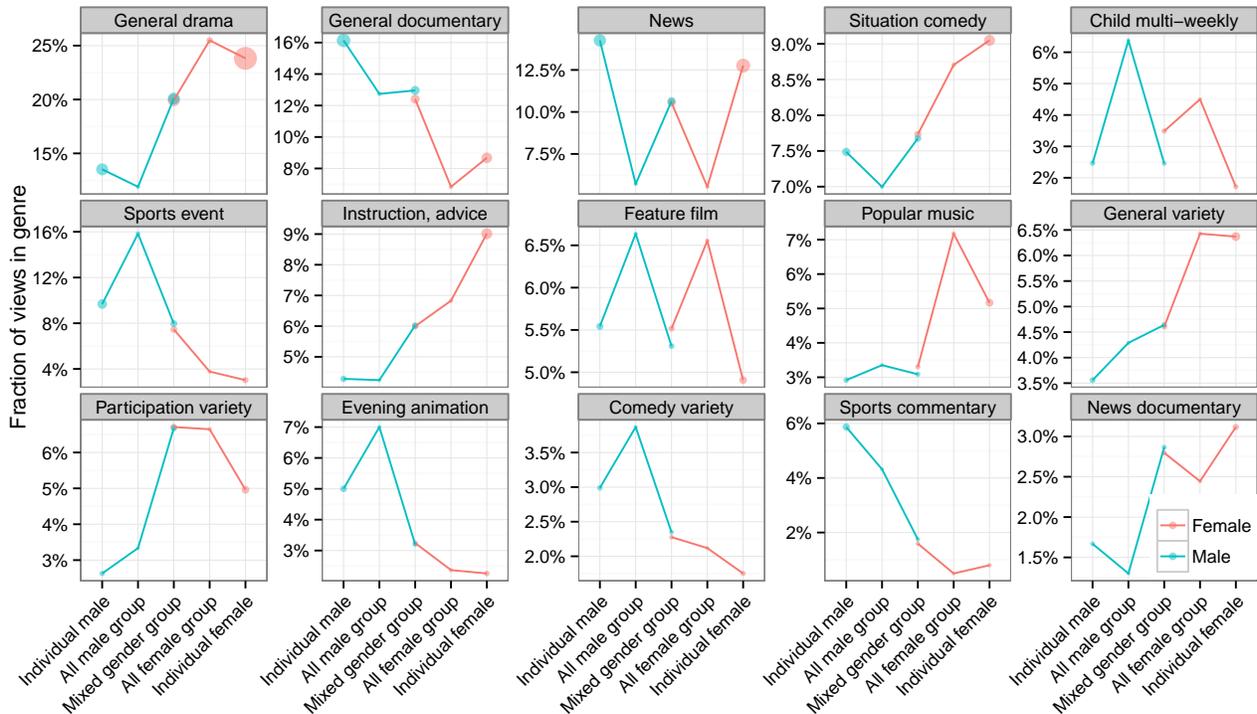


Figure 7: Distribution of views by genre for men and women in different group contexts.

Table 2: Ranked list of genres for individuals and groups. The *top* comparison is shown on varying age composition; the *bottom* comparison is pivoted on gender.

Individual adult	All adult group	Mixed age group	All child group	Individual child
General drama	General drama	General drama	Child multi-weekly	Child multi-weekly
News	General documentary	Child multi-weekly	Evening animation	Evening animation
General documentary	News	General documentary	Popular music	Popular music
Situation comedy	Situation comedy	Feature film	General drama	General drama
Instruction, advice	Sports event	Sports event	Feature film	Situation comedy
Sports event	Participation variety	Situation comedy	Situation comedy	General documentary
Feature film	Instruction, advice	Participation variety	General documentary	Feature film
General variety	Feature film	Popular music	Comedy variety	Comedy variety
Popular music	General variety	General variety	General variety	General variety
Participation variety	Popular music	Evening animation	Sports event	Sports event

