A State of the Art for Community-Driven Music Discovery

José Devezas
joseluisdevezas@gmail.com

November 22, 2013

Abstract

We surveyed existing characterization studies of how people listen to and discover music, either individually or in group. The goal was to encompass work not only from computer science, but also from other areas of research, like sociology or psychology, in order to build a technology-unbiased perception of how people interact with music. So, the idea was to gather relevant information that would encompass user studies based on data gathered from well-known music services, such as Last.fm, but also supported on real-world samples, which have the potential to add perspective on the unattended musical needs of the users. We also did an overview on recommender systems, introducing several information filtering techniques that contribute to the music discovery process. We then extended on this topic by focusing on a special kind of target audience: groups of people — as opposed to single individuals. Finally, we reviewed some of the most relevant bibliography specifically about music recommendation, focusing on context-aware applications, discovery based on the long tail, and the combination of social features with other available features, namely audio content. Our goal was to survey two main areas that we believe are useful when building a social music recommender system: user listening behavior and music discovery patterns, individually and in group; and music recommender systems, both targeted at individuals and groups. With this, we emphasize an *understand and explain first, to better predict and recommend later* and thus cover two main topics: music behavior characterization and music recommender systems. For either topic, we distinguish between two targets: isolated individuals, and interacting groups of individuals.

Keywords: User studies, music listening behavior, group interaction, recommender systems, recommendation to groups.

Contents

1 Introduction

Social information has been frequently used in collaborative filtering to recommend new content to users Bu et al. (2010); Fields (2011); Knijf et al. (2011); Ma et al. (2011). This approach consists of finding similar users based on their item ratings or, from a graph ranking perspective, to do a traversal from a starting user, through his/her rated items, and into new users and their items, ranking the target item nodes not only according to the number of paths that led there, but also to the rating that the starting user assigned to the item that led there.

While doing collaborative filtering is by itself a usage of social information as a collective intelligence, there are many types of social information that can be combined into the recommender engine, such as the social relationships and community structure of an artist's top friends network obtained from MySpace Fields (2011).

We are specially interested in using community structure to improve music recommendation to groups. While "community" and "group" frequently refer to the same concept, in this scenario we must distinguish between the two of them. We do so by using the time dimension. Thus, we propose that "community" should be defined as the set of social relationships that are stable over time, while "group" should be defined as a set of volatile relationships that haven't yet manifested stability over time. In this sense, communities are useful to position the individual into the different social contexts he/she belongs to, while groups might correspond to a sporadic gathering or an event, such as a convention or a party.

With the longterm goal of studying the usage of overlapping community structure in the characterization of the individual and the group, we present a survey focusing on the application of this idea to music information retrieval and recommendation, that brings together relevant bibliography on: (*i*) the characterization of the user as an individual listener versus the user as a group music listener (explaining); (*ii*) recommender systems, dealing with the multimodality of music data — namely the combination of audio content and social features into the recommendation process —, and understanding how recommendation to groups has been tackled (predicting).

This survey is organized as follows. In Section 2, we present an overview on available user studies in music information retrieval: in Section 2.1, we describe studies on music listening behavior, distinguishing between individual and group behavior; in Section 2.2, we describe a study on how music discovery happens in everyday life. In Section 3, we introduce recommender systems, covering collaborative filtering and evaluation, and expand on two subtopics of interest: recommendation to groups (Section 3.1) and music recommendation (Section 3.2). On Section 3.1, we present some uses cases and also cover existing rating aggregation techniques. On Section 3.2, we cover audio-based features (Section 3.2.1), which are relevant content features for music recommendation, and we also cover context-based methodologies (Section 3.2.2), the influence of long tail distributions in discovery (Section 3.2.3), and the usage of social features to improve recommendation (Section 3.2.4), including hypergraph-based algorithms. Finally, in Section 4, we comment on the covered topics and the open research paths in social music recommendation, focusing on the introduction of community-driven music discovery as a novel research direction.

2 User Studies in Music Information Retrieval

In this section, we present an overview of existing studies on the characterization of user behavior when engaging in music listening and discovery activities, either individually or in a group.

Lee and Cunningham (2012) have examined the impact of user studies in music information retrieval, by analyzing the growth, publication and citation patterns of 155 instances of this kind of work. They concluded that there are several barriers to the impact of user studies, namely a scattering of publications, weak connections among scholars, and dominance of small scale studies that aren't generalizable.

Lee and Waterman (2012) present the results of a large-scale user study regarding music needs, uses, seeking and management behavior. They show that there has been an increasing usage of streaming and mobile music services, the emergence of music identification and other cloud-based services, a centralization and increased appreciation of music video services, and an increased need for serendipitous music discovery. They point out that users value the social aspect, device interoperability and innovative ideas almost as much as the quality of the results. Given the artistic nature of music, it is not unexpected that users value innovative ways of interacting with and discovering new music.

