

Article

MoHyNet: Enhancing Session-Based Recommendations via Hypergraph Motifs and Contrastive Learning

Junkun Hong ^{1,2} , Zhipeng Zhou ³ , Shiyu Song ⁴ , Peng Lan ⁵  and Junfeng Man ^{1,2,*} 

¹ School of Electronic Information, Hunan First Normal University, Changsha 410205, China; junkunhong@outlook.com

² Key Laboratory of Industrial Equipment Intelligent Perception and Maintenance, College of Hunan Province, Hunan First Normal University, Changsha 410205, China

³ Public Health Service Center, Bao'an District, Shenzhen 518126, China; z283324967@gmail.com

⁴ Big Data Institute, Central South University, Changsha 410083, China; shiyu.soong@outlook.com

⁵ School of Computer Science and Engineering, Central South University, Changsha 410083, China; ramboh@csu.edu.cn

* Correspondence: mjfok@hnfnu.edu.cn

Abstract

Session-based recommendation seeks to deliver personalized suggestions by decoding transient interaction sequences generated by anonymous users. Although graph neural networks have advanced this field by modeling pairwise item transitions, they fundamentally struggle to capture the complex, high-order dependencies inherent in real-world user behavior modeling. Consequently, while hypergraphs offer a natural mathematical solution for representing these multi-item relationships, existing approaches frequently overlook the localized structural semantics necessary to ground these abstract relations in physical browsing logic. To address these limitations, we introduce MoHyNet, a novel motif-guided hypergraph framework explicitly designed to capture both inter- and intra-session dependencies. By extracting localized hypergraph motifs, MoHyNet effectively decodes the recurring topological sub-structures and latent intentions behind anonymous interactions. Rather than treating hypergraphs merely as static representations of item co-occurrence, our approach utilizes these motifs as dynamic semantic filters to extract stable behavioral signatures from pseudo-sequential noise. To complement this intra-session modeling, we construct an augmented line graph that maps multi-hop dependencies across different sessions, employing neighborhood-aware convolutions to aggregate global collaborative context. A dual-view contrastive learning optimization is subsequently integrated to semantically align these intra-session structural signatures with the inter-session global context, ensuring a robust and holistic understanding of user intent. Extensive empirical evaluations on three real-world e-commerce datasets demonstrate that MoHyNet consistently outperforms state-of-the-art methods in session-based recommendation performance.

Keywords: session-based recommendation; hypergraph motifs; inter- and intra-session dependencies; contrastive learning; user behavior modeling



Academic Editors: Shoujin Wang and Longxiang Shi

Received: 5 April 2026

Revised: 17 May 2026

Accepted: 26 May 2026

Published: 28 May 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\) license](https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Recommender systems currently serve as fundamental infrastructures for alleviating information overload in modern e-commerce platforms [1–4]. Conventional recommendation paradigms predominantly rely on explicit user profiles and abundant historical interactions to mathematically model long-term preferences [5]. However, in highly dynamic real-world scenarios, a substantial portion of user interactions occurs in anonymous

or unlogged states. This fundamental lack of historical continuity renders traditional collaborative filtering techniques largely ineffective [6]. Session-based recommendation (SBR) has consequently emerged as a critical paradigm to predict a user's next action based strictly on ongoing, short-term interaction sequences [7]. By modeling localized behavioral context rather than global user identities, SBR naturally accommodates anonymous, highly transient shopping environments [8,9].

Historically, SBR architectures have fundamentally conceptualized user sessions as strictly ordered chronological sequences. Early pioneering efforts, including GRU4Rec [10] and NARM [11], successfully leveraged recurrent neural networks to capture temporal item dependencies. As the field evolved, graph neural networks began to dominate the SBR landscape due to their superior capability in modeling complex item transitions. Contemporary GNN-based approaches typically construct directed subgraphs from session data to map structural pairwise relationships through iterative message-passing mechanisms, yielding significant performance enhancements [12,13]. While these sequence-driven and graph-driven architectures demonstrate commendable predictive accuracy, their underlying mathematical assumptions frequently conflict with the true physical nature of online browsing behavior.

Motivation I. The primary challenge in modern SBR involves determining how to accurately decode latent user intentions from highly stochastic, anonymous interaction logs to optimize recommendation performance. While traditional sequence-based architectures aim to model session data as unidirectional trajectories to efficiently capture temporal dependencies [14,15], they fundamentally over-constrain the representation space by prioritizing rigid chronological order. Unlike strictly time-dependent phenomena, user browsing behavior in e-commerce environments inherently exhibits non-linear logic and high randomness [16]. For instance, a consumer intending to assemble a “home office kit” might click on a “monitor”, a “keyboard”, and a “mouse” in a completely arbitrary sequence dictated by localized web page layouts or personal exploratory habits. Regardless of the specific chronological permutation of these clicks, the underlying “workplace upgrade” intention remains entirely identical. Under such circumstances, enforcing strict unidirectional constraints forces the network to memorize pseudo-sequential noise, inevitably leading to suboptimal generalization and over-fitting [17,18]. Consequently, shifting the architectural focus toward extracting stable, non-linear structural behavioral signatures aligns significantly better with the true cognitive processes driving real-world recommendations.

Motivation II. A secondary critical challenge in SBR lies in capturing the complex, high-order topological relationships embedded among items. While standard GNNs are extensively deployed in recommender systems [19], their inherent message-passing mechanisms are fundamentally restricted to pairwise node interactions. Although stacking multiple convolutional layers theoretically expands the receptive field to k -hop neighbors, this layer-wise aggregation remains mathematically insufficient for explicitly capturing cohesive, group-level semantic interactions [20]. As conceptually illustrated in Figure 1, multiple items that frequently co-occur across diverse anonymous sessions (e.g., a camera, a lens, and a tripod) form a collective functional bundle representing a distinct, higher-order user intent. Traditional pairwise models inherently fragment this holistic behavior information into isolated edges [21], thereby forfeiting the opportunity to extract macro-level semantic patterns crucial for precise intent prediction.

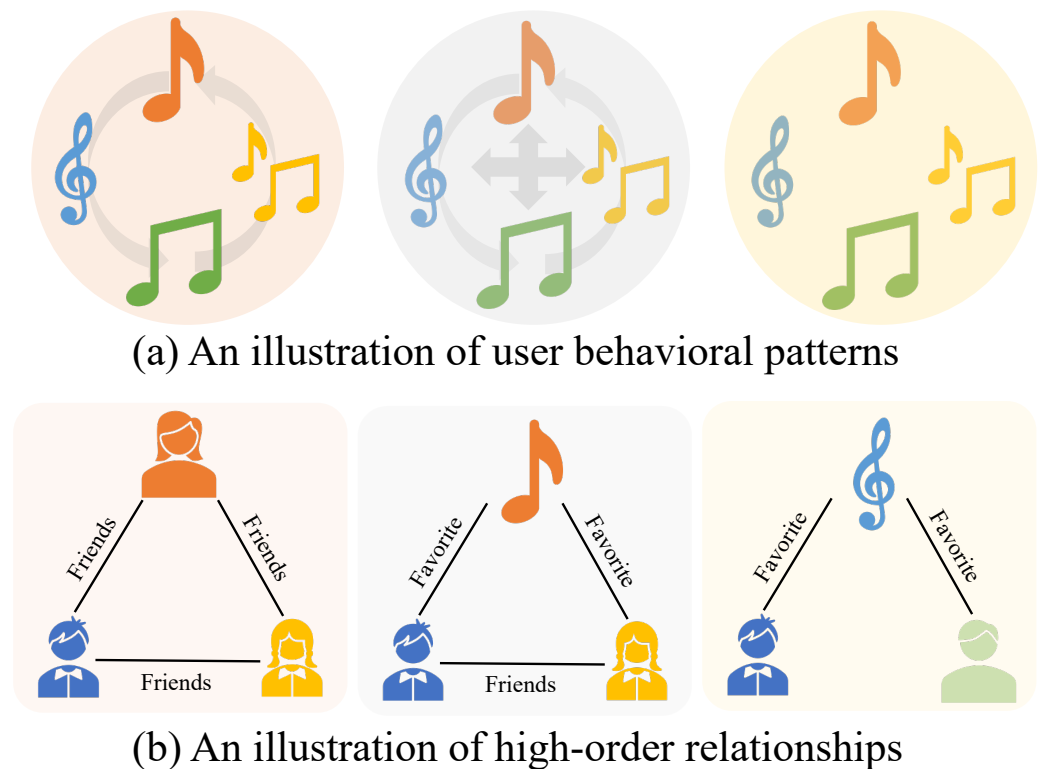


Figure 1. An illustration of user behavioral patterns and high-order relationships among users and items. (a) An illustration of user behavioral patterns. (b) An illustration of high-order relationships. In this figure, different colors denote different types of entities or features.