Individual male	All male group	Mixed gender group	All female group	Individual female
General documentary	Sports event	General drama	General drama	General drama
News	General documentary	General documentary	Situation comedy	News
General drama	General drama	News	Popular music	Situation comedy
Sports event	Evening animation	Sports event	General documentary	Instruction, advice
Situation comedy	Situation comedy	Situation comedy	Instruction, advice	General documentary
Sports commentary	Feature film	Participation variety	Participation variety	General variety
Feature film	Child multi-weekly	Instruction, advice	Feature film	Popular music
Evening animation	News	Feature film	General variety	Participation variety
Instruction, advice	Sports commentary	General variety	News	Feature film
General variety	General variety	Evening animation	Child multi-weekly	News documentary

vidual preferences, which we are able to estimate directly from our large-scale dataset. We propose a relatively simple model to account for these features that both provides further insight into group decision making and improves the quality of group recommendations.

6.1 Modeling Individuals

To examine how to best combine preferences of individuals in a group, we first need a means of determining each individual user’s interest in each telecast in our dataset. We use the Matchbox recommendation system [24], which fits a standard matrix factorization model to user’s individual viewing activity to approximate these preferences. We discuss specific details of this process, including the challenges of modeling “one-class” collaborative filtering data [21], below.

The Matchbox model is designed to work with binary rating data, e.g., whether or not a user likes a particular item, as well as other types of rating data, such as ordinal ratings, e.g., five-star ratings. Our dataset is composed only of program telecast views. Since we want to predict whether or not a user will view a telecast, we must also consider data regarding telecasts that a user did not watch. We call the set of telecasts that a user did not watch the “negative set”, and the set of telecasts that a user did watch is called the “positive set”. For a telecast viewing, the set of negatives is the set of telecasts in our dataset that the user could have watched but did not, i.e. the set of telecasts whose broadcast time overlaps with the viewed telecast. Since we do not have the list of available TV channels for each user in our dataset, when constructing the negative set for a user, we approximate this list of by only considering those channels that have been viewed by the user in our data.

Since the negative set is composed of the set of telecasts that a user could have watched but did not, it is much larger than the positive set. In our dataset, we observed that the negative set is approximately 16 times larger than the positive set. Therefore, to keep the size of our training

set manageable and to maintain balance between the positive and negative set sizes, we develop a scheme for sampling from this negative set that is similar to the negative sampling scheme presented in [21]. We sample negatives by popularity for each positive telecast, by constructing a categorical distribution for a positive telecast composed of the normalized total view counts for every corresponding negative telecast. In this categorical distribution, each category is a possible negative telecast. To maintain approximate balance between our positive and negative sets, we sample one negative telecast for each positive telecast. We use an interval tree to enable efficient queries for negative telecasts that overlap with a positive telecast and therefore efficient construction of each categorical distribution used in our negative sampling scheme; queries in this interval tree require $\mathcal{O}(\log n + m)$ time, where n is the number of telecasts and m is the number of reported telecasts in the query result.

Matchbox is trained with $K = 20$ latent trait dimensions on a randomly selected training set composed of 80% of the individual view data set, with the remaining 20% of individual views used for the test set. No user or telecast (item) features are used in the model. When training Matchbox, we set the user threshold prior variance to 0 and user threshold noise variance hyperparameters to 0, since there is only one threshold for our binary ratings and we assume that this threshold does not change over time in our data. We place flexible priors on users and items by setting the user trait variance and item trait variance hyperparameters to $\frac{1}{\sqrt{K}}$, and the user bias variance and item bias variance hyperparameters to 1.

The best-fit individual model found by Matchbox has an AUC of 88.3% on the held-out test set. Given this performance, we consider the model to be a reliable approximation to individual preferences and next investigate the group recommendation problem.