2.1 Listening Behavior

We present an overview on the literature covering user studies of music services and technology, separating surveys that study the user's behavior when individually discovering, organizing or listening to music, from surveys that study the behavior of users interacting with a music service as a group.

Individually Park and Kahng (2010) present a temporal study of music listening behavior on a Korean online music service. They did a basic user and song analysis, showing that the user listening events distribution almost follows a power law, similarly to the web usage distribution; that users repeat the same songs frequently and some even to an extreme; and that popular songs tend to be repeated more frequently than less popular songs. They analyzed periodicity, depicting the daily listening events over one year, as well as the aggregated weekly and daily activity, showing a decrease of activity on weekends and during the night-time. They analyzed temporal context, evaluating the influence of time of day in genre selection. They concluded that “kids” and “solo instrumental” music deviated particularly from the general behavior, peaking at noon and rapidly decreasing in the evening. A slight preference of “rhythm and blues” over “dance pop” was also noticeable at night. Next, they evaluated the influence of season in genre selection, noticing that “ballad songs” were preferred over “dance pop” during the winter, while “dance pop” peaked on July; they also noticed a peak of Christmas carols during the month of December. Finally, they analyzed popularity change by picking some songs and depicting their daily popularity over a year period. They showed a coherent pattern where songs achieved their popularity peak two weeks after release date, continuously decreasing after that and having a lifespan of 4–6 months. They closed the study by analyzing the popularity of songs over time as they age, and concluded that the distribution might fit into several diffusion models, such as the Bass diffusion model, describing how songs get adopted, which in turn helps predict the popularity of other songs with a similar behavior.

Carneiro (2012) also did a temporal analysis of music listening behavior, both from an individual and global perspective, but he used a data mining methodology based on association rules and circular statistics. Some of the conclusions shown in this work are congruent with the paper by Park and Kahng (2010), namely the existence of an association between genre and time of day, as well as genre and season, for individual users, but not from a global perspective. That is, while a particular user might listen more to rock at a given time of day or at summer, this doesn't happen globally for the users, who have distinct tastes. Studying the behavior at a community level might provide insightful and relevant results — even though a consensual global behavior was not found, there might be coherent association rules between the genre and the time of day at a mesoscale.

While not directly related with user studies in music information retrieval, Castro and Azevedo (2012) proposed a time series motifs (recurrent patterns) evaluation methodology based on statistical significance. They compute each motif's p-value and accept motifs with p-values smaller than an automatically derived significance level. The study of time series, and more specifically the identification of frequent patterns, is important to understand the individual listening behavior of users and can be an addition to circular statistics methodologies and the study of periodic behavior. Namely, extracting frequent listening patterns might be a good indicator of music discovery behavior and might aid in the modeling of systems that mimic or promote the behavior depicted by the time series motifs.

Baur (2011) has studied real-world music listening histories using visualization techniques and interpreting this type of content as a lifelog that captures the human experience and the past behavior of the individual. He proposed a design space for music listening histories based on three main dimensions: time (lifelogging and reminiscing), items (relational data and context) and listeners (social aspects). His work represents a contribution that is aware of musicology, human-computer interaction and music information retrieval, proposing to improve music recommendation through interaction and the integration of listening histories into existing systems.

In Group Cunningham and Nichols (2009) bring us one of the few studies focused on understanding group behavior regarding the collaborative nature of music selection and listening during a social gathering. Their dataset was based on the students of a third year university course on human-computer interaction, and they fundamentally focused on understanding user music selection and listening behavior during a party. They verified that some users simply skip through the songs on the active playlist until they find an improvement. Fewer people prefer search, which might be due to the nature of the event where people prioritize socialization. On one hand, this leaves room for music search to highly improve in the sense of becoming more intuitive and efficient, providing quicker answers to the user's queries. The authors fully describe the listening experience and the music selection process at this party, providing a valuable and insightful source of information for the music technology industry, when looking to invest in the development of social music software. As they refer in the paper, music helps lighten the mood by filling the gaps introduced by silent moments, promoting conversation through the reminiscence of the past. Researching music technology in the sense of improving existing, but also of creating novel interactions with music in a social context, is quite clearly a relevant topic that hasn't been sufficiently explored. Using the information available in social networks to automatically improve music recommendation and discovery in social gatherings is an area that is still open to research and has the potential of innovation.

2.2 Discovery Patterns

While explanation is different from prediction, as the first looks into the past and the latter tries to sense the future, understanding past behavior can be a great indicator or guide to help build predictive systems, such as recommender systems. Particularly, we find it crucial to understand in detail how people discovery new music, so that we can mimic or promote this behavior, taking music discovery one step further.

