Towards Scalable Metaverse Systems with Social-Aware VR Displays

Shaolong Guo, **Yuntao Wang***, **Qinnan Hu**, **Zhou Su*** and **Tom H. Luan**
School of Cyber Science and Engineering, Xi'an Jiaotong University, Xi'an, China
{shaolong.guo, rangerhu}@stu.xjtu.edu.cn, {yuntao.wang, tom.luan}@xjtu.edu.cn, zhousu@ieee.org

Abstract

The Metaverse is envisioned to support immersive, large-scale social interactions via virtual reality (VR) displays. However, scalability remains a major bottleneck: as the number of concurrent users grows, tracking and updating display content incurs quadratic overhead, often limiting a shared virtual space to only a few dozen participants. Our key observation is that user attention in social VR is highly selective, with users primarily focusing on socially relevant peers rather than all visible users. Motivated by this observation, we propose **SAGE**, a social-aware graph-based VR display framework that enables personalized displays based on inferred social relevance. SAGE introduces a dual-graph learning architecture to jointly model long-term social structures and short-term spatiotemporal co-presence patterns, generating complementary interest scores for display prioritization. Based on these scores, we formulate scalable VR display support as a multi-dimensional resource allocation problem and design a lightweight coordination mechanism with provable guarantees, including incentive compatibility and individual rationality. Experiments on Metaverse datasets show that SAGE improves interaction-relevance prediction by 11.64% and increases social welfare by up to 2.4× compared to state-of-the-art schemes. It scales to support up to 1,000 concurrent users and remains robust against strategic manipulation.

1 Introduction

The Metaverse envisions large-scale, immersive social interaction enabled by virtual reality (VR), where millions of users simultaneously interact, collaborate, and co-create within shared virtual spaces [Wang *et al.*, 2022; Ooi *et al.*, 2023]. In practice, however, today's commercial Metaverse platforms fall significantly short of this vision due to a fundamental scalability bottleneck. To sustain real-time responsiveness, mainstream systems such as Second Life, Fortnite, and Horizon Worlds partition users into isolated instances (i.e., shards), typically capping capacity at fewer than 100 participants [Haeberlen *et al.*, 2023]. As a result, large-scale social interaction is inherently fragmented, preventing the realization of a truly shared virtual space. This limitation stems from the prohibitive computational and communication overhead of continuously synchronizing global states and rendering high-fidelity avatars for dense crowds in real time.

We observe that user attention in the Metaverse is *inherently selective*: individuals rarely interact with or allocate visual attention uniformly to all surrounding entities. Instead, attention is predominantly focused on socially relevant peers, e.g., friends, collaborators, or co-participants in shared virtual activities. This observation motivates *social-aware VR displays*, which prioritize rendering and communication resources toward high-value social interactions, thereby avoiding unnecessary overhead while preserving the immersive experience [Chen *et al.*, 2024; Hsu *et al.*, 2024]. However, the efficacy of such prioritization relies on accurately inferring social relevance from complex and dynamic interaction patterns. Graph-based learning offers a powerful framework for modeling these relational structures, laying the foundation for the social-aware graph-based VR display (SAGE) paradigm.

Despite its promise, deploying SAGE in practical Metaverse systems faces two fundamental challenges. First, user attention is shaped not only by *long-term social relationships* (e.g., friendships), but also by *short-term spatiotemporal co-presence* arising from event-driven interactions (e.g., temporarily co-attending a virtual concert or activity). Such spatiotemporal dynamics introduce significant *volatility* into user interest patterns, as immediate shared events often trigger intense, short-lived interaction demands that can temporarily override long-term social preferences [Hu *et al.*, 2026]. However, existing approaches primarily focus on long-term social structures and overlook transient yet influential co-presence patterns, resulting in inaccurate or delayed interest modeling. Second, rendering resources in Metaverse systems are inherently limited and should be dynamically allocated among users with diverse and time-varying demands. This challenge is further exacerbated by *strategic user behavior*, in which selfish or adversarial users may misreport their preferences to gain disproportionate rendering priority, undermining both system efficiency and fairness.

To address these challenges, we propose SAGE, an adaptive VR display framework for enhancing Metaverse scala-

*Zhou Su and Yuntao Wang are the corresponding authors.

bility. SAGE comprises two tightly coupled components: *i) Social Interests Modeling (SIM)*, which captures both long-term social structures and short-term co-presence dynamics through a novel dual-graph learning architecture; and *ii) Socially Efficient Rendering Scheduler (SERS)*, which leverages the inferred interest scores to allocate multi-dimensional rendering resources through an incentive-compatible mechanism, maximizing social welfare while discouraging strategic manipulation under practical constraints. The main contributions are summarized as follows:

- We propose a novel *social-aware display* paradigm to tackle the scalability bottleneck in Metaverse systems and formulate the scalable VR display problem as a social welfare maximization task under multi-dimensional resource constraints with proved *NP-hardness*.
- To tackle this NP-hard problem, we develop an integrated framework consisting of: i) *SIM*, which prioritizes users' attention by jointly modeling long-term social structures and short-term co-presence patterns; and ii) *SERS*, a lightweight and low-cost rendering scheduler to achieve near-optimal social welfare while satisfying key properties, including *incentive compatibility (IC)* and *individual rationality (IR)*.
- Extensive experiments on Metaverse datasets demonstrate that SAGE outperforms state-of-the-art approaches, improving interest prediction accuracy by 14.18% and achieving up to 2.4× higher social welfare. Moreover, SAGE scales to over 1,000 concurrent users and remains robust against adversarial behavior.

2 Related Works

Collaborative Rendering in Metaverse. To reduce device-side computational burden, collaborative rendering pipelines have become a common solution. Representative works such as Furion [Lai *et al.*, 2020] and Coterie [Meng *et al.*, 2020] offload background rendering to the cloud while handling latency-sensitive foreground tasks locally. Subsequent research has further explored this paradigm through vertical edge-cloud collaboration [Li *et al.*, 2025a], hybrid rasterization-ray tracing techniques [Tan *et al.*, 2022], and novel latency metrics (e.g., Camera-to-Display) [Zhang *et al.*, 2024]. While effective in reducing latency, these systems typically treat all rendering tasks or users equally, limiting scalability in multi-user, interaction-rich environments.

Resource Management in Metaverse. Resource orchestration is critical for maintaining QoE under resource constraints [Su *et al.*, 2022]. Extensive research focuses on optimizing multi-dimensional resources (e.g., CPU frequency, transmission power) through decoupled optimization frameworks or utility-driven allocation schemes [Feng and Zhao, 2025; Jiang *et al.*, 2023; Zhao *et al.*, 2024; Guo *et al.*, 2024]. Furthermore, learning-based methods [Feng *et al.*, 2025; Yu *et al.*, 2024; Du *et al.*, 2023; Liu *et al.*, 2025] have been widely proposed to manage high-dimensional, dynamic orchestration states in real time. However, most approaches ignore social relevance and treat all users equally, leading to unnecessary resource waste and reduced service efficiency.

