AFTER: Adaptive Friend Discovery for Temporal-spatial and Social-aware XR

Bing-Jyue Chen[§]
Computer Science
University of Illinois, Urbana-Champaign
Illinois, United States
bjchen4@illinois.edu

Chiok Yew Ho
Electrical Engineering
National Taiwan University
Taipei, Taiwan
b08901125@ntu.edu.tw

De-Nian Yang
Institute of Information Science
Academia Sinica
Taipei, Taiwan
dnyang@iis.sinica.edu.tw

***Abstract*—Recent advancements in the field of extended reality (XR) have garnered significant interest in XR socialization. However, traditional social XR experiences often fall short of satisfying users' social expectations due to the negligence of the emerging opportunities in XR. In this paper, we propose a novel scenario of socializing in social XR, which has the potential to substantially enhance traditional social media through i) the recommendation of appropriate surrounding users that cater to users' individual preferences, ii) the adaptive avoidance of view occlusions to facilitate users in locating their friends, iii) the consideration of users' social presence, and iv) the development of cross-platform solutions to provide hybrid participation. To this end, we formulate Adaptive Friend Discovery for Temporal-spatial and Social-aware XR, a new NP-hard social recommendation problem aiming at satisfying social XR users. The proposed model, POSHGNN, is a deep temporal graph learning framework designed to provide efficient social recommendations for target users. Experimental results obtained from real-world social XR datasets and a user study that supports multiple XR interfaces demonstrate that the proposed method outperforms baseline approaches with an improvement of 18.5% in solution quality.**

***Index Terms*—extended reality, temporal recommendation, heterogeneous XR**

I. Introduction

As networking, computing, headsets, and new development tools evolve, extended reality (XR) has revolutionized the way people perform daily activities by offering them a shared 3D environment, either a fully virtual world (virtual reality, VR) or an ensemble of the real world and computer-generated contents (mixed reality, MR). The impact of XR extends to several industries, including healthcare [1], tourism [2], manufacturing [3], education [4], and social media [5]. P&S Intelligence [6] forecasts the XR market to reach a value of 1 trillion USD in 2030, demonstrating a compound annual growth rate of 48.3% from 2020 to 2030. Since human beings do not typically live in isolation, the growth of social XR is considered to be an indispensable development for the next paradigm of social media, as evidenced by the numerous social XR applications available today, such as Microsoft Mesh [7], Meta Horizon [8], Mozilla Hubs [9], Snapchat AR [10], VRChat [11], vTime XR [12], and Sansar [13].

Among the various applications of social XR, large-scale XR-based videoconferencing [14]–[16] holds a significant place as it enables crowds to gather and socialize in 3D social events, regardless of their physical locations. Bill Gates expressed his optimism by mentioning "the next two or three years, I predict most virtual meetings will move from 2D camera image grids to the metaverse."[1] For instance, ACM SIGGRAPH 2022 organized a virtual conference with 11700 attendees [17], KDD 2020 hosted a grand VR conference in vFair with 3950 participants [18], and the Global XR conference was held in Microsoft AltspaceVR with more than 1000 attendees [19].

Fig. 1: Illustrative example of a social XR application.

Compared to traditional videoconferencing platforms, social XR can provide additional user experience based on three unique XR features. **F1. Shared 3D environment**. Instead of seeing each other on a flat screen in conventional social media, social XR users can socialize with each other in a shared 3D environment, which is natively supported in various social XR applications. For example, Meta users can interact with others in Horizon via the Oculus headset[2]. **F2. Dynamically adaptive display**. Users and objects in a virtual environment can be rendered or hidden adaptively. This feature is supported in recent social XR applications[3]. For instance, [20] creates VR scenarios that adapt to each user's preferences in real time. **F3. Custom level of immersiveness**. Users may access the social environment in one of two modes (i.e., either a fully virtual environment (VR) or a hybrid environment that includes both real and virtual content (MR)). This feature can be realized by specific headsets or software with the support of heterogeneous user interfaces. Taking the newly released Apple Vision Pro as an example, users are allowed to switch

[§]The work is done in Academia Sinica.

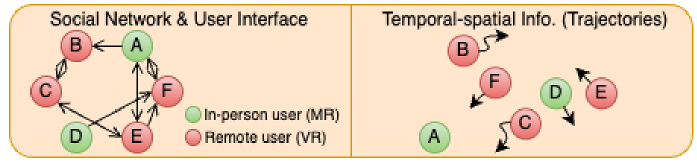
[1]https://bit.ly/3pONn0L
[2]https://bit.ly/43t6lYD
[3]https://bit.ly/3WAX1Pe

between the two modes by scrolling the Digital Crown on the headset[4].

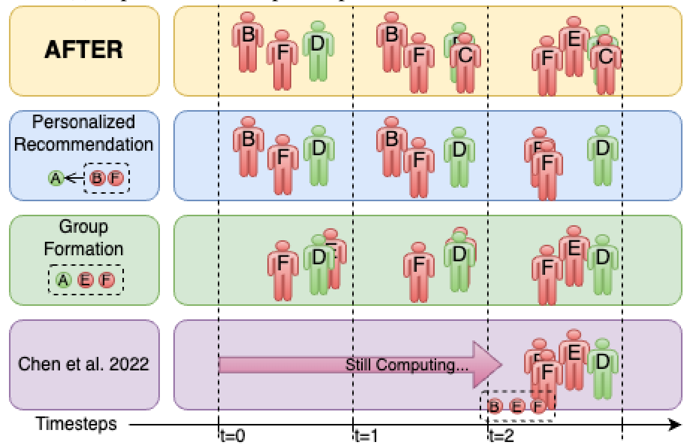
The above three unique features (i.e., **3F**) in social XR can correspondingly bring four untapped opportunities (i.e., **4P**) for XR-based videoconferencing. **P1. Personal Preference (PP).** Similar to traditional social media, social XR users expect fulfillment in personalized interactions. According to [21], 40% of customers are willing to pay more for customizable AR products. Moreover, 80% of VR users believe customizing the display of surrounding users is the most effective way to avoid interaction with irrelevant or offensive individuals [22]. These findings suggest that the displayed 3D scenes for XR users are not necessarily identical and can be tailored based on personal preferences **P2. Occlusion Prevention (OP)** In a crowded and active social XR conferencing room, it is desirable to prevent irrelevant users from blocking the view of target users during social interactions. Adaptive display (i.e., F2) enables users to make other users invisible if they occlude target users' favorite ones [23]. For instance, as depicted in Fig. 1, the MR user cannot see her co-located friends and remote friends unless the adaptive display is enabled. **P3. Social Presence (SP).** Social XR enables individuals to socialize with friends and interested users in a shared 3D environment (F1). According to [24], over two-thirds of users utilize XR for social needs such as meeting people in academic or industrial conferences and staying in contact, all of which promote continual social presence (i.e., the sense of being together). Hence, it is important to ensure users can consistently perceive their friends for socialization in a shared environment. **P4. Hybrid Participation (HP).** Owing to the support of a custom level of immersiveness (F3), social XR can be beneficial for both in-person participants and remote users in a hybrid conference. Specifically, the diverse spectrum of immersiveness enables users to participate in social XR events both physically and virtually, with in-person participants socializing through MR-enabled mobile phones or MR headsets and remote users socializing through VR headsets in a shared 1:1 scale virtual space. According to an XR survey [25], 53% of participants believe XR developers should provide cross-platform compatible content to cater to the diverse needs of all users. Hence, there is an increasing focus on heterogeneous settings of XR [26]–[28] to improve the accessibility of social XR. In Fig. 1, the in-person participant uses MR mode to interact with the girl in white and other remote users.

Motivated by the above three unique features (3F) and four inspiring opportunities (4P) in social XR, we aim to explore a novel scenario for socializing in social XR, where real-time user rendering can be customized adaptively to improve users' overall satisfaction (i.e., be 4P-aware) with the displayed scene. Therefore, we initiate the research by introducing a new temporal user recommendation problem for social XR, namely **AFTER** (*Adaptive Friend Discovery for Temporal-spatial and Social-aware Extended Reality*). The objective of AFTER is to adaptively recommend optimal surrounding users displayed to

[4]https://apple.co/3DfTkXE



(a) Input: social, temporal-spatial, and device information.



(b) Output: user A's view under the respective approach.

Fig. 2: Illustrative example for different approaches.

each user, such that their overall satisfaction can be maximized in the near future, based on their social, temporal-spatial, and device information (Fig. 2a). To assess the efficacy of AFTER, several conventional friend recommenders are evaluated, and their results are compared in Fig. 2b. The figure illustrates the rendered results for user A within consecutive time steps, where all users are remote except for User A and D who are co-located in the same place with MR headsets.

Existing methods might seize some opportunities in **4P**; however, none of them can consider all of them simultaneously. Below we take Fig. 2b as an example. **I. Conventional personalized recommendation [29]–[32]** prioritizes user preference by ranking all users based on their personal preferences and displaying the most preferred users. However, these methods are not designed for co-located XR friends in a 3D environment. Since personally preferred candidates (e.g., idols and celebrities) may not be suitable for socialization, these approaches may deteriorate users' social presence. Take Fig. 2b as an example; although user A's displayed users through personalized recommendation methods (blue shaded area) fulfill her individual preference in terms of her user attributes, user A can never perceive her friend E. **II. Traditional friend grouping approaches [33]–[36]** could improve social presence by assigning N users to k groups ($k << N$) and displaying the members in the same group for each user. While social presence can be improved in such approaches, personal preference may deteriorate. As seen in the green shaded area of Fig. 2b, although users A, E, and F are close friends (i.e., better social presence), A's favorite user may be assigned to other groups (i.e., user B is never recommended for A), which sacrifices users' personal interests. Besides, visibility is not guaranteed as occlusion may still impair the recommendation quality (i.e., the recommended users are occluded by other recommended ones). **III. Chen et al. 2022 [37]** is a method that considers user preference and view occlusion for user recommendations in a virtual world. However, this reinforcement learning-based approach views occlusion as a hard constraint and requires excessive steps to compute friend discovery. Therefore, the proposal cannot recommend users in real time (i.e., **practicality** fails). As shown in the purple shaded area of Fig. 2b, the recommendation at $t = 0$ is calculated after $t = 2$, and thus the results are no longer effective. Even if practicality holds, it fails to consider the continuity of recommendation between consecutive time steps. Consequentially, users are deprived of continual social presence, as this paper unrealistically assumes users' locations are fixed and the corresponding view occlusions are static, which is trivially false. Moreover, all the above approaches have not considered hybrid participation in social XR, making them insufficient in terms of hybrid participation. In Fig. 2b, the irrelevant in-person participant D usually occludes with the recommendation from the above approaches.