Cunningham et al. (2007) study how music discovery happens in real life, by documenting how a series of individuals came across, either on purpose or by chance, with new music during the course of three days. Each individual was tasked with maintaining a diary of unfamiliar songs discovered throughout the day. Each entry contained the time and date, the location, the activity the individual was carrying at the time, and comments on the music discovery event. The music discovery events were quantified hourly over a day, showing peaks of activity at 11:00, for users that are more immersed into music, and at 20:00, for those users with fewer free leisure time. The two most common places where music discovery happened was at home or during commute. Each discovery was classified by the subject as "positive", "negative", "indifferent", or "unknown". The authors also classified the discoveries as active or passive. Most reactions to new music were "positive", and most discoveries were done passively. For the analyzed sample, it was shown that the visual aspects of music (videoclips, movie soundtracks, etc.) were quite relevant to the amount of attention paid and to the final like or dislike decision. Nearly a quarter of discoveries led the user to search for more information. The authors suggest that players should have direct links to related information and that "laid back" search could be used in conjunction with a microphone when driving, for instance, while listening to the radio — with a press of a button, the microphone would record the song and search more information about it when a connection was available. This paper by Cunningham et al. is a good example of how to put explanation to work for prediction or, put differently, it's a good example of a useful user study for music information retrieval.

3 Recommender Systems

The goal of recommender systems is to enable the discovery of potentially interesting new items that the user hadn't previously considered. This can be achieved through many different types of input (e.g. a specific song, or a whole listener's profile) and with many different applications in mind (e.g. playlist generation, or the discovery of less publicized music). Recommender systems usually rely on three different methodologies to accomplish their goal: collaborative filtering, content-based filtering, and context-based filtering Jannach et al. (2010); Ricci et al. (2011). Currently, one of the most important research challenges lies within context-aware recommender engines, namely regarding the heterogeneity of the data Kaminskas and Ricci (2012). These systems aim at suggesting the best items (e.g. songs) by factoring information such as the surroundings of the user (location, weather, temperature, season, etc.) or other dynamic features such as the user's mood or current activity, or even the environment noise and user pacing.

Jannach et al. (2010) provide a practical, algorithm-oriented, introduction to recommender systems, while Ricci et al. (2011) extend on this topic by presenting a compilation of several relevant surveys and papers in the area, alongside real-world applications. They describe collaborative filtering, where new items are recommended by finding users with similar tastes and suggesting the items that these similar users like, that the source user doesn't know about yet. They describe content-based recommendation, where items are purely suggested based on their characteristics (e.g. songs with similar rhythm and harmony). They also describe knowledge-based recommender systems, where the user is frequently able to constraint the item features, progressively filtering the results, until a suitable list of items is obtained; and hybrid recommender systems, whose goal is to combine the other three alternatives in order to improve the quality of recommended contents. Finally, both books provide an overview on the evaluation of a recommendation: Jannach et al. emphasize criteria such as validity, reliability and sensibility, while Ricci et al. develop this topic a bit further by covering additional parameters such as coverage, acceptance, serendipity and scalability.

Collaborative Filtering As part of the winning team of the Netflix Prize competition[1], Koren et al. (2009) demonstrated that matrix factorization models are superior to nearest neighbor techniques, for collaborative filtering. Neighborhood methods tackle the problem of recommendation by either finding other users that rated items similarly to the user, or by finding items that were similarly rated by all users when compared to the user's highly rated items. This is usually done by comparing user or item vectors based on metrics such as the Pearson's correlation coefficient, the cosine similarity or the Jaccard index.

[1] http://www.netflixprize.com

On the other hand, latent factor models, which are usually based on matrix factorization, help uncover the hidden features that describe the items (e.g. in movies: comedy vs drama, amount of action, orientation to children), as well as how much the users like items based on each of the uncovered features. We can obtain these latent features by using for instance singular value decomposition (SVD), which, given a user-item matrix M, results in the user-features U matrix and the item-features V matrix, along with a matrix Σ that contains the singular values in the diagonal, which measure the importance of each feature Schutt (2012):

$$M = U\Sigma V^* \tag{1}$$

However, issues still arise from this technique, as the user-item matrix is sparse, i.e. most of the users haven't rated most of the items. To deal with sparsity, earlier systems used imputation, for instance by filling missing values with the mean of the row or column, thus making the matrix dense. This presents an issue, as it's more costly memory-wise to compute the SVD for a dense matrix than it is for a sparse matrix, but also because inaccurate imputation of values might distort the data. An alternative is to confine to available ratings, while avoiding overfitting through a regularization procedure. In Koren et al. (2009), two methods are described to solve the problem of minimizing the difference between known and predicted ratings: stochastic gradient descent and alternating least squares. The authors also describe temporal dynamics, modeling this problem based on item biases, user biases and user preferences, over time. While temporal features are certainly relevant to improve recommender systems, we are even more interested in the techniques used to integrate this information into the predicted ratings matrix, since one of our goals is to account for social, cultural or community bias in our approach to music recommendation.