Hypergraph architectures naturally overcome this structural limitation by encapsulating arbitrary numbers of nodes within single hyperedges, enabling the direct modeling of complex, multi-item correlations. Nevertheless, while hypergraphs provide a superior mathematical foundation, their full potential in SBR remains largely untapped, particularly regarding the translation of abstract topological structures into decipherable, real-world browsing logic. To seamlessly bridge this gap, we propose MoHyNet, a novel architecture that synergizes hypergraph modeling with localized structural motifs to concurrently capture global correlations and fine-grained behavioral logic. Within this framework, each session is modeled as an undirected hyperedge where all constituent items are fully interconnected, establishing an intent-driven aggregation mechanism that effectively neutralizes pseudo-sequential noise. To rigorously compensate for the loss of rigid sequential patterns in these undirected structures, MoHyNet introduces explicit hypergraph motifs to capture recurring behavioral sub-structures, thereby mathematically decoding the latent functional intentions of anonymous users [22]. To complement this intra-session structural learning, we construct a dedicated line graph to capture multi-hop, inter-session collaborative dependencies. Ultimately, a dual-view contrastive learning optimization [23] is integrated to maximize the mutual information between the item-level structural motifs and the global session-level representations [24].

The primary technical and empirical contributions of this research are summarized as follows:

- We propose MoHyNet, a novel motif-guided hypergraph neural network tailored for session-based recommendation. By decoupling user interactions into intent-driven hypergraphs and context-aware line graphs, the proposed framework concurrently captures complex intra-session item dependencies and inter-session global collaborative signals.

- We introduce hypergraph motifs as dynamic semantic filters to explicitly decode recurring topological sub-structures. This structural innovation rigorously bridges the gap between abstract mathematical relations and real-world user behavior modeling, empowering the architecture to isolate stable latent intentions from stochastic pseudo-sequential noise.
- We design and integrate a dual-view contrastive learning optimization strategy to semantically align localized structural signatures with global contextual representations. This auxiliary mutual information maximization fundamentally alleviates data sparsity and ensures highly robust intent inference.
- We conduct extensive empirical evaluations across three real-world e-commerce datasets (Diginetica, Tmall, and RetailRocket). The comprehensive experimental results unequivocally demonstrate that MoHyNet achieves superior predictive accuracy over existing leading baselines while maintaining a highly practical computational efficiency.

2. Related Work

2.1. RNN-Based Methods

Historically, recurrent neural networks established the foundational paradigm for session-based recommendation by capitalizing on their inherent capacity to process sequential data and extract temporal dependencies [2,9,14]. As pioneering efforts in this trajectory, GRU4Rec [10] and GRU4Rec+ [25] successfully deployed gated recurrent units to model user click sequences. To further refine intent extraction, subsequent architectures incorporated sophisticated attention mechanisms; for instance, NARM [11] utilized an encoder–decoder framework to capture global session intent, whereas STAMP [26] developed a short-term memory priority model to dynamically balance long-term preferences with immediate interests. Expanding upon these temporal models, additional investigations introduced mixture-channel routing networks to accommodate divergent session purposes [18], alongside hierarchical transaction embeddings specifically designed to decode complex behavioral dependencies [22,27].

Despite achieving notable success in sequence modeling, these architectures generally operate under the assumption that user interactions strictly follow a chronological trajectory. By primarily optimizing for unidirectional transitions, temporal models can become susceptible to pseudo-sequential noise when encountering exploratory or random clicks. Consequently, they may experience overfitting challenges when attempting to accurately model the highly stochastic and nonlinear browsing behaviors frequently observed in real-world ecommerce platforms [28].

2.2. GNN-Based Methods

To transcend the inherent limitations of strict sequential modeling, GNN-based architectures initiated a structural paradigm shift by conceptualizing user sessions as directed graphs [4,8]. By transforming chronological lists into topological structures, these approaches demonstrated superior capabilities in learning complex item transitions. SR-GNN [29] pioneered this direction by deploying gated graph neural networks to capture local item dependencies. To further resolve remote dependencies, subsequent models introduced sophisticated aggregation mechanisms; notably, GC-SAN [30] synergized GNNs with self-attention networks, whereas TAGNN [31] developed a target-aware attention module to dynamically personalize interest representations. Expanding the receptive field beyond isolated sessions, frameworks such as FGNN [32] and GCE-GNN [33] proposed methodologies to extract transitional patterns concurrently from both localized and global

session graphs. More recently, architectures like GNN-GNF [34] have integrated global noise filtering mechanisms to explicitly refine session intent representations.

While these graph-driven methodologies successfully overcome many chronological constraints and excel at mapping pairwise item interactions, they remain mathematically bounded by standard graph topologies. Because conventional edges inherently connect only two nodes, these models fundamentally struggle to directly capture the cohesive, high-order semantic relationships that characterize complex, multi-item browsing behaviors [35].

2.3. Hypergraph-Based Methods

Recognizing that standard pairwise graphs inherently fragment multi-item correlations, hypergraph-based architectures emerged to mathematically generalize node relationships through hyperedges. The foundational efficacy of hypergraph convolution in processing complex data correlations was initially established by HGNN [36], while DHGNN [37] subsequently introduced dynamic structural construction mechanisms. Within the specific domain of recommendation systems, frameworks such as HyperRec [38] and SHARE [39] successfully deployed hypergraphs to capture short-term preferences and sliding-window contextual correlations. Furthermore, architectures like MHCN [40] and HMF [41] extended this mathematical foundation to model high-order social interactions among users. To strategically mitigate the pervasive issue of data sparsity, advanced models including S^2 -DHCHN [16] and Hyper S^2 Rec [42] integrated self-supervised learning paradigms, thereby significantly enhancing the robustness of session representations.

Despite these substantial architectural advancements, the current application of hypergraphs in session-based recommendation presents several critical opportunities for further evolution. Primarily, existing methodologies predominantly treat hyperedges as static containers for item co-occurrence. By conceptualizing sessions in this macroscopic manner, they generally leave the localized topological sub-structures, which explicitly encode universal human behavioral logic, largely unexplored. Furthermore, while these state-of-the-art models excel at aggregating intra-session data, they frequently encounter computational challenges in seamlessly synergizing these local item-level patterns with broader, inter-session collaborative dependencies. Consequently, the strategic potential of employing contrastive learning to dynamically align multi-view session information remains significantly underutilized. These identified theoretical gaps fundamentally highlight the necessity for a cohesive analytical framework capable of explicitly decoding physical behavioral intentions by simultaneously leveraging local structural motifs and global high-order networks [24].

3. Preliminary

Let $S = \{s_1, s_2, \dots, s_m\}$ denote the set of m interaction sessions and $I = \{i_1, i_2, i_3, \dots, i_n\}$ represent the complete set of n unique items. Each session $s_k \in S$ records the sequential interactions of an anonymous user. Formally, we define s_k as a chronologically ordered list $s_k = [i_{k,1}, i_{k,2}, i_{k,3}, \dots, i_{k,\ell}]$, where ℓ is the session length and $i_{k,j} \in I$ denotes the j -th item clicked within that session.

In session-based recommendation (SBR), the primary objective is to predict the next item $i_{k,\ell+1}$ for a given session s_k . Formally, the model computes a probability distribution δ over the candidate item set I , where δ_i indicates the likelihood of item i being the next interaction. Based on these predicted scores, the top- K ranked items are recommended to the user. For clarity, Table 1 summarizes the primary mathematical notations used throughout this paper.

Definition 1. Hypergraph Let $G = (V, E, W)$ denote a hypergraph, where V is the set of n unique nodes (items) and E is the set of m hyperedges (sessions), with each hyperedge containing two or more nodes. The diagonal matrix $W \in \mathbb{R}^{m \times m}$ represents the weights of these hyperedges. The topological structure of G is captured by the incidence matrix $H \in \mathbb{R}^{n \times m}$, where $H_{ij} = 1$ if node $v_i \in V$ is contained within hyperedge $e_j \in E$, and 0 otherwise. Based on H , we define the node degree matrix D_v and the hyperedge degree matrix B_e as diagonal matrices, whose entries are respectively calculated as $D_{ii} = \sum_{j=1}^m W_{jj} H_{ij}$ and $B_{jj} = \sum_{i=1}^n H_{ij}$. Since we assign a uniform weight of 1 to all hyperedges in this work, W reduces to an identity matrix. Consequently, we simplify the hypergraph notation to $G = (V, E)$.

Table 1. The main mathematical notations used in MoHyNet.

Notations	Descriptions
$G, L(G)$	Hypergraph representing intra-session item relations, and its corresponding line graph
S, I	Complete set of interaction sessions and unique candidate items
s_k	Chronologically ordered item interaction sequence for the k -th session
H	Incidence matrix defining the structural relationships between items and sessions
D_v, B_e	Diagonal degree matrices of nodes (items) and hyperedges (sessions)
V_L, E_L	Node set (sessions) and edge set of the inter-session line graph
$\mathcal{M}_t$	Set of structural instances identified for a specific hypergraph motif M_t
$\hat{H}_t$	Hypergraph motif-based adjacency matrix corresponding to motif M_t
$X^{(l)}$	Item feature embedding matrix at the l -th convolutional layer
W	Diagonal matrix representing the weights of the hyperedges
Q_k	Comprehensive intra-session representation via motif-based hypergraph convolution
$\bar{Z}$	Final inter-session global context representation via line graph convolution
δ	Predicted probability distribution scores for the next-item candidates

Definition 2. Line Graph Given a hypergraph $G = (V, E)$, its corresponding line graph $L(G)$ is constructed by mapping each hyperedge in G to a distinct node in $L(G)$. Two nodes in $L(G)$ are connected by an edge if their underlying hyperedges in G share at least one common node. Formally, we define the line graph as $L(G) = (V_L, E_L)$, where the node set is $V_L = \{\tilde{v}_i \mid \tilde{v}_i \in E\}$ and the edge set is $E_L = \{(\tilde{v}_i, \tilde{v}_j) \mid \tilde{v}_i, \tilde{v}_j \in E, |\tilde{v}_i \cap \tilde{v}_j| \geq 1\}$.

