TastePaths: Enabling deeper exploration and understanding of personal preferences in recommender systems

ANONYMOUS AUTHOR(S)

ACM Reference Format:

Anonymous Author(s). 2021. TastePaths: Enabling deeper exploration and understanding of personal preferences in recommender systems. 1, 1 (October 2021), 21 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 ABSTRACT

Recommender systems are ubiquitous and influence the information we consume daily by helping us navigate vast catalogs of information like music databases. However, their linear approach of surfacing content in ranked lists limits their ability to help us grow and understand our personal preferences. In this paper, we study how we can better support users in exploring a novel space, specifically focusing on music genres. Informed by interviews with expert music listeners, we developed TastePaths: an interactive web tool that helps users explore an overview of the genre-space via a graph of connected artists. We conducted a comparative user study with 16 participants where each of them used a *personalized* version of TastePaths (built with a set of artists the user listens to frequently) and a *non-personalized* one (based on a set of the most popular artists in a genre). We find that participants employed various strategies to explore the space. Overall, they greatly preferred the personalized version as it helped anchor their exploration and provided recommendations that were more compatible with their personal taste. In addition to that, TastePaths helped participants specify and articulate their interest in the genre and gave them a better understanding of the system's organization of music. Based on our findings, we discuss opportunities and challenges for incorporating more control and expressive feedback in recommendation systems to help users explore spaces beyond their immediate interests and improve these systems' underlying algorithms.

2 INTRODUCTION

Recommender systems play an essential role in determining the information we consume in our daily lives; they provide suggestions on movies to watch, articles to read, music to listen to, and more [1, 21, 40]. The goal of these systems is to help users quickly find content they like in a vast library of information. As such, considerable work has been done to improve the algorithms that predict what unseen content the user is likely to consume [30, 57]. While improving these algorithms generally increases user satisfaction [29], there is a danger that users might get stuck in what is called the "filter bubble", an overly personalized area in the recommendation space that isolates users from other content [46]. These bubbles could in turn reduce users' creativity and individuality by leading many to similar content that is "easy" to consume and fulfills short-term goals, instead of content that helps them further explore and understand

Manuscript submitted to ACM

their interests [47]. Recommender systems are a vital aspect of our current and future digital world, and so they can be further leveraged to help users grow and develop their personal preferences.

Toward improving recommender systems from a user-centric point of view, a few avenues have been explored. Researchers have developed evaluation frameworks that not only consider the algorithm but also the user experience, like user-choice satisfaction and presentation of recommendations [28, 49]. Besides reconsidering their evaluation, researchers have also pushed the scope of their purpose, calling for recommender systems for "self-actualization" [25, 27]. They argue that instead of focusing on consumption, recommender systems should give agency back to the user: they should enable users to develop their individuality and preferences by helping them understand the item space, explore further, and learn about their interests.

Toward better supporting self-actualization in recommender systems, there are still many questions to answer. Existing informational retrieval systems tend to have linear interfaces, presenting a ranked list of items based on some relevance score that is often not clear to the user [56]. While this approach works well for finding relevant documents from keywords, it is unclear if it is just as effective for people subjectively exploring an unfamiliar topic. Past work has shown that personalization is useful for exploring new areas in the item-space [36, 37]. However, the role of personalization in further exploring a personal preference has yet to be fully investigated. Numerous interactive systems have been created for exploring a recommendation space [3, 7, 26]. While they make innovations in interactivity, they do not address how users explore and would like to explore within one of their interests. Finally, to help users learn about themselves and their habits, past work has focused on assisting them to understand what part of the item-space they consume content from at a high level [32, 44, 54]. However, the role of learning in exploration can be further examined.

To shed light on these areas, we study how to support users in exploring and understanding a music genre they commonly listen to. Like other information retrieval systems, the linear nature of these systems in the context of music limits them when they are being used as tools for exploration. We also focus on music because it is a common use case of recommender systems, with hundreds of millions of users [42]. At the same time, consuming music requires much less time than watching a movie or reading an article, making it a good medium to observe how people prefer to explore their interests in real-time. Finally, music genres are a natural way to convey and communicate an interest. Genres encapsulate information about the sounds, culture, and time-period a particular group of artists belongs to [19]. Because of this, people commonly use genres to communicate their musical interests to others [18]. For these reasons, we study how to help users deeply explore and understand their interests through music genres.

Accordingly, we pose the following research questions:

- **RQ1: Personalization.** What is the role of personalization in helping users explore and understand the music genres they listen to?
- **RQ2: Exploration.** If you remove the linear constraint of a music information retrieval system, what strategies do users employ to explore the genres that interest them? How can we better support users in their exploration processes?
- **RQ3: Learning.** How does learning about their music preferences help users? What would they like to learn?

To investigate how to best explore a music genre and what it means to understand one better, we conducted a formative study in which we observed music experts explore genres. We learned what experts look for when exploring a genre, which helped us formulate three design goals. Using those goals, we built TastePaths: an interactive genre-exploration web tool. TastePaths helps non-music experts explore and understand genres that interest them by presenting an overview of the genre landscape as a clustered graph of related artists. To help users quickly make sense of these clusters, each is labeled with its three most representative sub-genres. Finally, there are two versions of TastePaths: personalized and non-personalized. These versions provide different reference points for exploration. In the personalized version, the graph of artists is grown from three artists the user frequently listens to within that genre. In the non-personalized version, the graph is grown from the three most popular artists in the genre.

To answer our research questions, we conducted a within-subjects user study with 16 participants, where we compared the two versions of TastePaths. For RQ1, we found that participants greatly preferred the personalized version and wanted even more of their listening data reflected in the interface. Regarding RQ2, participants also had a variety of exploration strategies. They also wanted more control, desiring to prune and grow the graph to guide their exploration, and they made meaningful discoveries between or on the edge of clusters. Finally, with respect to RQ3, participants found that exploring with TastePaths improved their mental map of the underlying recommender system's organization; they felt better able to verbalize and search for the music that interested them within a genre.

Our work contributes in three ways: first, we derived three design goals for supporting interactive exploration of a music genre. These were identified from expert interviews with professional music curators who have years of experience exploring new music genres. Second, we present insights addressing our three research questions derived from our prototype-based study of how participants used TastePaths, our interactive genre-exploration tool. Lastly, we discuss opportunities for incorporating more control and expressive feedback in music recommender systems and for utilizing this feedback to improve their underlying algorithms. We also discuss incorporating finite, goal-based consumption into these systems to encourage meaningful and active exploration.

3 RELATED WORK

In the following section, we discuss the need for the intelligibility of recommendations and the understanding of personal preferences in recommendation systems and current tools that support this. Next, we situate TastePaths in recent work that supports interactive music exploration tools. Finally, we discuss TastePaths in the greater context of interactive systems which aim to support users in exploring and making sense of large datasets.

3.1 Understanding Personal Preferences in Recommender Systems

Recommender systems help users navigate a sea of content by showing them items close to their current preferences. While they effectively prevent information overload, there is a concern that recommender systems might be guiding users to an overly personalized space, called the "filter bubble" [47]. This concern has motivated research toward providing evidence that recommendation algorithms actually lead users to filter bubbles, and so far, the results are conflicting at best [43, 46, 59]. Regardless, users are concerned about the content they consume, and in response, there has been research to investigate how to help users understand the items that are recommended to them.

Toward helping users understand the content they consume, past work has mainly focused on making users aware of where their recommendations come from at a global level. In this vein, Tintarev et al. investigated how to help users understand their movie-genre consumption profiles and found that visualizing the distribution of genres helped users understand their broader "blind-spots" in the recommendation space [54]. Nagulendra and Vassileva extended these ideas to the social media domain and created an overview visualization that reveals what categories of content the system is and is not showing in the user's newsfeed, as well as the friends who shared that content [44]. This visualization increased users' awareness that they were viewing a small subset of the recommendation space, helping them feel more in control and more knowledgeable of the content they see. Finally, *NewsViz* also produces an overview of an entire recommendation space as a tree-map, with which users can interactively resize categories to change the distribution of their recommendations [32]. This interactive overview made the system more transparent to users and helped them feel more in control of their recommendations. While these works focus on helping users understand the recommendation space at a global level, there is still more to understand about how to support users in learning about a specific interest of theirs. Our work investigates how to help users learn more about the music genres they listen to frequently.

3.2 Supporting Music Discovery and Exploration

By understanding their own consumption profiles, users are better equipped to discover new content; they know where to look for less familiar content and where to go to dive deeper. Toward self-actualization, discovery is an essential component for supporting growth and development [27], and in music recommender systems, discovery has been identified as an important need for music listeners [20, 33–35]. Currently, however, popular music streaming platforms do not support discovery and exploration that well, but optimize search instead [23]. To fill these gaps, researchers have proposed many different solutions to support exploration of new content in music recommender systems.