Sarwar et al. (2002) proposed and validated an incremental technique for matrix factorization based on singular value decomposition, in the context of recommender systems. This technique, described as folding-in, consists of adding a new vector of user ratings to an existing factorization, which assumes that the item collection will be static after the initial factorization. To add a new user u, we first calculate its projection onto the latent factor space, using the formula $\hat{u} = uV\Sigma^{-1}$, and append the projected vector $\hat{u}$ to the user-features matrix U. While this same technique had been previously explored by Berry et al. (1995) in the context of information retrieval, Sarwar et al. not only proposed a new application to the area of recommender systems, but also provided an illustration to help clarify the folding-in process.

Evaluation Celma and Herrera (2008) have proposed a new approach to evaluate recommender systems by taking novelty into account. They experimented with item-centric and user-centric evaluation methodologies, comparing collaborative filtering, content-based filtering and hybrid algorithms. For collaborative filtering, they created a network based on the Last.fm social network, connecting each artist to the top 20 most similar other artists — similarity was obtained using the Audioscrobbler web services from Last.fm, given that its value accounted for social relationships. For content-based filtering, they created a network in a similar way, replacing the similarity measure by an audio-based metric, building the network based on the similarity of the 100 most representative tracks of each artist, for 30 second samples. The authors compared the assortativity of the similarity networks, concluding that, in the collaborative filtering network, the most connected artists tended to connect to other top connected artists, while in the content-based network there was no preferential connectivity. Next, they correlated artist indegree with artist total play counts, dividing the distribution into head, mid and tail. In the collaborative filtering network, similarity among artists was always higher in the head for head artists, in the mid for mid artists and in the tail for tail artists, while in the content-based network this did not happen. They verified that the number of clicks to go from the head to the tail was five for the collaborative filtering network, but only two for the content-based network. To evaluate the quality of the algorithms when providing novel songs, they used explicit user feedback. Based on the Last.fm profile of the surveyed users and using the top 20 most played artists, the system recommended new songs to the users, who evaluated them using a one to five scale, from "don't like" to "like it very much", also indicating whether they knew the artist name and the song title. This enabled the authors to assess the number of novel songs recommended with success. They concluded that there was a statistically significant difference among the ratings and the percentage of unknown songs of the three algorithms. While the average rating was higher for collaborative filtering, the percentage of unknown songs was lower. On the other hand, content-based filtering and the hybrid algorithm provided the largest amount of novel songs, although with a lower average rating.

Said (2013) addressed several issues regarding the evaluation schemes that are currently used to assess recommender systems quality. He discussed the utility of recommendation and the problems of current evaluations protocols regarding perceived quality and diversity. He then studied the effect of popularity bias, finding that the quality of recommendations can be boosted by up to 20%, after the cold start problem is surpassed, by using mitigation techniques. Next, he measured the quality and diversity of recommendations as perceived by the user, based on a recommendation algorithm he created aimed at delivering more diverse, novel and non-obvious recommendations. He concluded that, even though user feedback was similar for his algorithm and a standard baseline algorithm, the evaluation metrics reported a lower quality for his algorithm, validating the idea that current evaluation schemes are inappropriate to tune recommendation algorithms focused on novelty and diversity. He then obtained user feedback from recommendations given by k-nearest neighbors (kNN), k-furthest neighbors (kFN) and a random algorithm, verifying that kNN and kFN returned almost completely different results, both being positively evaluated regarding different aspects: kNN returned obvious or recognizable recommendations, while kFN excelled at recommending serendipitous and useful items. He extensively explored the correlations between factors such as rating, serendipity, usefulness, recognizability, or novelty, from the users perspective. Conclusions show that high ratings correlate with serendipity, usefulness and retention; obviousness and recognizability do not drive people away from reusing the system; and unknown/novel movies correlate with lower ratings and consumption. He closes his thesis by doing a mathematical characterization of the *magic barrier*, the point at which the performance and accuracy of the system cannot be further improved, without risking overfitting, because of noise in the data. He proposes using the principle of empirical risk minimization, from statistical learning theory, to reduce the magic barrier to the Root Mean Squared Error

(RMSE) of an optimal rating function — the magic barrier represents the lower bound that a rating function can achieve without overfitting. An experimental study was conducted based on the Moviepilot recommender system, where the users were asked to provide a movie opinion rating, in the same scale as their previously assigned rating available in the main system. Both values were then used to estimate the magic barrier by averaging the squared sample noise. It was shown that users are less consistent for lower ratings and more consistent for higher ratings, and that the Moviepilot system could still be improved. The author suggests that a poll should be conducted frequently to recalculate the magic barrier and audit the performance of the recommendation system.