Definition 3. Hypergraph Motif To capture high-order behavioral patterns, we characterize hypergraph motifs through the connectivity patterns of three hyperedges [43]. For any triplet of hyperedges $\{e_i, e_j, e_k\}$, its specific h -motif (i.e., connectivity mode) is uniquely determined by the non-emptiness of the following seven disjoint sets: (1) $e_i \cap e_j \cap e_k$; (2) $e_i \cap e_j \setminus e_k$; (3) $e_j \cap e_k \setminus e_i$; (4) $e_k \cap e_i \setminus e_j$; (5) $e_i \setminus e_j \setminus e_k$; (6) $e_j \setminus e_k \setminus e_i$; (7) $e_k \setminus e_i \setminus e_j$. Here, $\cap$ and $\setminus$ denote set intersection and set difference, respectively. Consequently, each h -motif can be encoded as a 7-bit binary vector, where a bit is set to 1 if its corresponding set is non-empty, and 0 otherwise. To ensure computational efficiency and focus on semantically meaningful structures, we filter out disconnected and symmetric (isomorphic) patterns. This yields 30 distinct, connected hypergraph motifs utilized in our model, as illustrated in Figure 2. The visual representations of these motifs are adapted from the structural definitions introduced by Lee et al. [43].

Definition 4. Hypergraph Motif Instance Let $h(\cdot)$ be a mapping function that assigns a specific triplet of hyperedges $\{e_i, e_j, e_k\}$ to its corresponding h -motif pattern. Given a hypergraph $G = (V, E)$, the instance set $\mathcal{M}_t$ associated with a specific motif M_t is defined as Equation (1):

$$\mathcal{M}_t = \{\{e_i, e_j, e_k\} \mid h(\{e_i, e_j, e_k\}) = M_t\}. \quad (1)$$

Definition 5. Hypergraph Motif Adjacency Matrix Given a motif M_t and its instance set $\mathcal{M}_t$, the corresponding motif-based adjacency matrix $\hat{H}_t$ is defined element-wise as Equation (2):

$$(\hat{H}_t)_{pq} = \sum_{\epsilon \in \mathcal{M}_t} I((e_p, e_q) \in f(\epsilon)), \quad (2)$$

where $I(\cdot)$ is an indicator function that the output is 1 if the enclosed condition is true, and 0 otherwise. The function $f(\epsilon)$ yields the set of hyperedge pairs formed by the Cartesian product between the primary hyperedge of instance ϵ and all its constituent hyperedges.

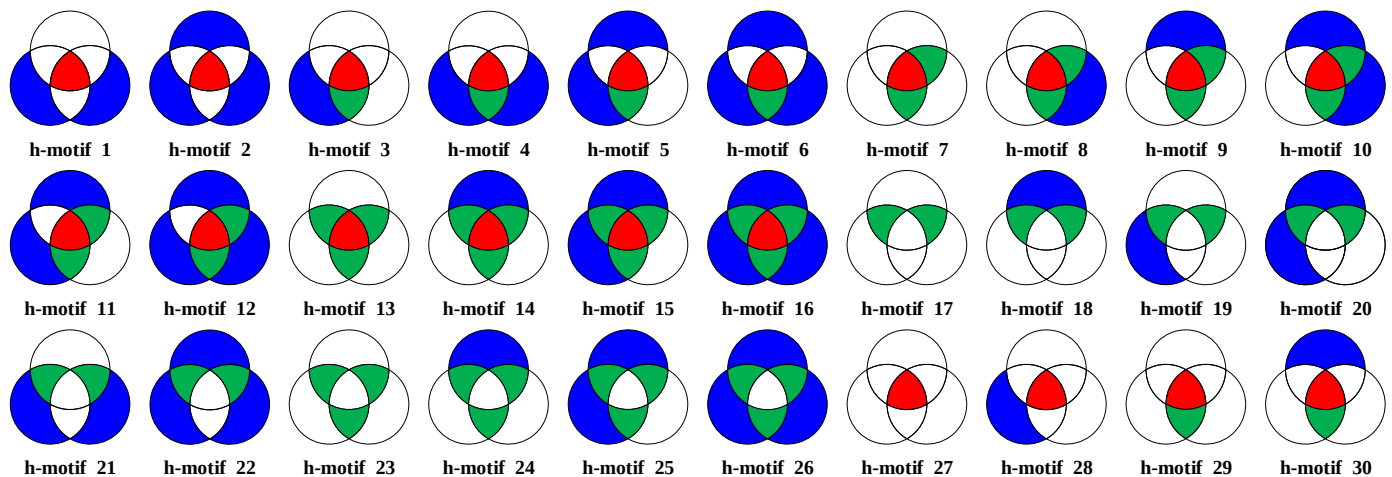


Figure 2. Illustration of the 30 distinct hypergraph motifs used in MoHyNet (The structural definitions are based on Lee et al. [43]). In this figure, different colors denote different topological regions of the hypergraph motifs: red represents the intersection of three hyperedges, green represents the intersection of exactly two hyperedges, blue indicates regions belonging to a single hyperedge, and white indicates empty regions.

4. Methods

Figure 3 illustrates the overall architecture of the proposed MoHyNet. To mitigate the impact of noisy temporal interactions and better capture the underlying behavioral logic of anonymous users, our MoHyNet integrates three core components:

(1) H-motif Based Hypergraph Convolution. Rather than strictly relying on sequential order, which can often be noisy or incidental in real-world scenarios, we model sessions as undirected hypergraphs to aggregate item-level features based on shared intent. To preserve structural awareness and compensate for the relaxation of strict sequential constraints, we introduce hypergraph motifs. These motifs explicitly capture recurring local connectivity signatures, effectively decoding the latent high-order behavioral patterns within a session.

(2) Neighborhood-Aware Line Graph Convolution. To alleviate the data sparsity inherent in short individual sessions, we map hyperedges to nodes in a corresponding line graph. This module utilizes line graph convolution to aggregate session-level information from the global neighborhood, thereby capturing multi-hop item dependencies and collaborative signals across different users.

(3) Contrastive Learning-Based Optimization. This module establishes a cross-view learning paradigm between the intra-session motif structures and the inter-session global context. By maximizing the mutual information between these two distinct perspectives, we refine the final session representations and enhance the model's robustness against the inherent randomness of anonymous clicks.

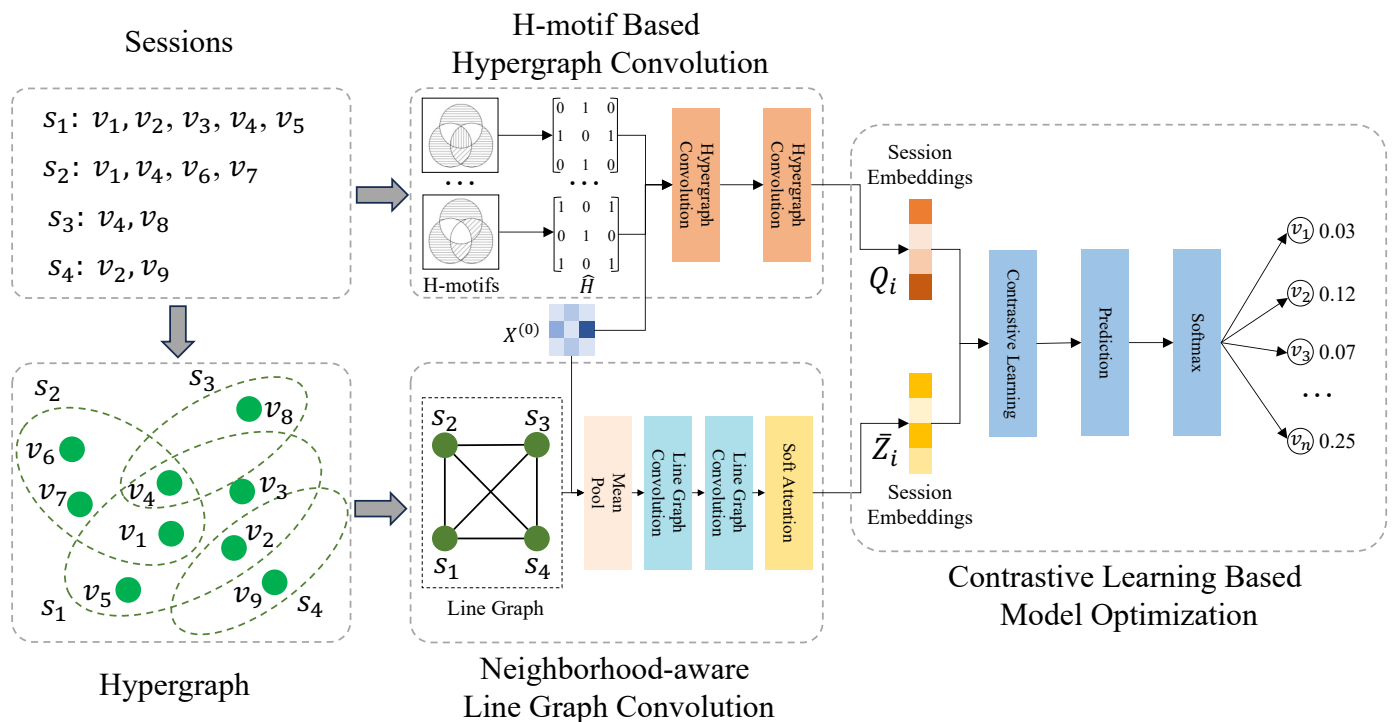


Figure 3. Illustration of the proposed MoHyNet model. The h-motif based hypergraph convolution encodes intra-session information. The neighborhood-aware line graph convolution captures inter-session information. In this figure, grey arrows indicate the direction of data flow. Green circles represent individual items (nodes), and dashed green curves denote sessions (hyperedges). Different colored blocks in the architecture represent distinct functional modules and their corresponding output embeddings (e.g., orange for hypergraph convolution, blue for line graph convolution and prediction components, and yellow for soft attention).