One approach, requiring minimal user effort, is to help users find new content through discovery playlists, which introduce users to new areas in the music-recommendation space. These playlists can be generated and sequenced in many different ways. To ease users into new content, Taramigkou et al. generated a playlist that gradually takes a user from their current listening preferences to a new desired genre [53]. Instead of taking users on a gradual path with the playlist, Liang and Willemsen experimented with playlists that immediately introduce the user to a new genre [36]. They found that discovery playlists should be personalized enough so that users can have a smooth entry into the genre but also need to be representative enough so that users can understand the genre's sound. In another study, Liang and Willemsen found that personalization can help nudge listeners towards new and more distant genres [38]. However, while discovery playlists can be an effective means of introducing new content with minimal effort, such playlists and other linear lists of recommendations in general are neither transparent nor controllable [56], which reduces users' acceptance of these recommendations.

To place users at the center of the recommendation process and increase intelligibility, researchers have developed many tools which incorporate interactivity to actively modify recommendations. One such system is *TagFlip* [26], which lets users specify social tags that are associated with the next song. Compared to the mobile Spotify interface, *TagFlip* was perceived to enable more control and transparency over recommendations. Supporting more fine-grained control, *Tasteweights* lets users adjust slider-weights for three different social and semantic web sources to generate new playlists [7]. *Tasteweights* also visualizes the connections between these sliders and the playlist to help explain the output; this combination of interactivity and visual explanation helped users understand how their recommendations were generated. Further exploring the interplay of interactivity and explanations, Millecamp et al. studied how users with different personalities perceived an interactive playlist generator, where sliders mapped to acoustic attributes [41]. They found that while explanations were beneficial for most users, they were less beneficial for others who felt that the explanation did not help them generate a better playlist. Thus, while interactivity paired with explanations generally helps users better understand the recommendation algorithm, it may not always help them find the music sub-space they most enjoy.

To help users easily locate and specify their preferences, researchers have used overview visualizations, which depict a larger portion of the music-space for users to explore. These overviews can be created and organized in many different ways. *Moodplay* organizes artists within a two-dimensional mood-space [3]. Users found navigating by mood to be fun and intuitive; they were generally able to find a sub-space that fit their current mood preference. These overview visualizations can also serve to situate and compare a user's preferences to a greater area in the music space, such as a genre. Liang and Willemsen created a mood-based contour-plot visualization that plots songs from a genre and the user's profile in the space [37]. Users found navigating a new genre with the contour plot more helpful than with a bar chart visualization which did not offer the same comparison. Finally, these overview visualizations can also be a more intuitive way to elicit feedback from users on their preferences. Kunkel et al. developed a 3D visualization that presents the entire item space in a map [31]. Users could delineate their preferences by either raising or lowering the elevation in certain regions of the map, and they generally found doing so natural and easy. In this work, we further study overview visualizations in the context of helping users deeply explore an established interest.

3.3 Exploratory Search and Sensemaking

TastePaths is highly related to and inspired by the broader areas of exploratory search and sensemaking. There are two interaction approaches in sensemaking [48]: (1) starting with an overview of the space, like *Recipescape* [13] and *Scatter/Gather* [16], or (2) starting with an example, like *Apolo* [14] and *SearchLens* [12]. When starting with an example, users can build up their own understanding of the underlying data with the help of the sensemaking system. For instance, as users manually organize articles into clusters, *Apolo* suggests related articles that fit the current organization. If the data is particularly complex or cumbersome to manually organize, some sensemaking systems offer a pre-computed overview for users to explore, like *Recipescape*, which generates clusters of different cooking methods for a particular dish. With this overview, users can spend their time understanding the space and finding interesting information. In Recipescape, users found unique recipes with less traditional methods and ingredients at the edges of clusters. In this work we extend these ideas from sensemaking to interactive recommender systems and provide a *personalized* overview to help users further understand their interest in a genre. At the same time, we study how they explore this overview and where they find meaningful discoveries, to improve future recommender systems.

4 METHOD OVERVIEW

To answer the research questions posed in this paper, we conducted two studies. The first one was a formative study, described in Section 5, during which we interviewed five professional music curators about their process of exploring a novel genre space. Based on the findings from those interviews, we defined three design goals that we used to inform the development of TastePaths. We illustrate the interface and implementation of TastePaths in Sections 6 and 7. Finally, in Section 8, we describe the procedure and findings from the comparative user study with 16 users who were interested in discovering new music.

5 FORMATIVE STUDY

To understand how to best help music listeners explore a novel music space, we wanted to gain insights from how expert music curators explore genres and what information they think is necessary to learn about a genre. Interviewing experts and distilling their process for design goals to help novices is a common practice and has been used to help novices generate compound icons [58] and explore recipes for dishes [13]. Professional music curators focus on labeling, organizing, and describing large volumes of music. Their job necessitates expertise in multiple genres and involves exploring genres daily. Because of their in-depth expertise, we observed music curators' process as they explored two genres of music they were interested in but less familiar with.

5.1 Procedure

We interviewed five professional music curators who all worked at the same large music streaming company and had between 2 to 10 years of experience (five male, average age 37.6). For the interview, we asked each expert to explore two genres of music for 15 minutes each. We encouraged them to use any tool or service they would normally use and to think aloud describing their process as they explored. For the first genre, they were asked to choose anything they wanted to explore. We were interested in seeing what kind of genre they picked and how granular or broad that genre would be. For the second genre, they were asked to pick from a set of the most popular genres from Rate Your Music[1], a large online music database: ambient, blues, classical music, country, electronic, experimental, folk, hip hop, jazz, metal, pop, punk, r&b, rock, and singer/songwriter. We wanted to see the experts' strategies for exploring these well-known genres. Finally, after exploring both genres, the curators were asked a series of questions, focusing on more details around their process, such as what information they were specifically looking for, why they used certain tools, and their overall strategy.

5.2 Findings

Overall, we found that the five experts we interviewed had a similar process for exploring a new genre. We describe their process in detail below, and building on it, we extract three design goals that we later applied to TastePaths.

When they first started exploring a genre, the experts generally looked for lists of representative artists and tried to identify ones they were already familiar with. This gave them some context around what the genre might sound like. For example, when P5 first searched for *outlaw country* on Google, he recognized Willie Nelson as one of the artists on the returned list. Related to that, P3 already knew a few of the notable artists in *drum and bass*, so his first step in exploring that genre was to search for them on Spotify. He explained that *"It's more you have an artist that interests you, and then you become interested in [the] genre once you find a collective of artists."* This implies that having a familiar artist within the genre provides a helpful starting point for exploration. This finding led us to our first design goal for TastePaths: **to anchor genre exploration in artists the user is already listening to**.

After listening to a few key artists to get a sense of the general sound, the experts focused on the genre at a higher level. They looked for information on its history and variety, such as the different sounds and subgenres that comprised it or its stylistic origins. For example, while reading about *outlaw country*, P5 examined its related genres because *"it helps contextualize this genre. I'm looking at these [related] genres, seeing if I recognize them and thinking about their musical or other types of qualities and trying to relate that back to what I just read about outlaw country."* Thus, he was using his prior knowledge to better understand how the new genre fits in the greater music landscape. In the same vein, P2 explained that knowing a genre better is being able to *"identify things about it that were different from other things."* Overall, this ability to contextualize a genre, be aware of its components, and know what it's related to and different from, is an essential part of understanding a genre. Accordingly, our second design goal for TastePaths was **to help users get an overview of the genre-space in order to give them an idea of what it contains**.