3.1 Recommendation to Groups

Jameson and Smyth (2007) have explored the challenges posed by the task of recommending for groups in the context of the adaptive web. They identified four steps in this recommendation process: (*i*) acquire information about the user's preferences, (*ii*) generate recommendations, (*iii*) explain recommendations to the users, and (*iv*) help the users settle on a final decision. This approach takes advantage of the collective intelligence of users, adapting results based on a collaborative process. We argue that, given the recent boom in ubiquitous computing, there are several open opportunities for contribution, specially regarding the contextual automatization of the recommendation adaptive process. In their survey, Jameson and Smyth cite some of the existing group recommender systems, such as PolyLens, by O'Connor et al. (2002), or Flytrap, an intelligent group music recommender system by Crossen et al. (2002). Both systems were developed in 2002, a time when smartphone usage was far from achieving its peak. In fact, that same year, BlackBerry launched what would be considered the first real smartphone. Masthoff (2004) studied user behavior in the selection of a sequence of television items for a group of viewers. She noticed that users instinctively took advantage of the *Average* strategy, the *Average without Misery* strategy, and the *Least Misery* strategy, thus having the whole group in mind when making their choice. In 2005, having noticed the rapid developments in mobile communications technologies, Zhiwen et al. (2005), proposed taking advantage of these advancements to build an in-vehicle adaptive multimedia recommender for groups of users. Their approach was based on connecting the mobile devices of the users and the automobile's multimedia system through a Wi-Fi LAN. They collected individual user preferences and fed this information to the multimedia system, which received content from available providers through a GPRS network and played the best content based on the tastes of the current passengers. Smyth et al. (2005) presented their work on adaptive web search, where, using a similar approach, they identified the implicit preferences for communities of searchers, in order to improve the results of future searches from users with the same characteristics and social context.

Deventer et al. (2013) have experimented with group recommendation in a television context. They merged the individual user profiles by using the *Least Misery* rating aggregation technique to model the group preferences, which is an effective strategy for small groups. The system's recommendations were based on a user-genre matrix, instead of a user-item matrix, which is quite smaller than the latter. They also experimented with user identification schemes, both using facial recognition and QR codes generated through a mobile application. They left some open challenges, namely discovering the correlation between the real group preferences and the *Least Misery* aggregation, providing reasons for the recommendations, and shielding the user's privacy given the shared environment with co-watching friends and family.

Piliponyte et al. (2013) proposed a new *Balancing* approach to music recommendation to groups, as an alternative to the traditional rating aggregation strategies: *Average*, *Average without Misery*, or *Least Misery* (see Ricci et al. (2011, Chapter 21, Table 21.3) for a complete list of rating aggregation techniques). They compare three approaches:

1. *Average* (the baseline), which consists of computing the predicted rating for each item and calculating the average over all the members of the group.

2. *Balancing without Decay*, which is done in two steps: first the "Average" aggregation is done and a section of the highest rated items is selected as the candidate set; then the individual satisfaction of each user in the group is calculated iteratively, for the candidate set, when adding a new track to the playlist, given the sequence of previously built recommendations, and, for each track, the sum of all the differences between user satisfactions is calculated, recommending the next track by selecting the lowest of these values. This means that each track is selected sequentially based on the satisfaction agreement of the group and the previously recommended tracks.

3. *Balancing with Decay*, which is similar to the previous, but the satisfaction score is calculated differently, assuming that user satisfaction is higher for tracks played more recently, but also that the tracks with the highest predicted rating for each user have a higher weight the largest their predicted rating for the user.

Based on the system they implemented to test their hypothesis, they conclude that the *Balancing* techniques can achieve better performance than the *Average*, as well as return results that are comparable to humanly generated playlists, as created by a randomly chosen group member.

Herr et al. (2012) present a selection of social psychological concepts for group recommender systems, including group identification, group norms and social roles. The authors describe several items within each topic, mapping their impact in group recommender systems. Within the group identification topic, they list: interpersonal attraction, self-categorization, and interdependence. Within the group norms, they list: communication rules, and attitude formation. And within the social roles, they list: cognitive centrality, individual characteristics, and expertise. This type of study is valuable to understand group dynamics and to better tune group recommender systems to account for real-world group behavior.

3.2 Music Recommendation

Music is a special type of content that contains audio features and that is potentially ubiquitous. We describe the different levels and types of audio features that are specific to music content. We then overview context-aware methodologies for music retrieval and recommendation, listing some of the types of context available. Finally, we summarize the work of two relevant doctoral theses in the area, where the authors have, respectively, tackled the problems of music discovery using the long tail, and music recommendation by integrating social and content-based features.