6.2 Preference Aggregation

As noted in Section 2, there are a number of approaches

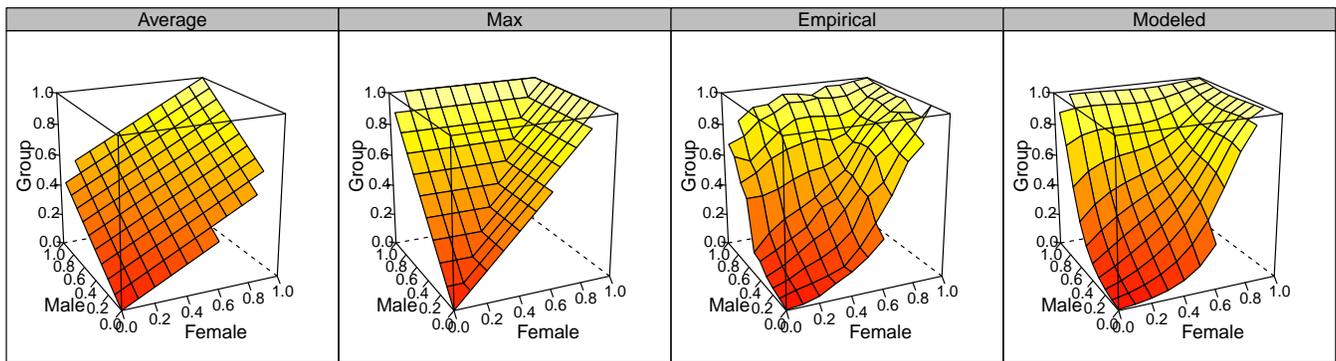


Figure 8: Modeled and actual probability of group viewing as a function of individual viewing for 2-person, mixed-gender adult couples.

to aggregating individual preferences, with the three most common being relatively simple functions: least misery, average satisfaction, and maximum satisfaction. Denoting individual preference that user u has for item i by p_{ui} , these methods predict group preferences for the item as follows:

$$\begin{aligned} \text{least misery} &: \min_{u \in G} p_{ui} \\ \text{average satisfaction} &: \frac{1}{|G|} \sum_{u \in G} p_{ui} \\ \text{max satisfaction} &: \max_{u \in G} p_{ui}. \end{aligned}$$

Least misery aims to minimize dissatisfaction of the least satisfied individual, maximum satisfaction to maximize enjoyment of the most satisfied, and average satisfaction takes an equal vote amongst all members.

We evaluate each of these methods across all group views in our dataset and find a strict ordering in terms of performance, with maximum satisfaction slightly outperforming average satisfaction, and both dominating least misery, across and within all group types. We find an overall AUC of 83.0% for maximum satisfaction, 82.6% for average satisfaction, and 79.7% for least misery. In further examining the quality of group predictions by group type, we see that mixed age and mixed gender group views are the most difficult to predict, with an AUC of 81.3%. Likewise, groups of all children are easiest to model, with performance on all male groups being considerably higher compared to all female groups (AUCs of 89.7% and 84.1%, respectively). We note that these results are largely consistent with the individual-to-group similarity comparison in Figure 5.

While work on preference aggregation has traditionally been constrained to these relatively simple functions over individual preferences, our large-scale dataset of hundreds of thousands of group views enables us to conduct a direct examination of group preference landscapes. For simplicity, we limit this analysis to groups of only two members (which comprise 80% of all group views). For each group viewing event in our dataset, we bin the individual predicted probability for each member of the group to the nearest ten percent and aggregate views to examine the empirical probability of a group view within each bin. Panel 3 of Figure 8 shows the result for adult mixed gender couples, with the binned male’s and female’s preference on the x- and y-axis, respectively, and the likelihood of a group view on the z-axis. The predicted landscape for average satisfaction and maximum satisfaction are shown in the first two panels for comparison, from which it is clear that these traditional ag-

gregation functions are too simple, missing crucial interactions and non-linearities in the group preference landscape.