Finally, throughout their process of exploring a genre, the experts enjoyed diving deeper into certain artists. For example, P2 explored a band's related artists in the "fans also like" feature on Spotify. From there, he selected a few artists and spent time looking at their artist photos, reading a few lines from their descriptions, and then would sample a few of their most popular songs. P4 also listened to a few unfamiliar artists in more detail and noted that *"what I'm learning, is I need to redefine my definition of tango, because what I'm hearing is not what I was expecting. I was expecting*

[1] https://rateyourmusic.com/genres/

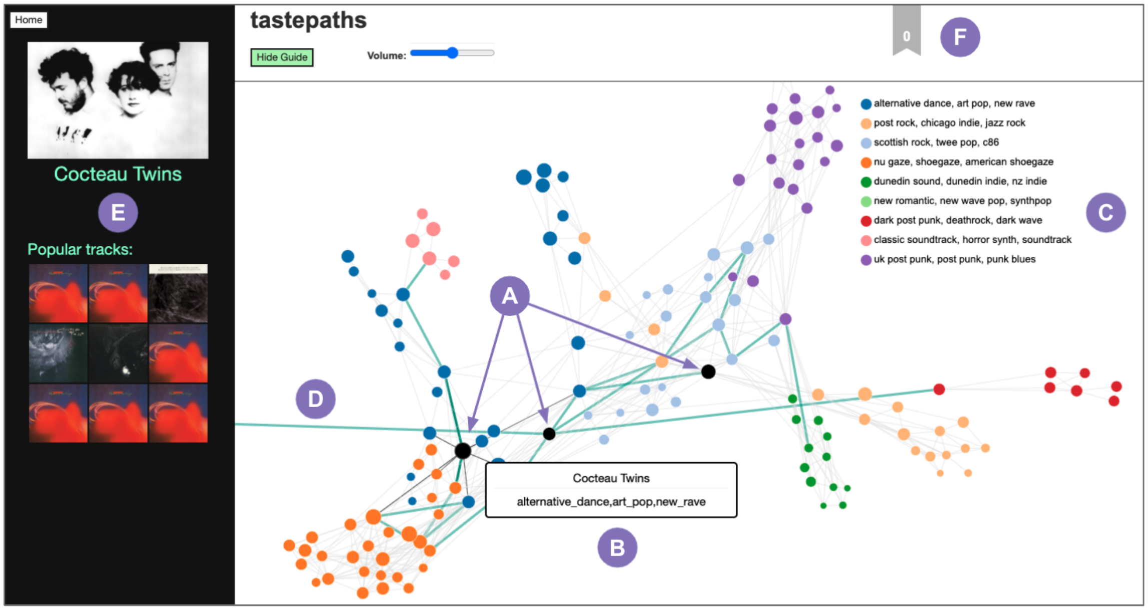

Fig. 1. **TastePaths interface for the genre *art pop*.** TastePaths displays a network of artists grown from the user's anchor artists, which are displayed as black nodes (A). When hovering over a node, the artist's name, here "Cocteau Twins", appears and a preview of their most popular track is automatically played (B). This network is clustered to capture different groups of artists within the genre. The three most representative genres for each cluster are shown in the legend (C). To help users navigate the network, there is a green path, called the "guide" (D), which connects the users' anchor artists to important artists within each cluster. By pressing a node, the sidebar displays the artist's cover art and their nine most popular tracks (E). To add a song to the playlist, the user has to click on one of these popular track images. The playlist itself is displayed when the ribbon is clicked (F). The ribbon also displays the number of songs currently in the playlist.

something more danceable. So that's super interesting." A common strategy among all the experts was to scan through an artist's most popular tracks on Spotify, which helped them evaluate if they wanted to explore this artist further. Being able to dive deeper and listen to an artist's music helped the experts direct their search towards a subsection of the genre they liked more. This inspired our third design goal for TastePaths: **to help users quickly and easily dive deeper into an artist's work**.

Design goals. In summary, we elucidated three design goals for TastePaths from the formative study:

D1: Anchor artists. To give users a meaningful starting point for exploring a genre, we should help anchor them with artists they already know and listen to.

D2: Genre-space overview. To contextualize the genre and help users understand it and its components, we should present an overview depicting the genre-space and its subgenres.

D3: Deep-dive. To allow users to easily assess what parts of the genre they like, we should have a quick and convenient way to deep-dive into an artist's work while being able to seamlessly go back to exploring.

6 TASTEPATHS INTERFACE

To address the design goals identified by the expert interviews, we created TastePaths: an interactive web tool, which enables users to better explore and understand a genre they are interested in (Figure 1). TastePaths visualizes a force-directed graph of related artists within a genre and assists the user in exploring and making sense of it. To address D1, TastePaths helps a user explore a genre by basing exploration from either three of their most frequently listened to artists in that genre or that genre's three most popular artists. We call these artists a user's "anchor artists"; they appear as black dots in the graph (Figure 1A). To address D2, TastePaths generates a graph consisting of 150 related artists stemming from the three anchor artists. To make this graph more of an overview, TastePaths clusters the artists and presents a legend, displaying each cluster's three most representative genres (1C). To help users navigate the graph, we relate each anchor artist to important artists, based on node-centrality, within each of the clusters through a highlighted green path called the "guide" (Figure 1D). Finally, to address D3, we enable a deeper dive into each artist by presenting an artist's image and top-9 tracks in the sidebar (Figure 1E) when a user clicks on their node in the graph. Users can listen to these tracks by hovering their mouse over the album covers, and they can add them to a playlist by clicking on them. Finally, users can view their current playlist by clicking on the ribbon on the top-right (Figure 1F).

7 TASTEPATHS IMPLEMENTATION

An essential part of exploring a genre is understanding its components and the different collections of artists that comprise it. To support this, TastePaths presents an overview of the genre-space via a graph of related artists within the genre. This graph is constructed from either the user's personalized anchor artists or the top-3 artists in the genre. To help users distinguish the different groups of artists in the graph, we cluster the graph and assign each cluster a label with its three most representative sub-genres. However, even with these labeled clusters, the graph is large and potentially difficult to explore. Therefore, to help users venture into these different clusters from their anchor artists, we highlight a path from their anchor artists to important artists within each cluster. Below, we describe each element of TastePaths' implementation.

Overall, TastePaths is implemented in the Flask web framework. To cluster the graph, calculate the betweenness centrality of each node, and create the Steiner tree for the guide, we use the python library NetworkX[2]. To get the song previews for each artist, the artist-to-genre data, and each artist's related artists, we use the Spotify Web API[3]. Finally, to visualize the force-directed graph of artists in the genre, we use D3.js[4] [8].

7.1 Identifying Genres to Explore and Anchor Artists

The first element of TastePaths is identifying what genre to visualize for a given user since TastePaths was built to help users deeply explore genres they have a demonstrated interest in. To fulfill D1 and anchor their exploration within the genre, we identify genres the user listens to often. Then we anchor their exploration with either (1) artists they listen to frequently within the genre (personalized version) or (2) the three most popular artists in that genre on a music streaming service (non-personalized version). To identify these genres, we access the list of tracks the user has listened to in the past 90 days on a popular music streaming service, and we retrieve the user's top-50 highest-affinity tracks. A user has a higher affinity toward a track if they have listened to it often and intentionally interacted with it (e.g. by adding it to a playlist or playing it). We then get the associated genres for each artist of these top-50 tracks and count

[2] https://networkx.org/
[3] https://developer.spotify.com/documentation/web-api/
[4] https://d3js.org/

the number of different artists and tracks that appear per genre. Any genre with less than three artists is removed from the resulting list; fewer than three artists might rather indicate an interest in those artists instead of the genre. Finally, we sort this list of genres by the number of tracks, since a greater number indicates a stronger interest in the artists and the genre. Any of these genres are suitable for exploration. In the personalized version, we take the three artists within a genre with the greatest number of tracks in the top-50 to serve as anchors. With these anchor artists, we can construct an overview through which the user can explore a genre.

7.2 Constructing the Related-artist Graph for a Genre

Next, we use the anchor artists to build a related-artist graph for the genre. To do so, we leverage a music streaming service's artist knowledge graph, where nodes are artists and edges between artists indicate that they share many listeners. We first construct an initial graph, which connects the three anchor artists (either personalized or non-personalized). To do this, we find the shortest path (via the bidirectional version of Dijkstra's algorithm) from the most popular anchor artist to each of the other anchor artists; all the intermediary nodes from these shortest paths are added to the initial graph. For each node in the initial graph, we add its two most related artists and their connections to the graph, creating a second layer of nodes. Next, from this second layer we do the same process and add two nodes for each artists, growing the graph layer by layer until there are 150 nodes in the graph, to ensure a reasonably large overview. By adding only two related artists per node, we grow the graph more deeply, capturing more groups of artists.

7.3 Clustering and Labeling the Graph

The graphs generated for a particular genre often revealed groups of densely connected artists that have many sub-genres in common, such as the dark orange *shoegaze* cluster in the bottom-left of Figure 1. To identify these densely connected groups, we clustered the graph using the Louvain graph clustering algorithm [5]. This algorithm clusters the graph hierarchically. We use the top-level clusters returned by the algorithm, which generally returned about 5-10 clusters for the graphs we generated.

To label these clusters in the overview, we select their top-3 most representative genres to include in the legend (Figure 1C). To do this, we access the set of genres associated with each artist in the graph, using the Spotify Web API[5]. Our first attempt to label a cluster was to pick the top-3 most common genres among all of the artists in the cluster. However, this top-3 would often include the name of the overarching genre, which would be shared across all the clusters, making them indistinguishable from each other. To find the genres that are common and unique to a particular cluster, we leverage a technique normally used in information retrieval called the term frequency-inverse document frequency (TF-IDF) [50]. This technique is used to determine how relevant a term is to a particular document in a collection of documents. Its output is a set of scores per term per document, where higher scores indicate that a term is specific to a document and lower scores indicate that the term appears often across all the documents. In our case, we treat each cluster as a document and its set of genres as its terms, and we select the top-3 genres with the highest TF-IDF scores to represent that genre. This method creates genre-labels that better distinguish clusters at the sub-genre level.

