Come Together: How Social Agents can Improve Music Discovery

Laura Triglia
laura.triglia@iit.it
Italian Institute of Technology, University of Genoa, Sony Brussels Laboratory
Genoa, Brussels, Italy, Belgium

Pietro Gravino
Sony Computer Science Laboratories Paris
Paris, France
pietro.gravino@sony.com

Thomas Carette
Sony Brussels Laboratory
Brussels, Belgium
thomas.carette@sony.com

Francesco Rea
Italian Institute of Technology
Genoa, Italy
francesco.rea@iit.it

Alessandra Sciutti
Italian Institute of Technology
Genoa, Italy
alessandra.sciutti@iit.it

Pablo Barros
Sony Brussels Laboratory
Brussels, Belgium
pablo.barros@sony.com

Abstract

Traditional recommendation interfaces often struggle to motivate users to explore beyond their established preferences. This study examines whether a conversational, LLM-powered social agent can encourage engagement with unfamiliar music during playlist co-creation. The agent first builds rapport over known preferences, then suggests both familiar and unfamiliar songs.

In a user study, participants interacted with both a chat-based agent and a traditional form interface, selecting songs "close to" or "far from" their usual tastes. We analyzed song choices, user experience, and perceived social dynamics.

While the agent did not significantly increase the selection of unfamiliar songs, it enhanced enjoyment of both familiar and unfamiliar tracks. Linguistic analysis indicates that longer engagement with the agent correlates with greater appreciation of novel recommendations. Despite some usability challenges, users found the agent more engaging than forms. These findings suggest socially aware recommendation systems can improve user experience and foster music exploration.

CCS Concepts

• **Artificial Intelligence (AI)** → **Social Agent**; • **Music Recommendation** → *Information Systems*; • **Information Retrieval** → Recommender Systems; • **User Engagement** → Computing methodologies.

Keywords

Music Recommendation, Social Agent, User Engagement, Conversational AI

ACM Reference Format:
Laura Triglia, Pietro Gravino, Thomas Carette, Francesco Rea, Alessandra Sciutti, and Pablo Barros. 2025. Come Together: How Social Agents can Improve Music Discovery. In *Proceedings of EARL Workshop - 19th ACM Conference on Recommender Systems (RecSys '25)*. ACM, New York, NY, USA, 5 pages. https://doi.org/XXXXXXX.XXXXXXX

RecSys '25, Prague, Czech Republic

ACM ISBN 978-1-4503-XXXX-X/2018/06
https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Music recommendation systems have advanced rapidly, leveraging AI and machine learning to personalize suggestions through collaborative filtering [17], content-based filtering [20], and hybrid methods [2]. Recent progress includes deep learning [21], reinforcement learning [5], and graph-based models [22], yet systems still struggle to encourage users beyond their comfort zones [18]. Strategies to improve serendipity and diversity include modeling user personality [8], context [19], and leveraging social influence [3, 13].

Conversational agents have emerged as a promising avenue for enhancing user engagement and trust in recommendations [6, 7, 10]. In music, dialogue-based agents allow users to express preferences and refine playlists naturally [4, 13], supporting more dynamic discovery.

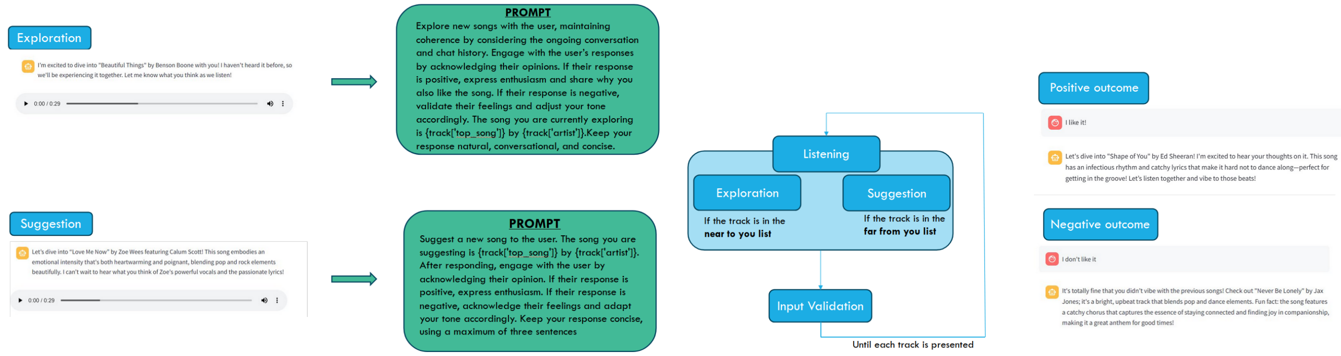
Building on this work, our study examines whether a chat-based social agent can guide users to explore music outside their usual tastes by fostering rapport and shared content consumption. We compare this approach to a traditional form-based interface, analyzing song selections, user experience (UEQ [11]), and perceived social presence (ASAQ [9]). Results show that, while the agent did not significantly increase unfamiliar song selection, it improved enjoyment and engagement, suggesting socially aware systems can enhance music exploration.