First, we note that the empirical landscape is reasonably close to maximum satisfaction, but considerably lower along the diagonal, where users share identical individual preferences. Thus it seems that when both individuals equally dislike a program, there is a lower probability that the group will view the show than traditional approaches would suggest. This difference is highlighted in Figure 9a, where the dotted line indicates the (identical) predictions made by average satisfaction, least misery, and maximum satisfaction, whereas the points show the empirical probabilities of group viewing. We see a similar deviation when matched preferences are large, showing a slightly higher likelihood of group viewing than naive methods predict. Additionally, we see that average satisfaction fails to deal well with the extremes, for example when one individual has a strong preference for a show while the other has a strong preference against it. One potential explanation for this behavior is that group viewing represents a repeated bargaining scenario where individuals iterate over time between which individual in the group is satisfied in each instance.

To capture these subtleties, we fit a simple logistic regression with interactions to determine the probability of a group view (p_G) from the individual probabilities:

$$\begin{aligned} \log \frac{p_G}{1 - p_G} &= \alpha_0 + \alpha_f p_f + \alpha_m p_m + \\ &\quad \beta_f p_f^2 + \beta_m p_m^2 + \gamma_f p_f^3 + \gamma_m p_m^3 + \delta p_f p_m, \end{aligned}$$

where p_f is the female’s probability of viewing the show individually and p_m is the male’s. The β and γ terms accommodate the non-linearities in the landscape, while the δ term accounts for interactions. The resulting model fit, shown in the fourth panel of Figure 8, provides an improved approximation to the empirical landscape, with an AUC of 83.1% compared to 82.9% and 82.7% for maximum satisfaction and average satisfaction, respectively, on a randomly selected 20% held-out test set. Importantly, we note that while the differences in these aggregate metrics may seem insignificant, the model performs substantially better in crucial portions of the landscape—for example, traditional methods overpredict in regions where both users dislike content (e.g., small individual values in Figure 9a), leading to potential dissatisfaction and possibly lost of trust in the recommender system. Aggregate metrics understate these improvements due to the

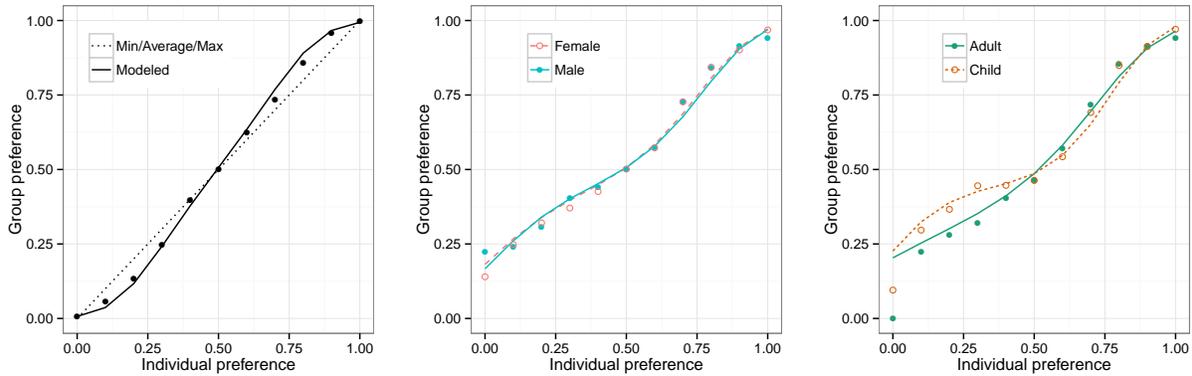


Figure 9: Modeled and actual probability of group viewing as a function of individual viewing for (a) mixed gender adult couples with identical preferences, (b) mixed gender adult couples where one member is indifferent, (c) mixed age pairs where one member is indifferent.

non-uniform density of group views along the landscape.

Figure 9b shows further details of the fitted model, taken along slices of the landscape where either the male or female is indifferent, corresponding to a individual preference of 0.5. Thus, for instance, the blue curve in Figure 9b shows how the probability of a group view changes as a function of the male’s individual preference with the female’s preference held fixed at 0.5, and vice versa for the pink curve. This highlights two key observations: first, the modeled curves are far from (piecewise) linear, as traditional aggregation functions would suggest, and second, we see strong symmetries between males and females, with no obvious signs of gender dominance. We contrast this with Figure 9c, which shows the model fit for mixed age groups. Here we see an asymmetry between adults and children, where the marginal increase in a child’s interest at low preference levels has a higher impact compared to an adult’s.