[5]https://developer.spotify.com/documentation/web-api/

7.4 Guiding Users with a Highlighted Path

Finally, the resulting graph can sometimes be densely connected, and it can be hard for users to see the connections from the anchor artists to each of the clusters. To make this overview easier to navigate, we highlight a simple path connecting the user's anchor artists to important artists in each cluster. This is visualized as a green path, called the "guide", in the graph (Figure 1D). We want the minimal number of edges connecting the anchor artists to each cluster to minimize visual complexity. This set of edges is called a Steiner tree [24], and we calculate it using an approximation algorithm in the NetworkX Python library[6].

To determine the important nodes in each cluster to include in the tree, we experimented with several node-centrality measures, including basic edge count, eigenvector centrality [45], and betweenness centrality [9]. Edge count and eigenvector centrality classified highly connected nodes at the center of clusters to be important. While accurate, the generated Steiner tree would then include many edges within each cluster to get to this central node. Meanwhile, betweenness centrality emphasized influential nodes at the edge of clusters that acted as "gateway artists" into that cluster. We decided to use this measure of centrality because the resulting Steiner tree was less visually complex.

8 EVALUATION

To answer our research questions and learn how TastePaths helps users understand and explore their interests, we conducted a within-subjects study, comparing a personalized version of TastePaths to a non-personalized one. To understand how users perceived the two versions of TastePaths, we analyzed data from multiple sources. This included a questionnaire that measured their perceived helpfulness of each version and user-logs containing their actions as they interacted with the systems. We also asked conducted semi-structured interviews and a thematic analysis on the resulting data. This gave us insights into how they perceived each version, explored the graphs, and what they learned about their interests.

8.1 Procedure

The general outline of the study was the following: (1) users were first interviewed on their music preferences and methods for finding new artists, (2) they then used the two versions of TastePaths to find new artists in two genres they commonly listen to, (3) after each version they filled out a questionnaire, rating their perceived engagement and helpfulness of the tool, (4) they were asked a series of questions on their thoughts of the tool in a semi-structured exit interview.

To set up the experiment, we selected two genres for each user to explore, one for each version of TastePaths. To do this, we followed the procedure outlined in Section 7.1 to get the user's top genres from their 90-day listening history. From this list of top genres, we selected their top two genres for them to explore. Sometimes these genres were similar and contained a few of the same anchor artists. In that case, we took their top genre-interest and then the next strongest genre-interest that featured no intersection with the top genre's anchor artists. To confirm that the participants were in fact interested in the genres we selected, we asked them to rate on a 7-point Likert scale how knowledgeable they are in that genre and how interested they are in exploring artists in that genre.

In the experiment phase of the study, participants were randomly assigned to a condition which determined which version they will interact with first: either personalized-anchors-first or popular-anchors-first. After being told the two genres they would explore, participants picked the order they explored them in. When interacting with TastePaths for

[6]https://networkx.org/

Metric	Statement (7-point Likert scale)
Engagement	Q1. It was entertaining and interesting to explore artists in the {first, second} network.
Interest	Q2. With the {first, second} network, I was able to find artists that matched my interest.
Serendipity	Q3. With the {first, second} network, I found artists that I had not considered in the first place but turned out to be a positive and surprising discovery.
Music Discovery	Q4. The {first, second} network helped me discover new artists.
Guidance	Q5. With the {first, second} network, it was easy to determine which artists I'd be interested in.
Confidence	Q6. I am confident I will like the songs in the playlist I made using the {first, second} network.
Learning	Q7. With the {first, second} network, I feel like I know more about the artists and sounds of the genre better than when I started.
Understanding	Q8. With the {first, second} network, I feel like I understand the artists and sounds of the genre better than before.

Table 1. Post-task questionnaire filled out by participants after they used a version of TastePaths to explore a genre. Each statement was rated on a 7-point Likert scale.

the first time, users were given a short explanation of the interface and its features. For each version, users were told if the anchor artists were either personalized or the most popular artists in that genre. They were then given ten minutes to find five new songs from five new artists. We wanted to see if either version of TastePaths would help users explore a genre more deeply and thus encourage them to find new artists; five artists seemed challenging enough but also doable given the time-limit. We also emphasized that this number was only to encourage them and that they should only add songs to their playlist if they were genuinely interested in that song. Participants were encouraged to talk and explain their process and actions as they explored the visualization.

After exploring each genre, participants were asked if they would like to save the playlist to their personal account associated with a music streaming service. They were also asked to fill out a questionnaire (Table 1) to understand their perception of the tool for the task. Finally, after experimenting with both versions of TastePaths, we asked them a series of questions that probed at their preference for each system, exploration strategies, and knowledge they gained from using the tool.

8.2 Questionnaire

The questionnaire we gave participants borrows ideas from a few common evaluation frameworks for recommender systems (Table 1). To understand how the system helped users find music, we measure user-perceived interest in the artists (Q2), music discovery (Q4), and user confidence (Q6), which are adapted from Cai et al. [11]. In the same vein, we also include a question on guidance (Q5), to see if one version made identifying interesting artists easier than the other. To understand if one version was more interesting to use than another, we also include Q1 to measure engagement [55]. Finally, to understand if one version of TastePaths helped users learn about the genre more than the other, we added Q7 and Q8.

	Personal	Non-personal	p-value
Engagement	6.56 (0.89)	6.75 (0.58)	.48
Interest	6.5 (0.89)	5.69 (1.7)	.047
Serendipity	6.25 (1.13)	5.31 (1.66)	.03
Music Discovery	6.75 (0.45)	5.94 (1.95)	.26
Guidance	5.94 (1.48)	5.69 (1.7)	.61
Confidence	5.94 (1.06)	6.25 (0.77)	.21
Learning	5.44 (1.55)	5.63 (1.54)	.46
Understanding	5.31 (1.7)	5.63 (1.75)	.27

Table 2. **Comparison of the Personal and Non-personal version of TastePaths for each category in the post-task questionnaire. 8 paired-sample Wilcoxon tests, with Bonferroni correction, show no significant differences for any metric in the questionnaire. Across the metrics the biggest difference is in *Interest* and *Serendipity*. TastePaths participants using the personalized version found more artists in their interest. In parenthesis is standard deviation.**

	Personal	Non-personal	p-value
Listening duration (minutes)	5.18 (1.4)	5.26 (1.5)	.82
Nodes deeply explored count	7.75 (2)	7.12 (2.6)	.39
Nodes hovered count	68.56 (24)	74.5 (28.2)	.25
Green-path nodes deeply explored count	1.63 (1.6)	1.93 (1.9)	.58
Clusters explored count	4.31 (1)	3.5 (1.2)	.036
Clusters count	7.31 (2.5)	6.13 (1.4)	.12
Playlist song count	7.25 (2.6)	5.68 (2.6)	.015
Playlist artist count	5.94 (1.5)	4.94 (2.4)	.079
Playlist saved	.875 (.33)	.625 (.48)	.045

Table 3. **Comparison of the Personal and Non-personal version of TastePaths across a number of measurements taken from the recorded user logs in the study. We conducted nine paired-sample Wilcoxon tests, with Bonferroni correction, and found no significant differences in each of these metrics. These logs indicate that in both conditions, users were highly engaged with the system, hovering over many artists and listening to music for over 50% of their time with each system. Participants using the personalized TastePaths made longer playlists on average, with more unique artists, and were more likely to save it. In parenthesis is standard deviation.**

8.3 Participants

We recruited 16 participants (P1-P16), 7 female, 8 male, and 1 non-binary person, from the dscout platform for remote studies[7]. Their ages ranged from 19 to 53, with the average being 31 years old, and they had diverse backgrounds (including diverse occupations, locations within the US, music interests, income levels). To be eligible for the study, they had to be over 18 years old, reside in the US and speak English, have a paid premium account on a music streaming service for at least a year, be interested in exploring new music, and listen to discovery-focused playlists at least once in the last three months on that music streaming service. The interviews were conducted remotely, and participants had to have a computer with a Google Chrome web browser. Consistent with internal guidelines, participants were reimbursed $100 for the 60-minute interview, paid via the dscout app.

[7]https://dscout.com/

8.4 Analysis

We performed inductive thematic analysis on the qualitative data from the semi-structured interviews [10, 51]. Through an iterative coding process, two of the authors coded the interview data and discussed any disagreements. Examples of codes included 'sonic comparison between different clusters' and 'helpful aspects of the legend.'