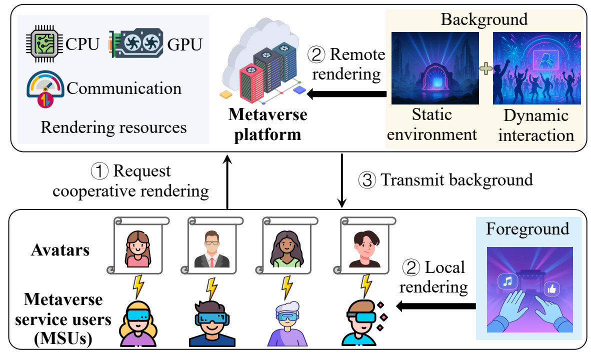


Figure 1: System model of cooperative rendering in Metaverse.

Social-Aware Metaverse Services. Users' social interactions critically influence the Metaverse experience, and various models have been proposed to capture social patterns. COMURNet [Chen and Yang, 2022] uses a dual-GNN to model viewports and mutual occlusions for friend recommendation in social VR. POSHGNN [Chen *et al.*, 2024] constructs spatio-social graphs from location histories and social ties for adaptive social discovery. Other studies either integrate social clustering with collaborative filtering for display recommendation [Li *et al.*, 2025b] or optimize virtual interface layouts based on semantic associations or social engagement [Cheng *et al.*, 2021; Hsu *et al.*, 2024]. Unlike prior work that focuses on user-level rendering or layout optimization, our approach addresses system-level scalability for accommodating multiple users with heterogeneous rendering demands under resource constraints.

3 System Model

As illustrated in Fig. 1, we consider a Metaverse system comprising a set of Metaverse service users (MSUs) $\mathcal{V} = \{1, \ldots, V\}$ and a cloud platform M. The platform orchestrates a finite pool of heterogeneous resources $\mathcal{R} = \{g, c, w\}$ (representing GPU, CPU, and bandwidth). For each resource type $r \in \mathcal{R}$, D^r denotes the total available capacity.

Cooperative Rendering Framework. Due to limited device capabilities, MSUs rely on the platform M for remote rendering. To alleviate device-side computational burden while preserving immersion, we adopt a *cooperative rendering model*. In this model, latency-sensitive foreground elements (e.g., UI elements and local gesture feedback) are rendered on devices, while computationally intensive background content (e.g., environmental details and other avatars' dynamic behaviors) is offloaded to the platform. Unlike static environmental content that can be preloaded, dynamic elements (e.g., avatar movements and interactions) require real-time updates due to their high spatiotemporal variability.

Definition 1 (View-Dependent Display)**.** *A view-dependent display is defined as a directed interaction (i, j) where MSU j lies within the viewport of MSU i. The global interaction set is denoted by $\mathcal{I} = \bigcup_{i \in \mathcal{V}} \{(i, j) \mid j \in \mathcal{I}_i\}$, with $\mathcal{I}_i$ representing the set of peers visible to i.*

Social-Aware Resource Model. For each display $(i, j) \in \mathcal{I}$, the SIM module (Sec. 4.2) infers an *interest score* $a_{i,j} \in$

$[0, 1]$, which quantifies the social relevance of peer j to MSU i. To capture heterogeneous overheads, let $\tilde{\mathbf{d}}_{i,j} = [\tilde{d}^g_{i,j}, \tilde{d}^c_{i,j}, \tilde{d}^w_{i,j}]^\top \in \mathbb{R}^3_+$ denote the full-fidelity resource requirement for display (i, j). We model the interest-aware resource consumption as $\mathbf{d}_{i,j} = a_{i,j}\tilde{\mathbf{d}}_{i,j}$, where each component $d^r_{i,j}$ scales linearly with $a_{i,j}$. It captures the trade-off between service fidelity and resource usage, ensuring high-interest peers receive prioritized rendering quality.

3.1 Utility Model and Problem Formulation

Utilities and Social Welfare. The utility of MSU i is defined as the net benefit obtained from allocated rendering services, i.e., the difference between its private valuation of the received services and the incurred payment:

$$U_i = \sum_{j\in\mathcal{I}_i} \Big(x_{i,j}\, v_i(a_{i,j}\tilde{\mathbf{d}}_{i,j}) - P_{i,j} \Big),$$

where $x_{i,j} \in \{0, 1\}$ indicates whether the resource bundle for interaction (i, j) is assigned. Here, $v_i(\cdot)$ denotes the private valuation of MSU i for the allocated resources, and $P_{i,j}$ is the corresponding payment. The platform utility is defined as total revenue minus resource provisioning costs:

$$U_M = \sum_{(i,j)\in\mathcal{I}} \Big(P_{i,j} - \sum_{r\in\mathcal{R}} \lambda^r a_{i,j}\tilde{d}^r_{i,j} x_{i,j} \Big),$$

where λ^r is the unit cost of resource r [Wang *et al.*, 2023]. Consequently, the system-wide *social welfare* is derived as the aggregate utility of MSU and platform, which is given by

$$W = \sum_{(i,j)\in\mathcal{I}} x_{i,j} \Big(v_i(a_{i,j}\tilde{\mathbf{d}}_{i,j}) - \sum_{r\in\mathcal{R}} \lambda^r a_{i,j}\tilde{d}^r_{i,j} \Big). \tag{1}$$

Problem Formulation. Given the social interest matrix $\mathcal{A}$ inferred by SIM, the platform jointly determines the allocation $\mathcal{X}$ and pricing $\mathcal{P}$ to maximize social welfare while ensuring truthful and voluntary participation:

$$\begin{aligned}
\mathbf{P1}:\ \max_{\mathcal{X},\mathcal{P}}\ & \sum_{(i,j)\in\mathcal{I}} x_{i,j} \Big(v_i(a_{i,j}\tilde{\mathbf{d}}_{i,j}) - \sum_{r\in\mathcal{R}} \lambda^r a_{i,j}\tilde{d}^r_{i,j} \Big) \\
\text{s.t.}\ \ & x_{i,j} v_i(a_{i,j}\tilde{\mathbf{d}}_{i,j}) - P_{i,j} \geq x'_{i,j} v_i(a_{i,j}\tilde{\mathbf{d}}_{i,j}) - P'_{i,j}, \\
& x_{i,j} v_i(a_{i,j}\tilde{\mathbf{d}}_{i,j}) - P_{i,j} \geq 0, \\
& \sum_{(i,j)\in\mathcal{I}} a_{i,j}\tilde{d}^r_{i,j} x_{i,j} \leq D^r, \quad \forall r \in \mathcal{R}, \\
& x_{i,j} \in \{0, 1\}, \quad P_{i,j} \geq 0, \quad \forall (i, j) \in \mathcal{I}.
\end{aligned}$$

Here, $x'_{i,j}$ and $P'_{i,j}$ denote the allocation and payment induced when MSU i misreports its valuation as v'_i. The first two constraints enforce IC and IR, respectively, ensuring that valuation misreporting cannot improve utility and truthful participation yields a non-negative payoff. The third constraint enforces resource feasibility under capacity D^r, and the last constraint specifies the variable domains. Finding the welfare-maximizing allocation in **P1** is NP-hard, as it requires selecting an optimal subset of display requests under multidimensional resource constraints [Luong *et al.*, 2017].