2 Inside Our Social Agent

Our proposed agent is created to enhance user interaction and foster rapport during the playlist co-creation process. It is powered by an OpenAI-based language model (ChatGPT-4) and manages the flow of the experiment using a state machine. To facilitate the visualization of the agent's responses and allow for user interaction, the state machine is implemented via a Streamlit[1] web app interface.

The state machine flow consists of two key phases: Profile Creation and Song Exploration.

[1]https://streamlit.io/



(a) Song Exploration Phase with Prompt (b) Song Exploration Phase Flow

Figure 1: Overview of the Song Exploration Phase: (A) Prompt Example; (B) Session Flow

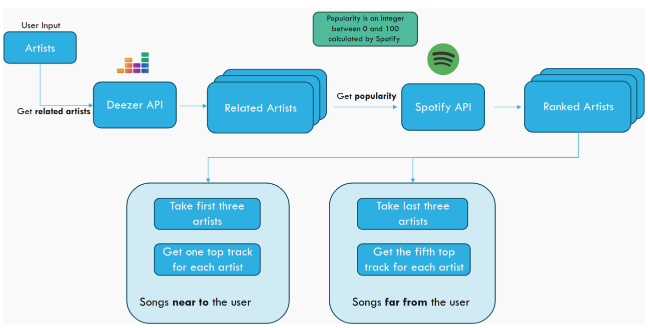


Figure 2: Flow of the creation of the Lists

2.1 Profile Creation

The Profile Creation phase comprises three states in the agent's state machine: collecting age, preferred music genres, and favorite artists. Each response is validated using the LLM, allowing users to reply naturally and helping establish rapport through flexible, conversational interaction.

The collected data is then used to generate two personalized playlists, ensuring the rapport built is maintained throughout the experience. To do this, the system queries the Deezer API (http://api.deezer.com/radio) for related artists based on user preferences and ranks them using popularity scores from the Spotify API(https://developer.spotify.com/documentation/web-api).

Two lists are constructed: the "Near to You" list features top tracks from the three most popular related artists, while the "Far from You" list presents the fifth most popular tracks from the three least popular artists. This approach exposes users to both familiar and novel music options while keeping the agent's engagement consistent (see Figure 2).

2.2 Song Exploration

After the Profile Creation phase, the agent enters the Song Exploration stage, where tracks from both curated lists are presented in a conversational manner. Each song is introduced with a tailored agent comment: "Near to You" tracks are positioned as shared interests, while "Far from You" tracks are framed as opportunities for discovery. This approach leverages distinct prompts to foster an engaging and social dialogue (see Figure 1a).

Users can listen to a 30-second preview of each song, after which the agent collects feedback on enjoyment, prior familiarity, and any personal associations. This feedback-driven loop continues until all tracks are explored, as depicted in Figure 1b. At the end of the session, users are invited to select three favorite songs for inclusion in a final playlist, encapsulating their preferences shaped by agent-guided exploration.

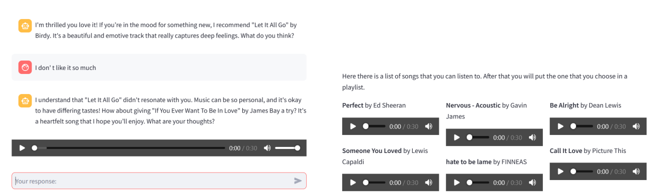
3 Experimental Methodology

The experiment was conducted via an online platform hosted on a dedicated server at the Italian Institute of Technology (IIT). Participants were randomly assigned to one of two conditions:

- **Chat-Agent**: Users interacted with the social agent through a conversational chatbot.
- **Form-No Agent**: Users completed a static form without any agent interaction.

Each participant experienced only one condition. The average session lasted about 20 minutes for the Chat-Agent and 14 minutes for the Form-No Agent.