To analyze the questionnaire data, we conducted paired-sample Wilcoxon tests with Bonferroni correction, since we compared two paired groups with ordinal data. We found no significant differences between the two versions of TastePaths for any of the metrics (Table 2). To analyze the user-log data, we also conducted paired-sample Wilcoxon tests for the same reasons and once again found no significant differences between the two versions for any of the metrics (Table 3).

9 FINDINGS

Through our analysis, we identified four main themes: (1) personalization is key, (2) best discoveries are between or on the edge of genres, (3) users want more control: human-in-the-loop growing and pruning of the graph, (4) improved recommendation explainability through mental map. The first theme helps to address RQ1 by explaining why users preferred the personalized-anchors version of TastePaths. The second and third themes help to address RQ2 by summarizing (1) what exploration strategies were most effective and (2) how users imagined themselves interacting with the graph. Finally, the fourth theme addresses RQ4 by summarizing what users learned and wished they had learned.

9.1 Personalization is Key

Twelve out of 16 said they preferred the version of TastePaths with personalized anchor artists. Overall, participants found the personalized version more interesting since it featured more artists they knew and liked. P5 explained that the "*personalized was more useful because it was based on [my] specific taste*." On average, participants collected more songs they liked for their playlists with the personalized version. They collected an average of 7.25 (stdev=2.6) songs per playlist with the personalized version and 5.68 (stdev=2.6) with the non-personalized one (Table 3). This aligns with the questionnaire results, where users rated the personalized version higher for both music discovery and interest, suggesting that the personalized version of TastePaths was more helpful to find new and interesting artists.

Participants wanted even more personalization to better guide their exploration in the graph. Beyond the three personal anchor artists, participants imagined more ways their data could be visualized in the graph to help them explore. One idea posed by P9, P5, and P6 was to have the graph indicate which artists the user has listened to before. This way, they could focus on exploring unknown artists nearby the ones they had interacted with before. Other participants imagined even more advanced ways the visualization could be more personalized. P7, for example, wanted TastePaths to prioritize clusters based on her affinity towards them: "*Maybe if there were a couple of artists it knew I liked [and I could] find more direct correlations... [I] want to see a heat map almost - this is the hot spot of what you might like*." Participants wanted TastePaths to incorporate more personal listening data to better guide them to new content they are likely to enjoy in the graph.

9.2 Best Discoveries are Between or on the Edge of Genres

Participants found new artists they really liked between genre clusters or at the outskirts of a cluster. Artists between two clusters captured essences of two musical styles, which led to exciting discoveries when these were styles

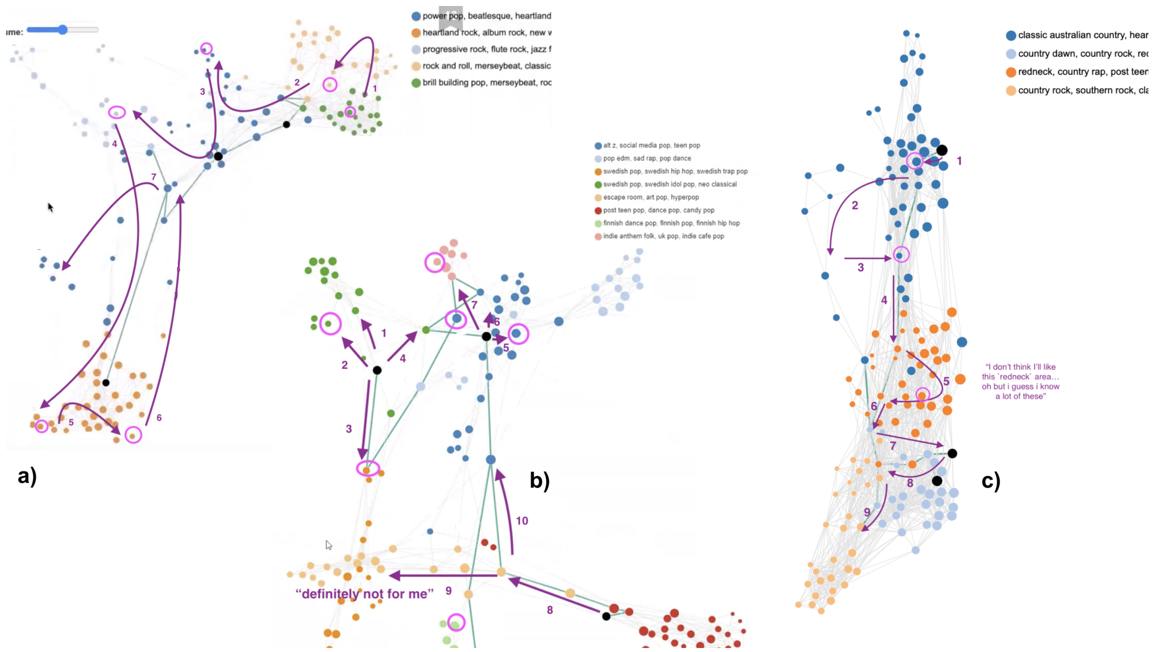

Fig. 2. **In this figure, we show three exploration strategies: a) systematic exploration, b) anchoring, and c) following the guide. In systematic exploration (a), users picked a cluster to start from and then systematically explored the other clusters one-by-one. In anchoring (b), users explored artists stemming directly from the three anchor artists. In following the guide (c), users followed the nodes along the green path guide. The numbered arrows indicate the order in which users explored the nodes.**

the participant enjoyed. For example, while exploring a personalized network for *dance pop*, P9 found an artist she had never heard of before between two clusters she typically enjoys: "*I've never heard of this person...and I like this because it's a blur between EDM and dance pop. This is definitely on the edge between those genres. I have a lot music that sits in the middle*." Similarly, while exploring a personalized network of *classic rock*, P8 found that he already knew most of the artists in the graph, and so the most interesting finds were at the edges of clusters: "*Outliers are most fun to discover because I haven't heard them in the past*." The richest discoveries for users when exploring a familiar genre were those that helped them dive deeper and united multiple aspects of that genre that they liked.

To make these meaningful discoveries, **participants employed several different strategies to explore the graph visualization**. The most common strategy was to use the legend to identify an interest in the graph (8 out of 16 participants). Since the legend summarized the three sub-genres that best describe each cluster, participants used their preexisting knowledge of sub-genres to pick a cluster that looked interesting. Along with the legend, participants also used the anchor artists to direct their exploration. A common strategy was to start with the anchor artists and explore their direct connections. This was more popular in the non-personalized condition, where participants would first inspect the popularity-based anchor artists, and if they liked them, explored nodes close to them (Figure 2B). A few participants also used the green path as a guide (Figure 2C). But this was generally an uncommon strategy; on average, participants clicked on less than two artists in the guide in both conditions (Table 3). Participants were less inclined to follow the guide to different clusters and more willing to jump around from cluster to cluster. Finally, a couple of the participants did not care for any guidance at all and started from one cluster and systematically explored the clusters one-by-one (Figure 2A). To them, it was more important to see all the clusters rather than to focus on any cluster in particular.

9.3 Users want more Control: Human-in-the-loop Growing and Pruning of the Graph

As participants explored, they expressed a desire for more control to shape and direct their search. They wanted to remove artists they had heard before and did not like. Also, a few participants mentioned removing or minimizing entire graph clusters that were less interesting. For example, P4 explained that while he would not want to completely remove a cluster, he wanted to "*de-emphasize it*" and interactively control which clusters to view by "*checking which of these genres to even see*." P7 echoed this idea and added that by pruning, she would be able to better explore other parts of the graph: "*It would be helpful [if i could prune this cluster], and get a better view of the pathways in Swedish pop*." By pruning, users would be able to create a graph that is (1) easier to explore and (2) more reflective of their interests.

In addition to removing portions of the graph, users also expressed a desire to grow the graph and imagined an adaptive guide that would lead them. For example, P6 discovered a cluster with a sub-genre she had not heard of called *soul flow* with many artists she liked; she wanted to expand the graph from this cluster and continue to explore it. Besides growing the graph, participants also wanted a more intelligent green-path guide that would change according to their input. P2 explained, "*I wish there was an adaptive guide: once you press on an artist, it would create one [and show] other things that might be on the same pace based on that song that you like*." P7 also imagined giving feedback on what she did not like along the path of the guide, which would redirect it to a sub-section she liked more. Overall, participants wanted even more interactivity to better explore the graph.