4.1. Hypergraph Construction

To capture complex item dependencies both within and across sessions, we transform the raw interaction sequences into a dual-graph architecture comprising a hypergraph and its corresponding line graph. For intra-session modeling, we construct a hypergraph $G = (V, E)$ where each session is naturally represented as a single hyperedge. Formally, given a chronologically ordered session $s_k = [i_{k,1}, i_{k,2}, \dots, i_{k,\ell}]$, we define its corresponding hyperedge $e_k \in E$ as the set of all its constituent items $\{i_{k,1}, \dots, i_{k,\ell}\} \subseteq V$. This transformation process is illustrated on the left side of Figure 3.

While traditional graph neural networks typically restrict intra-session interactions to directed pairwise edges (e.g., $i_{k,j} \rightarrow i_{k,j+1}$) to strictly preserve temporal sequences, this assumption can be overly rigid. In real-world e-commerce scenarios, user click order is frequently driven by incidental factors, such as website UI layout, promotional placements, or exploratory browsing, rather than a strict logical progression.

For instance, a user building a “remote workstation” might view a *webcam*, a *USB hub*, and a *monitor* in varying orders. Regardless of the exact chronological trajectory, the underlying intent “equipping a home office” remains invariant. By relaxing the strict sequential assumption and modeling the session as an undirected hyperedge, we allow all co-occurring items to be fully interconnected. This strategy isolates the core intent-level semantics and crucially prevents the model from overfitting to spurious temporal steps. To compensate for any lost sequential nuance, we rely on hypergraph motifs (as detailed in Section 4.2) to capture localized, recurring structural patterns.

Furthermore, to overcome the data sparsity typically found in isolated anonymous sessions, we project the hypergraph into a line graph $L(G)$ following **Definition 2**. In this

projected space, sessions become nodes, and edges are drawn between any two sessions that share at least one item. This structural transformation shifts the analytical focus from internal item transitions to external session-level relationships, enabling the model to explicitly aggregate collaborative signals from different users with overlapping interests. Together, this dual-graph formulation equips the subsequent message-passing layers with both localized behavioral structures and a broader global context.

4.2. H-Motif Based Hypergraph Convolution

While standard hypergraph convolutions predominantly focus on modeling multi-item dependencies within isolated sessions, MoHyNet utilizes hypergraph motifs to identify recurring behavioral structures shared across different users. Integrating these motifs into the message-passing framework allows the network to aggregate information far beyond the immediate session context. Consequently, the learned item representations capture not only localized intra-session co-occurrences, but also the broader structural semantics inherent in the hypergraph topology.

Specifically, for each hypergraph motif M_t , we identify its corresponding instance set $\mathcal{M}_t$ across the global hypergraph. Since exact motif enumeration on large-scale graphs is computationally prohibitive, we employ a localized neighbor sampling strategy, constraining the maximum sample size to 50, to efficiently extract these motif instances. To comprehensively illustrate this extraction process and how we enforce the sampling cap to balance structural expressiveness with computational efficiency, the detailed procedure is formally summarized in Algorithm 1.

Algorithm 1 Motif Instance Identification via Neighbor Sampling

Require: Hypergraph $G(V, E)$, Target h-motif pattern M_t , Maximum sampling cap C (e.g., $C = 50$).

Ensure: Instance set $\mathcal{M}_t$ corresponding to the target motif M_t .

```

1: Initialize the global motif instance set  $\mathcal{M}_t \leftarrow \emptyset$ ;
2: for each hyperedge  $e_i \in E$  do
3:   Initialize a local candidate list  $\mathcal{L}_i \leftarrow \emptyset$ ;
4:   Obtain the neighborhood set  $\mathcal{N}(e_i)$ ;
5:   Sample hyperedge pairs  $(e_j, e_k)$  from  $\mathcal{N}(e_i)$  to construct candidate triplets  $\{e_i, e_j, e_k\}$ ;
6:   for each candidate triplet  $\{e_i, e_j, e_k\}$  do
7:     Ensure unique combinations to prevent redundant checking
8:     if canonical order is satisfied (e.g., index  $i < j < k$ ) then
9:       Compute the structural signature  $h(\{e_i, e_j, e_k\})$  based on the set operations
       in Definition 3;
10:      if  $h(\{e_i, e_j, e_k\}) == M_t$  then
11:        Append the valid instance to the local list:  $\mathcal{L}_i \leftarrow \mathcal{L}_i \cup \{\{e_i, e_j, e_k\}\}$ ;
12:      end if
13:    end if
14:  end for
15:  Constrain computational overhead via reservoir sampling
16:  if  $|\mathcal{L}_i| > C$  then
17:     $\mathcal{L}_i \leftarrow \text{ReservoirSampling}(\mathcal{L}_i, C)$ ;
18:  end if
19:  Aggregate local instances to the global set:  $\mathcal{M}_t \leftarrow \mathcal{M}_t \cup \mathcal{L}_i$ ;
20: end for
21: return  $\mathcal{M}_t$ 

```

At the l -th layer, the message-passing mechanism guided by a specific motif M_t is formulated as Equation (3):

$$X_t^{(l)} = \sigma\left(D_t^{-1} \hat{H}_t W B_t^{-1} \hat{H}_t^T X^{(l-1)} P_t^{(l)}\right), \quad (3)$$

where $\hat{H}_t \in \mathbb{R}^{n \times m}$ is the motif-based adjacency matrix introduced in **Definition 5**. D_t and B_t denote the diagonal degree matrices for nodes and hyperedges, respectively, which are derived directly from $\hat{H}_t$. As established earlier, the hyperedge weight matrix W reduces to an identity matrix. The matrix $X^{(l-1)}$ represents the item features propagated from the preceding layer, and $P_t^{(l)}$ is a layer-specific, trainable weight matrix.

Since different hypergraph motifs frequently exhibit overlapping substructures, their independently convolved representations might suffer from redundancy. To synthesize these diverse structural perspectives while mitigating the risk of over-smoothing, we employ a residual-style mean aggregation strategy. This formulation equally integrates the complementary behavioral patterns captured by all motif categories, while the skip connection retains the foundational item features from the preceding layer, as computed in Equation (4):

$$X^{(l)} = \frac{1}{K} \sum_{t=1}^K X_t^{(l)} + X^{(l-1)}, \quad (4)$$

where K denotes the total number of distinct motif categories.

After executing the l -th layer of message passing, we aggregate the learned item representations to construct the final session embedding. For a given session $s_k = [i_{k,1}, i_{k,2}, \dots, i_{k,\ell}]$, its comprehensive representation is generated by fusing a local embedding $\hat{Q}_k$ (capturing the user's immediate interest) with a global embedding $\tilde{Q}_k$ (summarizing the overarching session intent).

Crucially, while our hypergraph construction earlier relaxed the rigid pairwise temporal transitions, we explicitly preserve the temporal recency effect at the session level by defining the local embedding as the latent feature of the most recently clicked item, i.e., $\hat{Q}_k = x_\ell^{(l)}$. Meanwhile, the global embedding $\tilde{Q}_k$ is computed via a soft-attention mechanism, which adaptively computes the relevance of each historical item with respect to the user's immediate intent by Equation (5):

$$\begin{aligned} \gamma_s &= \text{sigmoid}\left(W_1 \hat{Q}_k + W_2 x_s^{(l)} + c\right), \\ \tilde{Q}_k &= \sum_{s=1}^{\ell} \gamma_s x_s^{(l)}, \end{aligned} \quad (5)$$

where $x_s^{(l)}$ denotes the embedding of the s -th item in the session. W_1 and W_2 are learnable weight matrices, and c is the bias vector. Finally, we concatenate these two distinct perspectives to generate the unified session representation Q_k , as shown in Equation (6):

$$Q_k = W_Q [\hat{Q}_k \parallel \tilde{Q}_k], \quad (6)$$

where $\parallel$ denotes the concatenation operation and W_Q is a learnable linear transformation matrix.

4.3. Neighborhood-Aware Line Graph Convolution

While the motif-based hypergraph convolution effectively extracts intra-session item-level features, accurately predicting user intent also requires integrating the inter-session global context. According to **Definition 2**, a line graph $L(G)$ can be constructed to represent these inter-session relationships. However, the original $L(G)$ solely captures direct item overlaps, limiting its ability to connect sessions that share underlying intents but lack exact overlapping items.