The ideal solution for AFTER aims to capitalize on the four opportunities (i.e., **4P**) in social XR by recommending attractive and recognizable users that are compatible with users' viewports. The proposed solution should not only prioritize rendering users based on their personal preferences (**PP**) but also strive to preserve the social presence (**SP**) by continually showing users' friends. Additionally, it should minimize view obstructions (**OP**) and take the impact of heterogeneous XR interfaces (**HP**) into consideration. A comprehensive solution is illustrated in the yellow shaded area (AFTER) of Fig. 2b. It first discovers user B and F at $t = 0$. Then, at $t = 1$, it recommends user C to occlude the irrelevant co-located user D for user A. Finally, at $t = 2$, it finds user A's friend E is not occluded anymore and therefore displays her for user A while simultaneously disabling the ineffective rendering of the occluded user B. During the temporal recommendation process, the ideal solution adaptively recommends proper non-occluded users to cater to user A's personal preference in real time. It also preserves the continuity of recommendation results to ensure user A's social presence with her friends. Besides, it is aware of the co-located participants so that it never recommends users that will be occluded by these participants while also determining better recommendation results to occlude non-interested co-located participants for social XR users.

A proper solution of AFTER can ease users to socialize in crowded XR-based videoconferencing. However, to realize it is a challenging task. Indeed, seizing all four opportunities in real time poses three new challenges, as listed below. **C1. The complexity of multi-modal information.** The challenge is complicated by the coupling of **PP**, **OP**, and **HP**. They not only belong to multi-modal information (PP comes from users' social network, and OP comes from user trajectories) but also intertwine with each other (e.g., while it is improper to recommend remote users that would be occluded (OP) by co-located participants (HP) for MR users, it is feasible to find attractive remote users to occlude the non-interested co-located ones). Therefore, the research challenge remains: how to effectively integrate the multi-modal information. **C2. The dilemma between efficiency and effectiveness.** This challenge emerges when **PP** and **OP** are considered along with **practicality**. Particularly, it is inefficient to maximize personal preference and social presence with the constraint of view occlusion since it is an NP-hard optimization problem [38] (the proof will be provided in Sec. III). Consequently, the solution quality is impossible to be optimal if the model is required to find a solution in real time (i.e., practicality holds). Therefore, the question is how to design a model to recommend users both efficiently and effectively. **C3. The puzzle of continual de-occlusion.** This challenge results from the mutual consideration of **PP**, **OP**, and **SP**. In particular, the optimal solution for the NP-hard problem (**PP** and **OP**) at subsequent time steps may not comprise similar recommended users. Therefore, it leads to terrible social presence for users since their surrounding friends would consistently "flicker" on their viewports in a short period. However, if a continual recommendation is forced, the goal of occlusion prevention may fail. Therefore, the issue is how to properly recommend users and achieve continual de-occlusion.

To comprehensively solve the AFTER problem, we present a new deep temporal graph learning framework called **POSHGNN** (*PP, OP, SP, HP-aware Graph Neural Network*). The neural network comprises three submodules to address the above three challenges, respectively. Firstly (for **C1**), we designed a module called **MIA** (*Multi-modal Information Aggregator*) that can integrate pre-trained user social network embeddings and user trajectories, transforming the target user's information and fusing it into an *attributed dynamic occlusion graph*. This graph encapsulates occlusion relationships and user preferences. Furthermore, the above module can conduct node pruning based on the requirements of the HP (hybrid participation). Following this, the graph can be processed in the subsequent Graph Neural Network (GNN) modules for user recommendation.

Next (for **C2**), upon the fusion of the aforementioned attributed dynamic occlusion graph, our goal is to provide recommendation results at each time step t, striving to balance both efficiency and effectiveness. However, since finding the optimal solution within a short time frame is implausible (NP-hard), we take advantage of the gradual changes in the dynamic graph (as view occlusion does not drastically change within a specific time frame), and progressively resolve and refine the recommendation results at consecutive time steps. Adhering to this philosophy, we designed a GNN module named **PDR** (*Partial View De-occlusion Recommender*) that seeks partial solutions on the occlusion subgraph at time t, and forwards the remaining undetermined portions to time $t+1$ for further resolution. Therefore, the model can efficiently compute the recommendation results within a single time unit, with the recommendation results being a close approximation to the optimal solution, thereby attaining both practicality and effectiveness after a certain period.

Finally (for **C3**), if we aim to seek a balance between the continuity of recommendations for SP and unobstructed views for OP, we must learn to inherit some of the recommendations from time $t-1$ and identify parts that cannot be inherited (e.g., changes in the view occlusion) in resolving the recommended results at time t, so as to avoid occlusion (OP) and maximize individual preferences (PP) at the time step. Based on the aforementioned idea, to introduce the inter-temporal change and past information of dynamic occlusion graph to the overall model, we introduced a GNN module **LWP** (*Learning Which to Preserve*). The module calculates the recommendations to be inherited (selected nodes) based on inter-temporal dynamics, including the previous recommendations and the changes of view occlusion graphs over time. Then, it passes the non-inheritable parts that require additional resolution to the PDR. The module can ensure a certain level of recommendation continuity to guarantee SP, as well as re-examine parts where there are significant differences in the occlusion graph or where the recommendation results are inferior since it has learned how to inherit the recommendations from the previous time steps.

The main contributions of this paper are summarized below.

- We identify key features and opportunities for next-generation videoconferencing in crowded social XR. Accordingly, we introduce a novel NP-hard problem, AFTER, which considers several new factors in the recommendation process to ease the real-time socialization for XR-based videoconferencing users.
- We present a novel graph learning framework POSHGNN to address the three challenging issues in the AFTER problem. We introduce MIA to aggregate multi-modal information, PDR to achieve both solution effectiveness and model efficiency and LWP to balance the continuity of recommendation and quality of de-occlusion.
- We evaluate POSHGNN on three social XR datasets and build an XR user study system to collect real user feedback. Experimental results reveal that POSHGNN outperforms state-of-the-art recommendation algorithms for XR socializing by at least 18.5% in large-scale datasets and 43.9% in the user study in terms of solution quality.

II. Related work

Our work revolves around two lines of research interests: friend recommendation systems for social media users and human psychological effects in social XR environments.

A. Friend recommendation systems

To pinpoint our interest in the proposal, we surveyed numerous studies of user recommendation systems to outline our proposed framework. The main idea of a user recommendation system is to recommend similar or potentially close users to each user in the platform so that social activity can be promoted. We classify user recommendation systems (recommenders) into **personalized ranking-based** recommendations and **grouping-based** recommendations. **Ranking-based** recommenders [29]–[32], [39]–[45] consider user-wise features and give recommendations that aim to satisfy personal preferences. These recommenders are usually implemented in conventional social applications and services by recommending the top-k users/items that the users within the platforms are interested in. [29] modeled a context-aware location-based social network (LBSN) by a graph that contains the context of each user (e.g., social relations, locations, preferences) and proposed a random walk-based ranking of personalized recommendations on a graph. [30] integrated the concept of Bayesian Pairwise Ranking (BPR) with a Convolutional Neural Network (CNN) that captures deep representations from social network data. [42] based their social recommendation with an emphasis on contact strength by the hypothesis that user similarity and contact strength altogether play a vital role in explaining virtual social ties. [31] leveraged on GNNs to implement a multi-faced social recommendation, GraFrank, with multi-modal and user interaction features. [43] proposed a motif-based GNN to realize reciprocal recommendations for online dating recommendation tasks that take the form of bipartite graphs with sparse reciprocal links.

Grouping-based recommenders [33]–[36], [46]–[51] consider users as an aggregate and perform recommendations to fit a social group's mutual interest. Grouping-based social recommendations are usually achieved by clustering users such that the recommendations preserve the maximal co-existence of users within a group. [34] incorporated both content and social interests to perform group recommendations by a group consensus function that improves on suboptimal groupings. [47] leveraged a multi-attention network to extract preference interactions from the representations learned from user groups, such as group co-occurrence and social features. [48] achieved user-wise group recommendations by grouping users of similar interests. [49] performed community detection on user similarity network extracted via user-item collaborative filtering to provide recommendations.

To our best knowledge, [37] is the only work that considers view occlusion when performing recommendations in a 3D social environment. While these works all consider important aspects of digital socialization, they failed to address the challenges simultaneously introduced by currently existing social XR opportunities, taking dynamic occlusion, continual social presence, and hybrid participation as instances, and these discrepancies are further demonstrated and discussed in our experimental study.

B. User study in social XR

Given the advancement and widespread adoption of virtual reality ecosystems, researchers are motivated to investigate extended reality and its psycho-physiological effect on human users as virtually tangible interaction space exists among users in these ecosystems. MR (or AR) users enjoy their social presence based on the similarity and homogeneity of the virtual environment [52]. [53] demonstrated that users in a 3D VR environment exhibit a greater extent of emotional arousal than those in a 2D VR environment. [54] discovered that VR elicits a greater extent of physiological arousal compared to face-to-face (F2F) interaction. [55] pointed out spatial co-presence, transportation, and informational richness as aspects of investigation in the composite framework of asymmetric VR. [56] identified the importance of social presence that can be enhanced by non-verbal cues and self-embodiment within VR space. [26] validated the effectiveness of human pair work in an asymmetric VR-MR interface. [57] demonstrated that 120 frames per second (FPS) is an important threshold for VR. After exceeding the frame rate, users tend to feel lower simulator sickness.

These altogether account for the necessity to propose a social XR recommendation that is inclusive of the aforementioned characteristics, similar to other approaches that aim to enhance immersive virtual environment experiences in various aspects of interaction [27], [58]. Inspired and supported by the aforementioned studies, we hereby present our AFTER framework by considering the above insights.

III. Problem formulation and hardness result

In this section, we first define the notation of this paper and then formally introduce the AFTER problem by connecting the four opportunities (4P) addressed in Sec. I. Existing research [59]–[62] indicates that users' satisfaction in a group activity is influenced by two critical factors: personal preference and social utility, which allow them to interact with their favorite friends or celebrities. Therefore, we first consider both personal preferences and social utilities in the problem formulation below and subsequently incorporate concerns about dynamic view occlusion and heterogeneous interfaces in this section. Table. 1 summarizes the notation table.

