Prefer2SD: A Human-in-the-Loop Approach to Balancing Similarity and Diversity in In-Game Friend Recommendations

N.B. When citing this work, cite the original published paper.

(article starts on next page)

Prefer2SD: A Human-in-the-Loop Approach to Balancing Similarity and Diversity in In-Game Friend Recommendations

Xiyuan Wang
School of Information Science and
Technology
ShanghaiTech University
Shanghai, China
wangxy7@shanghaitech.edu.cn

Ziang Li
Tongji University
Shanghai, China
ziangli@tongji.edu.cn

Sizhe Chen
Chalmers University of Technology
Gothenburg, Sweden
sizhec@student.chalmers.se

Xingxing Xing
UX Center
Netease Games
Guangzhou, China
xingxingxing@corp.netease.com

Wei Wan
Netease Games UX Center
Hangzhou, Zhejiang, China
gzwanwei@corp.netease.com

Quan Li*
School of Information Science and
Technology
ShanghaiTech University
Shanghai, China
liquan@shanghaitech.edu.cn

Abstract

In-game friend recommendations significantly impact player retention and sustained engagement in online games. Balancing similarity and diversity in recommendations is crucial for fostering stronger social bonds across diverse player groups. However, automated recommendation systems struggle to achieve this balance, especially as player preferences evolve over time. To tackle this challenge, we introduce *Prefer2SD* (derived from **Prefer**ence **to** **S**imilarity and **D**iversity), an iterative, human-in-the-loop approach designed to optimize the similarity-diversity (SD) ratio in friend recommendations. Developed in collaboration with a local game company, *Prefer2D* leverages a visual analytics system to help experts explore, analyze, and adjust friend recommendations dynamically, incorporating players' shifting preferences. The system employs interactive visualizations that enable experts to fine-tune the balance between similarity and diversity for distinct player groups. We demonstrate the efficacy of *Prefer2SD* through a within-subjects study (N=12), a case study, and expert interviews, showcasing its ability to enhance in-game friend recommendations and offering insights for the broader field of personalized recommendation systems.

CCS Concepts

• **Human-centered computing**; • **Human computer interaction**; • **Interactive systems and tools**; • **User interface toolkits**;

Keywords

Friend Recommendation, Similarity and Diversity, Visual Analytics, Active Learning.

---

*Corresponding Author.

1 Introduction

In recommendation systems, the challenge of balancing similarity with diversity has long been a central research focus [29, 73, 98]. In this work, we introduce the Similarity-Diversity (SD) ratio, a metric designed to quantify the trade-off between these two factors. Striking an effective balance is critical, as users tend to prefer content that aligns with their existing interests while also benefiting from exposure to novel and diverse perspectives. This balance not only enhances user experience but also helps mitigate the "filter bubble" effect [62], where repetitive exposure to similar content limits users' access to diverse new experiences, thus encouraging exploration across a wider range of content domains.

The balance between similarity and diversity is equally important in friend recommendation systems within online video games, such as massively multiplayer online role-playing games (MMORPGs) and multiplayer online battle arenas (MOBAs) [85, 91]. In these environments, similarity reflects players' tendencies to form connections with others who exhibit similar gaming behaviors, while diversity emphasizes the advantages of engaging with individuals possessing complementary skills or gameplay. Maintaining this balance enriches players' social experience [40], facilitating interactions with like-minded companions while also encouraging the exploration of social circles and exposure to various gameplay styles. However, despite the recognized value of this equilibrium, most current game friend recommendation systems prioritize similarity as the primary metric [20, 93, 96, 99], often overlooking the intricate interplay between similarity and diversity.

In the recommendation domain, various machine learning (ML) techniques have been proposed to dynamically adjust the SD ratio

in recommendations. Common approaches include collaborative filtering and graph-based models, which aim to balance accuracy and diversity by optimizing user-item interactions and leveraging complex relationship structures in data [2, 12, 31, 34, 87, 88, 98], ensuring that users receive both personalized and varied content. Techniques such as multi-stage optimization and graph neural networks are widely used to achieve these goals, providing flexible frameworks that adapt to different scenarios. However, directly applying these methods to friend recommendations in online video games presents several unique challenges. **1) Rapid Environmental Changes.** The fast-paced nature of online games, driven by frequent updates and seasonal events, leads to rapid shifts in player interests, which traditional ML models struggle to adapt to. **2) Diverse Needs Across Player Phases.** Players at different stages of their gaming journey exhibit varying preferences regarding the SD ratio [78]. Novices often seek connections with those of similar experience, while more experienced players tend to value diversity in recommendations [56, 80]. Existing recommendation methods, however, commonly adopt a "one-size-fits-all" approach [4, 22, 49, 69, 72], limiting their ability to cater to these diverse needs. **3) Lack of User Control.** Many friend recommendation systems, particularly in gaming communities, lack transparency and user control [70, 93], which can lead to unpredictable and occasionally unfavorable outcomes, particularly in the complex and dynamic environments of online games. It raises concerns about user experience and system performance, highlighting the need for more transparent and human-in-the-loop approaches in game friend recommendation systems.

To bridge this gap, it is essential to involve algorithm experts who design and refine friend recommendation systems for specific game environments, integrating their domain-specific knowledge into the process. These experts have a deep understanding of the intricacies inherent in in-game social interactions, enabling them to navigate the rapidly evolving landscape and address the diverse needs of players. Their involvement brings subjective insights and detailed reasoning, enhancing recommendations to align with player expectations. Given this, developing an interactive framework with tailored visualizations for experts is essential to enhance transparency, interpretability, and recommendation effectiveness, ultimately increasing players' satisfaction. Several interactive visualization tools [14, 64, 81] have been designed to help end-users understand and participate in the recommendation process. However, these tools are primarily oriented towards end-users, offering guidance and additional input through interactions. They often lack features specifically designed for algorithm experts, which limits the experts' ability to conduct in-depth analysis, gain deeper insights, and effectively adjust the SD ratio in the recommendation process.

Building on the challenges and insights identified, we introduce *Prefer2SD* (**Prefer**ence **to S**imilarity and **D**iversity), a human-in-the-loop solution for enhancing in-game friend recommendations. This interactive visual workflow empowers algorithm experts to optimize the SD ratio for specific player cohorts they choose through a structured two-step process. Expanding upon the traditional multistage recommendation pipeline (Fig. 1), our workflow integrates several innovations. In the candidate generation phase (Fig. 1(A-B)), *Prefer2SD* captures players' multi-modal preferences, diversifying

the friend candidate pool. Moving to the ranking stage (Fig. 1(B-C)), expert knowledge is employed to adjust the composition of friend candidates by carefully mediating intra- and inter-preference ratios to achieve the ideal SD ratio at the group level. *Prefer2SD* was developed in collaboration with a leading local game company over a year, incorporating insights from both in-game algorithm and operation experts. It seamlessly aligns with backend friend recommendation processes while offering enhanced frontend interactivity through six coordinated views. To validate its effectiveness, we evaluate *Prefer2SD* through a within-subjects study (N=12), a case study, and expert interviews. Results demonstrate its success in achieving desired SD ratios and improving recommendation performance for targeted player groups. The main contributions of this work are as follows:

1) We present *Prefer2SD*, an innovative two-step human-in-the-loop workflow that enables algorithm experts to fine-tune the SD ratio of friend recommendations for specific player groups.

2) We design a highly interactive visual analytics system that aligns with the *Prefer2SD* workflow, allowing algorithm experts to dynamically explore, analyze, and interpret the recommendation process. It incorporates fine-grained preference ratio adjustments to give users more precise control over recommendation outcomes.

3) We validate the efficacy of *Prefer2SD* in achieving target SD ratios and enhancing recommendation performance for distinct player groups through a within-subjects study (N=12), a case study, and expert interviews.

2 Background

2.1 About MMORPGs

A MMORPGis an Internet-based video game where numerous players take on their roles in a fictional world and engage in a virtual gaming environment. This study focuses on MMORPGs due to their emphasis on sustained social interactions and long-term player relationships. Unlike MOBA games, where player interactions are often short-term and session-based, MMORPGs require friend recommendation systems that helps create enduring connections among players. Although contemporary MMORPGs exhibit variations, they typically share fundamental characteristics.

Social Interaction. Social interaction is a core aspect of MMORPGs, often facilitated through ***guilds***—organized groups of players working together to achieve common goals. Guilds promote teamwork and shared goals, especially in ***PvE*** settings, where cooperation is crucial to success. Working together on challenging objectives fosters a sense of camaraderie and enhances the overall social experience, as players rely on each other for strategy and support.

Role-playing. Role-playing lets players immerse themselves in the game world by adopting specific ***roles or schools*** that define their abilities and responsibilities. ***Avatars***, which represent characters, are highly customizable, reflecting each player's identity and style. With numerous avatar options, players can express their individuality. Role-playing deepens engagement by combining character development with strategic gameplay and storytelling.

In our study, we incorporate these shared characteristics with player preferences, and categorize into three distinct parts: *social*, *gameplay* and *avatar* aspects, which will be detailed in Section 5.1.
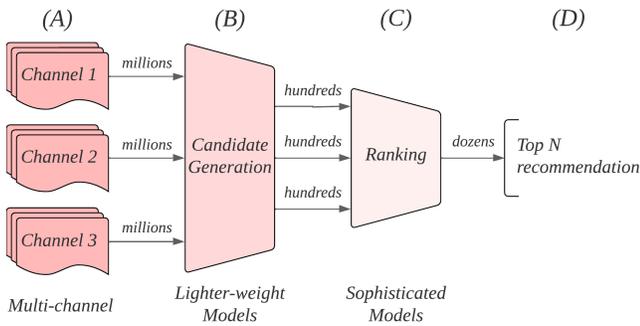
Figure 1: Workflow of multi-stage recommendation systems: Illustrated is an example of multi-channel matching for candidate generation. (A) Multi-channel classifies various items into specific channels, forming an extensive recommendation pool. (B) Candidate Generation efficiently narrows down items from millions to hundreds using models designed for large data handling. (C-D) The Ranking stage then sorts items by predicted user relevance, delivering a *top@N* list through sophisticated models.

2.2 Multi-stage Recommendation Systems

With the widespread availability of online services and content, recommendation systems play a crucial role in assisting users in discovering items of interest [102]. For MMORPGs, recommending friends based on player preferences is crucial to fostering long-term social interactions and enhancing player engagement within the game. Among the various recommendation strategies, multi-stage recommendation systems have been proven to be highly effective, especially in industries [3]. These systems leverage sequential filtering and refining techniques to enhance recommendation quality, providing users with more relevant and suitable suggestions. The architecture of multi-stage recommendation systems typically involves two main phases: *candidate generation* (Fig. 1(B)) and *ranking* (Fig. 1(C)). These stages enable the system to streamline the extensive item space (Fig. 1(A)) into a manageable set (Fig. 1(D)), paving the way for a more refined selection process.

Candidate Generation. In this phase (Fig. 1(B)), the objective is to narrow down the vast array of potential items (often in millions) to a more manageable subset of candidates (typically ranging from hundreds to thousands). Given the scale of this task, lighter-weight models such as matrix factorization, embedding-based retrieval, or shallow neural networks are commonly employed. Notably, ***multi-channel matching*** [89] proves to be an effective and common approach in the industry at this stage (Fig. 1(A-B)). By leveraging data from various channels (Fig. 1(A)) such as user browsing and purchase history, and social interactions, the system can attain a comprehensive understanding of user preferences. Each channel contributes a distinct perspective, and their integration can reflect the interests of users as a whole.

Ranking. Following the candidate generation phase, this stage (Fig. 1(C)) focuses on ranking the items according to their predicted relevance to the user, ultimately delivering *top@N* recommendations (Fig. 1(D)) that are expected to be most appealing. Advanced

models, including machine learning and deep learning architectures, are deployed in this stage, considering intricate user-item interactions, user and item metadata, context, and other nuanced factors.

3 Related Work

3.1 Friend Recommendation in Games

Friend recommendation strategies vary significantly across game genres due to differing player interaction and social goals. This section examines the unique challenges of MMORPGs and contrasts them with the approaches typically used in MOBA and First Person Shooter (FPS) games.

Friend recommendation strategies in social networks [8, 21, 23, 41, 84] commonly involve constructing user graphs and utilizing various sources, such as scholars' interactivity [90] and users' lifestyles [84], to gauge similarity between users. While graph neural network (GNN) approaches [20, 87] have attracted attention, they may not be well-suited for game friend recommendations, given the sparse nature of in-game user interactions and the challenge of designing suitable GNN architectures for various recommendation scenarios in large online game communities [50, 51, 92]. In contrast, integrating player portraits into the friend recommendation system has shown performance improvement [86, 97].

In MMORPGs, friend recommendations aim to encourage long-term relationships [68] by considering various and diverse player preferences, such as social behaviors and gameplay [21]. However, in games like MOBA and FPS, the focus shifts to short-term team formation [28] for match-based gameplay. FPS games typically rely on behavior-based approaches, analyzing metrics such as win rates [65] and recent performance [25] to balance team compositions. MOBA games, on the other hand, emphasize real-time compatibility, often relying on communication frequency [46] and skill-based matchmaking [66]. While these systems are efficient for forming transient teams, they may overlook mechanisms for building lasting social connections or promoting diversity. Unlike competitive games with ELO systems [35] designed for short-term compatibility, our work addresses the challenge of recommending lasting friendships, particularly in MMORPGs, where relationships often persist for extended periods.

3.2 Similarity and Diversity in Friend Recommendations

Balancing similarity and diversity has long been a central challenge in recommendation systems [29, 88]. Similarity reflects users' tendencies to connect with those who share common interests or behaviors [33], while diversity encourages exposure to novel and varied experiences [48]. Achieving this balance is particularly important in gaming contexts, as it directly influences player engagement and social satisfaction.

Previous studies have predominantly concentrated on suggesting similar friends, neglecting the diverse friend needs of players [55, 58, 101]. Enhancing recommendation diversity can address the long tail problem [63] and mitigate biased outcomes. Current efforts to enhance recommendation diversity can be categorized into

collaborative filtering-based, *content-based*, and *rank-based* methods [1]. Some researchers have explored improving recommendation diversity through interface design [77]. Nevertheless, the equilibrium between similarity and diversity in game friend recommendations remains unexplored, presenting an avenue for future research in this domain. We present an interactive and dynamic framework that leverages human expertise. It is crafted to aid professionals in exploring, analyzing, and fine-tuning recommendation process to achieve an optimal equilibrium.