In the Song Exploration phase, all songs were displayed at once in the form-based condition, allowing users to freely browse and listen within the interface (Figure 3). This ensured comparable exposure to content across both conditions, with the primary difference being the presence or absence of social co-consumption with the agent.



(a) Interface Chat-Agent (b) Interface Form-No Agent

Figure 3: Streamlit web app for the Chat-Agent and Form-No Agent conditions during Song Exploration.

3.1 Participants

A total of 156 participants were recruited via the Prolific platform [14], with 73 assigned to the chat-based agent condition and 83 to the form-based condition. All participants were fluent in English. Compensation followed Prolific's guidelines: £3 for the Chat-Agent group and £2.25 for the Form-No Agent group.

3.2 Experimental Sessions

Both conditions followed a three-phase structure: (1) Profile Creation, (2) Song Exploration, and (3) Questionnaire. Participants first provided demographic information and musical preferences, then listened to six songs and selected favorites for a final playlist (see Section 2 for interface details). Sessions concluded with a series of questionnaires.

3.3 Questionnaires

After completing the session, all participants filled out two questionnaires: a self-assessment of song familiarity and preference, and the User Experience Questionnaire (UEQ) [11]—measuring perspicuity, efficiency, novelty, stimulation, and dependability. Participants in the Chat-Agent group also completed the ASAQ [9], assessing ten social constructs (e.g., engagement, trust, alliance, and enjoyability) to evaluate the agent's social presence and its influence on music preferences.

4 Results

Our main finding is that users in the chat-agent condition who encountered at least one unfamiliar song were more likely to select a diverse range of tracks for their playlists compared to those in the form condition.

To accurately capture this effect, we analyzed song selection based on participants' self-reported familiarity (from the first questionnaire), rather than solely relying on system classifications. This approach accounts for individual differences in music knowledge and ensures that "unknown" songs reflect users' actual experience, as system-generated labels may not always align with true familiarity.

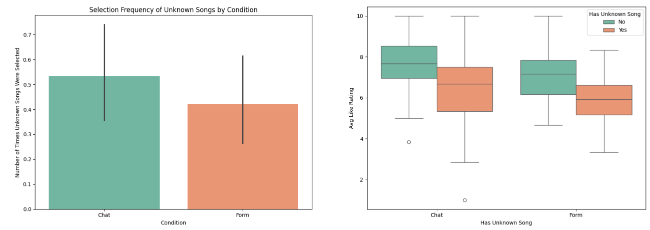
We focused our analysis on participants who reported receiving at least one "unknown" song. The average proportion of unfamiliar songs chosen for the final playlist was higher in the Chat-Agent group, as shown in Figure 4a. Statistical analysis using Boschloo's exact test [1, 16] confirmed that this difference was significant ($p = 0.03$), suggesting the effect was not due to chance.

These findings indicate that a social agent can encourage greater openness to unfamiliar content, reducing hesitation and promoting more diverse music choices. This underscores the potential for socially-aware recommendation systems to enhance user engagement and discovery.

4.1 What About Enjoyment?

Another key comparison examined how users rated recommended songs in each condition. Results show that participants in the Chat-Agent group generally rated songs higher than those in the Form-No Agent group.

As shown in Figure 4b, song ratings were higher and less variable in the chat-based condition, while ratings in the form condition were lower and more dispersed. This analysis was conducted for both users who received at least one unknown song and those who only received familiar ones, allowing us to see how unfamiliar content and interface type interact in shaping preferences.



(a) Selection of Unknown Songs Proposed

(b) Average Like Rating Across the Two Conditions

Figure 4: Exploration and appreciation of recommendations across conditions: (A) Proportion of unknown songs selected; (B) Average like rating.

One explanation is that the chat-based interaction, with its conversational and co-consumptive nature, encouraged users to reflect more deeply on their choices, reinforcing preferences and increasing satisfaction. In contrast, the form condition, being more static and isolated, may have resulted in less engagement and lower ratings.

A Mann–Whitney U test [12] confirmed the difference between conditions was highly significant ($p = 7.5 \times 10^{-6}$). The difference remained significant for "known" songs ($p = 4.3 \times 10^{-4}$), reinforcing the positive impact of the chat interface on song preference. For "unknown" songs, the difference was not statistically significant ($p = 0.090$), likely due to the smaller sample, but the trend persisted.

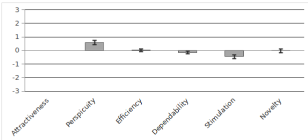
These results suggest that the interactive chat environment enhances users' involvement and confidence in their playlist choices, leading to greater enjoyment of recommended songs.