4 SAGE: Socially-Aware Graph-Based VR Display Scheme

4.1 Design Overview

To address problem $\mathbf{P1}$, SAGE determines the allocation $\mathcal{X}$ and pricing $\mathcal{P}$ via two integrated modules, as illustrated in Fig. 2: i) *Social Interests Modeling (SIM)* in Sec. 4.2, which employs a dual-graph learning architecture to capture long-term social topologies and short-term co-presence patterns, generating personalized interest scores; and ii) *Socially Efficient Rendering Scheduler (SERS)* in Sec. 4.3, which leverages these scores to allocate multi-dimensional resources and determine payments via a combinatorial auction mechanism, maximizing social welfare while satisfying IC and IR.

4.2 SIM: Social Interests Modeling

Graph Construction

To enable interest-aware scheduling, the platform constructs two complementary social graphs by integrating user-declared profiles with observed interaction logs. Upon joining the system, each MSU i provides prior social information for constructing an ego-centric graph $\mathcal{G}_i$. These ego graphs are fused with runtime interaction logs to build: a *global static graph* for long-term social structure, and a sequence of *temporal dynamic graphs* for short-term interactions.

- *Global static graph.* The static graph is defined as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where each node i represents an MSU with profile features $\mathbf{n}_i \in \mathbb{R}^{d_n}$. Each edge $e_{ij} \in \mathcal{E}$ denotes a persistent social relation (e.g., friendship or team membership), associated with an edge feature $\mathbf{l}_{ij} \in \mathbb{R}^{d_l}$ capturing tie strength or historical interaction intensity.
- *Temporal dynamic graphs.* The time horizon $\mathcal{T}$ is partitioned into K intervals $\{\Delta_1, \ldots, \Delta_K\}$. For each interval Δ_k, we construct a temporal graph $\mathcal{G}^{(k)} = (\mathcal{V}^{(k)}, \mathcal{E}^{(k)}, \bar{\mathcal{T}}^{(k)})$, where $\mathcal{V}^{(k)}$ denotes active MSUs, each edge $e^{(k)}_{ij}$ represents a co-occurrence event, and $\bar{\mathcal{T}}^{(k)} = \{t^{(k)}_{ij} \mid e^{(k)}_{ij} \in \mathcal{E}^{(k)}\}$ denotes the set of edge-level interaction durations.

Global Social Pattern Modeling

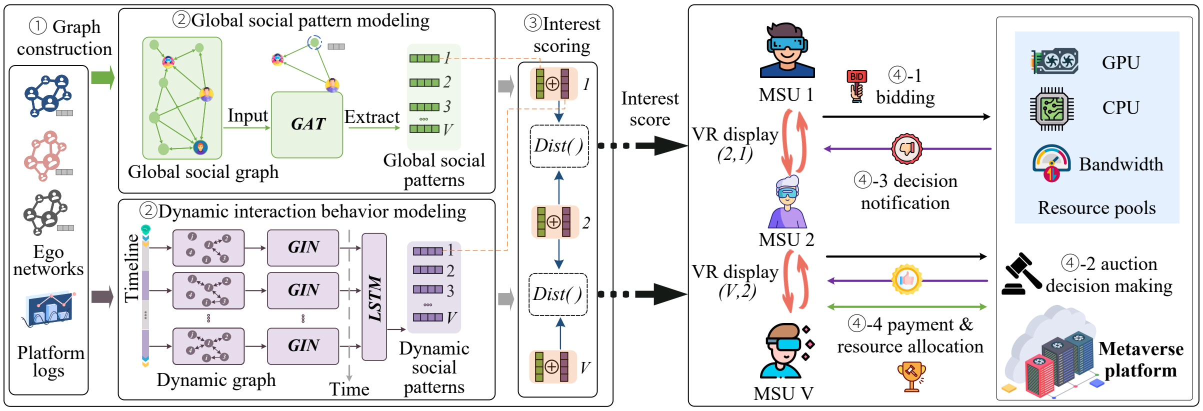
To capture long-term social patterns from a network-level perspective, we employ a social-aware Graph Attention Network (GAT) to propagate socially relevant representations over the global social graph $\mathcal{G}$. Given $\mathcal{G}$, the attention score between MSU i and its neighbor $j \in \mathcal{V}(i)$ is computed as

$$\hat{e}_{ij} = \mathrm{LeakyReLU}\left(\vec{a}_g^\top [W_1\mathbf{n}_i \,\|\, W_1\mathbf{n}_j] + \mathbf{b}_1^\top \phi(\mathbf{l}_{ij}) \right),$$

where $W_1 \in \mathbb{R}^{d'\times d_n}$ is a learnable projection matrix, $\vec{a}_g \in \mathbb{R}^{2d'}$ is the attention vector, $\mathbf{b}_1 \in \mathbb{R}^{d_l}$ models edge influence, and $\phi(\cdot)$ normalizes edge features. Then, the normalized attention weights are obtained via the softmax operation:

$$\alpha_{ij} = \frac{\exp(\hat{e}_{ij})}{\sum_{k\in\mathcal{V}(i)} \exp(\hat{e}_{ik})}.$$

To stabilize learning and capture diverse interaction subspaces, we use multi-head attention with H independent



a) SIM: Social Interests Modeling (§4.2) **b) SERS: Socially Efficient Rendering Scheduler (§4.3)**

Figure 2: The overview of the SAGE framework.

heads. The global social pattern of MSU i is computed as

$$\mathbf{z}_i^{\zeta} = \|_{h=1}^{H} \sigma\left(\sum\nolimits_{j \in \mathcal{V}(i)} \alpha_{ij}^{(h)} W_1^{(h)} \mathbf{n}_j \right),$$

where $\|$ denotes concatenation, $\sigma(\cdot)$ is a nonlinear activation function (e.g., ReLU), and $W_1^{(h)}$ denotes the projection matrix for the h-th head. The resulting embedding $\mathbf{z}_i^{\zeta}$ serves as the global social pattern of MSU i.

Dynamic Social Pattern Modeling

To model short-term social dynamics, we design a temporal encoding module that combines a social-aware Graph Isomorphism Network (GIN) [Xu *et al.*, 2019; Hu *et al.*, 2026] with a Long Short-Term Memory (LSTM) network [Greff *et al.*, 2017]. The GIN captures interaction patterns within each temporal graph, while the LSTM aggregates temporal dependencies across intervals.