We note that while we have discussed only mixed gender and age couples here, these same qualitative observations apply to other group types: a simple non-linear group model provides a better fit to the empirical group landscape compared to traditional aggregation functions, which translates to improved performance for group recommendations.

7. CONCLUSIONS

Throughout this study we have seen that groups of individuals are more complex than the sum of their parts. In particular, we saw that viewing habits shift substantially between individual and group contexts, and groups display markedly different preferences at the aggregate level depending on their demographic breakdowns. This led to a detailed investigation of preference aggregation functions for modeling group decision making. Owing to the unique nature of the large-scale observational dataset studied, we were able to directly estimate how individual preferences are combined in group settings, and observed subtle deviations from traditional approaches (e.g., least misery, average satisfaction, and maximum satisfaction).

While we were able to explain observed group behavior with a relatively simple model, these results raise nearly as many questions as they answer. For example, further investigation is required to understand *why* these preference

landscapes take the shape they do, with third-order non-linearities. Likewise, untangling the driving forces behind these observations requires more than simple observational data. On one hand, effects could be explained by direct influence of individuals on each other, while on the other hand these outcomes may be confounded with homophily, wherein individuals tend to preferentially participate in groups that share their tastes. We leave answers to these questions along with generalizations to arbitrary group settings as future work.

8. REFERENCES

- [1] Movielens data sets. <http://www.grouplens.org/node/73>.
- [2] G. Adomavicius and A. Tuzhilin. Context-aware recommender systems. In *RecSys*, pages 335–336, 2008.
- [3] S. Amer-Yahia, S. B. Roy, A. Chawla, G. Das, and C. Yu. Group recommendation: Semantics and efficiency. In *Proc. of VLDB 2009*.
- [4] L. Baltrunas, T. Makcinskas, and F. Ricci. Group recommendations with rank aggregation and collaborative filtering. In *Proceedings of the fourth ACM conference on Recommender systems, RecSys '10*, pages 119–126, New York, NY, USA, 2010. ACM.
- [5] S. Berkovsky, J. Freyne, and M. Coombe. Aggregation trade offs in family based recommendations. In *Proc. of the 22nd Australasian Joint Conference on Advances in Artificial Intelligence*, 2009.
- [6] C. Biancalana, F. Gasparrini, A. Micarelli, A. Miola, and G. Sansonetti. Context-aware movie recommendation based on signal processing and machine learning. In *Proceedings of the 2nd Challenge on Context-Aware Movie Recommendation, CAMRa '11*, pages 5–10, New York, NY, USA, 2011. ACM.
- [7] L. Boratto, S. Carta, A. Chessa, M. Agelli, and M. L. Clemente. Group recommendation with automatic identification of users communities. In *Proc. of WI-IAT '09*, 2009.
- [8] L. de Campos, J. Fernández-Luna, J. Huete, and M. Rueda-Morales. Managing uncertainty in group recommending processes. *Journal of User Modeling and User-Adapted Interaction*, 19, 2009.