TABLE I: Table of notations.

$V_{(\cdot)}$	A user set involving participants in a space of consideration with subsets denoted by (.) indicating the used interface.
N	Number of users in the social XR environment.
$G=(V,E)$	Users' social network with user vertices and social tie edges.
$\mathbf{W}$	Social XR space as a 3D Euclidean space.
τ_t^v	User v's trajectory at time t.
T	Total time steps of a traced trajectory.
$\mathbf{F}_t(\cdot)$	AFTER recommender.
$p(v,w)$	w's preference utility for the target user v.
$s(v,w)$	w's social presence utility for the target user v.
$\mathbb{1}[\text{condition}]$	Indicator function that returns 1 if condition is True, and 0, otherwise.
$v \overset{t}{\Rightarrow} w$	v sees w without view occlusion at time step t
β	User's importance of social presence s relative to personal preference p.
α	The penalty weight of view occlusion in the POSHGNN loss.
$u_t(v,w)$	w's AFTER utility for user v at time t.
$\mathcal{E}_t^v$	Edges in occlusion graphs implying occluded relations for v at time t.
$\mathcal{O}^v=(V,\mathcal{E}^v,T)$	Dynamic occlusion graph for the target user v
$\mathcal{O}_t^v=(V,\mathcal{E}_t^v)$	Static occlusion graph of a time instance of dynamic occlusion graph.
I_t^w	Arc that w occupies at time t in the target user's 360-degree view.
$W(v)$	The weight assigner function for node $v \in V$ in the MWIS problem.
$\mathcal{V}$	Connected and compact geometric objects.
$G_{\mathcal{V}}=(\mathcal{V},E(G_{\mathcal{V}}))$	Geometric intersection graph that include the intersections $E(G_{\mathcal{V}})$ among $\mathcal{V}$.
$\mathbf{p}$	A tensor that includes the target user's preference utilities $\forall w \in V$.
$\mathbf{s}$	A tensor that includes the target user's social presence utilities $\forall w \in V$.
$\hat{\mathbf{p}}_t$ and $\hat{\mathbf{s}}_t$	The normalized $\mathbf{p}$ and $\mathbf{s}$ by users' relative distances with target user at time t.
$\mathbf{x}_t$	The scene features at time t for the target user, which includes $\mathbf{p},\mathbf{s}$, relative distances between target user and all XR users, and their interfaces.
A_t	The adjacency matrix constructed from $\mathcal{E}_t$.
$\mathbf{e}^k$	Differences in the k^{th}-order propagation over all-one vector between A_t and A_{t-1}, where $k \geq 1$.
$\mathbf{m}_t$	Mask for hybrid participation at time t.
σ	The control parameter of preservation gate.
$\mathbf{r}_t$	Recommendation at time t. $\tilde{\mathbf{r}}_t$ denotes prototype recommendation.
$\mathbf{h}_t$	Hidden state of POSHGNN at time t.

A. Problem formulation

Given a user set V including all N participants in an XR-based videoconferencing virtual space, we use subscripts to indicate two subsets in V, which contains MR (i.e., V_{MR}) and VR (i.e., V_{VR}) players. Note that $V = V_{MR} \cup V_{VR}$. Given a social XR space $\mathbf{W}$, a 3D Euclidean space (i.e., $\mathbf{W} = \{(x,y,z) \in \mathbb{R}^3\}$), we leverage both users' social relations and trajectories in the explored social XR scenario to consider both social and temporal-spatial information in AFTER. Users' social networks are represented as $G=(V,E)$, where users form the vertex set V, and edge set E comprises the relations between them. The time label $T \in \mathbb{N}$ indicates all the time steps, and τ represents all users' trajectories, where $\tau_t^v \in \mathbf{W}, \forall v \in V, \forall t \in \{0,1,2,...,T\}$. An AFTER recommender can be defined as follows.

Definition 1 (AFTER recommender)**.** *An AFTER recommender is a function $\mathbf{F}_t(\cdot): V \to 2^V$. Given a target player $v \in V$, it can recommend a set of players to be rendered for v at time step t.*

The simplest approach is to return all surrounding players for any target player v, which is exactly the case in real-world scenarios; nevertheless, it satisfies a little social presence but misses out on the other three opportunities in social XR. Therefore, we require more robust AFTER recommenders. First, since user v's satisfaction with seeing w is related to her **personal preference** (PP), it is desirable to render as more attractive users as possible for v. The strength of the personal preference of user w for v is defined as *preference utility* $p(v,w) \in [0,1]$, which can be estimated from personalized recommenders [31], [32], [40], [44], [45].

Next, social XR players prefer to co-exist with friends when they are interacting in a virtual 3D environment. Accordingly, it is important to ensure that v can see her friends consistently to guarantee her **social presence**. Following [61], [62], we define the *social presence utility* as $s(v,w) \in [0,1]$ when v feels being together with her friend w. For v to perceive social presence with w requires w to continuously exist on v's viewport. Therefore, we acknowledge the necessity to couple the social presence with the recommendation continuity.

Apart from the fulfillment of personal preference and social presence, if w is occluded by others in v's view, the rendering would be ineffective. Therefore, we define the indicator function $\mathbb{1}[v \overset{t}{\Rightarrow} w]$ to be 1 when v can see w clearly at time t (i.e., w is recommended and non-occluded at time t), and 0 otherwise. Note that $\mathbb{1}[v \overset{t}{\Rightarrow} w] = 0$ when $t < 0$ since the XR videoconferencing has not yet been hosted.

Finally, if $v \in V_{MR}$, then she would co-locate with another in-person participant $w' \in V_{MR}$ in the hybrid conference. Even if an AFTER recommender $\mathbf{F}_t(\cdot)$ does not return w', the existence of w' on v's MR viewport is forced since w' is there physically. Therefore, w' may occlude or be occluded by other co-located participants or remote VR users for v, which would affect the visibility indicator function above.

Following [62], [63], which have demonstrated that a combination of both preferential and social factors with user-assigned or learned weights can effectively evaluate user satisfaction in group activities, here we define the overall satisfaction utility for v to see w at time t by integrating her *preference utility* with *social presence utility* as a weighted sum[5], namely *AFTER utility*.

Definition 2 (AFTER utility)**.** *A player v's AFTER utility $u_t(v,w)$ measures v's satisfaction toward the rendered player w at time t. The utility is a weighted combination of her preference utility and social presence utility with $\beta \in [0,1]$.*

$$\begin{aligned} u_t(v,w) &= (1-\beta) \cdot \mathbb{1}[v \overset{t}{\Rightarrow} w] \cdot p(v,w) \\ &\quad + \beta \cdot \mathbb{1}[v \overset{t\text{-}1}{\Longrightarrow} w] \cdot \mathbb{1}[v \overset{t}{\Rightarrow} w] \cdot s(v,w) \end{aligned}$$

The design of $\mathbb{1}[v \overset{\text{t-1}}{\Longrightarrow} w] \cdot \mathbb{1}[v \overset{\text{t}}{\Rightarrow} w]$ in the utility requires v to see w clearly in consecutive time steps to ensure consistent social presence. Below we formally introduce the AFTER problem:

Definition 3 (AFTER problem)**.** *Given a social XR space $\mathbf{W}$, users' social network $G = (V,E)$, their interfaces (i.e., MR or VR), their trajectories $\tau_t, \forall t \in \{0,1,2,...,T\}$, the goal of AFTER is to find an optimal AFTER recommender $\mathbf{F}_t^*$ that maximizes the total AFTER utility for the target user v in subsequent T time steps, which is:*

$$\max_{\{\mathbf{F}_t^*(v)\}} \sum_{t\in\{0,...,T\}} \sum_{w\in\mathbf{F}_t^*(v)} u_t(v,w)$$

[5]An alternative objective could be maximizing social utility while ensuring minimum personal preference. However, this formulation lacks the adaptability to accommodate diverse social XR scenarios. The AFTER formulation offers the advantage of balancing these two competing factors through adjustable trade-off parameters, thereby enabling customization for different scenarios.

AFTER problem is challenging because to recommend proper users for rendering in dynamic social XR, we need to consider individual interests, dynamic view obstruction caused by user mobility, heterogeneous interfaces, and potential social presence with friends. These factors intertwine and pose three challenges **C1-3** introduced in Sec. I.

B. Hardness result

Definition 4 (Dynamic Occlusion Graph (DOG))**.** *A dynamic occlusion graph for the target user v is a graph $\mathcal{O}^v = (V, \mathcal{E}^v, T)$, where vertices in V include all users in the shared XR-videoconferencing environment including v, edges in $\mathcal{E}^v \subseteq \binom{V}{2} \times \{0,1,2,...,T\}$ containing the occluded relations present only in a specified time step from v's perspective, and maximal time label $T \in \mathbb{N}$. The existence of an edge $e \in \mathcal{E}^v$ between two vertices w_1 and w_2 at time t indicates the overlapping between w_1's and w_2's images when they are both rendered on v's view at time t.*

A DOG $\mathcal{O}^v$ can represent a set of T successive static occlusion graphs $\mathcal{O}_t^v = (V, \mathcal{E}_t^v)$, where $\mathcal{E}_t^v = \{\{w_i,w_j\} | (\{w_i,w_j\},t) \in \mathcal{E}^v\}, \forall t \in \{0,1,2,...,T\}$. Consider a reduced case of $T = 0$, its DOG can be viewed as one static occlusion graph.

Given user trajectories τ in social XR, we may construct the corresponding DOG for each target user in social XR via an occlusion graph converter. Without loss of generality, here we provide a simple occlusion graph converter by first considering the environment of social XR is flat (i.e., $\tau_t^w \in \{(x,0,z) \in \mathbb{R}^3\}, \forall w \in V, t \in \{0,1,...,T\}$). Then, given a trajectory of a single time instance τ_t, we can first locate our target user v at a center of a circle and then calculate the radian that each surrounding user w occupies in the target user v's 360-degree view at the time instance t. Arcs $\{I_t^w | \forall w \in V - \{v\}\}$ can be created to form a circular-arc graph (nodes are arcs and an edge between $I_t^{w_i}$ and $I_t^{w_j}$ exists when $I_t^{w_i} \cap I_t^{w_j} \neq \varnothing$) according to the radians corresponding to the circular view angle. Since v's static occlusion graph $\mathcal{O}_v^t$ is exactly the circular-arc graph plus an isolated node v (i.e., herself) at t, we may obtain v's dynamic occlusion graph $\mathcal{O}^v$ by integrating all v's circular-arc graphs from $t = 0$ to $t = T$.