Motivated by the observation that longer co-participation in virtual activities can reflect stronger short-term social affinity, we use co-presence duration to quantify social interest. Accordingly, the normalized short-term social interest of MSU i toward MSU j in interval Δ_k is defined as

$$\beta_{ij}^{(k)} = \frac{\exp(t_{ij}^{(k)})}{\sum_{j' \in \mathcal{V}_i^{(k)}} \exp(t_{ij'}^{(k)})},$$

where $t_{ij}^{(k)}$ denotes the co-presence duration and $\mathcal{V}_i^{(k)}$ is the neighbor set of i in $\mathcal{G}^{(k)}$.

To extract the short-term social patterns of MSUs, we incorporate $\beta_{ij}^{(k)}$ into GIN message passing to emphasize duration-intensive interactions:

$$h_i^{(k,l)} = \mathrm{MLP}^{(l)}\left((1+\epsilon^{(l)}) h_i^{(k,l-1)} + \sum_{j \in \mathcal{V}_i^{(k)}} \beta_{ij}^{(k)} h_j^{(k,l-1)} \right),$$

where $\mathrm{MLP}^{(l)}$ is the learnable transformation in the l-th GIN layer, and $\epsilon^{(l)}$ controls the self-representation weight in the l-th neighborhood aggregation. After L layers, the short-term social pattern of MSU i in the interval Δ_k is given by $\mathbf{e}_i^{(k)} = \mathbf{h}_i^{(k,L)}$. Finally, to effectively model long-term temporal dependencies, the sequence of short-term embeddings $\{\mathbf{e}^{(k)}\}_{k=1}^{K}$ is processed by an LSTM network:

$$\{\mathbf{h}^{(k)}, \mathbf{c}^{(k)}\} = \mathrm{LSTM}\left(\mathbf{e}^{(k)}, \mathbf{h}^{(k-1)}, \mathbf{c}^{(k-1)} \right),$$

where $\mathbf{h}^{(k)}$ and $\mathbf{c}^{(k)}$ denote the hidden and cell states at k-th interval, and the final hidden state $\mathbf{z}^{(K)}$ represents the dynamic social pattern of MSUs.

Comprehensive Interests Scoring and Training Objective

To capture both persistent traits and temporally evolving behaviors of each MSU, we construct a comprehensive representation by concatenating each MSU i's global social pattern $\mathbf{z}_i^{\zeta}$ and dynamic social pattern $\mathbf{z}_i^{(K)}$, i.e., $\mathbf{z}_i = [\mathbf{z}_i^{\zeta} \,\|\, \mathbf{z}_i^{(K)}]$. Then, we compute a raw pairwise similarity score:

$$s_{ij} = \cos(\mathbf{z}_i, \mathbf{z}_j) = \frac{\mathbf{z}_i^{\top} \mathbf{z}_j}{\|\mathbf{z}_i\| \cdot \|\mathbf{z}_j\|}.$$

The final interest score is obtained by applying a normalization function, i.e., $a_{ij} = \mathcal{N}(s_{ij})$, where $\mathcal{N} : \mathbb{R} \to (0,1)$ is a monotonic normalization function, and a_{ij} measures the pairwise social relevance between MSUs i and j.

Training loss function. To train SIM, we use a binary interest prediction objective. Given a set of observed links $\mathcal{E}^{+}$ and a set of negative samples $\mathcal{E}^{-}$, the loss function is given by

$$\mathcal{L} = -\sum_{(i,j) \in \mathcal{E}^{+}} \log a_{ij} - \sum_{(i,j) \in \mathcal{E}^{-}} \log(1 - a_{ij}).$$

The negative samples are drawn from non-connected pairs, either uniformly or via hard negative sampling, e.g., from 3- or 4-hop neighborhoods, to improve training discriminability.

4.3 SERS: Socially Efficient Rendering Scheduler

To address the social welfare maximization problem **P1**, we propose SERS, a combinatorial auction mechanism for multi-dimensional resource allocation. Since partial resource allocation (e.g., insufficient bandwidth) severely compromises immersive quality, we adopt the *single-minded bidding* model [Blumrosen and Nisan, 2007]. Under this model, each MSU i submits a single-minded bid for each display (i, j), deriving positive utility only if the *entire* requested bundle $\mathbf{d}_{i,j}$ is allocated ($x_{i,j} = 1$); otherwise, the utility is zero.

Formally, for each (i, j), MSU i declares a unit valuation vector $\mathbf{b}_{i,j} = [b^g_{i,j}, b^c_{i,j}, b^w_{i,j}]^\top$ for the requested bundle $\mathbf{d}_{i,j}$, yielding a total bid $B_{i,j} = \mathbf{b}^\top_{i,j}\mathbf{d}_{i,j}$. The platform incurs a provisioning cost $C_{i,j} = \boldsymbol{\lambda}^\top \mathbf{d}_{i,j}$, where $\boldsymbol{\lambda} = [\lambda^g, \lambda^c, \lambda^w]^\top$ is the unit cost vector. The mechanism is defined as follows.

Definition 2 (Single-Minded Auction)**.** *Given the bid profile $\mathcal{B} = \{B_{i,j}\}_{(i,j)\in\mathcal{I}}$ and platform costs $\mathcal{C} = \{C_{i,j}\}_{(i,j)\in\mathcal{I}}$, the auction mechanism $\mathcal{M}$ is characterized by a tuple $(\mathcal{X}, \mathcal{P})$: i) an* allocation rule *$\mathcal{X}(\mathcal{B}) \to \{0, 1\}^{|\mathcal{I}|}$ that selects a feasible binary allocation $\mathcal{X}$ under capacity constraints; and ii) a* payment rule *$\mathcal{P}(\mathcal{B}) \to \mathbb{R}^{|\mathcal{I}|}_{\geq 0}$ that computes the payment $P_{i,j}$ for each winner.*

To ensure truthfulness and computational efficiency, SERS operates in two phases: 1) *welfare-oriented allocation*, and 2) *critical-value-based pricing*.

Welfare-Oriented Allocation

To prioritize displays with the highest economic efficiency, the platform first calculates the *net declared value* $\varpi_{i,j} = \sum_{r\in\mathcal{R}}(b^r_{i,j} - \lambda^r)d^r_{i,j}$. Candidates with non-positive net values are discarded, as they yield no social surplus, resulting in the feasible set $\mathcal{L} = \{(i, j) \mid \varpi_{i,j} > 0\}$. For each candidate in $\mathcal{L}$, we compute the *net welfare density* $\phi_{i,j}$, which measures the surplus per unit of aggregate resource consumption:

$$\phi_{i,j} = \frac{\varpi_{i,j}}{\sqrt{\sum_{r\in\mathcal{R}} d^r_{i,j}\eta^r}},$$

where $\eta^r > 0$ is a scarcity weight assigned to resource r.