4.2 Lessons from the User Experience (UEQ) questionnaire

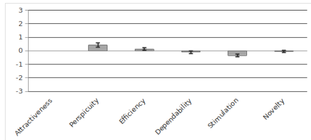
To better understand the differences between the Chat-Agent and Form-No Agent conditions, we compared user perceptions using the User Experience Questionnaire (UEQ) [11]. This analysis aimed to determine if higher song ratings in the chat condition could be explained by differences in interface usability or engagement.

The UEQ results, shown in Figure 5a and Figure 5b, indicate that users rated both interfaces similarly in overall user experience. This suggests that factors like usability, visual appeal, or intuitiveness did not drive the observed differences in song exploration or ratings. Instead, these differences are likely due to the interaction dynamics—the chat's conversational and iterative nature—rather than the interface design itself.

These findings reinforce that the chat environment, by allowing users to express and reflect on their preferences in dialogue, encouraged deeper engagement, more exploration, and higher confidence in playlist choices. The lack of significant differences in UEQ scores rules out interface bias and underscores the value of social, conversational interactions in recommendation systems.



(a) UEQ Chat-Agent Condition



(b) UEQ Form-No Agent Condition

Figure 5: User Experience Questionnaire (UEQ) results for (A) Chat-Agent and (B) Form-No Agent conditions.

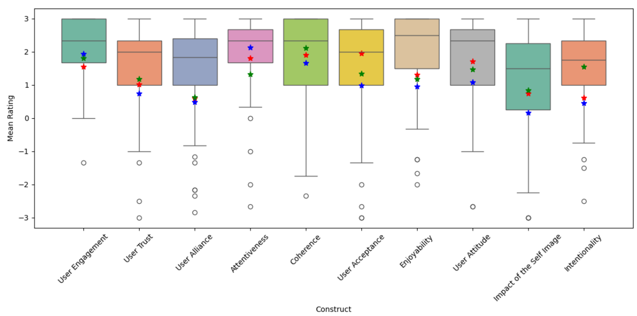


Figure 6: ASAQ construct results for the agent condition.

4.3 Building Rapport with the Agent

The ASAQ (Artificial Social Agent Questionnaire)[9] is a standardized tool designed to assess user interactions with artificial agents, focusing on aspects such as engagement, trust, and social presence. In our study, the ASAQ was used to evaluate how participants perceived our agent during the interaction.

Figure 6 presents a boxplot visualization of ASAQ scores, illustrating how our agent is rated relative to reference agents. All agents, including ours, are disembodied and designed to interact in similar ways. However, despite these structural similarities, notable differences in perception emerged. These variations can likely be attributed to scenario-dependent interactions—specific elements of the experimental context that influence user experiences and evaluations.

4.3.1 Sentiment. Beyond the questionnaire data, we conducted a sentiment analysis using the VADER (Valence Aware Dictionary and sEntiment Reasoner) library [15] to further explore the correlation between user sentiment during interactions and ASAQ constructs. We annotated the sentiment of every message in the conversation, also defining an average sentiment for the whole conversation. Our analysis revealed a significant correlation between the average participant sentiment and the Acceptance (p-value = 0.003) and Attitude (p-value = 0.004) constructs of the ASAQ. However, we did not find a notable positive or negative correlation between the user's sentiment and the agent's perceived behavior. Instead, our findings highlight a correlation concerning language neutrality—when a user expressed a more neutral sentiment, the responses of the agent also tended to be perceived as more neutral in terms of language.

Specifically, we observed the following results from our analysis of sentiment and neutrality:

- *User Average Sentiment vs. Agent Average Sentiment*: The statistical test showed no significant difference (statistic = 0.1406, p-value = 0.2355), suggesting no strong correlation between user and agent sentiment.
- *Single Message Sentiment (User vs. Agent)*: Excluding the first exchange, which was about age, the analysis showed no significant correlation (statistic = 0.0033, p-value = 0.9320).
- *User Average Neutrality vs. Agent Average Neutrality*: A significant correlation was found (statistic = 0.3977, p-value = 0.00049), suggesting that user neutrality in sentiment was closely aligned with the agent's neutral tone.
- *Single Message Neutrality (User vs. Agent)*: Excluding the first exchange, no significant correlation was found between user and agent neutrality (statistic = -0.0444, p-value = 0.2431).