9.4 Improved Recommendation Explainability through Mental Map

With both the non-personalized and personalized versions of TastePaths, users gained a better understanding of how much variance exists within a genre, were able to understand what they liked and disliked within the genre, and grew their vocabulary to describe their interests. For example, after exploring a non-personalized graph for *pop*, P2 learned about the many genres within it: "*I did not really know these genre names... and now I know what it's called. Hopefully after this I will explore them a little more... Genre is usually an afterthought, so it was nice to see what genre they were in*." After exploring a personalized graph for *alternative r&b*, P10 felt like she better understood which part of the genre she actually liked, pinpointing artists with "*hints [of] underground and with hints of jazz*" as her strongest interest. She felt it was important to understand these sub-genres to better reason about where her recommendations were coming from in a music streaming service. Overall, TastePaths helped users better understand the different sounds within a genre as well as their own preferences.

However, beyond becoming acquainted with its sub-genres, participants wanted to learn specific information about the genre itself, including its sonic characteristics, history, and influences. From the questionnaire results, on a scale of one to seven, participants rated "learning" on average 5.44 (stdev=1.55) with the personalized version and 5.63 (stdev=1.54) with the non-personalized version (Table 2). While they generally felt they had learned something, they wished they had learned more. For example, P15 wanted a greater understanding of the graph's organizational structure, including more information on why artists were grouped together and descriptors for each cluster's sonic characteristics. Meanwhile, P9 wanted specific information at the artist level: "*I know more artists, but I don't necessarily know more about them... a little bit more about the artists or their background, their process, how they make their music, things generally about the genre. Something about influence - how house music came along, the lineage from Detroit EDM to house music* *etc.*" While not implemented in the current version, adding more information could improve the visualization in future iterations.

Interestingly, a couple of participants felt that the experience with TastePaths made them reevaluate their knowledge of the genre, sometimes even confusing them. After exploring a non-personalized graph of *pop* music, P10 felt overwhelmed at the vastness of that genre: "*I know nothing about pop now. I feel like pop has just become more confusing and now I'm lost in a sea of subgenres... I feel like I was sitting on a step and now they invited me in the house, and I'm like 'what'.*" Without being alienated by the entire genre, some participants felt a disconnect between the sub-genre names of the cluster and how they perceived the music. For example, while exploring a *chopped and screwed* cluster in a non-personalized *hip hop* network, P6 noticed a few artists that did not have that quintessential sound of the sub-genre: "*[chopped and screwed] is kind of slowed down and altered in some way... but this one does not sound so slowed down. So maybe within an artist they have a different vibe.*" Across their discographies, artists can make music that touches multiple genres, perhaps not making them the perfect fit for a cluster. Overall, a couple of participants felt a disconnect with TastePaths, in some cases at the entire genre level and in other cases with the label of a cluster.

10 DISCUSSION

10.1 Informing Future Recommender Systems

From the user study, we learned that participants wanted even more control than what TastePaths already provided to provide in-depth feedback to the system. They viewed the graph as the system's representation of their taste, and they imagined more clusters they could explore outside of their local interests at the graph's boundaries. Participants wanted to extend the graph in directions they liked while also editing this representation by de-emphasizing or pruning certain artists and clusters. This willingness to provide richer feedback has been shown in prior work [31], and is in stark contrast to the current methods offered by recommender systems to elicit feedback. Currently, systems either collect implicit feedback, such as play length and skips, which are not transparent to the user, or explicit feedback like ratings, which are cumbersome to collect [15, 22, 52] and perhaps even misleading [2]. Future recommender systems can include support for more expressive and natural feedback from users to tailor how the system represents and understands their interests.

By enabling expressive feedback, we can better inform the algorithms powering popular recommendation systems. One finding from our user study was that particularly rich and interesting discoveries would lie either between two clusters or on the edge of a cluster. This information could be used as implicit feedback to generate discovery playlists to enable further exploration with minimal effort. In the future, larger studies can be conducted to collect this implicit feedback and understand where users are making discoveries to help design better recommender systems and discovery playlists.

Finally, in addition to being useful for understanding how novices explore, TastePaths can also support experts in generating better curated playlists to improve recommendations. User-created playlists and their metadata are often used to calculate the similarity of tracks and artists [6]. From the formative study interviews with experts, we learned that they often used many different resources to explore a less familiar genre and generate a playlist, which requires a lot of time and effort. To provide more and better training data for models using these manually created playlists, we can enable experts to explore faster and more easily with TastePaths. Future work can include understanding how experts use TastePaths, what they discover, and using their exploration results to power recommender systems.

10.2 Providing Users Closure to Facilitate more Responsible AI

Participants in our study appreciated that the graph was both expansive and finite. They felt a sense of accomplishment having explored most of the clusters and pride if they realized they knew most of the artists within them. For example, while exploring a personalized graph of *mathcore*, P15 stated: "*I'm very proud of myself, in my metal fandom. Having seen a lot of these bands, I'm happy with the amount I have been able to recognize.*" Currently, many recommender systems do not design for an end to the experience but instead aim to maximize their share of the user's time. Because of this, users often consume content to their detriment, neglecting their other plans and goals [17] and losing their sense of agency [4]. Recent work has shown that users prefer versions of recommender systems that promote active interaction and agency when they have a specific intention in mind. One promising way to support agency is through planning [39]. By setting and following goals, users feel more in control of their consumption, as they feel there is an end to the process, unlike in an endless feed of media. Future work can extend these principles to music recommender systems. During a specific task like exploring a genre, music services could help users form goals on how far or how long to explore during the session to encourage growth and agency as opposed to longer listening sessions.

10.3 Guidelines for Helping Users Deeply Explore their Interests

While recommender systems are very useful for helping users find content that closely aligns with their current preferences, they can be augmented further to support users in deeply exploring and expanding their interests. By doing so, we could limit the effects of the filter bubble and promote creativity and individuality instead. From the user study, we established two general guidelines for helping users interactively explore and understand their interests: (1) anchor exploration with content the user knows well and help them venture out in many different ways and (2) help users learn about their interests so that they can recognize and consciously interact with their bubble.

In the user study, participants' prior knowledge helped them navigate the space more confidently. From the artists they knew, they were able to identify opportunity spaces; participants were excited to see an unknown artist, or cluster of artists, connecting two other artists they already liked. Future systems can provide multiple ways for users to explore new content from what they currently enjoy. This could include suggesting what lies between two items they know well (either articles, movies, or artists), or suggesting "gateway" items that are connected but less similar to their current interests to introduce them to a new cluster or adjacent genre within the space. These kind of recommendations can help users confidently explore new content outside of their immediate bubble.

As well as anchoring exploration from content the user currently enjoys, recommender systems can also help users learn about this content to help them understand and interact with their filter bubble. Past work has shown that providing a broad overview of the user's consumption increases their awareness of the content they consume and their feeling of control over it [32, 44]. In addition to helping users acknowledge what content they consume at a high level, we show that overviews can also help users better understand what exactly they like about a sub-area in the space, such as a genre. Future systems can provide users with a more fine-grained understanding of their taste by specifying both the broader categories the user is interested in, and the sub-categories that better reflect the user's taste. By incorporating this information, users will be aware of the system's representation of their interests; they can then consciously choose to remain within this bubble or to explore elsewhere.

10.4 Limitations and Future Work

While we carefully designed the study, it is not without limitations. One factor that varied across participants and the two conditions in the evaluation was the edge density and number of clusters in the graphs. While each graph was created in the same way and included exactly 150 nodes, the resulting structure of the graph and number of edges was variable. Because of this, some networks were very densely connected with fewer clusters like Figure 2C, while others were more spread-out with more clusters like Figure 2B. From the interviews, we found that participants generally preferred to explore networks that were more spread-out and had more clusters, and so there might be graph attributes affecting how participants explore. Future work could investigate how different graph shapes and clusters affect how users explore them.

Another limitation in this work is that we recruited participants interested in exploring new music and who have done so in the past three months. Therefore, our results apply more towards those who are open to exploring rather than the general populace of music listeners. Future work can study how to support users who are less willing to explore in understanding and exploring their interests. In addition to that, in the formative study, we only interviewed expert music curators, but it would be valuable to also learn more about the tools used by people who are less experienced or less interested in music. For example, it might be less important for them to get a sense of the range of artists in a genre and more important to know what's popular, what the social connections are, etc.

11 CONCLUSION

This work presents TastePaths, an interactive web tool that helps users deeply explore and understand the music genres they listen to. We conducted a qualitative study where participants used a personalized and non-personalized version of TastePaths to explore two music genres they listen to often. Our study aimed to understand if TastePaths helps users explore their genres of interest and more broadly, how to better support users in exploring and understanding their preferences. We found that participants greatly preferred the personalized version and wanted even more personalization. They also wanted more control of the graph, including the ability to expand or prune sections of it to better reflect their interests. Finally, they also gained a better mental model of what they liked within their interests and desired to learn even more. Future tools in this space can investigate how to better incorporate learning into exploratory search, how to incorporate more closure and goal-fulfillment in recommendation systems, and how to support users in modifying the system's representation of their taste and interests.