In MMORPGs, diversity, as defined by the game company, refers to including players with different gameplay preferences to create balanced and engaging experiences. This study defines diversity along three main dimensions: social, gameplay, and avatar preferences. For example, in MMORPGs, where quests often require a mix of combat, support, and exploration skills, assembling teams with diverse play styles ensures a better chance of success and engagement for players [19]. This framework addresses these challenges by ensuring teams benefit from a variety of play styles rather than focusing solely on similarity, filling a gap in existing friend recommendation systems.

3.3 Interactive Recommendation Systems

Interactive recommendation systems prioritize user-centered control and interactivity, fostering a comprehensive understanding of the recommendation process. This approach aims to deliver a more engaging and personalized experience, promoting transparency, explanation, and trust [14, 30, 57, 82]. Recent efforts in this field leverage various interactive visualizations, employing flow charts [44], Venn diagrams [5, 64], graph-based representations [82], cluster maps [82], concentric circles [59], LineUp-like displays [14, 36], bubbles [6], maps [54], and more sophisticated visual analytics systems [37, 57]. For example, Parra et al. [64] empower users to manually blend recommender strategies and adjust their significance, using interactive Venn diagram to inspect the fusion results. TasteWeights [14], by combining techniques from traditional recommender system with a novel interactive interface, elucidates the recommendation process and facilitates user preference elicitation. While existing studies concentrate on enhancing end-user experience and trust through process control and flexibility, they often overlook assisting algorithm practitioners in understanding and debugging recommendation algorithms. In contrast, *Prefer2SD* reveals the entire process of multi-stage recommendations (Fig. 1). It aids algorithm experts in balancing similarity and diversity in friend recommendations, addressing the "filter bubble" [62] problem from a different perspective.

3.4 Human-AI Collaboration for Labeling

Human-AI collaboration in the labeling process [7, 15, 71, 94] is a critical area of research in Human-Computer Interaction (HCI), with approaches like active learning (AL) [16] being commonly used. In AL, human annotators label a small subset of data, which is then used to train models and improve performance. Recent advances emphasize the importance of combining both model-based and human-centered perspectives. For instance, Bernard et al.[11] introduced a unified approach that integrates AL with visualization techniques, while Sun et al.[75] explored "Label-and-Learn" interfaces

that improve developers' dataset comprehension. Choi et al. [24] took a further step with an interactive attention module that links human labelers with machine learning models through attention-based deep neural networks, enhancing document labeling. These studies highlight the power of combining human expertise with machine learning to create more effective labeling strategies. In addition, tools like KMTLabeler [83] and OneLabeler [100] have been designed to improve the efficiency and accuracy of labeling tasks. KMTLabeler introduces an interactive, knowledge-assisted framework for medical text classification, leveraging human feedback to refine the model while addressing the nuances of medical data labeling. Similarly, OneLabeler incorporates human judgment with machine learning, optimizing the process by assisting annotators with AI-generated suggestions. Both systems demonstrate how interactive human-AI collaboration can make labeling tasks more precise, especially in complex domains like medical text classification. In our study, human-AI collaboration enables efficient propagation of expert-defined ratios, which is critical for capturing customized player preferences and achieving the desired SD balance in friend recommendations. In addition, we use a combination of label propagation and active learning [27] to propagate expert-defined ratios across a selected group of participants, allowing for efficient human-AI collaboration.

4 Formative Study

4.1 About the Online Service and the Team

We formed a long-term collaboration with a leading MMORPG development team to explore friend recommendation algorithms. The team includes three male algorithm experts (**E1-3**), aged 31, 29, and 32, a female game designer (**E4**), aged 28, and a female product manager (**E5**), aged 27. **E1**, with five years of experience, focuses on game data analysis and algorithms for friend and gift recommendations. **E2-3** work on improving recommendation systems. **E4** creates in-game events and new gameplay, while **E5** connects the algorithm team with designers. Their goal is to develop more accurate friend recommendation algorithms to strengthen social ties among players.

4.2 Procedure and Analysis

Over the course of one year, we used participant observation [13] to embed ourselves in the company's routine operations, gaining direct insights into daily workflows and operational bottlenecks. Additionally, we conducted semi-structured interviews with key experts (Section 4.1), capturing qualitative understanding of their experiences with existing practices.

We synthesized insights from these methods using inductive thematic analysis [17], which helped us uncover key themes and patterns. Observational data were captured through detailed notes, while interviews, totaling around fifteen hours of audio, were recorded, transcribed via Otter.ai[1], and manually reviewed for accuracy. Two researchers immersed themselves in the transcripts and observation logs, repeatedly reviewing the data to identify emerging patterns. Each independently performed an initial coding of meaningful segments from the notes and transcripts, allowing

---

[1]http://otter.ai/

the data to guide the process without a predefined framework. They then met to compare and refine their codes, developing a unified coding framework. Potential themes were collaboratively identified and iteratively refined to ensure clarity and distinction. Finally, these findings were synthesized into *conventional practices* and *three key bottlenecks*, which informed the development of five design requirements aimed at addressing these bottlenecks.

4.3 Findings

4.3.1 Conventional Practice. The experts described the conventional method for developing friend recommendation systems, which prioritizes identifying players who are most likely to accept friend requests and build broader social connections within the game. The method typically follows two key stages: *candidate generation* and *ranking*.

Candidate Generation. According to **E1**, the process begins with generating a large pool of potential friend candidates, filtered down from millions of potential matches to a more manageable subset. **E1** explained, "*The standard practice is a multi-step process beginning with candidate generation and followed by matching, and finally ranking, which reduces the pool from millions to hundreds and ultimately to tens* (Fig. 1)." To further refine the candidate pool, *E2* and *E3* emphasized the role of **multi-channel matching** (Fig. 1(A-B)). This approach integrates various matching criteria to narrow the initial pool from millions to hundreds, leading to a more targeted set of matches.

Ranking System. E3 explained, "*After narrowing down the candidate pool, the remaining potential friends are processed through a ranking system* (Fig. 1(C))." This system prioritizes candidates based on metrics like cosine similarity and ranking precision, ultimately reducing the list to fewer than ten recommended friends per user.

4.3.2 Bottlenecks. Despite the strengths of these conventional methods, the experts highlighted several bottlenecks that hinder the overall effectiveness of friend recommendations.

B1. Comprehensive Preference Capture. The current friend recommendation system relies heavily on behavioral data, such as player actions and social interactions, to infer preferences. However, this approach cannot capture the full complexity of player preferences, particularly the more implicit and personalized aspects. Experts suggested that a more holistic approach was needed to understand players' deeper motivations. **E4** noted, "*Right now, we're mostly looking at what players do in the game—how they interact, who they connect with—but that only gives us part of the story. There's a whole lot [more] to their preferences that we're not capturing.*" **E5** added, "*Players show their personality in lots of different ways, like picking avatars or customizing their in-game look, but right now, our system doesn't really factor those in.*" While behavioral data is valuable, it is insufficient on its own. Incorporating additional data, such as symbolic choices like avatars, can improve the accuracy and personalization of recommendations. However, this remains an unaddressed limitation of the current system.

B2. Similarity Bias. The current system tends to prioritize players with similar in-game behaviors and interests, which can foster initial connections but also leads to a lack of diversity in social networks. **E5** observed, "*The system usually suggests people who are really similar, which works to a certain extent, but it [kind of] limits*

the range of connections players can make. It's not just about finding someone exactly like you." This bias restricts players' social circles, leading to repetitive recommendations. **E5** noted, "*Being constantly advised to associate with the same type of people can lead to a sense of stagnation in a cycle without any new experiences.*" While fostering connections based on similarity can be beneficial, it also reduces the diversity of interactions, limiting players' exposure to different perspectives. Balancing similarity and diversity in recommendations is a key challenge that the current system has yet to address.

B3. Group-Specific Needs. Different player groups—such as new and experienced players—have varying social needs when it comes to friend recommendations. New players often benefit from being matched with similar players to help them build memories of leveling up together, where experienced players seek more diverse connections to explore new content and strategies. The current system lacks the flexibility to adapt to these distinct needs. **E4** explained, "*For new [players], having friends with similar interests can help them feel more comfortable. But for veterans, they're often looking for more diverse connections to keep things new.*" **E1** added, "*There's no one-size-fits-all solution. We need a system that can adapt to where players are in their journey and connect them with friends appropriate for that stage.*" The inability to adjust the recommendations based on player group dynamics is a significant limitation of the existing system.

4.4 Design Requirements

Building on the conventional practice and identified bottlenecks, we derive five design requirements for enhancing the friend recommendation system.

DR1. Comprehensive Preference Capture. This design requirement addresses the challenge of incomplete player preference capture in the current system, which primarily relies on behavioral data (**B1**). **E4** highlighted that MMORPGs offer a variety of game experiences, such as *social interactions*, *monster hunting*, and *dungeon challenges*. Avatars also serve as significant forms of self-expression, allowing players to differentiate their characters. Therefore, to improve friend recommendations, the system must capture a wider range of player preferences, going beyond behavioral metrics to include symbolic choices like avatars and other personalized aspects of gameplay.

DR2. Balanced Friend Recommendation Algorithm. The current bias towards recommending similar friends has led to limited diversity in social connections (**B2**). Addressing this requires restructuring the algorithm to better balance similarity and diversity. Rather than relying solely on cosine similarity or behavioral overlaps, the recommendation system should incorporate additional factors that promote a wider range of social connections. **E1** noted that this shift will ensure the diversity of friendships is more reflective of the dynamic game world, fostering richer and more varied interactions among players.

DR3. Flexible Player Group Selection. The inability to adapt friend recommendations to the various social needs of distinct player groups—particularly new and experienced players—calls for a more flexible approach (**B3**). While new players often benefit from connecting with similar players, experienced players typically seek more diverse connections. Therefore, the system should be designed
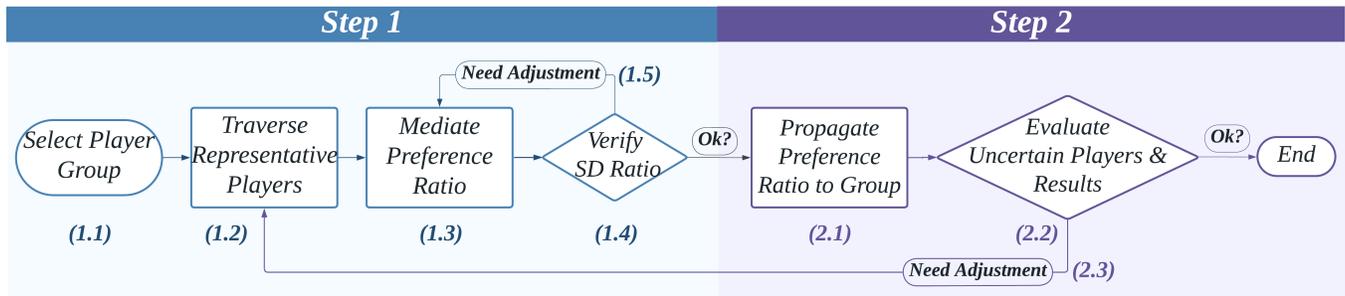
Figure 2: Framework of *Prefer2SD*. It introduces a two-step human-in-the-loop workflow. Step 1: Achieve an optimal SD ratio among representative players in a specific group by mediating their preference ratios. Step 2: Extend the preference ratios from Step 1 across the entire player cohort to achieve the desired group-level SD ratio.

to categorize players flexibly and adjust recommendation parameters based on their current stage and preferences. **E2** emphasized the importance of this adaptive approach to provide tailored friend recommendations that align with each player's evolving needs.

DR4. Dynamic Group-Specific Tuning. The system requires dynamic tuning capabilities to further refine friend recommendations for specific player groups (**B3**) while addressing the similarity bias (**B2**). This requirement involves developing an adaptive mechanism that adjusts recommendation parameters in real-time, ensuring the balance between similarity and diversity aligns with the unique preferences of both new and experienced players. As players progress through the game, their social needs evolve, and the system should adapt by offering friend recommendations that reflect these shifts. **E5** stressed the importance of continuously fine-tuning the balance to sustain player engagement throughout their journey.

DR5. Continuous Evaluation. To guarantee the effectiveness, a continuous evaluation mechanism should be established (**B2**). This would involve ongoing assessment of the system's adjustments, comparing the optimized recommendations against baseline results from traditional approaches. By regularly evaluating the impact of these adjustments, the system can ensure it optimizes for both similarity and diversity, ultimately overcoming the limitations of overly homogeneous recommendations. Experts agreed that this iterative assessment process is critical for maintaining a dynamic and engaging recommendation system for all player types.

5 Overview of Prefer2SD

The *Prefer2SD* framework is illustrated in Fig. 2, showing the integration of expert knowledge to mediate preference ratios, aiming to achieve an optimal SD ratio for a specific player group (**DR2**). The **preference ratio** refers to the proportion of friend candidates coming from different preference channels, representing how friend recommendations are distributed across various preference-based channels. The framework involves a two-step iterative procedure. Specifically, **Step 1** first guides experts to select a specific player group (Fig. 2(1.1)) for adjusting the SD ratio, identifying representative players within that group (Fig. 2(1.2)). Subsequently, experts mediate a suitable preference ratio (Fig. 2(1.3)) for each selected representative player to achieve the ideal individual-level SD ratio (Fig. 2(1.4-1.5)). In **Step 2**, *Prefer2SD* propagates the preference

ratios established in **Step 1** throughout the entire selected player group (Fig. 2(2.1)). Active learning displays a list of players with the most uncertain propagation results (Fig. 2(2.2)), allowing algorithm experts to adjust their preference ratios for further optimization of the group-level SD ratio (Fig. 2(2.3)).

The framework is powered by a backend engine and a frontend visualization interface, which are described in the following sections.

5.1 Backend Engine

The backend section is structured based on the two steps outlined in Fig. 2. It includes subsections covering *Data Description and Processing*, *Preference Extraction*, *Multi-stage Recommendation*, and *Evaluation Metrics* for **Step 1**. Meanwhile, **Step 2** includes the subsections on *Ratio Propagation* and *Active Learning*.