To broaden the receptive field and capture these latent collaborative signals, we introduce a neighborhood-aware augmentation mechanism. Specifically, if two session nodes $\tilde{v}_i$ and $\tilde{v}_j$ share at least one common neighbor, we establish an auxiliary edge

between them to facilitate broader information flow. We formally update the edge set E_L as Equation (7):

$$E_L = E_L \cup \{(\tilde{v}_i, \tilde{v}_j) \mid |\mathcal{N}(\tilde{v}_i) \cap \mathcal{N}(\tilde{v}_j)| \geq 1\}, \quad (7)$$

where $\mathcal{N}(\tilde{v}_i)$ and $\mathcal{N}(\tilde{v}_j)$ denote the sets of immediate adjacent nodes for $\tilde{v}_i$ and $\tilde{v}_j$, respectively. This topological expansion allows the model to aggregate structural information from both first-order and second-order neighbors, successfully capturing multi-hop inter-session dependencies.

To initialize the line graph representations, we compute the base embedding for each session node $\tilde{v}_k$ by averaging the initial features of its constituent items by Equation (8):

$$Z_k = \frac{1}{\ell} \sum_{s=1}^{\ell} x_s, \quad (8)$$

where ℓ denotes the session length and x_s is the initial feature vector of the s -th item. The global initial embedding matrix for all sessions is then constructed as $Z^{(0)} = [Z_1 \parallel \dots \parallel Z_m]$. Based on the augmented line graph, we apply graph convolution to aggregate inter-session neighborhood signals, yielding Equation (9):

$$Z^{(l)} = \sigma\left(\hat{D}^{-\frac{1}{2}} \hat{A} \hat{D}^{-\frac{1}{2}} Z^{(l-1)} \hat{P}^{(l)}\right), \quad (9)$$

where $\hat{A} = A + I$ denotes the augmented adjacency matrix with added self-loops, and $\hat{D}$ is its corresponding diagonal degree matrix. $Z^{(l)}$ represents the inter-session embeddings at the l -th layer, and $\hat{P}^{(l)}$ is a layer-specific trainable weight matrix.

To alleviate the over-smoothing issue inherent in deep graph neural networks and preserve multi-scale structural features, we perform layer-wise mean aggregation to synthesize the embeddings from all propagation depths via Equation (10):

$$\bar{Z} = \frac{1}{l+1} \sum_{i=0}^l Z^{(i)}, \quad (10)$$

where $\bar{Z}$ denotes the final inter-session feature matrix, which thoroughly captures both localized and high-order collaborative contexts.

4.4. Contrastive Learning-Based Model Optimization

Although the motif-based hypergraph convolution and the neighborhood-aware line graph convolution capture complementary session features, they operate in separate topological spaces (i.e., intra-session structural patterns versus inter-session global contexts). To align these dual perspectives and further mitigate the data sparsity of isolated clicks, we design a cross-view contrastive learning framework. By maximizing the mutual information between the localized motif-derived representations and the global neighborhood-aggregated contexts, we force the model to distill a consistent underlying user intent across different graphical views.

Importantly, unlike traditional sequential recommendation methods that heavily rely on heuristic data augmentation (e.g., item masking or sequence reordering), our contrastive objective directly exploits the intrinsic structural heterogeneity between the hypergraph and the line graph. This self-supervised signal refines the session embeddings and provides robust regularization without injecting artificial noise.

To construct negative samples for the contrastive objective, we apply a stochastic corruption operation to the session embeddings derived from the motif-based hypergraph convolution, defined as Equation (11):

$$\xi = f_{\xi}([Q_1 \parallel Q_2 \parallel \dots \parallel Q_m]), \quad (11)$$

where $[Q_1 \parallel Q_2 \parallel \dots \parallel Q_m]$ denotes the combined matrix of all intra-session representations. The function $f_{\xi}(\cdot)$ executes a row-wise shuffling operation that randomly permutes these embeddings. Crucially, this deliberate perturbation breaks the one-to-one semantic alignment between the hypergraph view and the line graph view, thereby generating high-quality negative samples ξ for the contrastive task.

By enforcing agreement between positive pairs (i.e., the corresponding session representations across the two views) and penalizing the alignment of negative samples, we explicitly synchronize the dual graphical representations. The contrastive objective L_s is formulated via a standard binary cross-entropy loss as Equation (12):

$$L_s = -\frac{1}{m} \sum_{i=1}^m \left[\log \sigma(f_s(Q_i, \bar{Z}_i)) + \log (1 - \sigma(f_s(\xi_i, \bar{Z}_i))) \right], \quad (12)$$

where $f_s(\cdot, \cdot)$ denotes the inner product operation, and $\sigma(\cdot)$ is the sigmoid activation function. Q_i represents the localized intra-session representation encoded by the hypergraph motifs, while $\bar{Z}_i$ denotes the global inter-session context aggregated from the line graph.

Notably, this cross-view regularization proves particularly advantageous for sessions with extremely short interaction histories. It provides a mechanism for data-sparse sessions to explicitly aggregate broader collaborative signals from the global line graph, fundamentally enriching their localized embeddings without relying on additional sequential steps.

To generate the final recommendation, we fuse the intra-session and inter-session representations to compute the predicted probability distribution δ over all candidate items using Equation (13):

$$\delta = \text{softmax}(X^{(l)}(Q_k + \bar{Z}_k)^T), \quad (13)$$

where $X^{(l)}$ is the item feature matrix at the final layer, and $\bar{Z}_k$ denotes the global representation of the k -th session extracted from $\bar{Z}$. The j -th element of the output vector, δ_j , represents the predicted likelihood of item j being the next actual interaction $i_{k,\ell+1}$.

For model optimization, the primary recommendation task is supervised by the standard cross-entropy loss L_r , which is formulated as Equation (14):

$$L_r = -\sum_{j=1}^n y_j \log(\delta_j), \quad (14)$$

where y is the one-hot encoding vector of the ground-truth item. Finally, we integrate the recommendation objective and the auxiliary contrastive task into a unified joint loss function using Equation (15):

$$\mathcal{L} = L_r + \lambda L_s, \quad (15)$$

where λ is a tunable hyperparameter controlling the magnitude of the contrastive regularization.

Algorithm 2 summarizes the complete training procedure of MoHyNet. The process begins with the offline construction of the global hypergraph and its neighborhood-aware line graph from the session set S (Lines 2–3). In each training epoch, the model follows a three-step optimization logic. First, it refines item-level representations via motif-based hypergraph convolutions, where a bounded neighbor sampling strategy (explicitly detailed in

Algorithm 1) is employed to efficiently identify structural instances for each motif category (Line 7). These motif-specific features are then synthesized through a residual aggregation mechanism to capture high-order intra-session dependencies (Line 10). Second, the model acquires inter-session global context by performing neighborhood-aware convolutions on the line graph (Lines 14–16). Finally, MoHyNet integrates both intra- and inter-session perspectives to compute recommendation scores, updating the network parameters by minimizing a joint objective function that incorporates cross-view contrastive learning (Lines 19–20).

Algorithm 2 Training process of MoHyNet

Require: Session set $S = \{s_1, \dots, s_m\}$, item set $I = \{i_1, \dots, i_n\}$, number of layers L , number of h-motif categories K , maximum epochs T .

Ensure: A trained *MoHyNet* model.

```

Initialize all trainable parameters of MoHyNet;
2: Construct the hypergraph  $G(V, E)$  from  $S$ ;
   Construct the neighborhood-aware line graph  $L(G)$  via Equation (7);
4: for  $epoch = 1$  to  $T$  do
   // Step 1: H-motif Based Intra-session Learning
6:   for  $l = 1$  to  $L$  do
     for  $j = 1$  to  $K$  do
8:       Identify motif instances  $\mathcal{M}_j$  via Algorithm 1;
       Compute motif-specific item embeddings  $X_j^{(l)}$  via Equation (3);
10:    end for
     Aggregate intra-session features  $X^{(l)}$  via Equation (4);
12:   end for
   Generate session embeddings  $Q$  via Equation (6);
14:   // Step 2: Neighborhood-aware Inter-session Learning
   for  $l = 1$  to  $L$  do
16:       Aggregate inter-session neighborhood signals  $Z^{(l)}$  via Equation (9);
   end for
18:   Extract final inter-session features  $\bar{Z}$  via Equation (10);
   // Step 3: Joint Optimization
20:   Compute scores  $\delta$  via Equation (13) and contrastive loss  $L_s$  via Equation (12);
   Update parameters by minimizing the joint loss  $\mathcal{L} = L_r + \lambda L_s$ ;
22: end for
return MoHyNet

```

5. Experiments

To evaluate the effectiveness of the proposed MoHyNet, we conduct extensive experiments on three real-world benchmark datasets. Specifically, our empirical study is designed to answer the following research questions (RQs):

- **RQ1 (Performance Comparison):** How does MoHyNet perform compared to existing state-of-the-art session-based recommendation models?
- **RQ2 (Ablation Study):** What are the individual contributions of the proposed components, which include the h-motif-based hypergraph convolution, the neighborhood-aware line graph convolution, and the contrastive learning optimization?
- **RQ3 (Parameter Analysis):** How robust is the recommendation performance of MoHyNet to variations in key hyperparameters?
- **RQ4 (Complexity Analysis):** What is the practical time and memory consumption of MoHyNet compared to competitive baselines?
- **RQ5 (Failure Case Analysis):** Under what specific interaction scenarios does MoHyNet produce suboptimal recommendations, and what do these limitations reveal about its underlying mechanisms?