The platform then sorts $\mathcal{L}$ in descending order of $\phi_{i,j}$ and executes an iterative allocation process, starting from the top-ranked display. For each candidate (i, j), the platform verifies whether the request fits within the remaining resources, i.e., checking if $d^r_{i,j} \leq D^r_{\text{res}}$ holds for all $r \in \mathcal{R}$. If feasible, the interaction is admitted (i.e., $x_{i,j} = 1$) and the residual capacity is updated as $D^r_{\text{res}} \leftarrow D^r_{\text{res}} - d^r_{i,j}$. Otherwise, the candidate is skipped due to insufficient resources. This procedure yields the final winner set $\mathcal{W} = \{(i, j) \in \mathcal{L} \mid x_{i,j} = 1\}$.

Pricing via Critical Value

We leverage a *critical-value-based pricing* rule to determine each winner's payment, defined as follows.

Definition 3 (Critical Value)**.** *For a winning interaction $(i, j) \in \mathcal{W}$, the* critical value *$\chi^{\text{crit}}_{i,j}$ is defined as the minimum total declared bid that MSU i should submit for its bundle $\mathbf{d}_{i,j}$ to be a winner, given that all other bids are fixed. Formally,*

$$\chi^{\text{crit}}_{i,j} = \min\{B'_{i,j} \mid x_{i,j}(B'_{i,j}, \mathbf{d}_{i,j}; \mathcal{B}_{-(i,j)}) = 1\},$$
$$\text{s.t.} \quad B'_{i,j} < \chi^{\text{crit}}_{i,j} \Rightarrow x_{i,j}(B'_{i,j}, \mathbf{d}_{i,j}; \mathcal{B}_{-(i,j)}) = 0,$$

where $\mathcal{B}_{-(i,j)}$ denotes the bid profile of all other interactions.

To compute $\chi^{\text{crit}}_{i,j}$, the platform identifies the *critical competitor* (i', j'): the first bidder in the sorted list after (i, j) that would be allocated when (i, j) was removed, but is denied due to (i, j)'s presence. If such a critical competitor exists, the payment for (i, j) is derived from (i', j')'s declared bid and cost structure; otherwise, the critical value is set to the platform's provisioning cost for the bundle. Specifically, the payment charged to interaction (i, j) is given by

$$P_{i,j} = \begin{cases} \varpi_{i',j'} \dfrac{\sqrt{\sum_r \eta^r d^r_{i,j}}}{\sqrt{\sum_r \eta^r d^r_{i',j'}}} + \sum_r \lambda^r d^r_{i,j}, & \text{if } x_{i,j} = 1 \text{ and } \exists\, (i', j'), \\ \sum_r \lambda^r d^r_{i,j}, & \text{if } x_{i,j} = 1 \text{ and } \nexists\, (i', j'), \\ 0, & \text{otherwise.} \end{cases}$$

Note that $P_{i,j} \geq \sum_{r\in\mathcal{R}} \lambda^r d^r_{i,j}$ for all winners, ensuring non-negative profit for the platform on each allocated bundle.

Remark. The above rule implements critical-value pricing: each winner pays the minimum total bid that guarantees allocation, which equals the platform cost plus the minimum net surplus needed to displace its tightest competitor. It ensures truthfulness while guaranteeing non-negative platform profit.

Algorithm 1: Workflow of **SERS**

```
Input: I; D; B; λ; η; D.
Output: X; P.
/* MSU: Bidding                                          */
foreach MSU i ∈ V and peer j ∈ I_i do
    Submit scalar bid B_{i,j} for bundle d_{i,j}
/* Platform:  Allocation & Pricing                       */
Initialize: L ← ∅, X ← 0, D_res ← D
foreach interaction (i, j) ∈ I do
    Calculate net value: ϖ_{i,j} ← B_{i,j} − λ^⊤ d_{i,j}
    if ϖ_{i,j} > 0 then
        ρ_{i,j} ← sqrt(d_{i,j}^⊤ η), φ_{i,j} ← ϖ_{i,j}/ρ_{i,j}
        Add to list: L ← L ∪ {((i, j), φ_{i,j})}
Sort L descending by φ_{i,j} (indexed as 1 . . . L)
foreach candidate ((i, j), φ_{i,j}) ∈ L do
    if d_{i,j} ⪯ D_res then
        D_res ← D_res − d_{i,j} , x_{i,j} ← 1
P ← CriticalPricing(X, L, D, D, λ, ϖ, ρ)
return X, P
```

Algorithm Design

Algorithm 1 outlines the SERS workflow. First, MSUs submit bids for their desired interactions. The platform then filters candidates by net value, sorts the remainder based on welfare density $\phi_{i,j}$, and executes allocation to sequentially admit candidates that satisfy residual capacity constraints. Finally, the CriticalPricing procedure (Algorithm 2) determines payments.

Property Analysis

Next, we formally analyze key economic properties of SERS.

Theorem 1 (Truthfulness)**.** *Under the SERS mechanism, truthfully reporting the private valuation $v_{i,j}$ (i.e., submitting a bid $B_{i,j}$ derived from $v_{i,j}$) is a dominant strategy.*

Proof. We prove truthfulness by verifying the four sufficient conditions for single-minded bidders as established in [Lehmann *et al.*, 2002]. *i) Exactness.* For each display (i, j), SERS either allocates the entire requested bundle $\mathbf{d}_{i,j}$

Algorithm 2: CRITICALPRICING

Input: $\mathcal{X}$; $\mathcal{L}$ (size L); $\mathcal{D}$; $\mathbf{D}$; $\boldsymbol{\lambda}$; $\boldsymbol{\varpi}$; $\boldsymbol{\rho}$.
Output: $\mathcal{P}$.

Initialize $\mathcal{P} \leftarrow \mathbf{0}$
foreach *winner* (i, j) *where* $x_{i,j} = 1$ **do**
  Let ℓ be the rank index of (i, j) in $\mathcal{L}$
  Initialize: $\mathbf{D}' \leftarrow \mathbf{D}$, $\texttt{crit} \leftarrow$ null
  for $t = 1$ **to** L **do**
    if $t == \ell$ **then**
      continue `// Skip self & next iter`
    Let (k, l) be the candidate at rank t
    if $\mathbf{d}_{k,l} \preceq \mathbf{D}'$ **then**
      `// Identify critical bidder`
      if $x_{k,l} == 0$ **then**
        $\texttt{crit} \leftarrow (k, l)$; **break**
      $\mathbf{D}' \leftarrow \mathbf{D}' - \mathbf{d}_{k,l}$
  $P_{\text{base}} \leftarrow \boldsymbol{\lambda}^\top \mathbf{d}_{i,j}$
  if $\texttt{crit} \neq$ null **then**
    Let $(i', j') \leftarrow \texttt{crit}$
    $P_{i,j} \leftarrow P_{\text{base}} + \varpi_{i',j'}(\rho_{i,j}/\rho_{i',j'})$
  else
    $P_{i,j} \leftarrow P_{\text{base}}$
return $\mathcal{P}$