3.2.1 Audio-Based Features

In April 2011, Fu et al. (2011) surveyed the state of the art in audio-based music classification and annotation for music information retrieval. They described three different levels of audio features — low-level, mid-level and top-level — and several classification techniques based on audio content – including k-Nearest Neighbor and Support-Vector Machines. They also identified some of the open problems in the area, including genre classification, which can still be accomplished much more efficiently by a human than automatically by a machine. Low-level features are easily extracted and are usually tied to timbre (short-term) and temporal characteristics (long-term). Mid-level features are usually long-term, meaning they can only be acquired by analyzing a large temporal window, and are frequently associated with rhythm, pitch and harmony. Top-level features result of an interpretation of the mid-level features for the classification of music according, for instance, to genre, mood or instrument. It is said that there is a semantic gap between mid-level and top-level features and that top-level labels are a human-readable representation of mid-level features.

3.2.2 Context-Based Methodologies

In April 2012, Kaminskas and Ricci (2012) published an overview on the state of the art of contextual music information retrieval and recommendation, pointing out some of the open challenges in the area. They focused on the context provided by the ubiquitous computing of the present, a time of smartphones and tablets, the Internet of Things and QR codes. The idea was to give the user the best music by taking into account: environment-related context, such as location, time of day, weather, traffic or noise; user-related context, such as current activity, demographical information or emotional state; multimedia context, such as text (e.g. lyrics) or image (e.g. album covers); and social context, for instance obtained from music social networks or music folksonomies.

3.2.3 Using the Long Tail to Improve Discovery

In his PhD thesis, Celma (2009) has focused on music recommendation by exploring the long tail of the music popularity distribution. He took advantage of the fact that there is as much value on a large group of less played songs as there is in a few highly played songs:

> *"The Long Tail is the realization that the sum of many small markets is worth as much, if not more, than a few large markets."*
>
> — Jason Foster (Anderson, 2005)

More importantly, Celma has taken into consideration an important characteristic of music recommendation: discovery. A problem with recommender systems is that they tend to be biased towards suggesting popular content, which easily gets a higher score simply due to the reason that most users like this content. Assuming that the process of discovery through a recommender engine should introduce a higher gain when faced with traditional methods, and knowing that popular content is easily spread through traditional media and word-of-mouth, focusing on the items in the long tail will indeed improve on the gain provided by a recommender system. Celma proposes that music recommendation should both take advantage of the user-item similarity network and the popularity of the item, decreasing the score of the recommended item alongside its popularity, in order to increase the number of recommendations in the long tail.

A similar approach has been taken by Park and Tuzhilin (2008), where they defined a cutting point α to divide the head from the tail in the item set distribution. They then used different recommendation methodologies for the head set and for the tail set: the *Each Item*, where they defined a custom data mining model for each item in the head set, and the *Clustered Tail*, where they defined a custom data mining model for each of the identified clusters in the tail.

3.2.4 Using Social Features to Improve Recommendation

Fields (2011) has also worked on music recommendation, however he has focused on the task of automated playlist generation for music recommendation. He proposed a novel multimodal similarity measure that integrates the social dimension with content-based features by taking advantage of the MySpace artists top friends network. He expands this network by using songs as nodes and weighting the edges of this socially induced network with the similarity between songs using audio-based features. Fields points out that, while it is obvious that playlist generation highly depends on the understanding of relationships between songs, no previous work [prior to 2011] has acknowledged this fact. Regarding community detection methodologies, which are of central interest to this survey, Fields also demonstrated a successful application to the artist-based song similarity network in order to cluster different musical genres successfully. While our problem is not directly related to musical genre discovery, community detection methodologies will be fundamental in the characterization of group members by identifying their individual social context.

Other work Knijf et al. (2011); Ma et al. (2011); Sen et al. (2009) has focused on improving recommender systems by taking advantage of the social context available in social networks and folksonomies. Ma et al. (2011) have established this context based on social trust networks, which they combined with the user-item rating matrix based on probabilistic matrix factorization. Sen et al. (2009) proposed several implicit and explicit tag-based algorithms to predict users ratings for movies, based on their inferred tag preferences. Knijf et al. (2011) proposed a recommendation methodology based on random walks with restart, to find related nodes in a bipartite artists-tags network, mined from social bookmarking services such as Last.fm[2].

Hypergraph-Based Algorithms Some work makes use of hypergraphs to combine social information with other types of music features. Bu et al. (2010)

[2] http://www.last.fm/

and Tan et al. (2011) proposed modeling music recommendation as a ranking problem in a hypergraph, combining users, groups of users, tags, tracks, albums and artists as nodes in a unified hypergraph model, and using a regularization framework to derive the ranking results for query vertices. Recommendations correspond to the track nodes with highest ranking, for a given user node query, and are limited to songs the user hasn't listened to before.

Theodoridis et al. (2013) extended on the work by Bu el al. and Tan et al. by also modeling music recommendation as a ranking problem on a unified hypergraph. However, they solved the ranking problem using a group sparse optimization approach. They showed that the non-overlapping group structure of the hypergraph significantly improved the accuracy of recommendations.