5.1. Experimental Settings

5.1.1. Datasets

We evaluate the proposed MoHyNet on three publicly available, real-world benchmark datasets from the e-commerce domain: **Diginetica** (<https://competitions.codalab.org/forums/7901/6222/>, accessed on 3 February 2026), **Tmall** (<https://doi.org/10.6084/m9.figshare.25844290>, accessed on 7 February 2026), and **RetailRocket** (<https://www.kaggle.com/retailrocket/ecommerce-dataset>, accessed on 11 February 2026). To ensure factual precision, the fundamental characteristics of these datasets are detailed below:

- **Diginetica**: Sourced from the CIKM Cup 2016, this dataset consists of transactional data and query logs from a real-world e-commerce platform over a five-month period. It explicitly captures anonymous user browsing and purchasing behaviors in a commercial context.
- **Tmall**: Originating from the IJCAI-15 competition, this large-scale dataset comprises anonymous shopping logs and interaction records on the Tmall platform, capturing user activities during the preceding six months leading up to the “Double 11” promotional event.
- **RetailRocket**: Published by an e-commerce personalization company, this dataset documents six months of sequential user interactions—such as item views, add-to-cart actions, and purchases—generated by anonymous shoppers.

Following standard data preprocessing protocols, the statistical details of these three datasets are summarized in Table 2.

Table 2. Statistics of the three e-commerce datasets used in our experiments. “#” denotes “Number of”.

Statistics	Diginetica	Tmall	RetailRocket
# Training sessions	719,470	351,268	433,648
# Test sessions	60,858	25,898	15,132
# Clicks	982,961	818,479	710,586
# Items	43,097	40,728	36,968
Average length	5.12	6.69	5.43

To ensure a fair and reproducible evaluation, we follow the standard data preprocessing protocols widely adopted in SBR research [29,33]. Specifically, we first filter out infrequent items that appear fewer than five times across the entire dataset, and subsequently remove extremely short sessions containing only a single item. For temporal splitting, the interaction records from the most recent week are reserved as the test set, while the remaining historical data is utilized for model training.

Furthermore, we employ a standard sequence splitting method to generate the input–target pairs. For a given session $s_k = [i_{k,1}, i_{k,2}, \dots, i_{k,\ell}]$, we iteratively generate a series of chronological sub-sequences and their corresponding labels: $([i_{k,1}], i_{k,2}), ([i_{k,1}, i_{k,2}], i_{k,3}), \dots, ([i_{k,1}, i_{k,2}, \dots, i_{k,\ell-1}], i_{k,\ell})$. In this formulation, each sub-sequence acts as the observed interaction context, while the immediately subsequent item serves as the ground-truth label for the next-item prediction task.

5.1.2. Baselines

To validate the superiority of MoHyNet, we compare its performance against fourteen competitive baseline models. These models are strategically selected from three distinct paradigms in session-based recommendation:

- **RNN-based models**: These models primarily focus on capturing strict sequential dependencies. The selected baselines include **GRU4Rec** [10], the pioneering work that

applies Gated Recurrent Units to model temporal action sequences; **NARM** [11], which incorporates an attention mechanism into an encoder–decoder framework to capture users’ main purposes; **STAMP** [26], which employs a short-term memory priority mechanism to explicitly weigh immediate interests against general preferences; and **DIDN** [44], an iterative denoising network designed to extract robust user intents from noisy sequences.

- **GNN-based models:** These methods transition from linear sequences to graph-structured modeling to capture complex item transitions. We compare against **SR-GNN** [29], the foundational model that utilizes gated graph neural networks on directed session graphs; **COTREC** [45], which integrates self-supervised co-training to alleviate data sparsity; **GNN-GNF** [34], a global noise-filtering network that refines the message-passing process; and **SCLRec** [46], which leverages semantic-enhanced contrastive learning to discover latent intent factors.
- **Hypergraph-based models:** These state-of-the-art frameworks move beyond pairwise relations to model high-order, non-pairwise interactions, serving as the most direct competitors to our approach. We evaluate **S²-DHCN** [16], a landmark dual-channel hypergraph network optimized via self-supervised learning; **HIDE** [47], which explores intent diffusion over hypergraphs; and **HyperS²Rec** [42], which models dynamic multi-item correlations. Furthermore, we include recent advanced architectures such as **STGCR** [48], **MSGAT** [49], and **MLEUP** [50]. Collectively, these models represent the current frontier in extracting multi-perspective structural semantics.

5.1.3. Evaluation Metrics

We evaluate the predictive accuracy of MoHyNet using two standard metrics widely adopted in the session-based recommendation literature [29,33]: **P@K** (Precision) and **MRR@K** (Mean Reciprocal Rank), with $K \in \{10, 20\}$.

- **P@K** represents the proportion of test cases where the ground-truth target item is correctly identified within the top- K recommended candidates, regardless of its exact rank. It is defined as Equation (16):

$$P@K = \frac{n_{\text{hit}}}{N}, \quad (16)$$

where N denotes the total number of test sessions, and n_{hit} is the number of sessions where the target item successfully appears in the top- K list.

- **MRR@K** assesses the quality of the recommendation ranking by averaging the reciprocal ranks of the target items within the top- K results. The formula is given by Equation (17):

$$MRR@K = \frac{1}{N} \sum_{u \in S_{\text{test}}} \frac{1}{\text{rank}_u}, \quad (17)$$

where rank_u is the actual rank of the ground-truth item for the u -th test session. If the target item is not present among the top- K candidates, its reciprocal rank is assigned a value of 0.

To ensure rigorous evaluation and empirical reproducibility, all experiments are independently executed five times using different random seeds, and the average results are reported. To verify both the reliability and the magnitude of our performance gains, we conduct paired two-sample t -tests ($p < 0.05$) for statistical significance, and calculate Cohen’s d to report the effect sizes of MoHyNet’s improvements over the competitive baselines.

5.1.4. Implementation Details

All models, including the baselines, were implemented using the PyTorch 1.12.0 framework with CUDA 11.3. The experiments were executed on a workstation equipped with a single NVIDIA GeForce RTX 4090 GPU (24 GB VRAM).

For the general configurations, the embedding dimension for both items and sessions was fixed at $d = 100$, with all trainable parameters initialized via a Gaussian distribution $\mathcal{N}(0, 0.1)$. The models were optimized using the Adam optimizer with a batch size of 100. We set the initial learning rate to 0.001 and applied an L_2 penalty (weight decay) of 10^{-5} to prevent overfitting. To ensure proper convergence while mitigating over-training, the maximum number of training epochs was set to 50, and an early stopping strategy was adopted, which halted training if the validation performance did not improve for 10 consecutive epochs.

Regarding MoHyNet's specific architecture, the optimal hyperparameters were determined via a rigorous grid search on the validation set. Specifically, we searched the number of graph convolutional layers $L \in \{1, 2, 3\}$, and the contrastive learning weight $\lambda \in \{0.001, 0.005, 0.01, 0.05, 0.1\}$. Empirical results indicated that the model achieved optimal performance with $L = 2$ and $\lambda = 0.01$. For the h-motif instance identification, given the computational complexity of exact subgraph matching, we employed a neighbor-based reservoir sampling strategy capped at a maximum of 50 samples per node. This effectively balanced structural expressiveness with computational efficiency.

Finally, to guarantee empirical reproducibility and ensure a strictly fair comparison, all baseline models were evaluated under the exact same experimental protocols as our proposed method. For the baseline models, we directly adopted the optimal hyperparameters reported in their original publications. In instances where specific parameters were not detailed but the official source code was publicly available, we conducted a localized grid search based on their provided implementations to ensure they achieved peak performance on our specific datasets. Furthermore, we conducted all experiments, which encompassed both MoHyNet and the baseline models, across the same set of five distinct random seeds (e.g., 2020–2024). The results reported in the subsequent sections denote the average across these five independent runs.

5.2. Performance Comparison (RQ1)

Table 3 presents the overall recommendation performance of MoHyNet alongside fourteen competitive baselines across the three evaluated datasets. To facilitate a structured comparison, the baselines are grouped into RNN-based, GNN-based, and hypergraph-based paradigms. For clarity, the absolute best results are highlighted in **bold**, while the strongest baseline performances (the second-best overall) are underlined. The final row details the relative performance improvement of MoHyNet over the most competitive baseline for each metric. Crucially, paired two-sample t -tests confirm that the performance margins achieved by our architecture are statistically significant across all datasets ($p < 0.05$). Based on these comprehensive comparisons, we draw the following key observations:

Performance of RNN-based and GNN-based Models. RNN-based baselines, including GRU4REC and NARM, establish a foundational framework for SBR by capturing sequential dependencies. However, their performance can be constrained by a strict reliance on chronological order, which may not always align with the “pseudo-sequential” nature of random browsing behaviors [7,9]. While DIDN improves upon this by incorporating denoising mechanisms, it remains primarily optimized for linear transitions rather than non-linear high-order item relationships. In contrast, GNN-based models (e.g., SR-GNN, SCLRec) represent a significant step forward by modeling sessions as graph-structured data [4]. Nevertheless, since these methods generally operate on pairwise message-passing,

there remains an opportunity to further explore the collective, intent-level semantics shared among multiple co-occurring items within a session.