Definition 5 (Maximum Weighted Independent Set (MWIS))**.** *Given an undirected graph $G = (V,E)$ with vertex set V and edge set E, where each vertex $v \in V$ is assigned a weight $W(v) \in \mathbb{R}$, the objective of the Maximum Weighted Independent Set (MWIS) problem is to find a subset $S \subseteq V$ such that no two vertices in S are adjacent (i.e., there is no edge between them), and the sum of the weights of the vertices in S is maximized.*

Definition 6 (Geometric Intersection Graph (GIG))**.** *Let $\mathcal{V}$ be a set of geometric objects in $\mathbb{R}^2$ that are connected and compact. A geometric intersection graph (GIG) $G_{\mathcal{V}}$ is an undirected graph defined as follows: the vertex set of $G_{\mathcal{V}}$ is $\mathcal{V}$, and the edge set $E(G_{\mathcal{V}})$ consists of all pairs of vertices v_i, v_j, where v_i and v_j are elements of $\mathcal{V}$ and their corresponding geometric objects have a non-empty intersection, i.e., $v_i \cap v_j \neq \varnothing$.*

Finding an MWIS on a GIG has been proved to be NP-hard even in the simplest case (i.e., MWIS on a GIG when $w(v) = 1, \forall v \in \mathcal{V}$ and the shape of v is a unit disk) [64], [65].

Lemma 1. *Any GIG can be transformed into a DOG.*

Proof. Given a GIG $G_{\mathcal{V}}$ with geometric objects $\mathcal{V}$ and their intersection relations $E(G_{\mathcal{V}})$, we can view the plane as a panoramic scene for a user u in social XR. Each geometric object w in $\mathcal{V}$ represents a user around the target user v. By setting $T = 0$, we can instantiate a DOG O^v with vertices $\mathcal{V} \cup \{v\}$ and edges $E(G_{\mathcal{V}})$, capturing the occlusion relations between users w_i and w_j within the 360-degree viewport of user v. $\square$

Theorem 1. *AFTER problem is NP-hard.*

Proof. We aim to show that any instance of MWIS problem on a GIG can be transformed into an instance of the AFTER problem. According to Lemma 1, a GIG $G_{\mathcal{V}} = (\mathcal{V}, E(G_{\mathcal{V}}))$ can be transformed into a special DOG denoted as $\mathcal{O}^v$, where v is the target user and $T = 0$. In the context of social XR, we consider the user set V as $\mathcal{V} \cup \{v\}$. The total utility in the AFTER problem can be expressed as:

$$\sum_{t\in\{0\}} \sum_{w\in\mathbf{F}(v)} u_t(v,w) = \sum_{w\in\mathbf{F}(v)} \mathbb{1}[v \overset{\text{t=0}}{\Longrightarrow} w] \cdot (1-\beta) \cdot p(v,w)$$

Let $W(w)$, for all $w \in \mathcal{V}$, be the weight assigned to each node in the MWIS problem. We define $W_{\min} = \min_w W(w)$ and $W_{\max} = \max_w W(w)$. By setting $W'(w) = \frac{W(w)+W_{\min}}{W_{\max}+W_{\min}} = (1-\beta) \cdot p(v,w) \in [0,1]$ for all $w \in \mathcal{V}$, the objective translates into maximizing the sum of the weighted nodes w in $\mathcal{V}$ without overlapping. This objective is equivalent to finding an MWIS on $G_{\mathcal{V}}$ since the occluded relations precisely correspond to the edges in $E(G_{\mathcal{V}})$. Assuming the existence of a polynomial-time algorithm $\mathbf{F}_{poly}$ for the AFTER problem, we can solve MWIS on GIG efficiently by employing $\mathbf{F}_{poly}$. However, since MWIS on GIG is known to be an NP-hard optimization problem, the existence of $\mathbf{F}_{poly}$ would imply that P = NP. $\square$

IV. THE PROPOSED FRAMEWORK: POSHGNN

In this section, we present POSHGNN to address the AFTER problem. To this end, three modules are designed to address each of its corresponding challenges (i.e., **C1-3**): 1) **MIA** (*Multi-modal Information Aggregator*) to prune and fuse users' social, temporal-spatial, and device information into a temporal attributed graph, 2) **PDR** (*Partial View De-occlusion Recommender*) to efficiently discover non-occluded users to cater to the target user's preference and social presence in her partial view, and 3) **LWP** (*Learning Which to Preserve*) to preserve recommendation continuity and ensure model effectiveness. Besides coping with challenges, POSHGNN is also designed to support the opportunities (i.e., **P1-4**) in social XR. Particularly, MIA can prune ineffective candidates that are occluded by co-located in-person participants (**HP**), and integrate the preference utility (**PP**), social presence utility (**SP**), and occlusion graph (**OP**) after preprocessing. Afterward, PDR and LWP serve to maximize AFTER utility in a cooperative manner by recommending some new users and preserving partial recommendations, respectively. This design enables POSHGNN to efficiently and consistently generate effective recommendations for real-time XR-based videoconferencing. The overall POSHGNN framework is illustrated in Fig. 3a.

A. Multi-modal Information Aggregator (MIA)

In the context of social XR, the Multi-modal Information Aggregator (MIA) plays a crucial role in preprocessing and extracting information from the social XR environment. This includes various aspects such as users' social networks, levels of immersiveness, and trajectories. To accomplish this, pre-trained personalized and social recommenders [31], [66] are utilized, along with the occlusion graph converter introduced in Section IIIB.

At each time step t, the raw data is transformed into the input for the POSHGNN model, represented by $\mathbf{x}_t$ and $\mathcal{E}_t$[6]. Here, $\mathcal{E}_t$ denotes the occlusion graph for the target user v, while $\mathbf{x}_t$ encompasses the target user's preference utility $\mathbf{p} \in \mathbb{R}^{|V|}$, social preference utility $\mathbf{s} \in \mathbb{R}^{|V|}$, relative distance, and other users' levels of immersiveness (with 1 denoting MR and 0 denoting VR).[7]

In a naive approach, these multi-modal inputs can be directly propagated to the downstream model, treating them as an attributed graph with $\mathbf{x}_t$ as node embeddings and $\mathcal{E}_t$ as the edge set. However, this approach does not adequately consider the gradual structural changes in consecutive occlusion graphs, thereby hindering the realization of consistent recommendations for the downstream modules. Additionally, it fails to account for users who may participate in XR-based videoconferencing physically or remotely, resulting in unexpected physical occlusion for MR users.

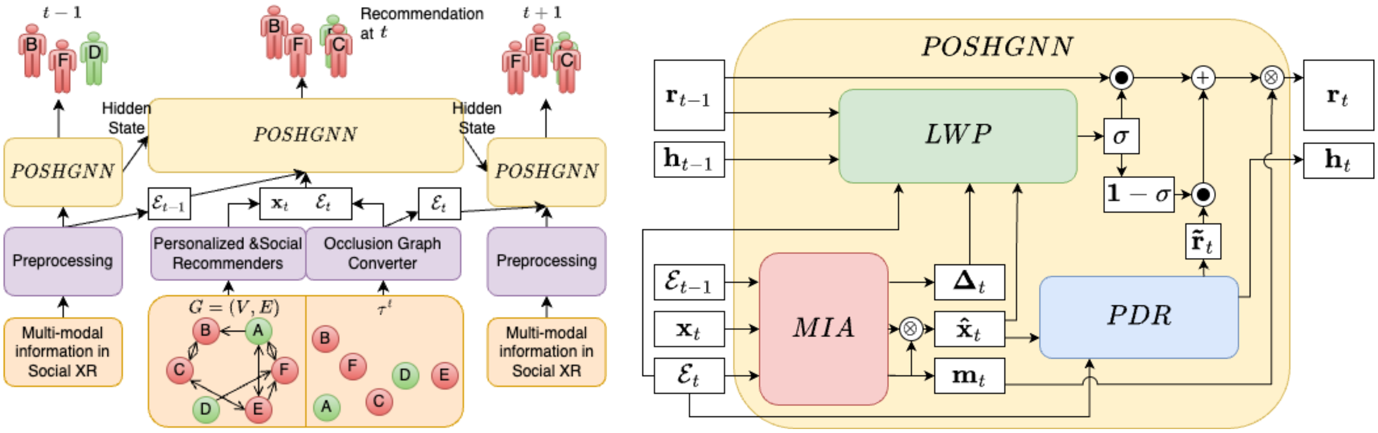
To address these challenges, MIA is designed to calculate the structural differences and prune the physically occluded users. The occlusion graphs from the previous time step $\mathcal{E}_{t-1}$ and the current time step $\mathcal{E}_t$ are transformed into adjacency matrices A_{t-1} and A_t respectively. By leveraging the all-one vector $\mathbf{e^0} \in \{1\}^{|V|}$ and differences in the first-order and second-order propagations, represented by $\mathbf{e^1} = (A_t - A_{t-1}) \cdot \mathbf{e^0}$ and $\mathbf{e^2} = (A_t^2 - A_{t-1}^2) \cdot \mathbf{e^0}$ respectively, a node embedding $\mathbf{\Delta}_t = [\mathbf{e^0}||\mathbf{e^1}||\mathbf{e^2}] \in \mathbb{Z}^{|V|\times 3}$ is created to capture the influence of structural changes for each node, where $||$ denotes the operator of concatenation.

To account for in-person MR-mode participants who may occlude other users, a mask $\mathbf{m}_t$ is applied[8]to select non-occluded users as recommendation candidates for downstream modules. Physically occluded users are pruned by setting the target user v's preference and social presence utility to zero. MIA returns a vector of normalized utilities, denoted as $\hat{\mathbf{x}}_t$, obtained by normalizing the preference and social presence

[6]For simplicity, we discard the superscript v (i.e., our target user) in the following notations.

[7]Note that the tensors $\mathbf{p}$ and $\mathbf{s}$ for the target user v include elements $p(v,w)$ and $s(v,w)$, $\forall w \in V$, respectively.

[8]An inter-user blocklist or allowlist could easily be achieved by a slight modification of the MIA mask (e.g., 0 is for the blocked users).



(a) The overall POSHGNN framework.

(b) Dataflows and submodules in POSHGNN.