4 Final Remarks

A great deal of focus has been given to using the community structure of music networks as a feature to improve music recommender systems. However, there has been little work done on community-aware music recommendation, even though it is different to recommend for people interacting as a group than it is to recommend for individuals. This is an interesting problem that remains an open challenge worth exploring in the future. The main goal is to recommend music for groups of people, by taking advantage of latent information found in the communities of the individual members of the group. Music recommendation for groups can be done by recommending the best songs for the group as a whole or by combining the best songs for the individuals that form the group. These individuals inherently belong to several communities that might overlap with each other or with the communities of other individuals in the group. We should use this to our advantage, in order to characterize the group as a whole based on the social characteristics of the individual members. Future work should be centered on the research and improvement of context-aware group recommender systems for music discovery, specially focusing on the social and cultural context harnessed from the community structure in social networks.

References

Anderson, C. (2005). The Long Tail – Definitions: Final Round! http://www.longtail.com/the_long_tail/2005/01/definitions_fin.html. (Accessed: Nov. 22, 2013).

Baur, D. (2011). *The Songs of Our Past: Visualizing Music Listening Histories*. Ph. D. thesis, Fakultät für Mathematik, Informatik und Statistik, der Ludwig-Maximilians-Universität München.

Berry, M. W., S. T. Dumais, and G. W. O'Brien (1995). Using Linear Algebra for Intelligent Information Retrieval. *SIAM Review 37*(4), 573–595.

Bu, J., S. Tan, C. Chen, C. Wang, and H. Wu (2010). Music recommendation by unified hypergraph: combining social media information and music content. In *Proceedings of the international conference on Multimedia (MM 2010)*, New York, NY, USA, pp. 391–400. ACM.

Carneiro, M. J. (2012). *Towards the discovery of temporal patterns in music listening using Last.fm profiles.* Master's thesis, Faculdade de Engenharia da Universidade do Porto.

Castro, N. C. and P. J. Azevedo (2012). Significant motifs in time series. *Statistical Analysis and Data Mining 5*(1), 35–53.

Celma, O. (2009). *Music recommendation and discovery in the long tail.* Phd thesis, Universitat Pompeu Fabra.

Celma, O. and P. Herrera (2008). A new approach to evaluating novel recommendations. In *Proceedings of the 2008 ACM conference on Recommender systems (RecSys 2008)*, New York, NY, USA, pp. 179–186. ACM.

Crossen, A., J. Budzik, and K. Hammond (2002). Flytrap: Intelligent Group Music Recommendation. In *Proceedings of the 7th international conference on Intelligent user interfaces (IUI 2002)*, New York, NY, USA, pp. 184–185. ACM.

Cunningham, S. J., D. Bainbridge, and D. McKay (2007). Finding new music: a diary study of everyday encounter with novel songs. In *Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR 2007)*, Vienna, AT, pp. 546. Österreichische Computer Gesellschaft.

Cunningham, S. J. and D. M. Nichols (2009). Exploring social music behaviour: An investigation of music selection at parties. In *Proceedings of the 10th International Society for Music Information Retrieval Conference (ISMIR 2009)*, Kobe, Japan, pp. 747–752. International Society for Music Information Retrieval.

Deventer, O. V., J. de Wit, J. Vanattenhoven, and M. Guelbahar (2013). Group recommendation in an Hybrid Broadcast Broadband Television context. In *In Proceedings of the Workshop on Group Recommender Systems: Concepts, Technology, Evaluation (GroupRS), at the 21st Conference on User Modeling, Adaptation and Personalization (UMAP 2013)*, Rome, Italy, pp. 12–18. Springer.

Fields, B. (2011). *Contextualize your listening: The playlist as recommendation engine.* Phd thesis, Goldsmiths, University of London.

Fu, Z., G. Lu, K. M. Ting, and D. Zhang (2011, April). A Survey of Audio-Based Music Classification and Annotation. *IEEE Transactions on Multimedia 13*(2), 303–319.

Herr, S., A. Rösch, C. Beckmann, and T. Gross (2012). Informing the design of group recommender systems. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems (CHI 2012)*, New York, New York, USA, pp. 2507–2512. ACM Press.

Jameson, A. and B. Smyth (2007). Recommendation to groups. In *The Adaptive Web*, pp. 596–627. Berlin: Springer.

Jannach, D., M. Zanker, A. Felfernig, and G. Friedrich (2010). *Recommender systems: an introduction.* Cambridge, UK: Cambridge University Press.

Kaminskas, M. and F. Ricci (2012, May). Contextual music information retrieval and recommendation: State of the art and challenges. *Computer Science Review 6*(2-3), 89–119.

Knijf, J. D., A. Liekens, and B. Goethals (2011). GaMuSo: graph base music recommendation in a social bookmarking service. In *Advances in Intelligent Data Analysis X*, New York, NY, USA, pp. 138–149. Springer.