5.1.1 Data Description and Processing. In this study, we analyze a two-month log dataset from an online martial arts-themed MMORPG, spanning from January 1st to March 1st, 2022, which includes data from $732,437$ players. Monthly updates introduce new social events, gameplay features, and avatar enhancements. Rigorous data masking techniques are employed to ensure user privacy. Daily game logs are extracted based on key characteristics outlined in Section 2.1.

5.1.2 Step 1 Preference Extraction. After consulting with the experts, we derive three types of preference embeddings from the origin game logs: **social**, **gameplay**, and **avatar** (**DR1**). This approach aims to holistically capture the various facets of players' needs in the game. Furthermore, we obtain a **baseline** preference derived from their behavioral activities for subsequent comparison (Fig. 3(A)).

Social Preference. Social networks in MMORPGs reflect players' emotional connections. We extract social preference embeddings from both long-term and short-term behaviors. "Long-term" denotes stable social patterns, represented by a cumulative undirected graph of player interactions over a week, generating a 64-dimensional embedding vector $v_{longterm}$ using *Node2vec* [38].
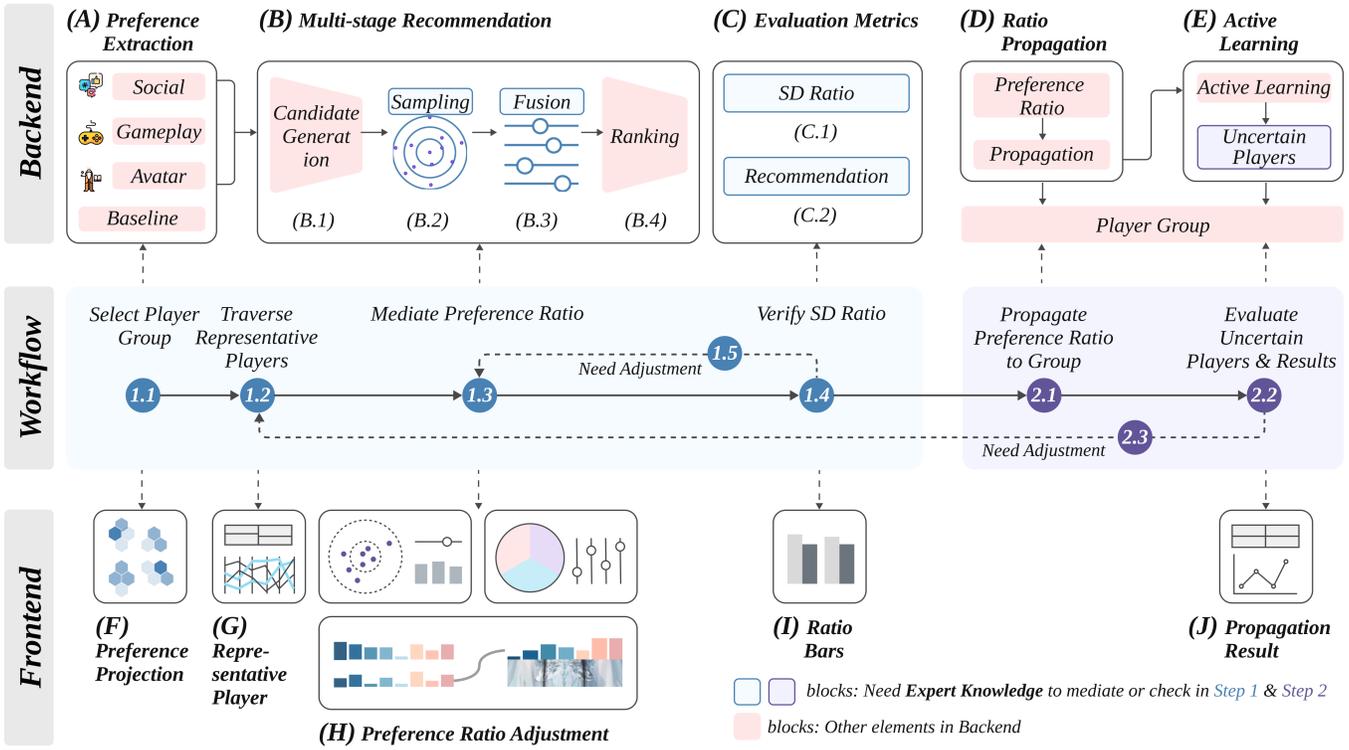
Figure 3: Framework of *Prefer2SD* supported by the Backend and the Frontend. (A-C) & (F-I) Backend and Frontend sections corresponding to Step 1. (D-E) & (J) Backend and Frontend sections for Step 2.

Short-term preferences are captured through daily interaction networks, reflecting transient social behaviors. Given the variability in daily networks, we incorporate three traditional social network metrics: *Page Rank* [10], *KCore* [26], and *Closeness Centrality* [60], assigning a default zero to absent nodes. These metrics form $v_{shortterm}$. The final social preference vector is $P_{social} = CONCAT(v_{longterm}, v_{shortterm})$.

Gameplay Preference. The variety of gameplay significantly influences MMORPG appeal, including options like *PVE*, *PVP*, and *Guild* (Section 2.1). We assess player engagement with each mode on a dichotomous scale, calculating statistical averages over one to seven days to create mean vectors ($v_{1d}, v_{3d}, v_{5d}, v_{7d}$), which are combined into the gameplay preference vector $P_{gameplay} = CONCAT(v_{1d}, v_{3d}, v_{5d}, v_{7d})$. Notably, this method does not rely on user behavioral data, which can be skewed by the game's storyline, leading to similar behavior across players. As preferences can only be expressed at specific points, behavioral measures may not fully capture true player preferences.

Avatar Preference. Customizing avatars with various costumes allows players to express personality [53]. We analyze game logs for avatar data and visual attire images. From logs, we extract data on owned avatars and acquisition sources, forming three vectors v_{num}, $v_{displayed}$, and $v_{acquisition}$. On the visual side, we use a database of 529 attire images and generate a 768-dimensional visual embedding v_{visual} with *CLIP* [67]. These vectors support tasks like clustering

or image similarity analysis. The final avatar preference vector is $P_{avatar} = CONCAT(v_{num}, v_{displayed}, v_{acquisition}, v_{visual})$.

Baseline Preference. Our baseline recommendation system relies on standard game practices for candidate generation (**DR5**). We extract attributes from players' behavioral records, forming a baseline preference vector for cosine similarity calculations. It allows comparison between original and expert-adjusted recommendation results, highlighting how expert knowledge can improve recommendation diversity and user satisfaction.

5.1.3   Step 1 Multi-stage Recommendation. In our MMORPG recommendation system, we use cosine similarity to generate candidate friends from four preference channels, ensuring a diverse pool (Fig. 3(B.1)). We then combine sampling (Fig. 3(B.2)) and fusion (Fig. 3(B.3)) techniques to balance algorithmic precision with expert insight. Finally, machine learning models (Fig. 3(B.4)) are employed to rank the top candidates, ensuring both technical robustness and practical applicability (**DR2**).

Candidate Generation. The four preferences serve as distinct channels during candidate generation. Given the large data volume, we use cosine similarity, computed as $sim_{a,b} = \cos(a, b) = \frac{a \cdot b}{|a| \cdot |b|}$, where a and b represent the players' preference embeddings. The system then identifies potential friends with the highest similarity in each channel.

Sampling and Fusion. To enhance recommendation diversity, we integrate expert knowledge to control the ratio of preference channels. *1) Sampling: Intra- Preference Candidate Selection*

Control. It involves organizing candidates in concentric circles based on their cosine similarity to the player [76, 95]. This structured approach improves diversity beyond random sampling. **2) Fusion: Inter-Preference Proportion Control.** It allows experts to dynamically adjust the proportion of each channel's contribution, ensuring the sum equals one (Fig. 3(B.3)). This method blends algorithmic efficiency with human judgment, adapting to varied user needs.

Ranking. The ranking phase (Fig. 3(B.4)) uses machine learning to produce accurate *top@N* candidates. We train on the first 40 days of a 60-day dataset of friend additions, reserving the last 20 days for testing. We compare two popular ranking algorithms: *GBDT* [32] and *LightGBM* [45]. For *GBDT*, we set a learning rate of 0.06, sub-sample rate of 0.56, and max depth of 9. *LightGBM* is configured with 31 leaves and the same max depth. The performance comparison is presented in Table 1. We ultimately select *GBDT* for the ranking phase.

Table 1: Performance of *GBDT* and *LightGBM*.

Model	Acc.	Recall	F1	Precision	AUC
GBDT	**76.3%**	**0.87**	**0.76**	0.66	**0.73**
LightGBM	75.5%	0.83	0.75	**0.67**	0.72

5.1.4 Step 1 Evaluation Metrics. We use **Diversity** and **Recommendation Quality** (Fig. 3(C)) to assess our friend recommendation system (**DR5**). **Diversity** (Fig. 3(C.1)) is measured using the SD ratio, consisting of *Content Diversity*, *Total_sim*, and *Fri_sim*. *Content Diversity* [18] ensures a wide range of potential connections. *Total_sim* calculates cosine similarity between candidates and the player, while *Fri_sim* measures similarity between recommended candidates and the player's friends. Lower values indicate higher diversity. Metrics like *item coverage* and *Gini index* are excluded as they suit item recommendations based on popularity [47]. **Recommendation Quality** (Fig. 3(C.2)) is assessed using *recall*, *precision*, *F1 score*, and *hit rate*. *F1 score* balances precision and recall. *Recall* evaluates the system's ability to identify relevant friends, while *precision* assesses the relevance. *Hit rate* shows how often a relevant friend is recommended.

5.1.5 Step 2 Ratio Propagation and Active Learning. Once ratios are assigned to representative players within a group, they are propagated to the entire group using a method similar to label propagation [42], where the expert-defined preference and SD ratios serve as labels for the group (Fig. 3(E)). For samples that are difficult to label automatically, an active learning approach [27] is employed (Fig. 3(E)). In this step, machine learning algorithms select a subset of difficult-to-label candidates for manual intervention by experts. Experts can then refine the ratios for these specific samples, and perform further iterations of ratio propagation. This human-ai collaboration allows for iterative fine-tuning of the recommendation system, improving the accuracy and adaptability of the recommendations by ensuring that hard-to-label cases are handled with expert input.

Once the ratios for the representative players are assigned, these ratios are propagated throughout the player group (Fig. 3(D)) using

a method akin to label propagation [42]. Label propagation, a semi-supervised machine learning algorithm, generally assigns labels to previously unlabeled data points. In this context, the predetermined expert-set ratios, i.e, preference ratio and SD ratio (Fig. 3(D)), serve as analogous to labels, applied across the player group via label propagation.

For samples challenging to label accurately through automated label propagation methods, active learning [27] is employed (Fig. 3(E)), which filters out the appropriate candidate set for manual labeling through machine learning methods. Expert knowledge is utilized for further secondary ratio setting, and additional iterations of ratio propagation are conducted to achieve targeted and optimal recommendation outcomes. This approach involves human expert intervention alongside machine learning methods, allowing for more precise labeling of these difficult samples. This collaboration enhances the accuracy and efficiency of the system in subsequent iterations.

5.2 Frontend Visualization

In line with the well-established mantra "*overview first, zoom and filter, then details on demand*" [74], we designed and developed an interactive system to facilitate the exploration of the human-in-the-loop workflow (Fig. 2).

We provide a *Guidance View* (Fig. 4(A)) in the interface, guiding users through each step, excluding adjustments (Steps 1.5 & 2.3). As users progress, active steps turn black, completed steps turn blue, and future steps remain gray. This dynamic navigation helps track users' progress and minimizes confusion.

In **Step 1**, users begin by navigating through the preference hexbins in the *Preference Projection View* (Fig. 4(B)) to identify group-of-interest. Subsequently, in the *Representative Player Selection View* (Fig. 4(C)), they analyze specific attributes of the selected group and choose a cohort of representatives for adjusting preference ratios. The *Preference Ratio Adjustment View* (Fig. 4(D)) offers two methods-"sampling" and "fusion"-for refining intra- and inter-preference candidate selection, which is vital for achieving the desired SD ratio. This process is repeated for each defined player group to ensure precision.

Step 2 focuses on refining ratio propagation through an active-learning-based iterative process. The expert-determined preference ratios are propagated across the remaining players within the same group. Propagation results are reviewed in the *Propagation with Active Learning View* (Fig. 4(E)) and *Result Comparison View* (Fig. 4(F)), helping identify the most optimal recommendations. Active learning is employed to tackle the challenge of accurately labeling complex, hard-to-replicate players that emerge during **Step 1**.

5.2.1 Preference Projection View. The *Preference Projection View* (Fig. 4(B)) offers an overview of the four preferences, aiding experts in selecting player groups (**DR3**) for targeted intervention in **Step 1.1** (Fig. 3(F)). Utilizing *t-SNE* [79], we project social, gameplay, avatar, and baseline preferences onto a two-dimensional space, where each point corresponds to a player and clusters represent similar preferences. A hexbin visualization overlays the data, with darker colors indicating higher player density. Tooltips display average preferences per hexbin, and hovering highlights related

Figure 4: The *Prefer2SD* interface incorporates six main views, each serving a distinct purpose. Color-coded labels enhance navigation: blue indicates **Step 1**, purple represents **Step 2**, and black highlights guidance. (A) The *Guidance View* provides an overview of progress and directs subsequent actions. (B) The *Preference Projection View* enables selection of specific player groups. (C) The *Representative Player Selection View* assists in choosing a representative player. (D) The *Preference Ratio Adjustment View* supports iterative mediation of preference ratios for ideal SD ratios. (E) The *Propagation with Active Learning View* helps identify uncertain players post-propagation. (F) The *Result Comparison View* summarizes recommendation results across iteration.

preferences, helping experts quickly identify target groups. Design alternatives are discussed in Appendix A.

center (Fig. 4(C.2)), offering a visual reference. Once a player is selected, experts proceed to adjust preference ratios in the *Preference Ratio Adjustment View* to achieve the target SD ratio.