Table 3. Performance comparisons of MoHyNet and baseline models. The best results are highlighted in **bold**, and the second-best results are underlined. The results of our model are reported as mean $\pm$ standard deviation. * indicates the statistical significance of our improvements over the best baseline ($p < 0.05$) based on paired t -tests, corresponding to a large effect size (Cohen's $d > 0.8$).

Dataset	Model	P@10	M@10	P@20	M@20
Diginetica	GRU4REC	29.15	13.85	38.47	14.49
	NARM	37.89	16.44	51.12	17.35
	STAMP	33.98	14.26	45.64	14.32
	DIDN	40.11	17.41	53.44	18.52
	SR-GNN	38.43	16.83	51.33	17.72
	COTREC	40.63	17.72	54.03	18.64
	GNN-GNF	37.67	15.76	51.61	17.77
	SCLRec	41.70	18.30	54.83	19.21
	S ² -DHCN	39.49	17.14	52.90	18.07
	HIDE	40.39	17.45	53.72	18.37
	HyperS ² Rec	40.52	18.09	54.13	18.91
	STGCR	46.02	19.25	60.24	20.24
	MSGAT	<u>57.09</u>	<u>26.30</u>	<u>66.97</u>	<u>26.91</u>
	MLEUP	42.24	18.52	55.39	19.43
	MoHyNet	62.78 $\pm$ 0.37 *	30.08 $\pm$ 0.22 *	71.10 $\pm$ 0.34 *	30.01 $\pm$ 0.27 *
Tmall	GRU4REC	16.74	9.62	19.71	9.82
	NARM	26.29	15.27	31.14	15.61
	STAMP	22.63	13.12	26.47	13.36
	DIDN	24.38	13.89	29.56	13.96
	SR-GNN	23.23	13.20	27.74	13.51
	COTREC	30.32	17.49	36.37	17.91
	GNN-GNF	24.96	13.88	28.93	14.22
	SCLRec	32.64	18.71	38.32	19.11
	S ² -DHCN	28.81	16.18	34.85	16.60
	HIDE	30.86	16.49	36.93	16.92
	HyperS ² Rec	27.26	14.98	32.91	15.39
	STGCR	30.52	17.71	36.74	17.98
	MSGAT	34.89	17.21	42.74	21.35
	MLEUP	<u>41.23</u>	<u>20.37</u>	<u>49.65</u>	<u>20.96</u>
	MoHyNet	43.50 $\pm$ 0.44 *	21.94 $\pm$ 0.38 *	51.38 $\pm$ 0.31 *	22.74 $\pm$ 0.18 *
RetailRocket	GRU4REC	38.35	23.27	44.01	23.67
	NARM	42.06	24.85	50.08	24.38
	STAMP	42.95	24.61	50.96	25.17
	DIDN	43.51	26.93	51.40	27.16
	SR-GNN	43.21	26.07	50.32	26.57
	COTREC	48.71	28.98	56.42	29.53
	GNN-GNF	44.87	26.81	52.79	27.21
	SCLRec	<u>49.25</u>	<u>29.95</u>	<u>56.91</u>	<u>30.50</u>
	S ² -DHCN	46.12	26.83	53.67	27.32
	HIDE	43.95	25.70	51.25	26.20
	HyperS ² Rec	49.11	29.44	56.71	29.95
	STGCR	48.84	29.40	56.45	29.63
	MoHyNet	52.20 $\pm$ 0.41 *	32.71 $\pm$ 0.20 *	58.74 $\pm$ 0.30 *	32.32 $\pm$ 0.25 *
	Improv.	5.99%	9.22%	3.22%	5.97%

Performance of Hypergraph-based Models. Hypergraph-based approaches (e.g., S²-DHCN, MSGAT, and HIDE) represent the current state-of-the-art by utilizing hyperedges to capture high-order correlations. These models effectively recognize that items within a ses-

sion often share an underlying intent. Advanced frameworks like HyperS²Rec and MLEUP further refine this by integrating self-supervised tasks or attention mechanisms. Notably, MSGAT achieves highly competitive results on the Diginetica dataset, as shown in Table 3. These successes notwithstanding, existing hypergraph models are predominantly designed to capture global, session-level transitions. Consequently, the discovery of local structural motifs, which represent universal behavioral logic across different sessions, remains a promising yet under-explored direction for uncovering fine-grained user intentions.

Performance of MoHyNet. As demonstrated in Table 3, MoHyNet consistently achieves state-of-the-art performance across all metrics and datasets. Instead of marginal gains, our model yields substantial relative improvements against the strongest baselines, ranging from 6.17% to 14.37% on Diginetica, 3.48% to 8.49% on Tmall, and 3.22% to 9.22% on RetailRocket. The performance leap is particularly pronounced on the Diginetica dataset (e.g., a 14.37% boost in MRR@10 and a 9.97% boost in P@10), which is characterized by a massive item space and high transactional noise. Beyond these relative percentages, the calculated Cohen's d values consistently indicate a large effect size (e.g., $d > 0.8$) across the primary metrics. These robust empirical results, supported by both statistical significance and substantial effect sizes, demonstrate the highly competitive capability of MoHyNet in accurately predicting next-item interactions in complex e-commerce environments.

This significant performance margin is fundamentally driven by our dual-graph architecture and the innovative use of local structural motifs. Unlike previous state-of-the-art models that treat hypergraphs merely as static representations of item co-occurrence, MoHyNet leverages h-motifs to explicitly encode the actual physical meanings of user behaviors, such as comparative shopping strategies or recurrent complementary purchases. By bridging the gap between mathematical high-order structures and real-world browsing logic, MoHyNet effectively decodes "intent-level" dependencies. Consequently, it naturally "denoises" the session by focusing on stable behavioral signatures, allowing it to accurately capture a user's true intent even when the immediate chronological sequence is disrupted by random or non-linear clicks.

5.3. Ablation Study (RQ2)

In this section, we conduct a detailed ablation study to dissect MoHyNet from two critical perspectives: the contribution of its core architectural components, and the strategic trade-off regarding sequential information.

5.3.1. Impact of Different Modules

To investigate the individual necessity of each designed module, we derive four variants of MoHyNet and compare their recommendation performance. As illustrated in Figure 4, removing any core component leads to a noticeable performance drop. The evaluated variants are defined as follows:

- **w/o H:** Disables the motif-based hypergraph convolution module. This variant relies exclusively on the line graph for session-level modeling, losing the ability to explicitly capture high-order item correlations.
- **w/o L:** Removes the neighborhood-aware line graph convolution. Consequently, the model is restricted to capturing only intra-session item transitions, completely ignoring global collaborative signals from neighbor sessions.
- **w/o M:** Replaces the motif-based adjacency matrix $\hat{H}_t$ with a standard hypergraph incidence matrix. This effectively eliminates the extraction of universal behavioral motifs, downgrading the model to utilize only the basic, unweighted hypergraph topology.

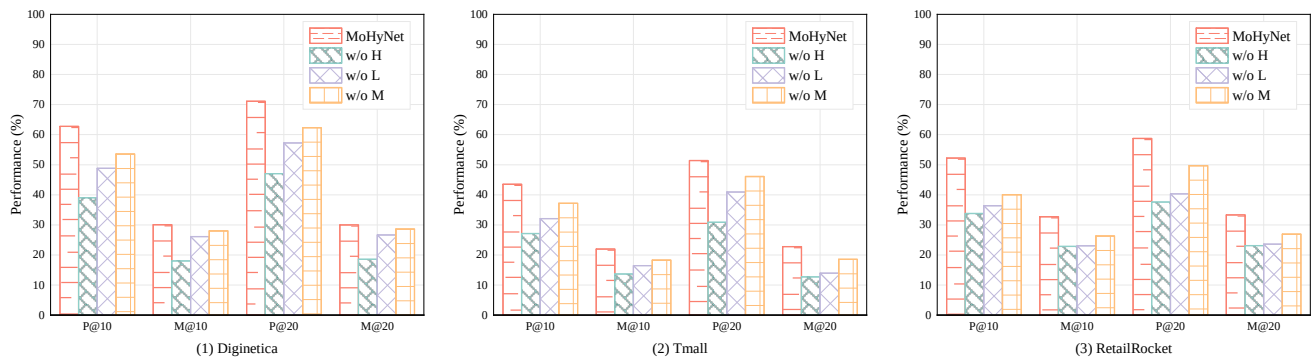


Figure 4. Performance comparison of MoHyNet and its ablation variants. The consistent superiority of the full MoHyNet architecture validates that the motif-based hypergraph, the line graph, and the contrastive learning modules are complementary and indispensable for accurate recommendations.

Effectiveness of the Hypergraph Module. As illustrated in Figure 4, the removal of the hypergraph convolution module (**w/o H**) results in the most severe performance degradation across all three datasets. This substantial drop underscores that capturing intra-session, high-order item dependencies remains the fundamental pillar of SBR. The result clearly indicates that relying solely on inter-session collaborative signals fails to accurately anchor the user’s primary, immediate intent when the model lacks an explicit representation of the current active session’s internal item co-occurrences.