Koren, Y., R. Bell, and C. Volinsky (2009). Matrix factorization techniques for recommender systems. *Computer 42*(8), 30–37.

Lee, J. H. and S. J. Cunningham (2012). The impact (or non-impact) of user studies in music information retrieval. In *Proceedings of the 13th International Society for Music Information Retrieval Conference (ISMIR 2012)*, Porto, Portugal, pp. 391–396. FEUP Edições.

Lee, J. H. and N. M. Waterman (2012). Understanding User Requirements for Music Information Services. In *Proceedings of the 13th International Society for Music Information Retrieval Conference (ISMIR 2012)*, Porto, Portugal, pp. 253–258. FEUP Edições.

Ma, H., T. C. Zhou, M. R. Lyu, and I. King (2011, April). Improving Recommender Systems by Incorporating Social Contextual Information. *ACM Transactions on Information Systems 29*(2), 1–23.

Masthoff, J. (2004, February). Group Modeling: Selecting a Sequence of Television Items to Suit a Group of Viewers. *User Modeling and User-Adapted Interaction 14*(1), 37–85.

O'Connor, M., D. Cosley, J. Konstan, and J. Riedl (2002). PolyLens: A recommender system for groups of users. In *Proceedings of the Seventh European Conference on Computer Supported Cooperative Work (ECSCW 2001)*, Bonn, Germany, pp. 16–20. Springer.

Park, C. H. and M. Kahng (2010, August). Temporal Dynamics in Music Listening Behavior: A Case Study of Online Music Service. In *2010 IEEE/ACIS 9th International Conference on Computer and Information Science*, Kaminoyama, Yamagata, Japan, pp. 573–578. IEEE Computer Society & Springer-Verlag.

Park, Y.-J. and A. Tuzhilin (2008). The long tail of recommender systems and how to leverage it. In *Proceedings of the 2008 ACM conference on Recommender systems - RecSys '08*, New York, New York, USA, pp. 11. ACM Press.

Piliponyte, A., F. Ricci, and J. Koschwitz (2013). Sequential Music Recommendations for Groups by Balancing User Satisfaction. In *In Proceedings of the Workshop on Group Recommender Systems: Concepts, Technology, Evaluation (GroupRS), at the 21st Conference on User Modeling, Adaptation and Personalization (UMAP 2013)*, Rome, Italy, pp. 6–11. Springer.

Ricci, F., L. Rokach, B. Shapira, and P. B. Kantor (2011). *Recommender Systems Handbook*. Boston, MA: Springer US.

Said, A. (2013). *Evaluating the Accuracy and Utility of Recommender Systems*. Phd thesis, Fakultät Elektrotechnik und Informatik, Technische Universität Berlin.

Sarwar, B., G. Karypis, J. Konstan, and J. Riedl (2002). Incremental Singular Value Decomposition Algorithms for Highly Scalable Recommender Systems. In *Proceedings of the 5th International Conference on Computer and Information Science*, Los Alamitos, CA, USA, pp. 27–28. IEEE Computer Society.

Schutt, R. (2012). Week 7: hunch.com, Recommendation Engines, SVD, Alternating Least Squares, Convexity, Filter Bubbles. http://columbiadatascience.com/2012/10/18/week-7-hunch-com-recommendation-engines-svd-alternating-least-squares-convexity-filter-bubbles/. (Accessed: Nov. 11, 2013).

Sen, S., J. Vig, and J. Riedl (2009). Tagommenders: connecting users to items through tags. In *Proceedings of the 18th international conference on World wide web*, New York, NY, USA, pp. 671–680. ACM.

Smyth, B., E. Balfe, J. Freyne, P. Briggs, M. Coyle, and O. Boydell (2005, April). Exploiting Query Repetition and Regularity in an Adaptive Community-Based Web Search Engine. *User Modeling and User-Adapted Interaction 14*(5), 383–423.

Tan, S., J. Bu, C. Chen, Bin Xu, C. Wang, and X. He (2011). Using Rich Social Information for Music Recommendation via Hypergraph Model. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP): Special section on ACM multimedia 2010 best paper candidates, and issue on social media 7S*(1), 22.

Theodoridis, A., C. Kotropoulos, and Y. Panagakis (2013). Music Recommendation Using Hypergraphs and Group Sparsity. In *Proceedings of the 38th International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2013)*, Los Alamitos, CA, USA, pp. 56–60. IEEE Computer Society.

Zhiwen, Y., Z. Xingshe, and Z. Daqing (2005). An adaptive in-vehicle multimedia recommender for group users. In *2005 IEEE 61st Vehicular Technology Conference (VTC 2005)*, Los Alamitos, CA, USA, pp. 2800–2804. IEEE Computer Society.