5.2.2 Representative Player Selection View. The *Representative Player Selection View* (Fig. 4(C)) presents detailed information on the four preferences from the previously selected group in the *Projection Hexbin View*, supporting the selection of representative players in **Step 1.2** (Fig. 3(G)). To effectively handle the high-dimensional preference data, each player's values are displayed through a parallel coordinates plot (Fig. 4(C.1)), with attribute names color-coded to correspond to their respective preferences for easy distinction. Experts can narrow down the selection by brushing attribute ranges, which filters the players into a smaller subset displayed in a table (Fig. 4(C.3)). Clicking on a row reveals the player's avatar in the

5.2.3 Preference Ratio Adjustment View. This view (Fig. 4(D)) streamlines the multi-stage recommendation process, facilitating experts in iteratively refining preference ratios to achieve the desired SD ratio (**DR4**) during **Step 1.3-1.5** (Fig. 3(H-I)). It is divided into the **upper part** and the **lower part**. The former integrates three views (Fig. 4(D.1-3)) to assist sampling, fusing preferences, and monitoring SD ratio compliance. The latter (Fig. 4(D.4)) provides individual-level insights, enabling experts to make informed adjustments to preference ratios based on detailed player data.

Upper Part. This part is organized into six columns from left to right: the first four columns handle sampling (Fig. 4(D.1)), the
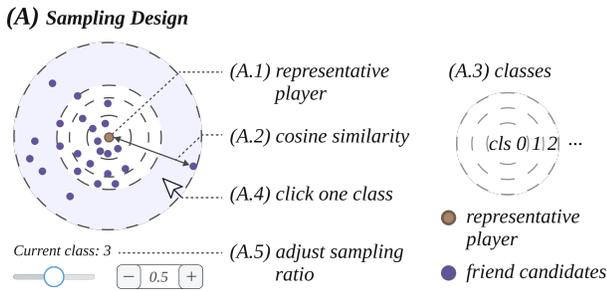
(A) *Sampling Design*

Figure 5: Sampling and Fusion Design. (A) Sampling: a scatter plot of candidates created by radial projection. Concentric circles divide the candidates into four classes.

fifth column (Fig. 4(D.2)) focuses on fusion, and the last column is dedicated to evaluating the ranking results (Fig. 4(D.3)).

Sampling. The sampling view (Fig. 4(D.1)) is divided into four columns, each representing a different candidate generation channel corresponding to a specific preference. At the top of each column, a radial projection scatter plot visualizes the candidates. Candidates are displayed around a ceteral player (Fig. 5(A)), with proximity to the center indicating higher cosine similarity (Fig. 5(A.2)). The candidates are arranged in four concentric circles (Fig. 5(A.3)), each circle representing a class of candidates. Experts can click on any class (Fig. 5(A.4)), to adjust the sampling ratio using a slider bar below (Fig. 5(A.5)), enabling precise control over intra-preference sampling. A bar chart beneath the slider provides feedback on content diversity for the selected candidates, aiding experts in tracking the SD ratio as sampling iterates. This approach offers a more refined sampling process compared to random selection, enhancing both precision and diversity.

Fusion. This view (Fig. 4(D.2)) adjusts inter-preference ratios by combining sampled candidates and projecting them onto a $2D$ plane using *t-SNE* [79]. Similar to the sampling columns, each point represents a candidate; however, in this view, pie charts (Fig. 6(B1)) are employed to depict candidates across multiple preferences. Since a player may be selected from various preferences, the pie charts use color coding to match each preference, ensuring a unified and intuitive visual expression. Each pie's radius represents the cosine similarity between the candidate and the central player, with larger radii indicating stronger similarity. This multidimensional visualization aids users in making nuanced inter-preference ratio adjustments, which can be fine-tuned using the four sliders below (Fig. 6(B2)). This interactive approach empowers experts to effectively mediate preferences, providing clarity and control.

Ranking and Interactions Among Columns. The upper section of the Ranking column (Fig. 4(D.3)) features a table recording intra- and inter-preference ratios, displaying sampling and fusion results. Users can initiate backend functions for targeted candidate selection by clicking the "Sample" and "Fusion" buttons (Fig. 6(A-B)) located below the table. To finalize the selections, experts can click the "Rank" button (Fig. 6(C)), which generates a ranked list of recommended friends. A bar chart beneath the table visualizes the SD ratio, calculated using *fri_sim* and *total_sim*, providing essential insights for informed decision-making. If a preference ratio leads

to an unsatisfying result, experts can click on the corresponding table row for further adjustment (Fig. 6(D)). These columns support **Step 1.3-1.5**, where the **upper part** facilitates group-level candidate selection, while the **lower part** focuses on individual-level adjustments. If modifications are necessary, users can click the corresponding row in the table to alter sampling, fusion, or ranking parameters. The updated SD ratio results are displayed in the bar charts for easy comparison, enabling experts to effectively refine their recommendation outcomes.

Lower Part. The lower part (Fig. 4(D.4)) provides both group- and individual-level recommendation metrics, along with detailed candidate information. This design supports experts in evaluating metrics and individual attributes that have been filtered during the sampling, fusion, and ranking stages (**DR5**). Following their analysis, experts can modify preference ratios in the upper part. To effectively illustrate multiple attributes, we adopt *Lineup* [36], organizing each column vertically (Fig. 4(D.1-3)), to display detailed candidate information. This arrangement allows users to examine candidate details within each column. An design alternative is discussed in Appendix B.

To facilitate comparisons, the top "Current Player" section (Fig. 7(A)) shows specific preference attributes of the chosen player (Fig. 4(C.1)) with an interactive feature to reveal attribute names when hovering over the bars. The middle "Group-level Metrics" section (Fig. 7(B)) displays average recommendation metrics and SD ratios via bar charts. The bottom *Lineup* design visualizes each candidate's detailed information, with rows representing candidates ranked by cosine similarity (Fig. 7(C.1)), and attributes encoded by color within the respective preference columns (Fig. 7(C.2)), facilitating direct comparison with the representative. Metrics like individual *cosine similarity*, *total_sim*, and *fri_sim* are included to indicate the SD ratio (Fig. 7(C.3)). For avatar preferences, they are shown beneath the bar chart (Fig. 7(C.4)). In the fusion and ranking columns, only the metrics are shown due to the complexity of the attributes, with predicted outcomes also included. When a candidate appears across multiple columns, the connected curves (Fig. 7(C.5)) help compare the rankings across different preferences.

5.2.4 Propagation with Active Learning View. This view (Fig. 4(E)) effectively highlights group players exhibiting the highest levels of uncertainty, encouraging adjustments to the preference ratios following propagation in **Step 2.1**. This organized presentation of uncertainty systematically decreases from top to bottom. In addition, the table displays the probability of each player being assigned to a specific preference ratio, which assists experts in implementing targeted interventions. For interactive analysis, users can select individual players by clicking on corresponding rows. This selection incorporates the chosen players into a dynamic table within the representative player selection interface, alongside previously selected representatives. Consequently, experts can iteratively refine the preference ratios, optimizing the SD ratio more effectively.

5.2.5 Result Comparison View. The *Result Comparison View* (Fig. 4(F)) employs line charts to illustrate the metric results of the entire player group's recommendations following each ratio propagation in **Step 2.2** (Fig. 3(J)). The evaluation results are plotted on the y-axis, with the x-axis indicates the iteration number.
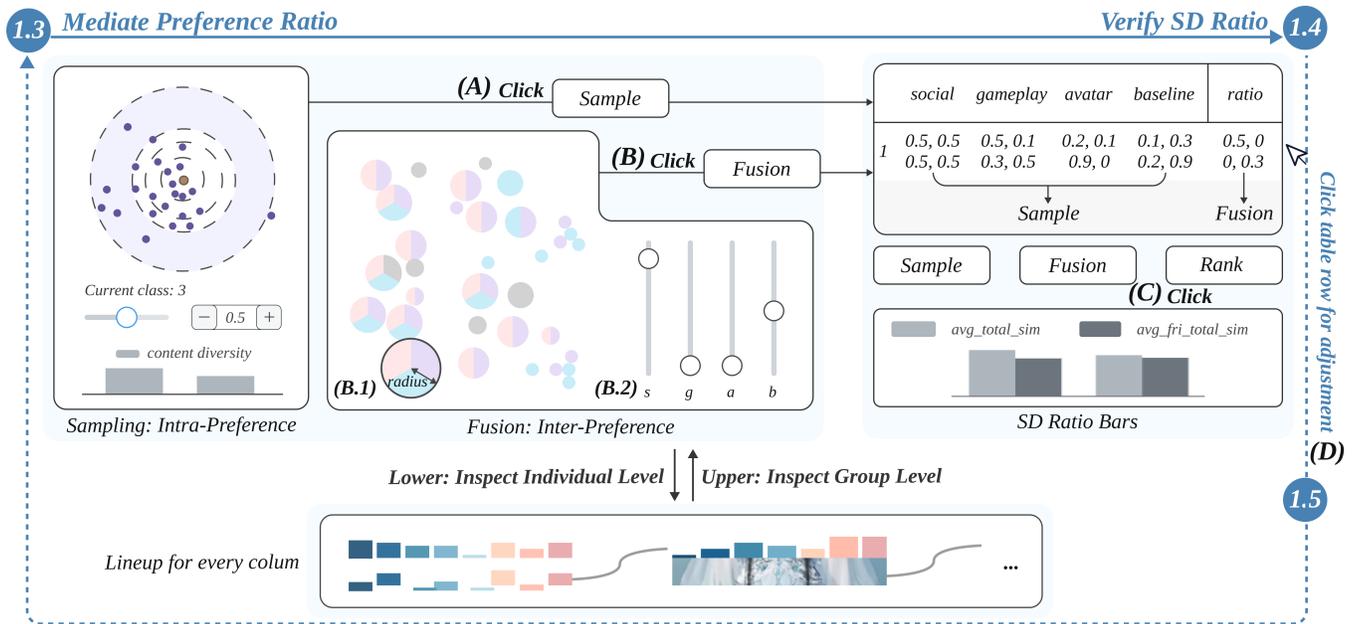
Figure 6: Interactions in Step 1.3-1.5: (A-B) Click "Sample" and "Fusion" button to record intra- and inter- preference ratios to the table. (C) Click "Rank" to trigger ranking function to verify SD ratio. (D) Click one table row for preference ratio re-adjustment.
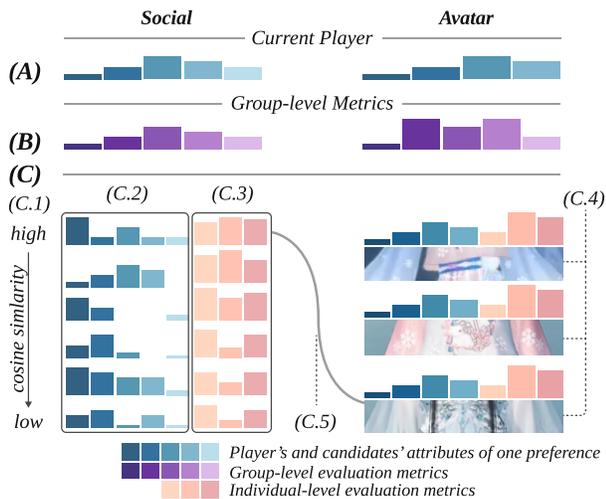

Figure 7: Lower part design. (A) Current player attributes. (B) Group-level evaluation metrics. (C) Lineup design showing candidates. (C.1) Row positioning based on cosine similarity. (C.2) Candidates' preference attributes. (C.3) Individual-level evaluation metrics of each candidate. (C.4) Candidates' avatars. (C.5) A line connecting the same candidate appearing in different preferences.

Users can hover over the charts to access detailed outcomes for each iteration, facilitating observation and comparison.

6 Evaluation

Upon obtaining institutional IRB approval, we conducted a within-subjects study with 12 participants to address three research questions aimed at balancing similarity and diversity in friend recommendations.

RQ1: How effective and usable is *Prefer2SD* in supporting friend recommendations?

RQ2: How does *Prefer2SD* impact users' cognitive load during the friend recommendation process?

RQ3: How is the interface design of *Prefer2SD* perceived by users?

6.1 Participants and Settings

We recruited 12 experts in friend recommendation algorithms from the collaborating local gaming company to participate in this study, comprising five females and seven males, with ages ranging from 22 to 40. Each participant received compensation of $10 per hour. Notably, three participants (**P1**–**P3**) were also involved in the formative study. For our baseline, we utilized the conventional method of filtering players through SQL and executing friend recommendations on the company's online platform. Regarding the data settings, we selected data corresponding to February 10th from the dataset (Section 5.1.1) to avoid potential data leakage. On February 10th, the dataset included 15, 339 players, with an average of 10.28 friends made in the preceding month and 9.05 friends made in the following month.
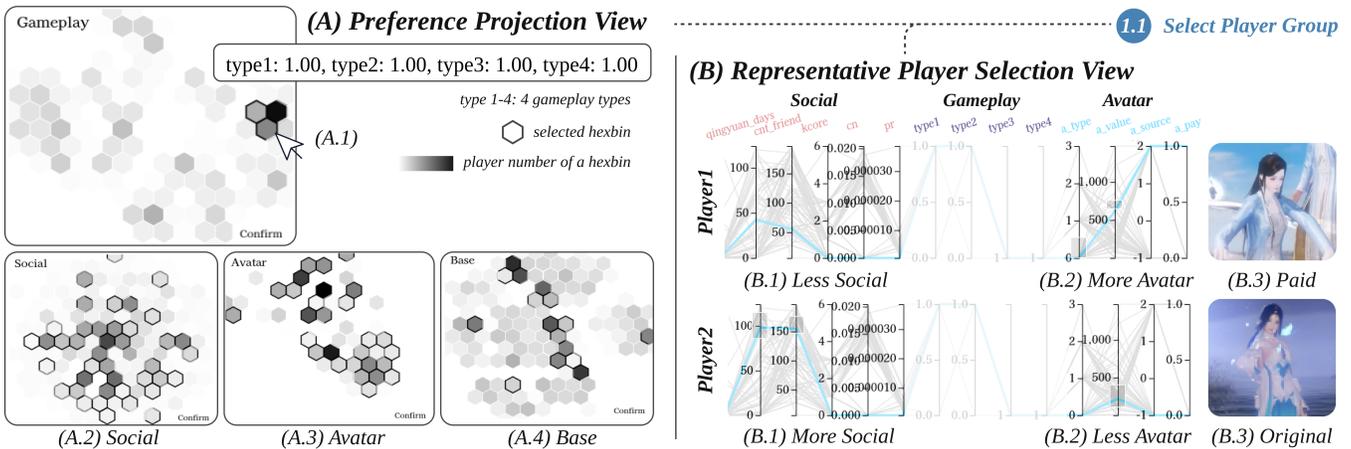
Figure 8: Step 1.1: (A) Selection of a cohort with gameplay preferences. (B) Identification of two representative players: one favoring avatars over social aspects, and the other favoring social aspects over avatars.