12 ACKNOWLEDGEMENTS

Anonymized for review.

REFERENCES

[1] Fernando Amat, Ashok Chandrashekar, Tony Jebara, and Justin Basilico. 2018. Artwork Personalization at Netflix. In *Proceedings of the 12th ACM Conference on Recommender Systems* (Vancouver, British Columbia, Canada) *(RecSys '18)*. Association for Computing Machinery, New York, NY, USA, 487–488. https://doi.org/10.1145/3240323.3241729

[2] Xavier Amatriain, Josep M. Pujol, Nava Tintarev, and Nuria Oliver. 2009. Rate It Again: Increasing Recommendation Accuracy by User Re-Rating. In *Proceedings of the Third ACM Conference on Recommender Systems* (New York, New York, USA) *(RecSys '09)*. Association for Computing Machinery, New York, NY, USA, 173–180. https://doi.org/10.1145/1639714.1639744

[3] Ivana Andjelkovic, Denis Parra, and John O'Donovan. 2019. Moodplay: Interactive music recommendation based on Artists' mood similarity. *International Journal of Human-Computer Studies* 121 (2019), 142–159. https://doi.org/10.1016/j.ijhcs.2018.04.004 Advances in Computer-Human Interaction for Recommender Systems.

[4] Eric P. S. Baumer, Rui Sun, and Peter Schaedler. 2018. Departing and Returning: Sense of Agency as an Organizing Concept for Understanding Social Media Non/Use Transitions. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 23 (Nov. 2018), 19 pages. https://doi.org/10.1145/3274292

[5] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment* 2008, 10 (2008), P10008.

[6] Geoffray Bonnin and Dietmar Jannach. 2014. Automated Generation of Music Playlists: Survey and Experiments. *ACM Comput. Surv.* 47, 2, Article 26 (Nov. 2014), 35 pages. https://doi.org/10.1145/2652481

[7] Svetlin Bostandjiev, John O'Donovan, and Tobias Höllerer. 2012. TasteWeights: A Visual Interactive Hybrid Recommender System. In *Proceedings of the Sixth ACM Conference on Recommender Systems* (Dublin, Ireland) *(RecSys '12)*. Association for Computing Machinery, New York, NY, USA, 35–42. https://doi.org/10.1145/2365952.2365964

[8] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. 2011. D3 Data-Driven Documents. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (Dec. 2011), 2301–2309. https://doi.org/10.1109/TVCG.2011.185

[9] Ulrik Brandes. 2001. A faster algorithm for betweenness centrality. *The Journal of Mathematical Sociology* 25, 2 (2001), 163–177. https://doi.org/10.1080/0022250X.2001.9990249 arXiv:https://doi.org/10.1080/0022250X.2001.9990249

[10] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[11] Wanling Cai, Yucheng Jin, and Li Chen. 2021. Critiquing for Music Exploration in Conversational Recommender Systems. In *26th International Conference on Intelligent User Interfaces* (College Station, TX, USA) *(IUI '21)*. Association for Computing Machinery, New York, NY, USA, 480–490. https://doi.org/10.1145/3397481.3450657

[12] Joseph Chee Chang, Nathan Hahn, Adam Perer, and Aniket Kittur. 2019. SearchLens: Composing and Capturing Complex User Interests for Exploratory Search. In *Proceedings of the 24th International Conference on Intelligent User Interfaces* (Marina del Ray, California) *(IUI '19)*. Association for Computing Machinery, New York, NY, USA, 498–509. https://doi.org/10.1145/3301275.3302321

[13] Minsuk Chang, Leonore V. Guillain, Hyeungshik Jung, Vivian M. Hare, Juho Kim, and Maneesh Agrawala. 2018. *RecipeScape: An Interactive Tool for Analyzing Cooking Instructions at Scale.* Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3174025

[14] Duen Horng Chau, Aniket Kittur, Jason I. Hong, and Christos Faloutsos. 2011. *Apolo: Making Sense of Large Network Data by Combining Rich User Interaction and Machine Learning.* Association for Computing Machinery, New York, NY, USA, 167–176. https://doi.org/10.1145/1978942.1978967

[15] Paolo Cremonesi, Franca Garzotto, and Roberto Turrin. 2012. User Effort vs. Accuracy in Rating-Based Elicitation. In *Proceedings of the Sixth ACM Conference on Recommender Systems* (Dublin, Ireland) *(RecSys '12)*. Association for Computing Machinery, New York, NY, USA, 27–34. https://doi.org/10.1145/2365952.2365963

[16] Douglass R. Cutting, David R. Karger, Jan O. Pedersen, and John W. Tukey. 2017. Scatter/Gather: A Cluster-Based Approach to Browsing Large Document Collections. *SIGIR Forum* 51, 2 (Aug. 2017), 148–159. https://doi.org/10.1145/3130348.3130362

[17] Liam Delaney and Leonhard K. Lades. 2017. Present Bias and Everyday Self-Control Failures: A Day Reconstruction Study. *Journal of Behavioral Decision Making* 30, 5 (2017), 1157–1167. https://doi.org/10.1002/bdm.2031 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/bdm.2031

[18] Marc JMH Delsing, Tom FM Ter Bogt, Rutger CME Engels, and Wim HJ Meeus. 2008. Adolescents' music preferences and personality characteristics. *European Journal of Personality: Published for the European Association of Personality Psychology* 22, 2 (2008), 109–130.

[19] John Frow. 2014. *Genre.* Routledge.

[20] Jean Garcia-Gathright, Brian St. Thomas, Christine Hosey, Zahra Nazari, and Fernando Diaz. 2018. Understanding and Evaluating User Satisfaction with Music Discovery. In *The 41st International ACM SIGIR Conference on Researc & Development in Information Retrieval* (Ann Arbor, MI, USA) *(SIGIR '18)*. Association for Computing Machinery, New York, NY, USA, 55–64. https://doi.org/10.1145/3209978.3210049

[21] Carlos A. Gomez-Uribe and Neil Hunt. 2016. The Netflix Recommender System: Algorithms, Business Value, and Innovation. *ACM Trans. Manage. Inf. Syst.* 6, 4, Article 13 (Dec. 2016), 19 pages. https://doi.org/10.1145/2843948

[22] F. Maxwell Harper, Xin Li, Yan Chen, and Joseph A. Konstan. 2005. An Economic Model of User Rating in an Online Recommender System. In *User Modeling 2005*, Liliana Ardissono, Paul Brna, and Antonija Mitrovic (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 307–316.

[23] Christine Hosey, Lara Vujović, Brian St. Thomas, Jean Garcia-Gathright, and Jennifer Thom. 2019. *Just Give Me What I Want: How People Use and Evaluate Music Search.* Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300529

[24] Frank K Hwang and Dana S Richards. 1992. Steiner tree problems. *Networks* 22, 1 (1992), 55–89.

[25] Dietmar Jannach and Gediminas Adomavicius. 2016. Recommendations with a Purpose. In *Proceedings of the 10th ACM Conference on Recommender Systems* (Boston, Massachusetts, USA) *(RecSys '16)*. Association for Computing Machinery, New York, NY, USA, 7–10. https://doi.org/10.1145/2959100.2959186

[26] Mohsen Kamalzadeh, Christoph Kralj, Torsten Möller, and Michael Sedlmair. 2016. TagFlip: Active Mobile Music Discovery with Social Tags. In *Proceedings of the 21st International Conference on Intelligent User Interfaces* (Sonoma, California, USA) *(IUI '16)*. Association for Computing Machinery, New York, NY, USA, 19–30. https://doi.org/10.1145/2856767.2856780

[27] Bart P. Knijnenburg, Saadhika Sivakumar, and Daricia Wilkinson. 2016. Recommender Systems for Self-Actualization. In *Proceedings of the 10th ACM Conference on Recommender Systems* (Boston, Massachusetts, USA) *(RecSys '16)*. Association for Computing Machinery, New York, NY, USA, 11–14. https://doi.org/10.1145/2959100.2959189

[28] Bart P. Knijnenburg and Martijn C. Willemsen. 2015. *Evaluating Recommender Systems with User Experiments.* Springer US, Boston, MA, 309–352. https://doi.org/10.1007/978-1-4899-7637-6_9

[29] Bart P Knijnenburg, Martijn C Willemsen, Zeno Gantner, Hakan Soncu, and Chris Newell. 2012. Explaining the user experience of recommender systems. *User Modeling and User-Adapted Interaction* 22, 4 (2012), 441–504. https://doi.org/10.1007/s11257-011-9118-4