Fig. 3: The architecture of POSHGNN.

utility with the square of the current distance between the target user v and other surrounding users $w \in V$. This normalization is crucial to ensure that POSHGNN focuses on preference and social presence rather than the users' relative distance.

In summary, MIA preprocesses and extracts information from the social XR environment to generate input vectors $\mathbf{x}_t$ and occlusion graphs $\mathcal{E}_t$. It distills structural differences, prunes physically occluded users, and provides normalized utility $\hat{\mathbf{x}}_t$ for the downstream POSHGNN model. This process enables effective recommendations in the context of social XR while addressing the challenges posed by occlusion and immersive interactions.

B. Partial view De-occlusion Recommender (PDR)

With the processed user utility vector $\hat{\mathbf{x}}_t$ and the current occlusion graph $\mathcal{E}_t$ (or the corresponding adjacency matrix A_t), the subsequent objective is to recommend appropriate and non-occluded users for the target users v. However, referring to the dilemma stated in **C2**, the model efficiency and effectiveness cannot be both achieved ideally. If a complicated GNN model (e.g., [37]) is adopted here, it is impossible to recommend users in real time. On the other hand, a lightweight GNN finds it difficult to obtain approximated solutions that come close to optimality.

Fortunately, the occlusion graphs in two consecutive time steps do not exhibit dramatic changes in general. Therefore, we can solve the problem by the technique of intertemporal optimization. To ensure model efficiency in every time step, we employ a light two-layer GNN as the Partial view De-occlusion Recommender (PDR) here to perform a de-occlusion recommendation on a partial occlusion graph. PDR not only returns the recommendation $\tilde{\mathbf{r}}_t \in [0,1]^{|V|}$ but also returns its hidden states $\mathbf{h}_t$ (with shape $|V| \times k$, k being a hyperparameter of POSHGNN). The hidden states embody and inherit information reflecting the recommendation uncertainty, which could be leveraged in the next time step $t+1$. Accordingly, given trainable matrices $\mathbf{M}_1^l$ and $\mathbf{M}_2^l$, PDR updates each node w in the l-th layer embedding as follows, where $\mathbf{h}_t^0 = \hat{\mathbf{x}}_t$, $\mathbf{h}_t^1 = \mathbf{h}_t$, $\mathbf{h}_t^2 = \tilde{\mathbf{r}}_t$, and δ is the ReLU activation function.

$$\mathbf{h}_{t,w_i}^{l+1} = \delta(\mathbf{M}_1^l \cdot \mathbf{h}_{t,w_i}^l + \mathbf{M}_2^l \cdot \sum(\{\mathbf{h}_{t,w_j}^l | \forall (w_i, w_j) \in \mathcal{E}_t\})) \quad (1)$$

We may train POSHGNN subsequently under a given recommendation $\tilde{\mathbf{r}}_t$. However, during the training process, $\tilde{\mathbf{r}}_t$ is a probability vector (i.e., $\tilde{\mathbf{r}}_t \in \mathbb{R}^{|V|}$) instead of a 0-1 vector (i.e., $\tilde{\mathbf{r}}_t \in \{0,1\}^{|V|}$) described in Definition 1. Therefore, a modified optimization goal called POSHGNN loss is required to realize neural execution:

Definition 7 (POSHGNN loss)**.** *Given recommendation logits* $\mathbf{r}_t$ *from AFTER recommender, a POSHGNN loss* L *measures the loss of POSHGNN, which approximates the negative temporal mean of the total AFTER utility over time. Note that* $\otimes$ *denotes Hadamard product.*

$$\sum_{t \in \{0,\dots,T\}} L(\hat{\mathbf{p}}_t, \hat{\mathbf{s}}_t, \mathbf{r}_t, \mathbf{r}_{t-1}, A_t) = \sum_{t \in \{0,\dots,T\}} [-(1-\beta) \cdot \mathbf{r}_t \cdot \hat{\mathbf{p}}_t - \beta \cdot \mathbf{r}_t \otimes \mathbf{r}_{t-1} \cdot \hat{\mathbf{s}}_t + \alpha \cdot \mathbf{r}_t^{\intercal} \cdot A_t \cdot \mathbf{r}_t + \gamma]$$

In POSHGNN loss, the first and second terms are used to include the recommendation gains for choosing proper users with good preference and social presence utility. The third term is used to penalize occlusion: if an edge exists in occlusion graph at time t (i.e., w_i and w_j are recommended when $(w_i, w_j) \in \mathcal{E}_t$), the model will be penalized a measure of occlusion (by $\alpha \cdot \mathbf{r}_t^{\intercal} \cdot A_t \cdot \mathbf{r}_t$) with a factor α being a hyperparameter to assign importance. Finally, $\gamma = \sum_w [(1-\beta) \cdot \hat{\mathbf{p}}_t + \beta \cdot \hat{\mathbf{s}}_t]$ is used to ensure the loss is positive.

C. Learning Which to Preserve (LWP)

One of the objectives of POSHGNN is to let the model learn which previously recommended users should be preserved and which ones should be re-considered. Therefore, we present Learning Which to Preserve (LWP) in POSHGNN. The decision-making process of LWP requires three additional pieces of information other than $\hat{\mathbf{x}}_t$ and $\mathcal{E}_t$ in PDR. 1) $\mathbf{\Delta}_t$ from MIA encapsulates node influence from changes in dynamic occlusion relations, hence it is designed to help LWP to determine parts of PDR that require re-evaluation. 2) the previous hidden state $\mathbf{h}_{t-1}$ is utilized to inject awareness of uncertainty in intertemporal recommendations because LWP should know which recommended users are certain for PDR in the previous time step $t-1$ and which ones are not 3) the previous recommendation $\mathbf{r}_{t-1}$ of POSHGNN for preserve consistent social presence. The three additional tensors are then concatenated with $\hat{\mathbf{x}}_t$ as the initial node embedding for LWP. Following similar graph convolution procedure described in Eq. 1, LWP returns a single vector $\sigma \in [0,1]^{|V|}$ for preservation where each value indicates the preserved proportion of the previous recommendations. Specifically, given the previous recommendation $\mathbf{r}_{t-1}$, the current prototype recommendation $\tilde{\mathbf{r}}_t$ from PDR, and the preservation vector σ from LWP, a *preservation gate* is designed to return the final recommendation $\mathbf{r}_t$ of POSHGNN:

$$\mathbf{r}_t = \mathbf{m}_t \otimes [(\mathbf{1}-\sigma) \cdot \tilde{\mathbf{r}}_t + \sigma \cdot \mathbf{r}_{t-1}]$$

With the refined recommendation obtained from LWP, the final recommendation $\tilde{\mathbf{r}}_t$ is replaced by $\mathbf{r}_t$. Therefore, POSHGNN loss (Definition. 7) is used naturally to train itself as the final recommendation remains as a probability vector.

V. Experimental results and user study

A. Experimental setup

Following the settings of [37], [61], [62], [67], we conduct a series of experiments, which will be discussed in detail below. We also release the source code and the complexity analysis in the link: https://github.com/Vimalakirti/POSHGNN.

1) Datasets: In order to establish the validity of different AFTER recommenders, we conduct evaluations on three real-world datasets. The datasets include Timik, SMM, and Hubs. **Timik** [68] is a social metaverse dataset collected from the platform Timik.pl that contains information on 850k users and 12M relationships. **SMM** [69] is a *Nintendo* multi-player game social network, which contains 115k game maps, 880k Super Mario players with their nationalities, photos of 3D avatars, and 7M social interactions, including likes and plays. The **Hub** dataset [70] contains 17k user trajectory points and their social interactions within a VR workshop in Mozilla Hubs, with the inclusion of headset users. Besides, we leverage the RVO2 library [71] to simulate crowd trajectories for Timik and SMM.

2) Baselines: While no exact algorithm is found to be suitable for AFTER problem, we compare the proposed POSHGNN with the following baselines, which include both static and dynamic, learning and algorithmic approaches. The first two methods are of greatest simplicity and least runtime cost albeit neglecting users' social features.

- **Random** randomly selects the surrounding users as the recommended users.
- **Nearest** only considers user locations at time t, and recommends the top-k nearest users.

The second group consists of static approaches, including grouping and personalized recommendation approaches for traditional social media without the consideration of users' temporal-spatial information.

- **MvAGC** (Grouping) [66] leverages graph filtering techniques and refines high-order neighborhood information for user clustering.
- **GraFrank** (Personalized Ranking) [31] adopts multi-faceted social features and aggregates them through a cross-facet attention module for personalized friend recommendation.

The third group contains two recurrent graph neural network kernels, which are capable of performing dynamic friend recommendations as POSHGNN does. For a fair comparison, they are all set to share similar parameters with POSHGNN and are also trained by POSHGNN loss in Definition. 7.

- **DCRNN** [72] captures the spatial dependency using bidirectional random walks on the graph, and the temporal dependency using the encoder-decoder architecture with scheduled sampling.
- **TGCN** [73] leverages a graph convolutional network to learn complex topological structures for capturing spatial dependence and the gated recurrent unit to learn dynamic changes for capturing temporal dependence.

The last baseline Chen et al. 2022 [37] considers maximizing user preferences without any view occlusion within metaverse user recommendations. It performs recommendations at each independent time step, while POSHGNN optimizes recommendation quality throughout inter-temporal optimization.

- **COMURNet** [37] is a reinforcement learning network that exploits an actor-critic structure and several pre-defined actions to perform occlusion-aware recommendations.

3) Evaluation Plan: The implementation of POSHGNN is carried out using the PyTorch framework, and its performance is evaluated on the aforementioned datasets, including Timik, SMM, and Hubs. In addition, to obtain the subjective evaluation of the user's satisfaction, a user study involving 48 participants was conducted using a self-constructed XR-based videoconferencing prototype built on the Unity3D framework.

4) Evaluation Metrics: As the pursuit of opportunities in AFTER recommendations cannot be comprehensively reflected in existing metrics, we adopt the AFTER utility as the primary metric. This metric is calculated and accumulated based on the resultant recommendation matrices generated by our POSHGNN and the baseline algorithms. Besides, we also report the corresponding preference, social presence utility, and the achieved view occlusion rate. Moreover, we demonstrate the running time of POSHGNN and baselines per time step to evaluate and validate their practicality.