6.2 Procedure

One member of our research team conducted a pilot study to assess the completeness and feasibility of the study design. The formal study was structured into four stages. Initially, participants received a brief **introduction** covering the research background, *Prefer2SD's* workflow, and the interface's visual design and interactions. It included a 5-minute overview of the research context, followed by a 25-minute discussion of *Prefer2SD*. Participants then engaged in a phase of **free exploration**, where they familiarized themselves with *Prefer2SD*, working collaboratively in pairs under a co-discovery learning protocol [52] for 30 minutes. During this stage, we observed a notable case that will be elaborated later in Section 6.4. Following exploration, participants completed a **task** using *Prefer2SD* and the company's platform to recommend suitable friends to players interested in avatar; this task averaged 100 minutes. Afterward, participants filled out a questionnaire for evaluation. Finally, **semi-structured interviews** were conducted to gather feedback and suggestions regarding *Prefer2SD*.

6.3 Measurement

We employed a 7-point Likert scale (1: strongly disagree, 7: strongly agree) to assess *Prefer2SD*'s effectiveness, usability, cognitive load, and interface design. For assessing **effectiveness and usability**, we adapted several items from the *System Usability Scale* [9], including integration of functions, ease of learning, ease of use, willingness to use again, and confidence in using the system. Additionally, we measured the system's ability to balance similarity and diversity, its efficiency in generating recommendations, and participants' willingness to recommend it to others. To measure **cognitive load**, we utilized the *NASA-TLX questionnaire* [39], which evaluates cognitive demands across six dimensions: mental demand, physical demand, temporal demand, performance, effort, and frustration. For the **interface design**, we evaluated both the visual the interaction aspects of *Prefer2SD*.

During task completion, we recorded participants' actions via screen and audio capture while they engaged in a think-aloud

protocol [43], verbalizing their thoughts and reasoning. We also monitored **task completion time** for both the baseline and *Prefer2SD*, as well as **recommendation quality** and **diversity**. Task quality was measured using several metrics, including overall similarity with ground truth (GT) friends, which indicated how closely the recommended friends matched those that participants actually added in the subsequent month. In terms of diversity, we considered the total similarity between the player and the recommended friends (*total_sim*), as well as the total similarity between the players' friends and the recommended friends (*total_fri_sim*). This approach prioritized the long-term social connections between users and their recommended friends, rather than relying solely on traditional metrics like *hit rate* or *recall*, which primarily focus on short-term interactions.

6.4 Case Study: Suggesting a Varied Set of Friends

This case study involved collaboration with **P1** and **P9**, focused on using *Prefer2SD* to suggest diverse friends for core players to enhance their satisfaction and sense of belonging.

Goal: Enhancing Social Connectivity for Core Players. The goal was to address the needs of core players, a group with strong gameplay preferences. However, recommending friends to this group was challenging due to the limited pool of skilled players. Continuous recommendations of similar individuals could create a "filter bubble" [62], where social interactions are limited to those with comparable skills and attributes. To solve it, **P9** proposed promoting diverse acquaintances to enrich their social experience.

Step 1.1 Exploring a Cohort with Gameplay Preference. **P1** and **P9** imported three months of data (Section 5.1.1), focusing on a weekend period. **P1** explained, "*Weekends usually have more players active, that's why we're focusing on them for our study*." They then directed to the hexbins in the *Preference Projection View* (Fig. 8(A)), quickly identifying a cluster of skilled players with gameplay type values near 1, spanning three distinct hexbins(Fig. 8(A.1)).
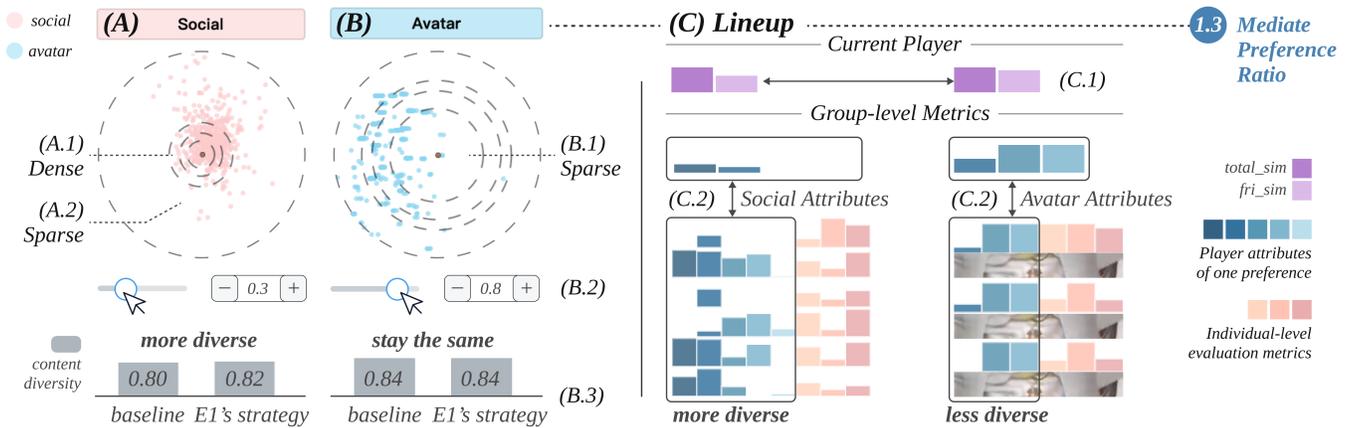
Figure 9: Step 1.3 (A-B) Conducting Sampling for both social and avatar preferences and evaluating metric values. **(C)** Delving deeper into candidate attributes from both preferences and comparing metrics at both group and individual levels.

Further analysis revealed that these players had diverse preferences in social, avatar, and base attributes, as highlighted in each corresponding view (Fig. 8(A.2-4)). **P9** remarked, "*It's completely natural since everyone has their own preferences.*" They then integrated the players in the selected hexbins into the *Representative Player Selection View* (Fig. 8(B)) for further analysis. Using the parallel coordinate plot, **P1** and **P9** evaluated the group across different preference attributes, selecting two representative players—one favoring avatar over social, and the other favoring social over avatar (Fig. 8(B.1-2)). Additionally, they noted that the avatar-preferring player had chosen a paid version, while the social-preferring player opted for the original version (Fig. 8(B.3)), reinforcing their earlier observations.

Step 1.2 Examining two representative players. P1 and **P9** began to examine two chosen representative players and chose the player with the avatar preference first. To increase recommendation diversity, they decided to recommend candidates from two preferences, **social** and **avatar**.

Step 1.3 Mediating intra-preference ratio: Sampling. In the *Preference Ratio Adjustment View*, they first explored the sampling columns (Fig. 9). **P1** observed that social preference candidates were tightly clustered within the inner three circles (Fig. 9(A.1)), indicating higher similarity, while the outermost circle showed more diversity due to the sparsity of the candidates (Fig. 9(A.2)). To standardize the process, they initially set the sampling frequency for each class to 1, feeding all candidates to ranking and click the "Sample" button. They then adjusted the frequency of each concentric circle, reducing it for the inner circles and increasing it for the sparse outer circle (Fig. 9(A.3)). Specifically, the intra-preference ratio for **social** was adjusted to 0.3, 0.3, 0.3 and 0.8 from inner to outer classes, they then evaluated the results. Content diversity improved slightly, from from 0.80 to 0.82. For the **avatar** preference, they started with a baseline sampling frequency of 1 (Fig. 9(B.2)) and noted that the points were more spread out compared to social (Fig. 9(B.1)). Thus, they increased the sampling frequency to
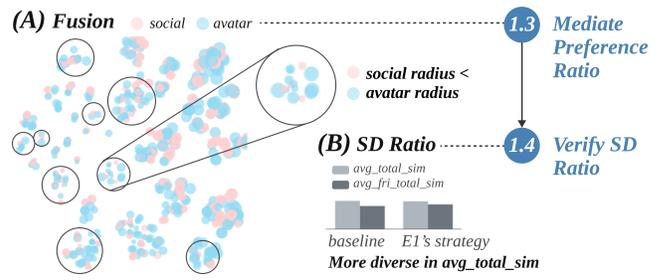

Figure 10: Step 1.3-4 (A) Exploring pies of social and avatar candidates in Fusion. **(B)** Comparing SD ratio between results of baseline and experts' own strategy.

0.8 for each class but found no significant change in *content diversity* (Fig. 9(B.3)), suggesting that adjusting the candidate sampling frequency had minimal impact.

To delve deeper into the candidates distribution for the two strategies, **P1** and **P9** shifted to the lineup view below (Fig. 9(C)). At the group level, the *total_sim* values were nearly identical, yet social candidates showed broader diversity in *friend_sim* (Fig. 9(C.1)). Upon reviewing individual attributes, they confirmed that social candidates exhibited more diversity compared to avatar ones (Fig. 9(C.2)), consistent with their earlier findings during sampling (Fig. 9(A-B)).

Step 1.3 Mediating inter-preference ratio: Fusion. During fusion, the experts analyzed the projected players in the fusion column. They noted that while social and avatar candidates had nearly identical distributions, social candidates had smaller radii overall. In Fig. 10(A), two key observations merged: 1) Few candidates were selected from both preferences, as most pies were dominated by a single color; 2) Social candidates had smaller radii than avatar candidates, suggesting greater diversity. Based on these findings, they adjusted the fusion ratio to prioritize social candidates, setting it at 0.7 for social and 0.3 for avatar to ensure a more diverse candidate pool.

Step 1.4 Verify SD ratio. After fusion, the results were fed into the ranking model to assess the SD ratio. For comparison, they used the baseline preference incorporated, representing their conventional method. The intra- and inter-preference ratios for baseline candidates were set to 1, and all others to 0, ensuring only baseline candidates entered the ranking stage. The results showed that their strategy achieved a lower *total_sim* compared to the baseline, while *fri_sim* remained unchanged (Fig. 10(B)). This indicated a more diverse recommendation. Consequently, they assigned the intra- and inter-preference ratio as Preference Ratio 1 for this player (Fig. 11(A)).

Back to Step 1.2 Traverse to the second representative. **P1** and **P9** applied the iterative mediation process to the second representative, who favored social over avatar preferences. Through experimentation, they found that setting the intra-preference ratios of both social and avatar classes to 0.5, and the inter-preference ratios of social and avatar to 0.3 and 0.7, respectively, resulted in a more diverse set of friend recommendations. Consequently, they identified this ratio configuration as Preference Ratio 2 (Fig. 11(A)). At this stage, the experts had successfully assigned two preference ratios to the selected player cohort based on their expertise.

Step 2.1-2 Preference Ratio Propagation and Evaluation. After assigning two preference ratios, the experts activated the *Propagation* button (Fig. 11(A.1)) to apply these ratios across the player cohort. Players flagged through active learning for re-mediation were displayed in the *Propagation with Active Learning View* (Fig. 11(B)). **P9** observed that the probability of propagating the Preference Ratio 1 was extremely low, with most values under 0.2 (Fig. 11(B.1)). As a result, they concluded that this ratio should be considered an outlier and unsuitable for propagation.

Step 2.3 Iterative Adjustment. The experts then selected the player with lowest possibility (0.11 for Preference Ratio 1) for re-mediation (Fig. 11(B.2)). After observations in the relevant views from **Step 1.3-5**, they assigned this player an intra-preference ratio of 0.5 for all classes, and an inter-preference ratio of 0.5 for both social and avatar classes, designating it as Preference Ratio 3. Upon executing propagation again (Fig. 11(B.2)), they observed a significant increase in propagation probability (Fig. 11(C.1)). The recommended outcomes were more diverse (Fig. 11(C.2)), with *total_sim* decreased from 0.65 to 0.62 while *total_fri_sim* remained constant. Based on the iterative process, **P1** and **P9** concluded that the Preference Ratio 2 & 3 were appropriate for the player group.

Summary. The experts recognized the value of *Prefer2SD* for friend recommendations, noting that visualizing candidate information at different granularities made analysis and adjustments more intuitive, thereby enhancing their efficiency.

7 Results and Analysis

In our quantitative analysis, we conducted the Wilcoxon signed-rank test [61] to evaluate the differences in participants' ratings across various factors between the two systems. For the qualitative analysis, we utilized inductive thematic analysis [17] on the interview transcripts. Two researchers independently coded the data, followed by a cross-checking of themes to ensure consistency and reliability. The findings were reported in relation to the three research questions.

7.1 RQ1: How effective and usable is *Prefer2SD* in supporting friend recommendations?

Effectiveness. *Prefer2SD* showed significant improvements in balancing similarity and diversity as well as improving recommendation efficiency ($W = 0.00$, $p < 0.01$ for both). Additionally, participants perceived that the system's functions were well integrated into the interface ($W = 7.00$, $p < 0.05$) compared to the baseline system in Table 2. Most participants (11/12) found the framework of *Prefer2SD* very familiar and easy to grasp. **P12** expressed confidence that introducing mediation to intra- and inter-preference ratios in Step 1 would effectively control the SD ratio, a crucial factor in achieving more balanced recommendation outcomes. "*But it requires a deep understanding of each player's personal desires.*" remarked **P9**. Additionally, **P2** emphasized the significance of Step 2, stating, "*It allows us to refine recommendations, enhancing the system's accuracy and flexibility.*"

As shown in Fig. 12, the task completion time was slightly reduced when using *Prefer2SD* compared to the baseline, indicating some improvement in efficiency (Fig. 12(A)). More importantly, recommendation quality significantly increased (Fig. 12(B)), with a higher match between recommended friends and actual friends added by the player ($W = 0.00$, $p < 0.01$). This suggests that *Prefer2SD* can provide more accurate and sustainable friend recommendations, which contribute to the development of long-term social networks. Several participants acknowledged these improvements. **P3** mentioned, "*Prefer2SD's ability to balance similarity and diversity in the recommendations is very noticeable. It's an improvement compared to typical systems that usually lean too much towards similarity.*" **P7** added, "*It is often difficult to achieve in our conventional practice.*"

As for diversity (Fig. 12(C)), the results between the two systems were relatively similar. This indicated that the player group who strongly prefer avatar tended to choose friends who were more similar to themselves. Several participants offered their insights. **P6** pointed out, "*It's important to recognize that while diversity is a key goal, too much diversity without sufficient similarity might lead to user disengagement. The system must find a balance that satisfies both diversity and personal relevance*".