[30] Joseph A Konstan and John Riedl. 2012. Recommender systems: from algorithms to user experience. *User modeling and user-adapted interaction* 22, 1 (2012), 101–123. https://doi.org/10.1007/s11257-011-9112-x

[31] Johannes Kunkel, Benedikt Loepp, and Jürgen Ziegler. 2017. A 3D Item Space Visualization for Presenting and Manipulating User Preferences in Collaborative Filtering. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces* (Limassol, Cyprus) *(IUI '17)*. Association for Computing Machinery, New York, NY, USA, 3–15. https://doi.org/10.1145/3025171.3025189

[32] Johannes Kunkel, Claudia Schwenger, and Jürgen Ziegler. 2020. NewsViz: Depicting and Controlling Preference Profiles Using Interactive Treemaps in News Recommender Systems. In *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization* (Genoa, Italy) *(UMAP '20)*. Association for Computing Machinery, New York, NY, USA, 126–135. https://doi.org/10.1145/3340631.3394869

[33] Audrey Laplante and J. Stephen Downie. 2011. The utilitarian and hedonic outcomes of music information-seeking in everyday life. *Library & Information Science Research* 33, 3 (2011), 202–210. https://doi.org/10.1016/j.lisr.2010.11.002

[34] Jin Ha Lee, Hyerim Cho, and Yea-Seul Kim. 2016. Users' music information needs and behaviors: Design implications for music information retrieval systems. *Journal of the Association for Information Science and Technology* 67, 6 (2016), 1301–1330. https://doi.org/10.1002/asi.23471 arXiv:https://asistdl.onlinelibrary.wiley.com/doi/pdf/10.1002/asi.23471

[35] Jin Ha Lee and Nichole Maiman Waterman. 2012. Understanding User Requirements for Music Information Services.. In *ISMIR*. Citeseer, 253–258.

[36] Yu Liang and Martijn C. Willemsen. 2019. Personalized Recommendations for Music Genre Exploration. In *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization* (Larnaca, Cyprus) *(UMAP '19)*. Association for Computing Machinery, New York, NY, USA, 276–284. https://doi.org/10.1145/3320435.3320455

[37] Yu Liang and Martijn C. Willemsen. 2021. Interactive Music Genre Exploration with Visualization and Mood Control. In *26th International Conference on Intelligent User Interfaces* (College Station, TX, USA) *(IUI '21)*. Association for Computing Machinery, New York, NY, USA, 175–185. https://doi.org/10.1145/3397481.3450700

[38] Yu Liang and Martijn C Willemsen. 2021. The role of preference consistency, defaults and musical expertise in users' exploration behavior in a genre exploration recommender. In *Fifteenth ACM Conference on Recommender Systems*. 230–240.

[39] Kai Lukoff, Ulrik Lyngs, Himanshu Zade, J. Vera Liao, James Choi, Kaiyue Fan, Sean A. Munson, and Alexis Hiniker. 2021. *How the Design of YouTube Influences User Sense of Agency*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3411764.3445467

[40] James McInerney, Benjamin Lacker, Samantha Hansen, Karl Higley, Hugues Bouchard, Alois Gruson, and Rishabh Mehrotra. 2018. Explore, Exploit, and Explain: Personalizing Explainable Recommendations with Bandits. In *Proceedings of the 12th ACM Conference on Recommender Systems (RecSys '18)*. 31–39.

[41] Martijn Millecamp, Nyi Nyi Htun, Cristina Conati, and Katrien Verbert. 2019. To Explain or Not to Explain: The Effects of Personal Characteristics When Explaining Music Recommendations. In *Proceedings of the 24th International Conference on Intelligent User Interfaces* (Marina del Ray, California) *(IUI '19)*. Association for Computing Machinery, New York, NY, USA, 397–407. https://doi.org/10.1145/3301275.3302313

[42] MusicAlly. 2021. *IFPI report reveals 7.4% growth in global recorded music revenues*. https://musically.com/2021/03/23/ifpi-report-reveals-7-4-growth-in-global-recorded-music-revenues/

[43] Judith Möller, Damian Trilling, Natali Helberger, and Bram van Es. 2018. Do not blame it on the algorithm: an empirical assessment of multiple recommender systems and their impact on content diversity. *Information, Communication & Society* 21, 7 (2018), 959–977. https://doi.org/10.1080/1369118X.2018.1444076 arXiv:https://doi.org/10.1080/1369118X.2018.1444076

[44] Sayooran Nagulendra and Julita Vassileva. 2014. Understanding and Controlling the Filter Bubble through Interactive Visualization: A User Study. In *Proceedings of the 25th ACM Conference on Hypertext and Social Media* (Santiago, Chile) *(HT '14)*. Association for Computing Machinery, New York, NY, USA, 107–115. https://doi.org/10.1145/2631775.2631811

[45] Mark EJ Newman. 2008. The mathematics of networks. *The new palgrave encyclopedia of economics* 2, 2008 (2008), 1–12.

[46] Tien T. Nguyen, Pik-Mai Hui, F. Maxwell Harper, Loren Terveen, and Joseph A. Konstan. 2014. Exploring the Filter Bubble: The Effect of Using Recommender Systems on Content Diversity. In *Proceedings of the 23rd International Conference on World Wide Web* (Seoul, Korea) *(WWW '14)*. Association for Computing Machinery, New York, NY, USA, 677–686. https://doi.org/10.1145/2566486.2568012

[47] Eli Pariser. 2011. *The filter bubble: How the new personalized web is changing what we read and how we think*. Penguin.

[48] Peter Pirolli and Stuart Card. 2005. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of international conference on intelligence analysis*, Vol. 5. McLean, VA, USA, 2–4.

[49] Pearl Pu, Li Chen, and Rong Hu. 2011. A User-Centric Evaluation Framework for Recommender Systems. In *Proceedings of the Fifth ACM Conference on Recommender Systems* (Chicago, Illinois, USA) *(RecSys '11)*. Association for Computing Machinery, New York, NY, USA, 157–164. https://doi.org/10.1145/2043932.2043962

[50] Anand Rajaraman and Jeffrey David Ullman. 2011. *Mining of massive datasets*. Cambridge University Press.

[51] Johnny Saldaña. 2015. *The coding manual for qualitative researchers*. Sage.

[52] E. Isaac Sparling and Shilad Sen. 2011. Rating: How Difficult is It?. In *Proceedings of the Fifth ACM Conference on Recommender Systems* (Chicago, Illinois, USA) *(RecSys '11)*. Association for Computing Machinery, New York, NY, USA, 149–156. https://doi.org/10.1145/2043932.2043961

[53] Maria Taramigkou, Efthimios Bothos, Konstantinos Christidis, Dimitris Apostolou, and Gregoris Mentzas. 2013. Escape the Bubble: Guided Exploration of Music Preferences for Serendipity and Novelty. In *Proceedings of the 7th ACM Conference on Recommender Systems* (Hong Kong,

China) *(RecSys '13)*. Association for Computing Machinery, New York, NY, USA, 335–338. https://doi.org/10.1145/2507157.2507223

[54] Nava Tintarev, Shahin Rostami, and Barry Smyth. 2018. Knowing the Unknown: Visualising Consumption Blind-Spots in Recommender Systems. In *Proceedings of the 33rd Annual ACM Symposium on Applied Computing* (Pau, France) *(SAC '18)*. Association for Computing Machinery, New York, NY, USA, 1396–1399. https://doi.org/10.1145/3167132.3167419

[55] Pontus Wärnestål. 2005. User evaluation of a conversational recommender system. In *Proceedings of the 4th Workshop on Knowledge and Reasoning in Practical Dialogue Systems*.

[56] Michael Yeomans, Anuj Shah, Sendhil Mullainathan, and Jon Kleinberg. 2019. Making sense of recommendations. *Journal of Behavioral Decision Making* 32, 4 (2019), 403–414.

[57] Shuai Zhang, Lina Yao, Aixin Sun, and Yi Tay. 2019. Deep Learning Based Recommender System: A Survey and New Perspectives. *ACM Comput. Surv.* 52, 1, Article 5 (Feb. 2019), 38 pages. https://doi.org/10.1145/3285029

[58] Nanxuan Zhao, Nam Wook Kim, Laura Mariah Herman, Hanspeter Pfister, Rynson W.H. Lau, Jose Echevarria, and Zoya Bylinskii. 2020. ICONATE: Automatic Compound Icon Generation and Ideation. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376618

[59] Frederik Zuiderveen Borgesius, Damian Trilling, Judith Möller, Balázs Bodó, Claes H De Vreese, and Natali Helberger. 2016. Should we worry about filter bubbles? *Internet Policy Review. Journal on Internet Regulation* 5, 1 (2016).