5) Implementation Details: The baseline models were trained using the default training parameters and architectures. The experiments were conducted on a Ubuntu server equipped with two Intel Xeon Platinum CPU 2.7GHz, 1TB RAM, and 1 NVIDIA GeForce RTX3060 GPU. The datasets were split into 80% for training and 20% for testing. The PDR and LWP networks are light-weight, which consisted of 2 and 3 layers, each with a hidden dimension of 8, and were trained using the Adam optimizer with a learning rate of $1e^{-2}$. For the datasets SMM and Timik with a 10 square meter virtual conferencing room, we set $T = 100, N = 200, \beta = 0.5, \alpha = 0.01$, and the proportion of VR users as 50% for fair comparisons if not explicitly stated otherwise. Note that β and α can be set based on individuals' preferences in real applications.

B. Quantitative results

1) Comparisons with baselines: The performance of POSHGNN is evaluated and compared against the baselines using the AFTER utility as the prime indicator of recommendation quality. As shown in Tables II, III, and IV, the results demonstrate that POSHGNN outperforms the baselines substantially, with an improvement of 18.5% and 19.1% on the Timik and SMM datasets, respectively. While the performance of POSHGNN on the Hub dataset is only slightly better (0.3%) than the second-best one since only dozens of candidates exist in a Hub conferencing room, POSHGNN still achieves a low view occlusion rate (0.7%), which translates practically into a clearer view for the target users. The differences between the results of POSHGNN and the baselines are statistically significant with a p-value less than 0.05 (p-value ≤ 0.0003).

The simplest method (Random) usually performs worse than all the other baselines in three datasets because target users' preferences and social needs are neglected by randomly displaying players. On the other hand, Nearest delivers optimistic performances compared to the average performance of baselines. The reasons are two-fold. First, the nearer the surrounded players, the more attractive and easier to socialize they usually are (i.e., AFTER utilities 160.6, 190.9, and 295.8 on Timik, SMM, and Hub). Second, the nearest players usually are non-occluded by others for the target users, thereby resulting in a better view occlusion rate than Random.

Although the conventional personalized recommendation approach [31] and grouping-based approach [66] are important for traditional web-based social media, they could not contribute to a comprehensive XR-videoconferencing experience encapsulated in the AFTER utility compared to randomly recommend users since both of these approaches fail to consider spatial information, such as users' trajectories, one of the dominating aspects of POSHGNN where it outperforms them by a margin of at least 25.3%.

Two recurrent GNNs (i.e., TGCN [73] and DCRNN [72]) are also adopted to learn how to reduce the POSHGNN loss. While they usually achieve sub-optimal performances (underlined results in tables) among the compared methods, their performances are unstable across all datasets: uneven emphases are observed across their performances in various datasets, exposing their weaknesses in balancing between occlusion reduction and optimization of preferential and presence utilities. The performance of POSHGNN is also compared with that of COMURNet [37], which seeks to maximize user preference with hard constraints: no view occlusion allowed in the recommendation results. The results demonstrate that POSHGNN outperforms the COMURNet by a margin of at least 60.4%. This is due to the fact that COMURNet, while being effective in providing a clear viewport for users, is limited in its ability to take into consideration both hybrid participation and consistent social presence. In contrast, POSHGNN leverages MIA to summarize the dynamic social environment into an attributed graph, and utilizes PDR and LWP to solve for recommendations that preserve hybrid participation and consistent social presence simultaneously, thereby comprehensively optimizing the users' total satisfaction utility.

Besides, we also report the average view occlusion rate and the running time per time step in the last two rows of Tables II, III, and IV. The best occlusion rate is 0%, attained by COMURNet. However, it cannot optimize the AFTER utility since it neglects hybrid participation, thereby omitting effective recommendations as a result of mandating occlusion-free solutions. Moreover, COMURNet is unpractical for real-world dynamic social XR applications since it requires ~ 22 seconds to perform de-occlusion recommendation at each time step for the Timik and SMM datasets, and 0.4 seconds for the Hub dataset. In contrast, view occlusion is relaxed as a soft constraint in POSHGNN, which allows it to recommend better VR participants to occlude the irrelevant MR users. On top of that, although the view occlusion rate of POSHGNN is greater than that of COMURNet, the light-weight design enables POSHGNN to practically perform recommendations with an update frequency of ~ 150 Hz, which allows users to experience XR environments without a significant negative effect on their experience according to [57].

2) Ablation study: The ablation studies, as depicted in Table. V, evaluate the importance of each module in POSHGNN. These studies involve two variants that aim to assess: 1) the role of the Multi-modal Information Aggregator (MIA) in improving the performance of Partial View de-occlusion Recommender (PDR), and 2) the contribution of the Learning Which to Preserve (LWP) in ensuring the recommendation quality over time. The results demonstrate that MIA enables the downstream PDR network to better leverage information in social XR, which in turn enables the robust and comprehensive encoding of both personal, social, and temporal-spatial information, resulting in the improvement of the overall AFTER utility from 301.7 to 302.3 in the Hub dataset.

Moreover, the results of the study showed that when LWP was removed from POSHGNN (Full), the performance of POSHGNN decreased by 2.0%. It is worth noting that while the PDR with MIA can be used as a means of recommending users from the social XR scene, while also achieving comparable AFTER utility with the full POSHGNN, it could not capture structural changes and perform consistent recommendations. Consequently, the pursuit of optimality in recommendation quality falls apart, which suggests the importance of addressing all challenges in social XR applications to achieve optimal performance.

3) Sensitivity test: Here we conduct sensitivity tests on different settings of social XR environments, including the user number N and the proportion of VR users (i.e., remote participants) in the total users.

TABLE II: Results of POSHGNN and baselines on Timik dataset.

Metrics	POSHGNN (Ours)	Random	Nearest	MvAGC	GraFrank	DCRNN	TGCN	COMURNet
AFTER Utility ↑	**192.5**	104.0	160.6	103.5	101.8	<u>162.4</u>	101.8	113.2
Preference ↑	**183.6**	101.4	152.6	99.2	97.0	<u>154.8</u>	96.8	112.4
Social Presence ↑	**201.2**	130.0	168.6	108.0	106.8	<u>169.8</u>	106.8	112.4
View Occlusion (%) ↓	44.3%	83.5%	74.1%	83.9%	84.3%	92.2%	<u>13.3%</u>	**0.0%**
Running Time (ms) ↓	7.5	**0.075**	<u>0.091</u>	9.0	8.9	10.3	9.7	22195.0

TABLE III: Results of POSHGNN and baselines on SMM dataset.

Metrics	POSHGNN (Ours)	Random	Nearest	MvAGC	GraFrank	DCRNN	TGCN	COMURNet
AFTER Utility ↑	**229.8**	122.8	190.9	120.8	120.8	<u>192.9</u>	120.8	134.2
Preference ↑	**228.6**	123.6	188.4	119.4	119.4	<u>191.2</u>	119.4	138.2
Social Presence ↑	**230.8**	122.2	193.2	122.2	122.2	<u>194.6</u>	122.2	13.0
View Occlusion (%) ↓	<u>52.1%</u>	75.8%	62.1%	84.3%	84.3%	95.0%	75.5%	**0.0%**
Running Time (ms) ↓	6.7	**0.071**	<u>0.083</u>	9.0	10.1	8.0	7.8	21677.3

TABLE IV: Results of POSHGNN and baselines on Hub dataset.

Metrics	POSHGNN (Ours)	Random	Nearest	MvAGC	GraFrank	DCRNN	TGCN	COMURNet
AFTER Utility ↑	**323.6**	220.0	295.8	223.3	258.2	194.0	<u>322.5</u>	201.7
Preference ↑	**321.7**	234.1	293.6	222.0	257.0	192.9	<u>321.2</u>	204.0
Social Presence ↑	**325.4**	205.9	297.9	224.7	259.5	195.0	<u>325.0</u>	199.4
View Occlusion (%) ↓	<u>0.7%</u>	25.9%	6.9%	50.8%	43.2%	27.7%	68.1%	**0.0%**
Running Time (ms) ↓	5.0	**0.045**	<u>0.053</u>	2.5	2.5	4.9	5.0	470.9

TABLE V: Ablation study for POSHGNN on Hub.

Metrics	Full	PDR w/ MIA	Only PDR
AFTER Utility ↑	**308.5**	302.3	301.7
Preference ↑	**306.9**	301.1	299.9
Social Presence ↑	**310.2**	303.5	303.5
View Occlusion (%) ↓	**19.9%**	42.4%	43.7%
Running Time (ms) ↓	8.3	5.1	**2.0**

First, the sensitivity test results on user number N during the training procedure of POSHGNN are shown in Table. VI using the SMM dataset. The greater the number N is, the more choices an AFTER recommender has; however, when a large number of N permits excessive users to exist in the interaction space, tremendous in-person participants will also occlude suitable candidates. In the previous comparisons, N was set as a constant 200, but this test demonstrates the user number in social XR may also affect user satisfaction. As shown in the table, the total AFTER utility reaches its greatest peak 664.7 when $N = 20$. As N increases, the total satisfaction utility decays because excessive in-person players occlude the suitable recommendation candidates. Besides, when $N = 10$, the performance also deteriorates since scarcity of candidates hinders friend discovery in this case. It is worth noting that although the total AFTER utility decreases as N increases, POSHGNN can consistently dominate other baselines, as demonstrated in Tables II, III, and IV.

Next, Table. VII shows the sensitivity test on the proportion of remote (VR) users in the social XR conferencing room. It is reflected that the greater the proportion is, the more flexible an AFTER recommender can perform because no physical (MR) participants would obstruct the recommendation results. Given $N = 200$ sampled from the SMM dataset, we adjust this proportion. Since a higher proportion of VR users allows more attractive users to be discovered in front of our target users, increasing the proportion, thereby resulting in a larger overall AFTER utility. As an empirical demonstration, POSHGNN gains 250.2 AFTER utility when the proportion is 75% and lower scores (229.8 and 214.9) as the proportion decreases.

C. User study

We constructed an XR-based videoconferencing room and conducted a user study with 48 participants to explore our proposed adaptive friend discovery for temporal-spatial and social-aware XR. We set the user number following other user studies of social XR (e.g., 55 users in [74] and 18 users in [75]). Our participants consisted of a diverse range of professionals, including students, government officials, technicians, civil engineers, bankers, and artists in Taipei and Kaohsiung. Prior to the experiment, we collected data on the participants' social networks and preferred β through questionnaires, following the methods in previous studies such as [61], [62], [67]. We assess the satisfaction of 48 participants, including 25 males and 23 females, regarding the adaptive display results recommended by five different methods: POSHGNN, GraFrank, MvAGC, COMURNet, and Rendering All Users (Original). The participants' satisfaction was gathered through Likert scores, ranging from 5 to 1, indicating a level of satisfaction from ”very satisfactory” to ”very unsatisfactory.” Additionally, we evaluate the proposed metrics' effectiveness by examining the correlation between the participants' satisfaction feedback and the metrics. For hybrid participation, the participants accessed the heterogeneous

TABLE VI: Sensitivity test on user number N in social XR, where half of them are MR (in-person) participants.