These results suggest that when users maintained a balanced, neutral tone in their expressions, they were more likely to perceive the agent's responses as similarly neutral, regardless of the agent's behavior. These insights emphasize the role of linguistic neutrality in shaping user perceptions of the agent, underscoring the importance of sentiment alignment in human-agent interactions.

4.3.2 Engagement. Engagement during interactions was measured through the length of the sentences written by users. To better understand the relationship between engagement and user perception of the agent, we performed a correlation analysis. Our findings indicate a positive correlation between the total length of user messages and the User Engagement construct from the ASAQ questionnaire (p-value = 0.00096). This confirms that participants who were more actively engaged in the conversation also perceived the agent as more engaging, reinforcing the validity of the message lengths as an engagement measure.

Additionally, we observed a positive correlation between user engagement and the agent's engagement across both aggregated conversations and individual messages. This means that when the agent's responses were longer and more involved, users also tended to write longer messages in return. Since the agent's messages always preceded the user's responses, this implies a causal relationship—indicating that the agent's level of engagement influenced how much users engaged in the conversation. This finding highlights the agent's role in shaping the dynamics of the interaction and suggests that its communicative style can directly impact user involvement.

Specifically, we observed the following results from our analysis:

- *User Sum Length vs. ASAQ-User Engagement*: A significant positive correlation was found (statistic = 0.3786, p-value = 0.00096), indicating that longer user messages were associated with higher engagement ratings on the ASAQ.
- *User Sum Length vs. Agent Average Length*: A significant positive correlation was found (statistic = 0.4736, p-value = 2.32e-05), suggesting that longer user messages correlated with longer agent responses.
- *Single Message Length*: User vs. Agent: Excluding the first exchange, which was about age, a significant positive correlation was found (statistic = 0.2820, p-value = 4.05e-14). This supports the causal relationship, as the agent's message always precedes the user's response.

These results underscore the agent's influence on user engagement, suggesting that its communicative style directly impacts the length of response and involvement of users.

5 Conclusion

This study examined the role of a social agent in music recommendation, specifically investigating whether conversational interaction could influence users to explore music beyond their existing preferences more effectively than a traditional form-based approach. Our findings suggest that the chat-based agent played a significant role in enhancing user engagement, shaping decision-making, and ultimately increasing the likelihood of selecting unfamiliar songs.

Through a controlled experiment with 165 participants, we found that users in the Chat-Agent condition exhibited a greater willingness to explore new music compared to those in the Form-No Agent condition. The statistical analysis, using Boschloo's exact test, revealed a significant difference in song selection behavior ($p = 0.03$), indicating that the conversational interface facilitated a more exploratory approach. This suggests that social interaction, even in an artificial setting, can foster openness to novel experiences and reduce reliance on familiar preferences.

Furthermore, participants in the chat-based condition rated their selected songs higher on average than those in the form-based condition. This finding, reinforced by a highly significant p-value (7.498e-6), highlights the potential of interactive recommendation systems to enhance user satisfaction. Interestingly, even when analyzing users who exclusively received known songs, the chat-based interaction still led to significantly higher ratings ($p = 4.254e-4$). These results support the notion that engagement in a dynamic, conversational exchange strengthens users' confidence in their choices and positively influences their perception of the recommendations.

Importantly, results from the User Experience Questionnaire (UEQ) demonstrated that users did not perceive one interface as significantly superior to the other in terms of usability. This indicates that the observed differences in exploratory behavior and song preference were not due to interface design but rather stemmed from the nature of the interaction itself. The chat-based interaction provided a more immersive recommendation experience, allowing users to reflect on their choices dynamically and feel more involved in the selection process.

Taken together, our findings suggest that integrating social interaction into music recommendation systems can enhance exploration, enjoyment, and engagement. While traditional recommendation methods primarily focus on optimizing algorithmic accuracy, this study underscores the potential of social agents to influence decision-making in ways that go beyond static preference matching. Future research could explore how different conversational strategies, levels of agent personalization, and long-term interactions affect user behavior in music discovery. Additionally, integrating multimodal feedback mechanisms—such as voice or gesture-based interactions—could further enhance the experience of interacting with social recommendation agents.

Overall, our research contributes to the growing body of work on human-AI interaction in recommendation settings by showing that a social agent can meaningfully shape users' music discovery processes—even when it is not responsible for generating the recommendations. By facilitating interaction and dialogue around suggested content, conversational agents can help users reflect on their preferences, break habitual listening patterns, and become more open to exploring unfamiliar music—ultimately enriching their listening experiences.