Usability. The usability analysis revealed notable improvements with *Prefer2SD* over the baseline system across several key areas. Although no significant differences were observed in terms of ease of learning ($W = 15.50$, $p > 0.05$) and ease of use ($W = 14.00$, $p > 0.05$), participants became increasingly adept as they familiarized themselves with *Prefer2SD*'s interface. While the initial design presented some challenges, participants adapted over time. For instance, **P3** highlighted that the *Guidance View* (Fig. 4(A)) played a pivotal role in this adaptation, connecting the familiar framework with each view and enhancing overall system comprehension.

Prefer2SD demonstrated clear strengths in usability metrics. Participants were significantly more likely to recommend the system to others ($W = 2.00$, $p < 0.01$) and showed a notably higher willingness to use it again ($W = 0.00$, $p < 0.01$), indicating a strong preference for *Prefer2SD*. Many participants (8/12) praised the system's versatility, highlighting its adaptability to other game friend recommendation scenarios. For instance, **P9** noted that *Prefer2SD*'s
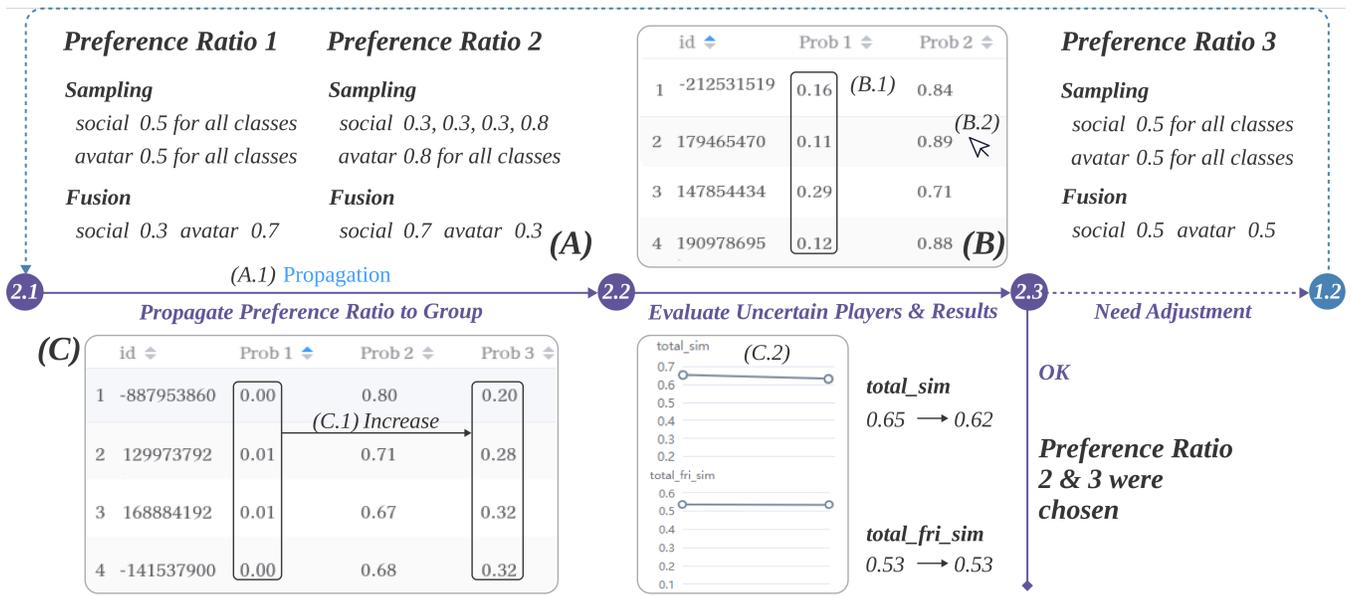
Figure 11: Step 2.1-3 (A) Propagate Preference Ratios 1 & 2 to the player group. (B) Identify the most uncertain player for adjusting the preference ratio and propagate again. (C) Adopt Preference Ratios 2 & 3 based on the favorable probabilities and metric values.

Table 2: The statistical feedback with *Baseline* and *Prefer2SD*, where the *p*-values (n.s: not significant, *: $p < .050$, **: $p < .010$, ***: $p < .001$) is reported.

Category	Factor	Baseline Mean/S.D.	Prefer2SD Mean/S.D.	W	p	$Eff.Size$
	Balancing similarity and diversity	2.00/0.95	5.91/0.99	0.00	0.002**	0.88
Effectiveness	Improving efficiency	3.25/1.22	5.33/1.23	0.00	0.008**	0.88
	Functions	4.17/1.40	5.83/1.02	7.00	0.020*	0.69
	Easy to learn	4.25/1.14	3.41/1.37	15.50	0.124n.s	0.46
	Easy to use	3.33/1.55	4.42/0.99	14.00	0.174n.s	0.42
Usability	Willing to recommend	3.25/1.22	5.50/1.31	2.00	0.006**	0.82
	Willing to use again	3.33/1.61	5.67/1.23	0.00	0.003**	0.88
	Confident to use	4.17/1.47	6.00/0.95	4.50	0.011*	0.75
	Less Mental	3.17/1.34	5.00/1.47	1.50	0.014*	0.81
	Less Physical	4.17/0.94	5.58/1.08	4.00	0.009**	0.79
Cognitive Load	Less Temporal	3.91/1.31	4.92/1.08	20.00	0.132n.s	0.43
	Performance	2.42/1.08	5.50/1.17	0.00	0.003**	0.88
	Less Effort	4.00/1.21	5.17/1.27	6.00	0.020*	0.68
	Less Frustration	3.33/1.49	4.75/1.54	26.00	0.060*	0.55
Design	Visual design	-	6.10/0.86	-	-	-
	Interaction design	-	6.28/0.63	-	-	-

flexible design could seamlessly integrate into different game environments, making it highly adaptable across various platforms. Additionally, they reported a significant boost in confidence when using *Prefer2SD* ($W = 4.50$, $p < 0.05$) compared to the baseline system. All participants (12/12) pointed out that the system's diversity indicators facilitated a more intuitive assessment of the variety in recommendations, simplifying the evaluation of its outputs.

These findings indicate that although both systems exhibited comparable learning curves and ease of use, *Prefer2SD* outperformed in fostering user satisfaction and engagement. Participants
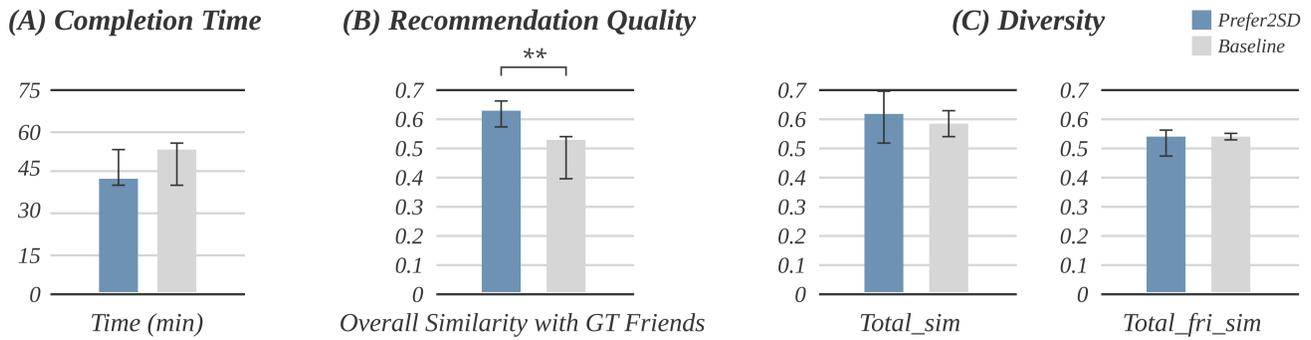
Figure 12: *Prefer2SD* outperformed the baseline in three dimensions: (A) Completion time; (B) Recommendation quality; (C) Diversity. (*: $p < .050$, **: $p < .010$, ***: $p < .001$)

were more inclined to recommend and reuse the system, demonstrating increased confidence in its application.

7.2 RQ2: How does *Prefer2SD* impact users' cognitive load during the friend recommendation process?

Prefer2SD significantly reduced cognitive load across multiple dimensions compared to the baseline system. Participants reported experiencing less mental ($W = 1.50$, $p < 0.05$) and physical load ($W = 4.00$, $p < 0.01$) when using *Prefer2SD* , indicating that the system required less cognitive processing for decision-making during the friend recommendation process. **P5** specifically noted that, compared to the company's node-based recommendation platform, *Prefer2SD*'s visual interface provided a clearer understanding of the recommendation process. "*Having a visual representation of the entire recommendation flow makes it much easier to grasp what's happening at each step,*" said **P5**.

The results also showed an improvement in temporal load, although the difference between *Prefer2SD* and the baseline system was not statistically significant ($W = 20.00$, $p > 0.05$). **P1** explained, "*The system provides more detailed information and gives me more control over the recommendation process, but it does take a bit longer to explore and fully understand everything.*" Participants, however, reported a significant improvement in performance ($W = 0.00$, $p < 0.01$), suggesting that they felt more effective and successful in completing tasks using *Prefer2SD*. As **P11** noted, "*With Prefer2SD, I can dynamically adjust the granularity of the recommendations and tailor strategies for different player groups. This level of control is invaluable for reducing player churn and enhancing the core user experience.*" *Prefer2SD* also required significantly less effort from users ($W = 6.00$, $p < 0.05$), further reducing cognitive strain during the friend recommendation process.

Lastly, *Prefer2SD* led to a significant reduction in frustration levels compared to the baseline system ($W = 26.00$, $p < 0.05$). Several participants appreciated how *Prefer2SD*'s visual tools enabled them to conduct in-depth analyses without the need to switch between multiple platforms. "*The ability to perform deeper analysis within a single interface saves us from jumping between different systems,*

which often causes confusion and delays," remarked **P4**. This integration streamlined their workflow, helping them avoid the frustration of managing multiple tools during the analysis process.

In summary, *Prefer2SD* significantly reduced participants' overall cognitive load, particularly in terms of mental, physical, and effort-related dimensions. By improving performance while simultaneously lowering effort and frustration, the system effectively streamlined the friend recommendation process, making the task more manageable and less cognitively taxing for users.

7.3 RQ3: How is the interface design of *Prefer2SD* perceived by users?

The interface design of *Prefer2SD* received highly positive feedback from participants, as reflected in the strong ratings for both visual and interaction design. The visual design achieved an impressive mean score of 6.10 ($SD = 0.86$), while the interaction design received an even higher rating of 6.28 ($SD = 0.63$), indicating that participants found the system not only visually appealing but also highly intuitive and user-friendly.

A key aspect of the interface that resonated with participants was the integration of hexbins and tooltips within the *Preference Projection View*. **P1** particularly praised this feature, emphasizing its effectiveness in quickly identifying player groups that required specialized intervention. This visual tool allowed experts to efficiently navigate large datasets, allowing them to quickly pinpoint areas requiring focused actions, thereby streamlining the decision-making process.

Additionally, participants valued the system's ability to provide in-depth and interactive feedback on the recommendations. The design elements, such as the sampling and fusion functions, allowed users to iteratively explore and modify the recommendation outcomes. The seamless interaction among different views enabled a smooth analytical flow, making the system highly adaptable to users' needs. This iterative examination of recommendation outcomes was particularly noted as a strength, empowering experts to dynamically adjust and refine their strategies.

In summary, the interface design of *Prefer2SD* received high praise from users, with both visual and interaction elements achieving top ratings. The effective integration of visualization tools,

along with the system's interactive and flexible design, allowed participants to conduct detailed and dynamic analyses, significantly enhancing the overall user experience.

8 Discussion and Limitations

In this section, we discuss *Prefer2SD* regarding three design considerations, generalizability and scalability, and limitations and future work.

8.1 Design Considerations

DC1: Tailored Focus on Target Users. Previous research on recommendation systems has primarily focused on end-users, emphasizing the importance of trust-building and enhancing user experience through transparent recommendation processes [14, 30, 37, 57, 82]. In contrast, *Prefer2SD* specifically targets algorithm experts responsible for designing these recommendation systems. This strategic shift aims to enhance user experience in friend recommendations by refining the effectiveness and adaptability of recommendation algorithms. By doing so, algorithm experts can optimize user experiences for end-users by offering more diverse and flexible recommendation models. Algorithm experts should leverage through the provision of more diverse and flexible recommendation models. Algorithm experts should utilize the system's adaptability to experiment with various recommendation models and game scenarios, allowing them to refine their strategies in response to differing user behaviors and needs. Regularly reviewing user interaction data will enable them to identify patterns that inform ongoing adjustments, ensuring that the system remains responsive to evolving user requirements.

DC2: Enhanced Granularity in Candidate Adjustment. Previous approaches have typically allowed for candidate filtering across various channels [14, 64], but often lacked the fine-tuned control necessary for more nuanced recommendations. *Prefer2SD* addresses this limitation by enabling experts to adjust both intra- and inter-preference ratios. This enhanced granularity provides precise control over how recommendations are customized, thereby improving flexibility and optimization potential across diverse recommendation scenarios.

For algorithm experts, it is highly recommended to utilize these granular adjustment options to fine-tune the balance between similarity and diversity, particularly in situations where subtle shifts in preferences can significantly affect recommendation quality. By closely monitoring the results of these adjustments, experts can identify the most effective combinations of ratios that resonate with different user groups.

DC3: Facilitation of Iterative Group-Level Adjustments. In contrast to prior work, *Prefer2SD* integrates ratio propagation and active learning techniques, enabling experts to iteratively refine recommendations at the group level. By establishing preference ratios for a select few representative players, experts can efficiently influence recommendation outcomes for entire groups. This iterative approach not only streamlines group-level recommendations but also enhances the system's adaptability and accuracy, facilitating ongoing refinement over time. Experts are encouraged to adopt this iterative strategy, gradually adjusting group-level preferences while monitoring the collective impact on the system. Through ratio

propagation techniques, they can implement changes across large groups more effectively, saving time while maintaining consistent recommendation quality across diverse user segments.