Metrics	$N = 10$	$N = 20$	$N = 50$	$N = 100$	$N = 200$	$N = 500$
AFTER Utility ↑	431.0	664.7	371.6	260.4	229.8	108.3
Preference ↑	438.3	648.7	370.5	256.4	228.8	108.8
Social Presence ↑	423.8	680.7	372.7	264.4	230.8	107.9
View Occlusion (%) ↓	15.0%	17.7%	39.8%	74.6%	52.1%	34.0%
Running Time (ms) ↓	4.9	5.5	7.1	6.1	8.1	11.1

TABLE VII: Sensitivity test on the proportion of VR users.

Metrics	VR= 75%	VR= 50%	VR= 25%
AFTER Utility ↑	**250.2**	229.8	214.9
Preference ↑	**241.5**	228.8	217.8
Social Presence ↑	**258.9**	230.8	211.9

TABLE VIII: Correlation analysis of utilities.

Correlation	Preference	Social Presence	AFTER util. (satisfaction)
Pearson	0.921	0.912	0.928
Spearman	0.700	0.899	0.700

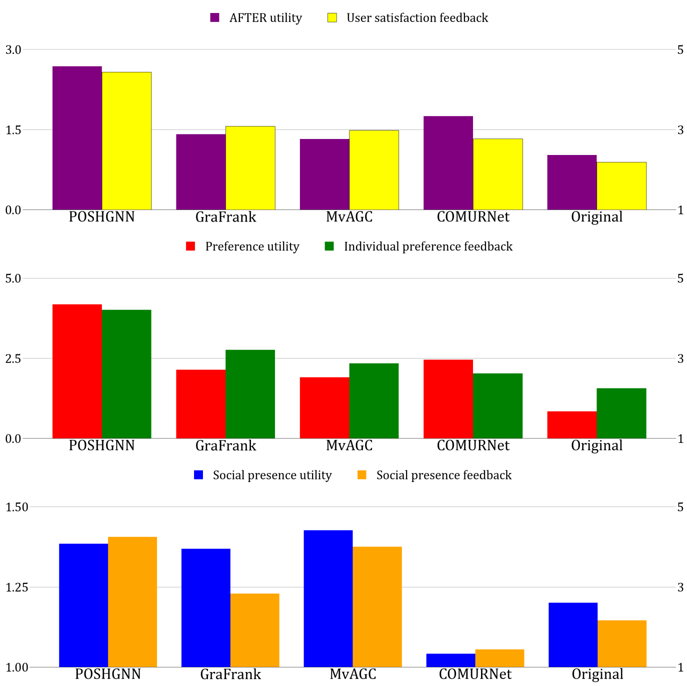


Fig. 4: Utility and user feedback in the user study.

social XR environment through iPhone (MR), or Oculus Quest 2 (VR).

The data on the average utilities per time step and the recorded user feedback are presented in Fig.4. The upper chart in Fig.4 contrasts the overall utility of POSHGNN and the user satisfaction feedback received for each of the five representative methods, namely POSHGNN, GraFrank, MvAGC, COMURNet, and Original. POSHGNN outperforms the baselines with a minimum increment of 53.4% in average AFTER utility and 43.9% in average user satisfaction. It is noteworthy that the differences between AFTER and the other baselines are statistically significant, with a p-value of $\leq 0.004 < 0.05$. The middle chart in Fig.4 depicts the average personal preference utility and the corresponding user feedback regarding the display of their viewports. Although GraFrank (3.208) performs better than MvAGC (2.875) in terms of customization, neither of them takes into account view obstruction. Conversely, while COMURNet can prevent view obstruction, it fails to consider users' continual social presence, resulting in a poor social presence utility (1.422) and user feedback score (2.458). In contrast, POSHGNN, by finding the most non-occluded and appealing users for XR users, attains both good personal preference utility (4.182) and user feedback (4.208). The bottom chart in Fig.4 displays the average social presence utility and the corresponding user feedback on the feeling of being in the company of friends in the heterogeneous social XR. Although MvAGC performs best in terms of social presence utility (1.497), its user feedback (4) is not of superiority to POSHGNN (4.25) since some participants reported that their favorite friends were assigned to different groups, and they were never observed during the experience. Furthermore, the correlations between the defined utilities and the actual user feedback are high, as shown in Table VIII, and are consistent with the findings of previous studies [61], [62]. For example, the Pearson correlation and Spearman correlation between AFTER utility and user satisfaction are 0.928 and 0.700, respectively, indicating that AFTER utility is a reliable estimation of overall satisfaction in the novel social XR recommendation problem. The results of a questionnaire on user opinions towards the new social XR interfaces identified that 89.6% of the participants preferred the use of an adaptive display with social recommendation algorithms, and 100% of the participants believed that support for hybrid participation is crucial for social XR.

VI. Conclusion

Few studies have investigated the dynamic and adaptive display scenarios of social XR videoconferencing and its associated friend recommendation that supports hybrid participation. This paper presents a novel problem, referred to as the AFTER problem, which considers personalization, dynamic view occlusion, and social presence in the context of hybrid social XR. To tackle this problem, we propose a deep temporal GNN framework, POSHGNN, which maximizes users' overall satisfaction in the dynamic non-occlusion scenario. The experimental results indicate that POSHGNN outperforms existing baseline methods by a minimum of 18.5% in terms of achieved utility when evaluated on large-scale real-world datasets.

ACKNOWLEDGMENT

This work is supported in part by MOST under grants 110-2221-E-001-014-MY3, and 109-2221-E-001-017-MY2. We thank to National Center for High-performance Computing (NCHC) of National Applied Research Laboratories (NARLabs) in Taiwan for providing computational and storage resources.

REFERENCES

[1] S. A. Mubin, V. Thiruchelvam, and Y. W. Andrew, "Extended reality: How they incorporated for asd intervention," in *2020 8th International Conference on Information Technology and Multimedia (ICIMU)*, 2020.

[2] N. Feld and B. Weyers, "Mixed reality in asymmetric collaborative environments: A research prototype for virtual city tours," in *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 2021.

[3] S. Doolani, C. Wessels, V. Kanal, C. Sevastopoulos, A. Jaiswal, H. Nambiappan, and F. Makedon, "A review of extended reality (xr) technologies for manufacturing training," *Technologies*, 2020.

[4] A. Vasilchenko, J. Li, B. Ryskeldiev, S. Sarcar, Y. Ochiai, K. Kunze, and I. Radu, "Collaborative learning & co-creation in xr," in *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020.

[5] H. Cabezas, N. Casanova, L. Cortés-Rico, and W. J. Sarmiento, "Home-made xr: Enriching conversations and shortening social distancing," in *2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 2021.

[6] P. Intelligence, "Extended reality (xr) market to generate revenue worth $1,005.9 billion by 2030," https://www.psmarketresearch.com/press-release/extended-reality-market, 2020, accessed: 2022-09-22.

[7] Microsoft, "Introducing microsoft mesh," https://www.microsoft.com/en-us/mesh, 2022, accessed: 2022-09-22.

[8] Meta, "Meta horizon worlds," https://www.oculus.com/horizon-worlds/, 2022, accessed: 2022-01-20.

[9] J. Williamson, J. Li, V. Vinayagamoorthy, D. A. Shamma, and P. Cesar, "Proxemics and social interactions in an instrumented virtual reality workshop," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 2021.

[10] Snap, "Build and share augmented reality for snapchat," https://ar.snap.com/, 2022, accessed: 2022-09-22.

[11] D. Saffo, C. Yildirim, S. Di Bartolomeo, and C. Dunne, "Crowdsourcing virtual reality experiments using vrchat," in *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020.

[12] vTime, "vtime - reality reimagined," https://vtime.net/, 2022, accessed: 2022-09-22.

[13] Sansar, "Sansar - the world's leading social virtual reality platform," https://www.sansar.com/, 2022, accessed: 2022-09-22.

[14] S. N. Gunkel, M. D. Dohmen, H. Stokking, and O. Niamut, "360-degree photo-realistic vr conferencing," in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2019.

[15] A. Abdullah, J. Kolkmeier, V. Lo, and M. Neff, "Videoconference and embodied vr: Communication patterns across task and medium," *Proc. ACM Hum.-Comput. Interact.*, 2021.

[16] S. Shirmohammadi, S.-Y. Hu, W. T. Ooi, G. Schiele, and A. Wacker, "Mixing virtual and physical participation: The future of conference attendance?" in *2012 IEEE International Workshop on Haptic Audio Visual Environments and Games (HAVE 2012) Proceedings*, 2012.

[17] Hubb, "Acm siggraph 2022 virtual conference," https://bit.ly/453nk4L, 2022, accessed: 2022-12-15.

[18] vFair, "Kdd 2020 virtual conference," (https://kdd.org/exploration_files/1._CR._36._kdd_2020_GC_notes-2.pdf, 2020, accessed: 2022-10-07.

[19] A. Meijers, "Global xr conference: The largest global xr community event in the world," https://techcommunity.microsoft.com/t5/mixed-reality-blog/global-xr-conference-the-largest-global-xr-community-event-in/ba-p/3043539, 2021, accessed: 2022-12-15.

[20] J. Kritikos, G. Alevizopoulos, and D. Koutsouris, "Personalized virtual reality human-computer interaction for psychiatric and neurological illnesses: A dynamically adaptive virtual reality environment that changes according to real-time feedback from electrophysiological signal responses," *Frontiers in Human Neuroscience*, 2021.

[21] threekit, "20 augmented reality statistics you should know in 2022," https://www.threekit.com/20-augmented-reality-statistics-you-should-know-in-2020, 2022, accessed: 2022-09-27.

[22] P. VR and T. E. Mind, "Survey of social vr users," https://www.extendedmind.io/2018-survey-of-social-vr-users, 2022, accessed: 2022-09-02.

[23] A. E. M. Wolf, J. Reinhardt, M. Kurzweg, and K. Wolf, "Socialslider: Changing the transparency of avatars," in *Augmented Humans 2022*, 2022.