References

[1] R.D. Boschloo. 1970. Raised conditional level of significance for the 2 × 2-table when testing the equality of two probabilities. *Statistica Neerlandica* 24, 1 (1970), 1–9. doi:10.1111/j.1467-9574.1970.tb00104.x

[2] Robin Burke. 2002. Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction* 12 (2002), 331–370.

[3] Òscar Celma Herrada. 2011. Music recommendation and discovery in the long tail. (2011).

[4] Arun Tejasvi Chaganty, Megan Leszczynski, Shu Zhang, Ravi Ganti, Krisztian Balog, and Filip Radlinski. 2023. Beyond single items: Exploring user preferences in item sets with the conversational playlist curation dataset. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2754–2764.

[5] Xiaocong Chen, Lina Yao, Julian McAuley, Guanglin Zhou, and Xianzhi Wang. 2023. Deep reinforcement learning in recommender systems: A survey and new perspectives. *Knowledge-Based Systems* 264 (2023), 110335.

[6] Konstantina Christakopoulou, Filip Radlinski, and Katja Hofmann. 2016. Towards conversational recommender systems. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 815–824.

[7] Stephan Diederich, Alfred Brendel, Stefan Morana, and Lutz Kolbe. 2022. On the Design of and Interaction with Conversational Agents: An Organizing and Assessing Review of Human-Computer Interaction Research. *Journal of the Association for Information Systems* (01 2022). doi:10.17705/1jais.00724

[8] Bruce Ferwerda, Markus Schedl, and Marko Tkalcic. 2015. Personality & emotional states: Understanding users' music listening needs. In *Posters, Demos, Late-breaking Results and Workshop Proceedings of the 23rd Conference on User Modeling, Adaptation, and Personalization (UMAP 2015) Dublin, Ireland, June 29-July 3, 2015*, Vol. 1388. CEUR-WS. org.

[9] Siska Fitrianie, Merijn Bruijnes, Amal Abdulrahman, and Willem-Paul Brinkman. 2025. The Artificial Social Agent Questionnaire (ASAQ) - Development and evaluation of a validated instrument for capturing human interaction experiences with artificial social agents. *International Journal of Human-Computer Studies* (2025), 103482. doi:10.1016/j.ijhcs.2025.103482

[10] Dietmar Jannach and Gediminas Adomavicius. 2021. Recommender systems: Challenges, open questions, and research opportunities. *AI Magazine* 42, 4 (2021), 76–85. doi:10.1609/aimag.v42i4.16313

[11] Bettina Laugwitz, Theo Held, and Martin Schrepp. 2008. Construction and evaluation of a user experience questionnaire. In *Symposium of the Austrian HCI and usability engineering group*. Springer, 63–76.

[12] Patrick E McKnight and Julius Najab. 2010. Mann-Whitney U Test. *The Corsini encyclopedia of psychology* (2010), 1–1.

[13] Martin Pichl and Eva Zangerle. 2021. User models for multi-context-aware music recommendation. *Multimedia Tools and Applications* 80 (2021), 22509–22531. doi:10.1007/s11042-020-09890-7

[14] Prolific. 2014. https://www.prolific.com.

[15] Katherine Roehrick. 2020. *vader: Valence Aware Dictionary and sEntiment Reasoner (VADER)*. https://CRAN.R-project.org/package=vader R package version 0.2.1.

[16] Lise M. Saari. 2004. Employee attitudes and job satisfaction. *Human Resource Management* (2004).

[17] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl. 2001. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web*. 285–295.

[18] Markus Schedl, Emilia Gómez, Julián Urbano, et al. 2014. Music information retrieval: Recent developments and applications. *Foundations and Trends® in Information Retrieval* 8, 2-3 (2014), 127–261.

[19] Markus Schedl, Hamed Zamani, Ching-Wei Chen, Yashar Deldjoo, and Mehdi Elahi. 2018. Current challenges and visions in music recommender systems research. *International Journal of Multimedia Information Retrieval* 7 (2018), 95–116.

[20] George Tzanetakis and Perry Cook. 2002. Musical genre classification of audio signals. *IEEE Transactions on speech and audio processing* 10, 5 (2002), 293–302.

[21] Aaron Van den Oord, Sander Dieleman, and Benjamin Schrauwen. 2013. Deep content-based music recommendation. *Advances in neural information processing systems* 26 (2013).

[22] Rex Ying, Ruining He, Kaifeng Chen, Pong Eksombatchai, William L Hamilton, and Jure Leskovec. 2018. Graph convolutional neural networks for web-scale recommender systems. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*. 974–983.