8.2 Generalizability and Scalability

This study introduces a two-step iterative and interactive framework (Fig. 2), aims at attaining the optimal group-level SD ratio by mediating preference ratios based on expert insight. *1) Generalizability.* Initially, the framework is tailored for friend recommendations within MMORPGs. Extending it to similar gaming contexts may require only minor adjustments, depending on the similarity of player behaviors and available data channels. For example, another MMORPG with comparable social structures would mainly involve modifying preference channels. However, in MOBA games, where player interactions are typically short-term and match-based, adapting the framework would require reconfiguring preference channels to focus on transient gameplay dynamics and immediate team composition needs. Additionally, this approach represents an advancement over multi-stage recommendation systems (Fig. 1), theoretically adaptable to analogous recommendation processes in other domains, such as friend recommendations on social media. Nevertheless, adaptations might necessitate alterations in the channels (Fig. 1(A)) utilized for candidate generation, replacing or modifying the preference channels employed in this study. *2) Scalability.* The visualizations within the *Prefer2SD* system are designed to accommodate a large volume of players. For instance, the utilization of hexbins (Fig. 13) in the *Preference Projection View* ensures system efficiency in processing extensive player datasets. Furthermore, as our framework propagates assigned preference ratios from a few key players to the entire cohort, the efficiency and quality of *Prefer2SD*'s results remain unaffected, even with larger player groups. However, *Prefer2SD* may struggle with managing a multitude of preferences because visualizing each of them can lead to information overload. Thus, when utilizing *Prefer2SD*, it becomes imperative to initially select a few influential preferences for analysis. Additionally, the high dimensionality of combined preference vectors in multi-modal preference extraction may increase computational costs and reduce scalability, especially as the number of preferences grows.

8.3 Limitations and Future Work

While our approach demonstrates promising capabilities, it is important to acknowledge several limitations that warrant attention. *1) Lack of online A/B testing:* Online A/B testing plays a pivotal role in evaluating system performance by directly measuring players' preferences for friends recommended through expert-customized strategies via *Prefer2SD*. However, challenges in integrating multiple channels and navigating the complexities of the game release process have hindered the implementation of online A/B testing, resulting in its abandonment. Furthermore, the lack of long-term A/B testing restricts our ability to evaluate how effectively the system maintains recommendation quality over time, including its impact on user engagement and satisfaction. To address this gap, we conducted an quantitative within-subjects user study (Section 6), partially compensating for the lack of A/B testing. By analyzing recommendation results from two perspectives - similarity to past

friends and similarity to friends added in the subsequent month - we observed a significant improvement compared to the baseline, thereby partially validating the effectiveness of *Prefer2SD*. **2) Inability to handle a large number of preferences:** As discussed in "Scalability" (Section 8.2), the current system struggles to manage an extensive array of preference settings. Consequently, future work should explore the development of algorithms and visualization designs capable of accommodating dozens or even hundreds of preference channels. **3) Dependence on human knowledge for recommended results:** The current approach heavily relies on expert-selected representative players, potentially resulting in inadequate ratio propagation if the chosen players exhibit excessive similarity. To mitigate this issue, future iterations will integrate statistical visualization techniques to support representative selection, ensuring a more comprehensive coverage of player distribution. **4) Absence of mining for other multi-modal preferences:** *Prefer2SD* currently does not incorporate other multi-modal preference data within the game, such as text and voice data from chats on public in-game channels. While including this data may yield more accurate recommendation results, it also raises privacy concerns for players, necessitating careful consideration and addressing in future endeavors. **5) Data privacy and open-sourcing:** Finally, while the current system remains proprietary, future work will consider open-sourcing the system after resolving data privacy concerns through discussions with the collaborating company. **6) Lack of direct player involvement in the formative study.** While expert feedback provided valuable insights into the systemic limitations, directly involving players is essential for understanding their personal preferences, which are key to the social dynamics in MMORPGs. Future work should engage a diverse range of players, including both new and experienced players, to ensure the design requirements are closely aligned with their needs. **7) Cold-start problem.** The current design does not explicitly tackle the cold-start problem for new players with limited or no historical data. Further iterations of *Prefer2SD* could address this issue by incorporating methods such as demographic-based initialization or collaborative strategies for small datasets. **8) Computational expense of ratio propagation.** Ratio propagation can become computationally expensive when scaling to tens of thousands of players due to increased computational load. To address this, we plan to implement parallel computing techniques to enhance computational efficiency. **9) Limited sample size in the user study.** The user study included only 12 participants, which may limit the generalizability of the findings. Future research should involve a larger and more diverse group of experts to ensure the system's design is applicable across different domains.

9   Conclusion

This study introduces *Prefer2SD*, a two-step human-in-the-loop workflow empowered by an interactive visual analytics system. This system aids algorithm experts in navigating, evaluating, and harmonizing the similarity and diversity in recommended friends for specific player groups. Our method integrates a range of player preferences and provides carefully designed interactive visualizations to assist experts in iteratively adjusting preference ratios, striving for an ideal equilibrium between similarity and diversity.

The effectiveness of *Prefer2SD* is substantiated through a within-subjects study, a case study, and expert interviews, affirming its practicality and relevance.

References

[1] Gediminas Adomavicius and YoungOk Kwon. 2011. Improving aggregate recommendation diversity using ranking-based techniques. *IEEE Transactions on Knowledge and Data Engineering* 24, 5 (2011), 896–911.

[2] Vinti Agarwal and Kamal Kant Bharadwaj. 2013. A collaborative filtering framework for friends recommendation in social networks based on interaction intensity and adaptive user similarity. *Social Network Analysis and Mining* 3 (2013), 359–379.

[3] Ronay Ak, Benedikt Schifferer, Sara Rabhi, and Gabriel De Souza Pereira Moreira. 2022. Training and Deploying Multi-Stage Recommender Systems. In *Proceedings of the 16th ACM Conference on Recommender Systems.* 706–707.

[4] Shrooq Algarni and Frederick Sheldon. 2023. Systematic Review of Recommendation Systems for Course Selection. *Machine Learning and Knowledge Extraction* 5, 2 (2023), 560–596.

[5] Ivana Andjelkovic, Denis Parra, and John O'Donovan. 2016. Moodplay: Interactive mood-based music discovery and recommendation. In *Proceedings of the 2016 conference on user modeling adaptation and personalization.* 275–279. doi:10.1145/2930238.2930280

[6] 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.

[7] Zahra Ashktorab, Michael Desmond, Josh Andres, Michael Muller, Narendra Nath Joshi, Michelle Brachman, Aabhas Sharma, Kristina Brimijoin, Qian Pan, Christine T Wolf, et al. 2021. Ai-assisted human labeling: Batching for efficiency without overreliance. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–27.

[8] Hakan Bagci and Pinar Karagoz. 2016. Context-aware friend recommendation for location based social networks using random walk. In *Proceedings of the 25th international conference companion on world wide web.* 531–536.

[9] Aaron Bangor, Philip Kortum, and James Miller. 2009. Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of usability studies* 4, 3 (2009), 114–123.

[10] Pavel Berkhin. 2005. A survey on PageRank computing. *Internet mathematics* 2, 1 (2005), 73–120.

[11] Jürgen Bernard, Matthias Zeppelzauer, Michael Sedlmair, and Wolfgang Aigner. 2018. VIAL: a unified process for visual interactive labeling. *The Visual Computer* 34, 9 (2018), 1189–1207.

[12] Li Bian and Henry Holtzman. 2011. Online friend recommendation through personality matching and collaborative filtering. *Proc. of UBICOMM* 5, 2011 (2011), 230–235.

[13] Ann Blandford, Dominic Furniss, and Stephann Makri. 2016. *Qualitative HCI research: Going behind the scenes.* Morgan & Claypool Publishers.

[14] 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.* 35–42.

[15] Michelle Brachman, Zahra Ashktorab, Michael Desmond, Evelyn Duesterwald, Casey Dugan, Narendra Nath Joshi, Qian Pan, and Aabhas Sharma. 2022. Reliance and Automation for Human-AI Collaborative Data Labeling Conflict Resolution. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–27.

[16] Cynthia Brame. 2016. Active learning. *Vanderbilt University Center for Teaching* (2016).

[17] Virginia Braun and Victoria Clarke. 2012. *Thematic analysis.* American Psychological Association.

[18] Serena Carpenter. 2010. A study of content diversity in online citizen journalism and online newspaper articles. *New media & society* 12, 7 (2010), 1064–1084.

[19] Shan-Mei Chang and Sunny SJ Lin. 2014. Team knowledge with motivation in a successful MMORPG game team: A case study. *Computers & Education* 73 (2014), 129–140.

[20] Yaomin Chang, Lin Shu, Erxin Du, Chuan Chen, Ziyang Zhang, Zibin Zheng, Yuzhao Huang, and Xingxing Xing. 2022. GraphRR: A multiplex Graph based Reciprocal friend Recommender system with applications on online gaming service. *Knowledge-Based Systems* 251 (2022), 109187. doi:10.1016/j.knosys.2022.109187

[21] Liang Chen, Yuanzhen Xie, Zibin Zheng, Huayou Zheng, and Jingdun Xie. 2020. Friend recommendation based on multi-social graph convolutional network. *IEEE Access* 8 (2020), 43618–43629.

[22] Li Chen, Wei Zeng, and Quan Yuan. 2013. A unified framework for recommending items, groups and friends in social media environment via mutual resource fusion. *Expert Systems with Applications* 40, 8 (2013), 2889–2903.

[23] Shulin Cheng, Bofeng Zhang, Guobing Zou, Mingqing Huang, and Zhu Zhang. 2019. Friend recommendation in social networks based on multi-source information fusion. *International Journal of Machine Learning and Cybernetics* 10 (2019), 1003–1024.

[24] Minsuk Choi, Cheonbok Park, Soyoung Yang, Yonggyu Kim, Jaegul Choo, and Sungsoo Ray Hong. 2019. Aila: Attentive interactive labeling assistant for document classification through attention-based deep neural networks. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–12.

[25] Kajal T Claypool and Mark Claypool. 2007. On frame rate and player performance in first person shooter games. *Multimedia systems* 13, 1 (2007), 3–17.

[26] Sergey N Dorogovtsev, Alexander V Goltsev, and Jose Ferreira F Mendes. 2006. K-core organization of complex networks. *Physical review letters* 96, 4 (2006), 040601. doi:10.1007/978-1-4614-0754-6_9

[27] Gregory Druck, Burr Settles, and Andrew McCallum. 2009. Active learning by labeling features. In *Proceedings of the 2009 conference on Empirical methods in natural language processing*. 81–90. doi:10.3115/1699510.1699522

[28] Adam Duell. 2014. From team play to squad play: The militarisation of interactions in multiplayer FPS video games. *Press Start* 1, 1 (2014), 59–78.

[29] Farzad Eskandanian and Bamshad Mobasher. 2020. Using stable matching to optimize the balance between accuracy and diversity in recommendation. In *Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization*. 71–79.

[30] Boi Faltings, Pearl Pu, Marc Torrens, and Paolo Viappiani. 2004. Designing example-critiquing interaction. In *Proceedings of the 9th international conference on Intelligent user interfaces*. 22–29.

[31] Wenqi Fan, Yao Ma, Qing Li, Yuan He, Eric Zhao, Jiliang Tang, and Dawei Yin. 2019. Graph neural networks for social recommendation. In *The world wide web conference*. 417–426.

[32] Ji Feng, Yang Yu, and Zhi-Hua Zhou. 2018. Multi-layered gradient boosting decision trees. *Advances in neural information processing systems* 31 (2018).

[33] Fethi Fkih. 2022. Similarity measures for Collaborative Filtering-based Recommender Systems: Review and experimental comparison. *Journal of King Saud University-Computer and Information Sciences* 34, 9 (2022), 7645–7669.

[34] Chen Gao, Yu Zheng, Nian Li, Yinfeng Li, Yingrong Qin, Jinghua Piao, Yuhan Quan, Jianxin Chang, Depeng Jin, Xiangnan He, et al. 2023. A survey of graph neural networks for recommender systems: Challenges, methods, and directions. *ACM Transactions on Recommender Systems* 1, 1 (2023), 1–51.

[35] Mark E Glickman. 1995. The glicko system. *Boston University* 16, 8 (1995), 9.

[36] Samuel Gratzl, Alexander Lex, Nils Gehlenborg, Hanspeter Pfister, and Marc Streit. 2013. Lineup: Visual analysis of multi-attribute rankings. *IEEE transactions on visualization and computer graphics* 19, 12 (2013), 2277–2286. doi:10.1109/tvcg.2013.173

[37] Brynjar Gretarsson, John O'Donovan, Svetlin Bostandjiev, Christopher Hall, and Tobias Höllerer. 2010. Smallworlds: visualizing social recommendations. In *Computer graphics forum*, Vol. 29. Wiley Online Library, 833–842.

[38] Aditya Grover and Jure Leskovec. 2016. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*. 855–864.

[39] Sandra G Hart. 2006. NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 50. Sage publications Sage CA: Los Angeles, CA, 904–908.

[40] Pamela J Hinds, Kathleen M Carley, David Krackhardt, and Doug Wholey. 2000. Choosing work group members: Balancing similarity, competence, and familiarity. *Organizational behavior and human decision processes* 81, 2 (2000), 226–251.

[41] Shangrong Huang, Jian Zhang, Lei Wang, and Xian-Sheng Hua. 2015. Social friend recommendation based on multiple network correlation. *IEEE transactions on multimedia* 18, 2 (2015), 287–299.

[42] Ahmet Iscen, Giorgos Tolias, Yannis Avrithis, and Ondrej Chum. 2019. Label propagation for deep semi-supervised learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 5070–5079.

[43] Riitta Jääskeläinen. 2010. Think-aloud protocol. *Handbook of translation studies* 1 (2010), 371–374.

[44] Yucheng Jin, Karsten Seipp, Erik Duval, and Katrien Verbert. 2016. Go with the flow: effects of transparency and user control on targeted advertising using flow charts. In *Proceedings of the international working conference on advanced visual interfaces*. 68–75.

[45] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. 2017. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems* 30 (2017).

[46] Ji Eun Kim, Lingyao Li, and Libby Hemphill. 2024. Communication strategies for improving performance in virtual teams: Lessons from Dota 2. *Authorea Preprints* (2024).