or nothing. Partial allocations are disallowed, satisfying the exactness requirement. *ii) Monotonicity.* For a fixed bundle $\mathbf{d}_{i,j}$, the sorting metric is the net welfare density $\phi_{i,j}$. Since the cost $C_{i,j} = \sum_r \lambda^r d^r_{i,j}$ and $\rho_{i,j}$ are constants from the bidder's perspective, $\phi_{i,j}$ is strictly increasing in the total declared bid $B_{i,j}$. Therefore, increasing $B_{i,j}$ or decreasing the resource size cannot cause a winning bidder to lose, which establishes monotonicity. *iii) Critical-value payment.* For each winning bidder (i, j), the payment is set to its critical value $\chi^{\text{crit}}_{i,j}$, defined as the minimum of total declared bids that secure allocation. As shown in Sec. 4.3, this value is computed via the critical competitor (i', j') and equals

$$\chi^{\text{crit}}_{i,j} = \left(\frac{B_{i',j'} - C_{i',j'}}{\rho_{i',j'}} \right) \rho_{i,j} + C_{i,j},$$

where $\rho_{i,j} = \sqrt{\sum_{r \in \mathcal{R}} d^r_{i,j}\, \eta^r}$ denotes the weighted resource norm of the bundle (i, j). $\chi^{\text{crit}}_{i,j}$ is precisely the minimum total declared bid required to maintain allocation under the greedy density rule. Hence, the payment rule satisfies the critical-value condition. *iv) Individual rationality.* Losers are charged zero and obtain zero utility. For each winner, since it wins with bid $B_{i,j}$, its critical payment satisfies $P_{i,j} = \chi^{\text{crit}}_{i,j} \leq B_{i,j}$. Under truthful bidding, $B_{i,j} = v_i(\mathbf{d}_{i,j})$, and thus

$$U_{i,j} = v_i(\mathbf{d}_{i,j}) - P_{i,j} \geq v_i(\mathbf{d}_{i,j}) - B_{i,j} = 0,$$

which ensures voluntary participation. Hence, all four properties are satisfied. According to Theorem 9.6 in [Lehmann *et al.*, 2002], the proposed SERS mechanism is truthful. ☐

Remark. Theorem 1 establishes two key economic properties: *1) SERS satisfies IC*: for each display request, reporting the true valuation is the best strategy for the corresponding MSU. *2) SERS satisfies IR*: winners pay no more than their truthful valuations, and losers pay zero.

Dataset	Source	Type	#Nodes	#Edges
D1	Simulated VRChat[1]	Temporal	1000	8743
D2	Copenhagen Study[2]	Temporal	800	34362

Table 1: Statistics of the experimental datasets.

Scheme	D1			D2		
	AUC	AP	F1 Score	AUC	AP	F1 Score
GCN	0.861	0.847	0.788	0.820	0.816	0.742
GAT	0.851	0.836	0.789	0.837	0.831	0.761
GIN	0.805	0.795	0.725	0.703	0.701	0.634
GRU	0.744	0.693	0.704	0.799	0.787	0.728
LSTM	0.740	0.686	0.707	0.756	0.752	0.677
Ours	**0.940**	**0.931**	**0.891**	**0.955**	**0.953**	**0.888**

Table 2: Performance comparison between baselines and our SIM.

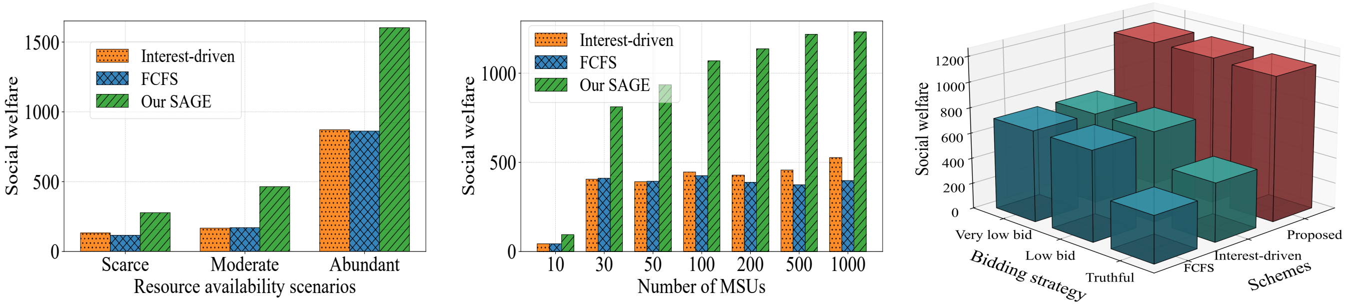
5 Experimental Evaluation

5.1 Evaluation Setup

Datasets & Settings. We conduct experiments on two temporal social interaction datasets, as summarized in Table 1. We consider a centralized Metaverse platform serving from 10 to 1,000 MSUs. The platform manages a multi-dimensional resource pool with capacities of 2,000 *pixels/ms* (GPU), 1,000 *cycles/ms* (CPU), and 1,500 *bits/ms* (bandwidth). For GPU, CPU, and bandwidth, the resource importance weights are set to 0.4, 0.2, and 0.4, and the unit costs are set to 0.05, 0.02, and 0.03. Each display request renders a virtual object with a normalized surface size of 500 pixels. MSUs' unit valuations for GPU, CPU, and bandwidth are independently sampled from uniform distributions $[0.1, 0.5]$, $[0.05, 0.2]$, and $[0.06, 0.25]$, respectively. We evaluate three valuation-reporting strategies: *truthful bidding*, *low bidding* with 20% underreporting, and *very low bidding* with 50% underreporting. All graph-based models are implemented using PyTorch Geometric and trained with the Adam optimizer for 100 epochs, using a learning rate of 1×10^{-2}.

Comparative Methods. We evaluate SAGE against representative baselines, spanning graph modeling paradigms for interest prediction and rendering decision strategies for resource allocation. *1) Static graph models* include GCN [Kipf and Welling, 2017], GIN [Xu *et al.*, 2019], and GAT [Veličković *et al.*, 2018]. These models operate on static graph topologies without temporal modeling, making them inadequate for capturing the dynamic interaction patterns among MSUs. *2) Dynamic models* include LSTM [Greff *et al.*, 2017] and GRU [Cho *et al.*, 2014]. These models capture temporal dependencies or temporal graph dynamics, yet lack a unified treatment of persistent social ties and short-term co-presence patterns. *3) Interest-driven rendering* [Hsu *et al.*, 2024] ranks display requests by social interest and admits them in descending order until the resource capacity is

[1] Based on prior knowledge in [Steinert *et al.*, 2024], we simulate social interactions in the VRChat Metaverse to construct D1.
[2] Copenhagen Networks Study [Sapiezynski *et al.*, 2019].



(a) Effect of resource availability
(b) Effect of network scale
(c) Effect of bidding behavior

Figure 3: Social welfare comparison of different schemes under varying experimental conditions.