[24] P. Sykownik, L. Graf, C. Zils, and M. Masuch, "The most social platform ever? a survey about activities & motives of social vr users," in *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, 2021.

[25] P. Coie and X. association, "Xr industry insider: 2021 xr survey," https://www.perkinscoie.com/content/designinteractive/xr2021/, 2021, accessed: 2022-09-27.

[26] J. G. Grandi, H. G. Debarba, and A. Maciel, "Characterizing asymmetric collaborative interactions in virtual and augmented realities," in *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2019.

[27] P. Jansen, F. Fischbach, J. Gugenheimer, E. Stemasov, J. Frommel, and E. Rukzio, "Share: Enabling co-located asymmetric multi-user interaction for augmented reality head-mounted displays," in *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, 2020.

[28] K. M. Ouverson and S. B. Gilbert, "A composite framework of co-located asymmetric virtual reality," *Proc. ACM Hum.-Comput. Interact.*, 2021.

[29] H. Bagci and P. Karagoz, "Context-aware friend recommendation for location based social networks using random walk," in *Proceedings of the 25th International Conference Companion on World Wide Web (WWW '16 Companion)*, 2016.

[30] D. Ding, M. Zhang, S.-Y. Li, J. Tang, X. Chen, and Z.-H. Zhou, "Baydnn: Friend recommendation with bayesian personalized ranking deep neural network," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management (CIKM '17)*, 2017.

[31] A. Sankar, Y. Liu, J. Yu, and N. Shah, "Graph neural networks for friend ranking in large-scale social platforms," in *Proceedings of the Web Conference 2021 (WWW '21)*, 2021.

[32] H. Ning, S. Dhelim, and N. Aung, "Personet: Friend recommendation system based on big-five personality traits and hybrid filtering," in *IEEE Transactions on Computational Social Systems (2019)*, 2019.

[33] Z. Lin and Z. Kang, "Graph filter-based multi-view attributed graph clustering," in *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*, 2021.

[34] M. Gartrell, X. Xing, Q. Lv, A. Beach, R. Han, S. Mishra, and K. Seada, "Enhancing group recommendation by incorporating social relationship interactions," in *Proceedings of the 16th ACM International Conference on Supporting Group Work*, 2010.

[35] D. Cao, X. He, L. Miao, Y. An, C. Yang, and R. Hong, "Attentive group recommendation," in *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, 2018.

[36] Y. Zhao, S. Liang, Z. Ren, J. Ma, E. Yilmaz, and M. de Rijke, "Explainable user clustering in short text streams," in *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2016.

[37] B.-J. Chen and D.-N. Yang, "User recommendation in social metaverse with vr," in *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, 2022.

[38] S. Ahn, Y. Seo, and J. Shin, "Learning what to defer for maximum independent sets," in *Proceedings of the 37th International Conference on Machine Learning*, 2020.

[39] D. Song, D. A. Meyer, and D. Tao, "Efficient latent link recommendation in signed networks," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '15)*, 2015.

[40] Z. Wang, J. Liao, Q. Cao, H. Qi, and Z. Wang, "Friendbook: A semantic-based friend recommendation system for social networks," in *IEEE Transactions on Mobile Computing (2015)*, 2015.

[41] N. B. Silva, I.-R. Tsang, G. D. C. Cavalcanti, and I.-J. Tsang, "A graph-based friend recommendation system using genetic algorithm," in *IEEE Congress on Evolutionary Computation*, 2010.

[42] H.-Y. Liao, K.-Y. Chen, and D.-R. Liu, "Virtual friend recommendations in virtual worlds," in *Decision Support Systems (2015)*, 2015.

[43] ——, "A motif-based graph neural network to reciprocal recommendation for online dating," in *Neural Information Processing*, 2020.
[44] L. Gou, F. You, J. Guo, L. Wu, and X. L. Zhang, "Sfviz: Interest-based friends exploration and recommendation in social networks," in *Proceedings of the 2011 Visual Information Communication - International Symposium (VINCI '11)*, 2011.
[45] M. Jiang, Y. Fang, H. Xie, J. Chong, and M. Meng, "User click prediction for personalized job recommendation," in *World Wide Web (2019)*, 2019.
[46] P.-Z. Li, L. Huang, C.-D. Wang, and J.-H. Lai, "Edmot: An edge enhancement approach for motif-aware community detection," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019.
[47] Z. Huang, X. Xu, H. Zhu, and M. Zhou, "An efficient group recommendation model with multiattention-based neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, 2020.
[48] L. Boratto and S. Carta, "State-of-the-art in group recommendation and new approaches for automatic identification of groups," in *Information Retrieval and Mining in Distributed Environments*, A. Soro, E. Vargiu, G. Armano, and G. Paddeu, Eds., 2011.
[49] C. Cao, Q. Ni, and Y. Zhai, "An improved collaborative filtering recommendation algorithm based on community detection in social networks," in *Proceedings of the 2015 Annual Conference on Genetic and Evolutionary Computation*, 2015.
[50] H. Chen, H. Yin, X. Li, M. Wang, W. Chen, and T. Chen, "People opinion topic model: Opinion based user clustering in social networks," in *Proceedings of the 26th International Conference on World Wide Web Companion*, 2017.
[51] X. Liu and K. Aberer, "Soco: A social network aided context-aware recommender system," in *Proceedings of the 22nd International Conference on World Wide Web*, 2013.
[52] M. R. Miller, H. Jun, F. Herrera, J. Y. Villa, G. Welch, and J. N. Bailenson, "Social interaction in augmented reality," *PLOS ONE*, 2019.
[53] F. Tian, M. Hua, W. Zhang, Y. Li, and X. Yang, "Emotional arousal in 2d versus 3d virtual reality environments," *PLOS ONE*, 2021.
[54] F. Bolinski, A. Etzelmüller, N. A. De Witte, C. van Beurden, G. Debard, B. Bonroy, P. Cuijpers, H. Riper, and A. Kleiboer, "Physiological and self-reported arousal in virtual reality versus face-to-face emotional activation and cognitive restructuring in university students: A crossover experimental study using wearable monitoring," *Behaviour Research and Therapy*, 2021.
[55] K. M. Ouverson and S. B. Gilbert, "A composite framework of co-located asymmetric virtual reality," *Proc. ACM Hum.-Comput. Interact.*, 2021.
[56] A. Yassien, P. ElAgroudy, E. Makled, and S. Abdennadher, "A design space for social presence in vr," in *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*, 2020.
[57] J. Wang, R. Shi, W. Zheng, W. Xie, D. Kao, and H. Liang, "Effect of frame rate on user experience, performance, and simulator sickness in virtual reality," *IEEE Transactions on Visualization and Computer Graphics*, 2023.
[58] J. Lee, M. Kim, and J. Kim, "RoleVR: Multi-experience in immersive virtual reality between co-located HMD and non-HMD users," *Multimedia Tools and Applications*, 2019.
[59] S. Gao, Z. Zhang, S. Su, and M. Zia, "Multi-role event organization in social networks," *Information Sciences*, 03 2018.
[60] X. Wang, W. Zhu, C. Chen, and M. Ester, "Joint user- and event- driven stable social event organization," in *Proceedings of the 2018 World Wide Web Conference*, 2018.
[61] S.-H. Ko, H.-C. Lai, H.-H. Shuai, W.-C. Lee, P. S. Yu, and D.-N. Yang, "Optimizing item and subgroup configurations for social-aware vr shopping," *Proc. VLDB Endow.*, 2020.
[62] H.-C. Lai, H.-H. Shuai, D.-N. Yang, J.-L. Huang, W.-C. Lee, and P. S. Yu, "Social-aware vr configuration recommendation via multi-feedback coupled tensor factorization," in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, ser. CIKM '19, 2019.
[63] X. Wang, W. Zhu, C. Chen, and M. Ester, "Joint user- and event- driven stable social event organization," in *Proceedings of the 2018 World Wide Web Conference*, ser. WWW '18, 2018.
[64] G. Das, G. da Fonseca, and R. Jallu, "Efficient independent set approximation in unit disk graphs," *Discrete Applied Mathematics*, 2018.
[65] B. Clark, C. Colbourn, and D. Johnson, "Unit disk graphs," *Discrete Mathematics*, 1990.
[66] Z. Lin and Z. Kang, "Graph filter-based multi-view attributed graph clustering," in *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, ser. IJCAI '21, 2021.
[67] H.-H. Shuai, C.-Y. Shen, D.-N. Yang, Y.-F. C. Lan, W.-C. Lee, P. S. Yu, and M.-S. Chen, "A comprehensive study on social network mental disorders detection via online social media mining," *IEEE Transactions on Knowledge and Data Engineering*, 2018.
[68] J. Jankowski, R. Michalski, and P. Bródka, "A multilayer network dataset of interaction and influence spreading in a virtual world," *Scientific Data*, 2017.
[69] L. M. P. Moraes and R. L. F. Cordeiro, "Smmnet: A social network of games dataset," in *Brazilian Symposium on Databases (SBBD) 2019 - Dataset Showcase Workshop (DSW)*, 2019.
[70] H. Hecht, R. Welsch, J. Viehoff, and M. R. Longo, "The shape of personal space," *Acta Psychologica*, 2019.
[71] J. Snape, S. J. Guy, J. P. van den Berg, and D. Manocha, "Smooth coordination and navigation for multiple differential-drive robots," in *International Symposium on Experimental Robotics*, 2010.
[72] Y. Li, R. Yu, C. Shahabi, and Y. Liu, "Diffusion convolutional recurrent neural network: Data-driven traffic forecasting," in *International Conference on Learning Representations (ICLR '18)*, 2018.
[73] L. Zhao, Y. Song, C. Zhang, Y. Liu, P. Wang, T. Lin, M. Deng, and H. Li, "T-gcn: A temporal graph convolutional network for traffic prediction," *IEEE Transactions on Intelligent Transportation Systems*, 2020.
[74] M. Murcia-López, T. Collingwoode-Williams, W. Steptoe, R. Schwartz, T. J. Loving, and M. Slater, "Evaluating virtual reality experiences through participant choices," in *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2020.
[75] J. Rantala, J. Kangas, O. Koskinen, T. Nukarinen, and R. Raisamo, "Comparison of controller-based locomotion techniques for visual observation in virtual reality," in *Multimodal Technologies and Interaction*, 2021.