[47] Anastasiia Klimashevskaia, Mehdi Elahi, Dietmar Jannach, Christoph Trattner, and Lars Skjærven. 2022. Mitigating popularity bias in recommendation: Potential and limits of calibration approaches. In *International Workshop on Algorithmic Bias in Search and Recommendation*. Springer, 82–90.

[48] Matevž Kunaver and Tomaž Požrl. 2017. Diversity in recommender systems–A survey. *Knowledge-based systems* 123 (2017), 154–162.

[49] Doron Levit and Anton Tsoy. 2022. A theory of one-size-fits-all recommendations. *American Economic Journal: Microeconomics* 14, 4 (2022), 318–347.

[50] Shanshan Li, Xinzhuan Hu, Jingfeng Guo, Bin Liu, Mingyue Qi, and Yutong Jia. 2024. Popularity-Debiased Graph Self-Supervised for Recommendation. *Electronics* 13, 4 (2024). doi:10.3390/electronics13040677

[51] Xiao Li, Li Sun, Mengjie Ling, and Yan Peng. 2023. A survey of graph neural network based recommendation in social networks. *Neurocomputing* 549 (2023), 126441.

[52] Kai H Lim, Lawrence M Ward, and Izak Benbasat. 1997. An empirical study of computer system learning: Comparison of co-discovery and self-discovery methods. *Information Systems Research* 8, 3 (1997), 254–272.

[53] Paul R Messinger, Xin Ge, Eleni Stroulia, Kelly Lyons, Kristen Smirnov, and Michael Bone. 2008. On the relationship between my avatar and myself. *Journal For Virtual Worlds Research* 1, 2 (2008).

[54] Afshin Moin. 2014. A unified approach to collaborative data visualization. In *Proceedings of the 29th Annual ACM Symposium on Applied Computing*. 280–286.

[55] Ruihui Mu. 2018. A survey of recommender systems based on deep learning. *Ieee Access* 6 (2018), 69009–69022.

[56] Bonnie Nardi and Justin Harris. 2006. Strangers and friends: Collaborative play in World of Warcraft. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*. 149–158.

[57] Sidra Naveed and Jürgen Ziegler. 2020. Featuristic: An interactive hybrid system for generating explainable recommendations-beyond system accuracy.. In *IntRS@ RecSys*. 14–25.

[58] Joseph Noel, Christopher Monterola, and Daniel Stanley Tan. 2024. Improving recommendation diversity without retraining from scratch. *International Journal of Data Science and Analytics* (2024), 1–10.

[59] John O'Donovan, Barry Smyth, Brynjar Gretarsson, Svetlin Bostandjiev, and Tobias Höllerer. 2008. PeerChooser: visual interactive recommendation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1085–1088.

[60] Kazuya Okamoto, Wei Chen, and Xiang-Yang Li. 2008. Ranking of closeness centrality for large-scale social networks. *Lecture Notes in Computer Science* 5059 (2008), 186–195.

[61] Ikewelugo Cyprian Anaene Oyeka, Godday Uwawunkonye Ebuh, et al. 2012. Modified Wilcoxon signed-rank test. *Open Journal of Statistics* 2, 2 (2012), 172–176.

[62] Eli Pariser. 2011. *The filter bubble: What the Internet is hiding from you*. penguin UK.

[63] Yoon-Joo Park and Alexander Tuzhilin. 2008. The long tail of recommender systems and how to leverage it. In *Proceedings of the 2008 ACM conference on Recommender systems*. 11–18.

[64] Denis Parra, Peter Brusilovsky, and Christoph Trattner. 2014. See what you want to see: visual user-driven approach for hybrid recommendation. In *Proceedings of the 19th international conference on Intelligent User Interfaces*. 235–240. doi:10.1145/2557500.2557542

[65] Yulin Peng. 2024. The Application of Machine Learning in Predicting the Results of Popular eSports Games: Win Rate Prediction in MOBA and FPS Games. *Highlights in Science, Engineering and Technology* 85 (2024), 1150–1156.

[66] Muhammad Farrel Pramono, Kevin Renalda, and Harco Leslie Hendric Spits Warnars. 2018. Matchmaking problems in MOBA Games. *Indonesian Journal of Electrical Engineering and Computer Science* 11, 3 (2018), 908–917.

[67] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*. PMLR, 8748–8763.

[68] Fanny Anne Ramirez. 2018. From good associates to true friends: An exploration of friendship practices in massively multiplayer online games. *Social interactions in virtual worlds: An interdisciplinary perspective* (2018), 62–79.

[69] Shaina Raza, Mizanur Rahman, Safiullah Kamawal, Armin Toroghi, Ananya Raval, Farshad Navah, and Amirmohammad Kazemeini. 2024. A Comprehensive Review of Recommender Systems: Transitioning from Theory to Practice. *arXiv preprint arXiv:2407.13699* (2024).

[70] Md Riyadh, Ali Arya, Gerry Chan, and Masud Imran. 2020. Enhancing Social Ties Through Manual Player Matchmaking in Online Multiplayer Games. In

HCI International 2020–Late Breaking Papers: Cognition, Learning and Games: 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings 22. Springer, 708–729. doi:10.1007/978-3-030-60128-7_52

[71] Mike Schaekermann. 2020. Human-ai interaction in the presence of ambiguity: From deliberation-based labeling to ambiguity-aware ai. (2020).

[72] Young-Duk Seo, Young-Gab Kim, Euijong Lee, and Doo-Kwon Baik. 2017. Personalized recommender system based on friendship strength in social network services. *Expert Systems with Applications* 69 (2017), 135–148.

[73] Lei Shi. 2013. Trading-off among accuracy, similarity, diversity, and long-tail: a graph-based recommendation approach. In *Proceedings of the 7th ACM Conference on Recommender Systems.* 57–64.

[74] Ben Shneiderman. 2003. The eyes have it: A task by data type taxonomy for information visualizations. In *The craft of information visualization.* Elsevier, 364–371.

[75] Yunjia Sun, Edward Lank, and Michael Terry. 2017. Label-and-learn: Visualizing the likelihood of machine learning classifier's success during data labeling. In *Proceedings of the 22nd International Conference on Intelligent User Interfaces.* 523–534.

[76] Yun Tian, He Wang, Laixin Xie, Xiaojuan Ma, and Quan Li. 2024. VFLens: Co-design the Modeling Process for Efficient Vertical Federated Learning via Visualization. In *Proceedings of the Tenth International Symposium of Chinese CHI* (<conf-loc>, <city>Guangzhou, China and Online</city>, <country>China</country>, </conf-loc>) *(Chinese CHI '22).* Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3565698.3565765

[77] Chun-Hua Tsai and Peter Brusilovsky. 2017. Leveraging interfaces to improve recommendation diversity. In *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization.* 65–70. doi:10.1145/3099023.3099073

[78] Jukka Vahlo, Johanna K Kaakinen, Suvi K Holm, and Aki Koponen. 2017. Digital game dynamics preferences and player types. *Journal of Computer-Mediated Communication* 22, 2 (2017), 88–103.

[79] Laurens Van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of machine learning research* 9, 11 (2008).

[80] Kellie Vella, Madison Klarkowski, Selen Turkay, and Daniel Johnson. 2020. Making friends in online games: gender differences and designing for greater social connectedness. *Behaviour & Information Technology* 39, 8 (2020), 917–934.

[81] Katrien Verbert, Denis Parra, and Peter Brusilovsky. 2014. The effect of different set-based visualizations on user exploration of recommendations. In *CEUR Workshop Proceedings,* Vol. 1253. University of Pittsburgh, 37–44.

[82] Katrien Verbert, Denis Parra, Peter Brusilovsky, and Erik Duval. 2013. Visualizing recommendations to support exploration, transparency and controllability. In *Proceedings of the 2013 international conference on Intelligent user interfaces.* 351–362.

[83] He Wang, Yang Ouyang, Yuchen Wu, Chang Jiang, Lixia Jin, Yuanwu Cao, and Quan Li. 2024. KMTLabeler: An Interactive Knowledge-Assisted Labeling Tool for Medical Text Classification. *IEEE Transactions on Visualization and Computer Graphics* (2024).

[84] Zhibo Wang, Jilong Liao, Qing Cao, Hairong Qi, and Zhi Wang. 2014. Friendbook: a semantic-based friend recommendation system for social networks. *IEEE transactions on mobile computing* 14, 3 (2014), 538–551. doi:10.1109/tmc.2014.2322373

[85] Michael R Ward. 2022. Network engagement from learning friends' preferences: evidence from a video gaming social network. *Electronic Markets* 32, 3 (2022), 1239–1255.

[86] Bryan Watson, Thomas Watson, and Jun Zheng. 2019. A study of friend recommendations for gaming communities. *International Journal of Web Based Communities* 15, 4 (2019), 292–314. doi:10.1504/ijwbc.2019.10023720

[87] Shiwen Wu, Fei Sun, Wentao Zhang, Xu Xie, and Bin Cui. 2022. Graph neural networks in recommender systems: a survey. *Comput. Surveys* 55, 5 (2022), 1–37.

[88] Ruobing Xie, Qi Liu, Shukai Liu, Ziwei Zhang, Peng Cui, Bo Zhang, and Leyu Lin. 2021. Improving accuracy and diversity in matching of recommendation with diversified preference network. *IEEE Transactions on Big Data* 8, 4 (2021), 955–967. doi:10.1109/tbdata.2021.3103263

[89] Ruobing Xie, Zhijie Qiu, Jun Rao, Yi Liu, Bo Zhang, and Leyu Lin. 2020. Internal and Contextual Attention Network for Cold-start Multi-channel Matching in Recommendation.. In *IJCAI.* 2732–2738. doi:10.24963/ijcai.2020/379

[90] Yunhong Xu, Duanning Zhou, and Jian Ma. 2019. Scholar-friend recommendation in online academic communities: an approach based on heterogeneous network. *Decision Support Systems* 119 (2019), 1–13. doi:10.1016/j.dss.2019.01.004

[91] Liangwei Yang, Zhiwei Liu, Yu Wang, Chen Wang, Ziwei Fan, and Philip S Yu. 2022. Large-scale personalized video game recommendation via social-aware contextualized graph neural network. In *Proceedings of the ACM Web Conference 2022.* 3376–3386.

[92] Yang Ye and Shihao Ji. 2021. Sparse graph attention networks. *IEEE Transactions on Knowledge and Data Engineering* 35, 1 (2021), 905–916.

[93] Kexin Yin, Xiao Fang, Bi-Yi Chen, and Olivia R. Liu Sheng. 2022. Diversity Preference-Aware Link Recommendation for Online Social Networks. *ArXiv* abs/2205.10689 (2022). doi:10.2139/ssrn.4158074

[94] Heyu Zhang, Yan He, Xiaomin Wu, Peixiang Huang, Wenkang Qin, Fan Wang, Juxiang Ye, Xirui Huang, Yanfang Liao, Hang Chen, et al. 2023. PathNarratives: Data annotation for pathological human-AI collaborative diagnosis. *Frontiers in Medicine* 9 (2023), 1070072.

[95] Li Zhang and Jun Guo. 2006. A method for the selection of training samples based on boundary samples. *Journal of Beijing University of Posts and Telecommunications* 29, 4 (2006), 77.

[96] Lin Zhang and Rui Li. 2022. A Large-scale Friend Suggestion Architecture. *ArXiv* abs/2212.12773 (2022). doi:10.1109/dsaa54385.2022.10032355

[97] Lin Zhang and Rui Li. 2022. A Large-scale Friend Suggestion Architecture. *arXiv preprint arXiv:2212.12773* (2022). doi:10.1109/dsaa54385.2022.10032355

[98] Liang Zhang, Quanshen Wei, Lei Zhang, Baojiao Wang, and Wen-Hsien Ho. 2020. Diversity balancing for two-stage collaborative filtering in recommender systems. *Applied Sciences* 10, 4 (2020), 1257.

[99] Shiqi Zhang, Jiachen Sun, Wenqing Lin, Xiaokui Xiao, and Bo Tang. 2022. Measuring Friendship Closeness: A Perspective of Social Identity Theory. *arXiv preprint arXiv:2208.09176* (2022). doi:10.1145/3511808.3557076

[100] Yu Zhang, Yun Wang, Haidong Zhang, Bin Zhu, Siming Chen, and Dongmei Zhang. 2022. Onelabeler: A flexible system for building data labeling tools. In *Proceedings of the 2022 CHI conference on human factors in computing systems.* 1–22.

[101] Yuying Zhao, Yu Wang, Yunchao Liu, Xueqi Cheng, Charu C Aggarwal, and Tyler Derr. 2023. Fairness and diversity in recommender systems: a survey. *ACM Transactions on Intelligent Systems and Technology* (2023).

[102] Wen Zhou and Wenbo Han. 2019. Personalized recommendation via user preference matching. *Information Processing & Management* 56, 3 (2019), 955–968.

A Design Alternative of Preference Projection View
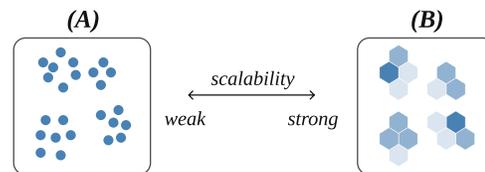

Figure 13: An alternative design of hexbin plot.

An alternative design involves employing a scatter plot (Fig. 13(A)), but it may encounter performance issues when dealing with a large number of players. The rationale behind selecting the hexbin plot (Fig. 13(B)) is its robust scalability, guaranteeing a seamless operation even with a substantial volume of player data.
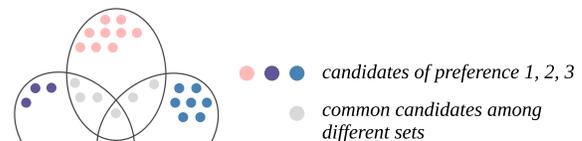
B Design Alternative of LineUp

Figure 14: An alternative design of Lineup View.

We proposed an alternative design for the candidate generation process that combines a Venn diagram and a recommendation list (Fig. 14). The Venn diagram visualizes each set of generated candidates based on a specific aspect, with overlapping areas representing common candidates among different sets. The recommendation list

displays each candidate colored consistently with its corresponding source in the Venn diagram. However, this alternative design can result in a complex color-coding scheme, especially when the number of sets in the Venn diagram is large, making it challenging to efficiently compare and interpret the recommendations. The lineup visualization method allows for more intuitive and efficient inspection and comparison of candidates from various aspects.