SIM Configuration	AUC	AP	F1 Score
w/o Static Branch	0.873	0.875	0.708
w/o Dynamic Branch	0.935	0.936	0.810
Full SIM Module	**0.966**	**0.963**	**0.894**

Table 3: Ablation study of the SIM in predicting social interest.

exhausted. *4) First-Come-First-Served (FCFS)* sequentially admits display requests based on arrival order, subject to remaining resource capacity.

Evaluation Metrics. We evaluate the interest prediction performance of SIM using standard evaluation metrics, including Area Under the ROC Curve (AUC), Average Precision (AP), and F1 Score. The adaptive rendering component, i.e., SERS, is evaluated using social welfare, which measures the aggregate utility of MSUs and the platform under resource constraints, as defined in Eq. (1).

5.2 Experimental Results

Performance of Interest Prediction. We evaluate the efficacy of SIM in predicting social interest among MSUs on D1 and D2. As detailed in Table 2, SIM consistently achieves the best performance across all metrics. Compared with the best-performing baseline on each dataset, SIM improves AUC, AP, and F1 Score by 11.64%, 12.30%, and 14.81% on average, respectively. These gains suggest the effectiveness of SIM's dual-graph learning architecture, which jointly models long-term social topology and short-term co-presence patterns for context-aware interest inference.

Ablation Study of SIM Module. We conduct an ablation study to assess the contribution of the static and dynamic branches in SIM. As shown in Table 3, removing the static graph branch causes a significant performance drop (e.g., AUC drops from 0.966 to 0.873). It indicates the importance of persistent social ties for interest prediction. Similarly, removing the dynamic branch reduces AUC to 0.935, suggesting that dynamic social interaction patterns provide complementary predictive signals. Overall, the full SIM achieves the best results, demonstrating the benefit of jointly modeling static social topology and temporal co-presence dynamics.

Impact of Resource Availability. Fig. 3(a) compares social welfare under scarce, moderate, and abundant resource settings. SAGE consistently outperforms the interest-driven and FCFS baselines, achieving up to 2.4× and 2.3× higher social welfare, respectively, under scarce conditions. Although the performance gap narrows as resources become abundant, SAGE maintains a consistent advantage because its welfare-oriented allocation rule prioritizes high-surplus display requests under resource constraints, while its critical-value pricing rule preserves truthful bidding incentives.

Scalability Analysis. Fig. 3(b) shows social welfare as the number of MSUs scales from 10 to 1,000. The proposed scheme exhibits superior scalability, achieving a 13× increase in social welfare, while the baselines yield only 5.6× and 9.2× improvements. The widening performance gap demonstrates the effectiveness of competitive bidding and efficient pricing in optimizing resource allocation in high-density Metaverse environments.

Robustness to Strategic Manipulation. Fig. 3(c) compares social welfare under truthful bidding and strategic bidding (i.e., low bidding and very low bidding). SAGE maintains stable social welfare across different valuation-reporting strategies, whereas the baselines show more pronounced welfare degradation under underbidding. This indicates that incentive-compatible pricing is important for mitigating the impact of strategic valuation misreporting, which is robust resource allocation in non-cooperative settings.

6 Conclusion

In this paper, we propose SAGE, a social-aware VR display framework that improves Metaverse scalability by prioritizing rendering fidelity toward socially relevant avatars. SAGE introduces a dual-graph learning architecture to jointly model long-term social relationships and short-term behavioral patterns, enabling fine-grained estimation of social interests across users. Guided by the inferred interest scores, we formulate VR display scheduling as a combinatorial optimization problem and develop an efficient coordination mechanism for socially efficient VR display scheduling. Experiments show that SAGE outperforms representative baselines, improving interest prediction accuracy by 11.64% and increasing social welfare by up to 2.4×, with strong scalability and robustness. Future work will extend SAGE to support cross-platform interoperability and consistent rendering decisions across heterogeneous Metaverse infrastructures.

Acknowledgments

This work was supported in part by NSFC (Nos. U24A20237, U22A2029, 62302387, U23A20276) and the Youth Innovation Team of Shaanxi Universities.

References

[Blumrosen and Nisan, 2007] Liad Blumrosen and Noam Nisan. Combinatorial auctions. *Algorithmic Game Theory*, 267:300, 2007.

[Chen and Yang, 2022] Bing-Jyue Chen and De-Nian Yang. User recommendation in social metaverse with VR. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management (CIKM)*, pages 148–158, 2022.

[Chen *et al.*, 2024] Bing-Jyue Chen, Chiok Yew Ho, and De-Nian Yang. AFTER: Adaptive friend discovery for temporal-spatial and social-aware XR. In *Proceedings of the IEEE 40th International Conference on Data Engineering (ICDE)*, pages 2639–2652, 2024.

[Cheng *et al.*, 2021] Yifei Cheng, Yukang Yan, Xin Yi, Yuanchun Shi, and David Lindlbauer. SemanticAdapt: Optimization-based adaptation of mixed reality layouts leveraging virtual-physical semantic connections. In *Proceedings of the 34th Annual ACM Symposium on User Interface Software and Technology (UIST)*, pages 282–297, 2021.

[Cho *et al.*, 2014] Kyunghyun Cho, Bart Van Merriënboer, Çaglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. In *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1724–1734, 2014.

[Du *et al.*, 2023] Hongyang Du, Jiazhen Liu, Dusit Niyato, Jiawen Kang, Zehui Xiong, Junshan Zhang, and Dong In Kim. Attention-aware resource allocation and QoE analysis for metaverse xurllc services. *IEEE Journal on Selected Areas in Communications*, 41(7):2158–2175, 2023.

[Feng and Zhao, 2025] Jie Feng and Jun Zhao. Resource allocation for augmented reality empowered vehicular edge metaverse. *IEEE Transactions on Communications*, 73(3):1987–2001, 2025.

[Feng *et al.*, 2025] Lei Feng, Xiaoyi Jiang, Yao Sun, Dusit Niyato, Yu Zhou, Shiyi Gu, Zhixiang Yang, Yang Yang, and Fanqin Zhou. Resource allocation for metaverse experience optimization: A multi-objective multi-agent evolutionary reinforcement learning approach. *IEEE Transactions on Mobile Computing*, 24(4):3473–3488, 2025.

[Greff *et al.*, 2017] Klaus Greff, Rupesh K. Srivastava, Jan Koutník, Bas R. Steunebrink, and Jürgen Schmidhuber. LSTM: A search space odyssey. *IEEE Transactions on Neural Networks and Learning Systems*, 28(10):2222–2232, 2017.

[Guo *et al.*, 2024] Beizhang Guo, Juntao Bao, Baili Chai, Di Wu, and Miao Hu. Lumos: Optimizing live 360-degree video upstreaming via spatial-temporal integrated neural enhancement. In *Proceedings of the 32nd ACM International Conference on Multimedia (MM)*, pages 7210–7219, 2024.