- [9] M. Gartrell, X. Xing, Q. Lv, A. Beach, R. Han, S. Mishra, and K. Seada. Enhancing group recommendation by incorporating social relationship interactions. In *Proceedings of the 16th ACM international conference on Supporting group work*, GROUP '10, pages 97–106, New York, NY, USA, 2010. ACM.
- [10] G. Gim, H. Jeong, H. Lee, and D. Yun. Group-aware prediction with exponential smoothing for collaborative filtering. In *Proceedings of the 2nd Challenge on Context-Aware Movie Recommendation*, CAMRa '11, pages 11–14, New York, NY, USA, 2011. ACM.
- [11] G. J. Goodhardt, A. S. C. Ehrenberg, M. A. Collins, et al. *The television audience: patterns of viewing. An update*. Number Ed. 2. Gower Publishing Co. Ltd., 1987.
- [12] J. Gorla, N. Lathia, S. Robertson, and J. Wang. Probabilistic group recommendation via information matching. In *Proceedings of the 22nd international conference on World Wide Web*, WWW '13, pages 495–504, Republic and Canton of Geneva, Switzerland, 2013. International World Wide Web Conferences Steering Committee.
- [13] X. Hu, X. Meng, and L. Wang. Svd-based group recommendation approaches: an experimental study of moviepilot. In *Proceedings of the 2nd Challenge on Context-Aware Movie Recommendation*, CAMRa '11, pages 23–28, New York, NY, USA, 2011. ACM.
- [14] A. Jameson and B. Smyth. Recommendation to groups. In *The adaptive web: methods and strategies of web personalization*, 2007.
- [15] H.-N. Kim, M. Rawashdeh, and A. El Saddik. Tailoring recommendations to groups of users: a graph walk-based approach. In *Proceedings of the 2013 international conference on Intelligent user interfaces*, IUI '13, pages 15–24, New York, NY, USA, 2013. ACM.
- [16] J. Masthoff. Group modeling: Selecting a sequence of television items to suit a group of viewers. *Journal of User Modeling and User-Adapted Interaction*, 14(1), 2004.
- [17] J. A. McCarty and L. Shrum. The role of personal values and demographics in predicting television viewing behavior: Implications for theory and application. *Journal of Advertising*, 22(4):77–101, 1993.
- [18] A. Moreno, H. Castro, and M. Riveill. Simple time-aware and social-aware user similarity for a knn-based recommender system. In *Proceedings of the 2nd Challenge on Context-Aware Movie Recommendation*, CAMRa '11, pages 36–38, New York, NY, USA, 2011. ACM.
- [19] E. Ntoutsi, K. Stefanidis, K. Nørvgå, and H.-P. Kriegel. Fast group recommendations by applying user clustering. In *Conceptual Modeling*, pages 126–140. Springer, 2012.
- [20] M. O'Connor, D. Cosley, J. A. Konstan, and J. Riedl. Polylens: A recommender system for groups of users. In *European Conference on Computer-Supported Cooperative Work*, 2001.
- [21] U. Paquet and N. Koenigstein. One-class collaborative filtering with random graphs. In *Proceedings of the 22nd international conference on World Wide Web*, WWW '13, pages 999–1008, Republic and Canton of Geneva, Switzerland, 2013. International World Wide Web Conferences Steering Committee.
- [22] A. Said, S. Berkovsky, and E. W. De Luca. Group recommendation in context. In *Proceedings of the 2nd Challenge on Context-Aware Movie Recommendation*, CAMRa '11, pages 2–4, New York, NY, USA, 2011. ACM.
- [23] D. Sprague, F. Wu, and M. Tory. Music selection using the partyvote democratic jukebox. In *Proc. of AVI 2008*, 2008.
- [24] D. H. Stern, R. Herbrich, and T. Graepel. Matchbox: large scale online bayesian recommendations. In *Proceedings of the 18th international conference on World wide web*, WWW '09, pages 111–120, New York, NY, USA, 2009. ACM.
- [25] E. Vildjiounaite, V. KyllÄänen, and T. Hannula. Unobtrusive dynamic modelling of tv programme preferences in a finnish household. *Journal of Multimedia Systems*, 15, 2009.
- [26] J. G. Webster and J. J. Wakshlag. The impact of group viewing on patterns of television program choice. *Journal of Broadcasting & Electronic Media*, 26(1):445–455, 1982.
- [27] Z. Yu, Z. Yu, X. Zhou, and Y. Nakamura. Handling conditional preferences in recommender systems. In *Proc. of IUI 2009*.
- [28] Z. Yu, X. Zhou, Y. Hao, and J. Gu. Tv program recommendation for multiple viewers based on user profile merging. *Journal of User Modeling and User-Adapted Interaction*, 16(1), 2006.