[Haeberlen *et al.*, 2023] Andreas Haeberlen, Linh Thi Xuan Phan, and Morgan McGuire. Metaverse as a service: Megascale social 3D on the cloud. In *Proceedings of the 2023 ACM Symposium on Cloud Computing (SoCC)*, pages 298–307, 2023.

[Hsu *et al.*, 2024] Bay-Yuan Hsu, Chih-Ya Shen, Hao Shan Yuan, Wang-Chien Lee, and De-Nian Yang. Social-Aware Group Display Configuration in VR Conference. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 38, pages 8517–8525, 2024.

[Hu *et al.*, 2026] Qinnan Hu, Yuntao Wang, Zhou Su, Shaolong Guo, Yuan Gao, Nan Liu, and Tom H. Luan. BlockAthena: A scalable approach for long-term blockchain crimes analysis. *IEEE Transactions on Information Forensics and Security*, pages 1–16, 2026. Early Access.

[Jiang *et al.*, 2023] Yuna Jiang, Jiawen Kang, Xiaohu Ge, Dusit Niyato, and Zehui Xiong. QoE analysis and resource allocation for wireless metaverse services. *IEEE Transactions on Communications*, 71(8):4735–4750, 2023.

[Kipf and Welling, 2017] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *Proceedings of International Conference on Learning Representations (ICLR)*, pages 1–14, 2017.

[Lai *et al.*, 2020] Zeqi Lai, Y. Charlie Hu, Yong Cui, Linhui Sun, Ningwei Dai, and Hung-Sheng Lee. Furion: Engineering high-quality immersive virtual reality on today's mobile devices. *IEEE Transactions on Mobile Computing*, 19(7):1586–1602, 2020.

[Lehmann *et al.*, 2002] Daniel Lehmann, Liadan Ita Oćallaghan, and Yoav Shoham. Truth revelation in approximately efficient combinatorial auctions. *Journal of the ACM*, 49(5):577–602, 2002.

[Li *et al.*, 2025a] Weimin Li, Qin Li, Weihong Tian, Jie Gao, Fan Wu, Jianxun Liu, and Ju Ren. MUCVR: Edge computing-enabled high-quality multi-user collaboration for interactive MVR. *IEEE Transactions on Parallel and Distributed Systems*, 36(10):2058–2072, 2025.

[Li *et al.*, 2025b] Xinping Li, Yan Hong, Meng Zhang, Xiangpeng Xie, Jie Zhang, and Song Guo. Social group oriented recommendation systems in wireless edge-empowered metaverse. *IEEE Network*, 39(3):270–276, 2025.

[Liu *et al.*, 2025] Nan Liu, Tom H. Luan, Yuntao Wang, Yiliang Liu, and Zhou Su. QoE-oriented cooperative vr rendering and dynamic resource leasing in metaverse. *IEEE Transactions on Mobile Computing*, 24(10):10247–10263, 2025.

[Luong *et al.*, 2017] Nguyen Cong Luong, Ping Wang, Dusit Niyato, Yonggang Wen, and Zhu Han. Resource management in cloud networking using economic analysis and pricing models: A survey. *IEEE Communications Surveys & Tutorials*, 19(2):954–1001, 2017.

[Meng *et al.*, 2020] Jiayi Meng, Sibendu Paul, and Y. Charlie Hu. Coterie: Exploiting frame similarity to enable high-quality multiplayer VR on commodity mobile devices. In *Proceedings of the Twenty-Fifth International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)*, pages 923–937, 2020.

[Ooi *et al.*, 2023] Beng Chin Ooi, Gang Chen, Mike Zheng Shou, Kian-Lee Tan, Anthony Tung, Xiaokui Xiao, James Wei Luen Yip, Bingxue Zhang, and Meihui Zhang. The metaverse data deluge: What can we do about it? In *Proceedings of the IEEE 39th International Conference on Data Engineering (ICDE)*, pages 3675–3687, 2023.

[Sapiezynski *et al.*, 2019] Piotr Sapiezynski, Arkadiusz Stopczynski, David Dreyer Lassen, and Sune Lehmann. Interaction data from the copenhagen networks study. *Scientific Data*, 6(1):315, 2019.

[Steinert *et al.*, 2024] Patrick Steinert, Stefan Wagenpfeil, Ingo Frommholz, and Matthias L. Hemmje. 256 metaverse records dataset. In *Proceedings of the 32nd ACM International Conference on Multimedia (MM)*, page 4256–4263, 2024.

[Su *et al.*, 2022] Zhou Su, Yuntao Wang, Qichao Xu, and Ning Zhang. LVBS: Lightweight vehicular blockchain for secure data sharing in disaster rescue. *IEEE Transactions on Dependable and Secure Computing*, 19(1):19–32, 2022.

[Tan *et al.*, 2022] Yu Wei Tan, Alden Tan, Nicholas Nge, and Anand Bhojan. DHR: Distributed hybrid rendering for metaverse experiences. In *Proceedings of the 1st Workshop on Interactive eXtended Reality (IXR)*, pages 51–59, 2022.

[Veličković *et al.*, 2018] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. In *Proceedings of International Conference on Learning Representations (ICLR)*, pages 1–12, 2018.

[Wang *et al.*, 2022] Yuntao Wang, Zhou Su, Ning Zhang, Rui Xing, Dongxiao Liu, Tom H Luan, and Xuemin Shen. A survey on metaverse: Fundamentals, security, and privacy. *IEEE Communications Surveys & Tutorials*, 25(1):319–352, 2022.

[Wang *et al.*, 2023] Yuntao Wang, Zhou Su, Qichao Xu, Ruidong Li, Tom H. Luan, and Pinghui Wang. A secure and intelligent data sharing scheme for uav-assisted disaster rescue. *IEEE/ACM Transactions on Networking*, 31(6):2422–2438, 2023.

[Xu *et al.*, 2019] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In *Proceedings of International Conference on Learning Representations (ICLR)*, pages 1–17, 2019.

[Yu *et al.*, 2024] Jiadong Yu, Ahmad Yousef Alhilal, Tailin Zhou, Pan Hui, and Danny HK Tsang. Attention-based QoE-aware digital twin empowered edge computing for immersive virtual reality. *IEEE Transactions on Wireless Communications*, 23(9):11276–11290, 2024.

[Zhang *et al.*, 2024] Ziliang Zhang, Zexin Li, Hyoseung Kim, and Cong Liu. BOXR: Body and head motion optimization framework for extended reality. In *Proceedings of the 2024 IEEE Real-Time Systems Symposium (RTSS)*, pages 70–82, 2024.

[Zhao *et al.*, 2024] Jun Zhao, Liangxin Qian, and Wenhan Yu. Human-centric resource allocation in the metaverse over wireless communications. *IEEE Journal on Selected Areas in Communications*, 42(3):514–537, 2024.