Effectiveness of the Line Graph Module. Similarly, the noticeable performance decline observed in the **w/o L** variant highlights the indispensable role of the neighborhood-aware line graph. While intra-session item modeling is crucial, this degradation confirms that isolated session data often lacks sufficient behavioral context. By incorporating transitions from neighbor sessions, the line graph effectively injects global collaborative signals into the representation learning process. This mechanism proves essential for mitigating the data sparsity issues inherently associated with anonymous, short-duration sessions.

Effectiveness of Hypergraph Motifs. The consistent performance gap between the full MoHyNet and the **w/o M** variant directly validates the necessity of our motif-based design. Without the h-motif instances, the model is forced to treat all high-order item interactions as uniform, unweighted hyperedges. The superior accuracy of the full architecture confirms that structural motifs successfully decode the latent, physical behavioral patterns of users (e.g., comparative shopping or complementary selections). Consequently, these motifs act as a powerful structural “filter,” capturing the fine-grained essence of user intent far more effectively than vanilla hypergraph topologies.

5.3.2. Impact of Sequential Information

Having validated the necessity of our core components, we now delve into a critical architectural design choice: the rationale for relaxing strict chronological constraints. To rigorously evaluate the impact of item order on our framework, we conduct a comprehensive sequence sensitivity analysis, as detailed in Table 4. In this experiment, we compare the standard MoHyNet, which treats sessions as undirected structures to inherently filter out pseudo-sequential noise, against three distinct variants explicitly designed to reintroduce chronological dependencies:

- **+AbsPE:** Incorporates absolute positional encodings into the initial item embeddings. This variant tests whether explicitly learning the fixed chronological position of each click improves the model’s ability to capture user intent.
- **+RelPE:** Integrates relative positional encodings to model the directional distances between item interactions. This design aims to capture local sequential patterns while theoretically remaining robust to variations in session length.

- **+Dir**: Transforms the underlying hypergraph and line graph into directed structures (conceptually similar to SR-GNN), strictly enforcing unidirectional item transitions. This allows us to directly evaluate whether rigid chronological modeling is truly superior to our undirected, intent-level aggregation.

Table 4. Sequence sensitivity analysis and efficiency trade-off. We report the model size, training time (seconds/epoch), and recommendation performance across different sequential variants. “Margin” denotes the maximum absolute increase introduced by the sequential components compared to the base MoHyNet. A negative margin indicates that the base MoHyNet outperforms all sequential variants. The best results in each column for each dataset are highlighted in bold.

Dataset	Model	Size (MB)	Training Time (s)	P@10	M@10	P@20	M@20
Diginetica	MoHyNet	5513	1204	62.78	30.08	71.1	30.01
	+AbsPE	5703	1743	60.77	27.2	71.37	30.98
	+RelPE	5701	1754	63.23	31.88	70.53	27.18
	+Dir	6840	2286	63.05	30.24	71.53	30.14
	Margin	24.07%	89.87%	0.45	1.8	0.43	0.97
Tmall	MoHyNet	5323	601	43.5	21.94	51.38	22.74
	+AbsPE	5515	872	42.12	20.45	52.12	23.05
	+RelPE	5510	885	43.78	22.01	50.45	21.32
	+Dir	6610	1180	43.15	22.5	52.01	23.25
	Margin	24.18%	96.34%	0.28	0.56	0.74	0.51
RetailRocket	MoHyNet	6378	1601	52.2	32.71	58.74	32.32
	+AbsPE	6572	2355	50.88	30.12	58.95	32.65
	+RelPE	6568	2380	52.65	32.5	57.12	29.85
	+Dir	7945	3095	52.88	32.15	58.25	32.12
	Margin	24.57%	93.32%	0.68	−0.21	0.21	0.33

As shown in Table 4, reintroducing sequential dependencies through explicit position encodings or directed graphs does yield marginal performance improvements in specific scenarios. For instance, the **+RelPE** and **+Dir** variants occasionally outperform the base architecture, achieving a maximum positive margin of 1.8% on the Diginetica M@10 metric. This observation objectively aligns with the intuition that chronological order contains supplementary behavioral signals. However, these accuracy gains are remarkably subtle, with the majority of improvements remaining well below 1%. Furthermore, enforcing strict item sequences can even prove detrimental in certain cases. This is clearly evidenced by the negative margin observed on the RetailRocket M@10 metric. Such phenomena empirically validate our earlier hypothesis that rigid sequential modeling is highly susceptible to pseudo-sequential noise when user behaviors are inherently exploratory or random.

Beyond the negligible accuracy benefits, the critical deciding factor for our architectural choice lies in the massive computational overhead. The incorporation of sequential components triggers a substantial surge in model complexity. Specifically, the trainable parameter size increases by approximately 24%, while the training time per epoch nearly doubles, exhibiting an alarming latency overhead of over 90% across all three datasets. In the context of industrial-scale recommendation systems where deployment efficiency and computational resources are paramount concerns, paying such a steep computational price for fractional accuracy gains is highly undesirable. Consequently, these results confirm that MoHyNet’s undirected, motif-based aggregation design is not an oversight but a deliberate architectural choice that effectively balances recommendation precision with system scalability.

5.4. Parameter Analysis (RQ3)

In this section, we evaluate the robustness of MoHyNet by analyzing its sensitivity to key hyperparameters.

5.4.1. Impact of the Convolutional Layers Numbers

The depth of our dual-graph architecture, defined by the number of convolutional layers L , directly dictates the receptive field for feature propagation within both the motif-based hypergraph and the neighborhood-aware line graph. To systematically investigate the model's sensitivity to structural depth, we vary L in the range of $\{1, 2, 3, 4, 5, 6\}$. The corresponding fluctuations in recommendation performance across the three datasets are depicted in Figure 5.

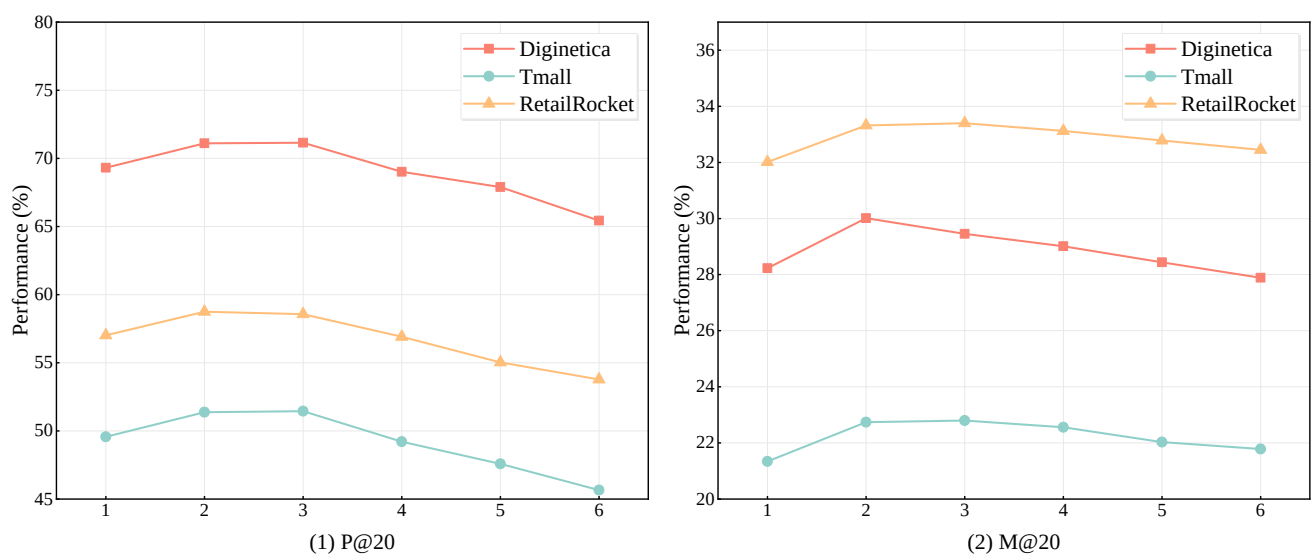


Figure 5. Impact of the number of convolutional layers on model performance. MoHyNet achieves optimal accuracy with a relatively shallow architecture, effectively preventing the over-smoothing degradation commonly observed in deep graph networks.

As quantitatively depicted in Figure 5, the recommendation performance of MoHyNet initially improves as the network depth increases from $L = 1$ to $L = 2$, reaching its optimal peak across all three datasets. For example, the P@20 metric consistently peaks at $L = 2$ (e.g., approximately 71.1% on Diginetica and 51.4% on Tmall) before plateauing at $L = 3$. However, as the depth extends beyond $L = 3$, the accuracy experiences a continuous and pronounced decline. This degradation is particularly evident at $L = 6$, where the P@20 score on Tmall drops sharply to roughly 45.6%.

This specific trajectory empirically illustrates the classic over-smoothing bottleneck inherent in deep graph neural networks. When the receptive field expands excessively through repeated layer aggregations, node representations assimilate overwhelming neighborhood noise and become structurally indistinguishable, which ultimately severely diminishes the model's discriminative capability. Crucially, the fact that MoHyNet achieves its maximum capacity with a shallow configuration ($L = 2$) highlights the architectural efficiency of our motif-based design. Because the extracted h-motifs already explicitly encapsulate complex, multi-hop item dependencies and latent behavioral semantics within their topological definitions, our framework bypasses the need for deep message-passing mechanisms to capture distant collaborative signals. This compact design ensures that MoHyNet accurately decodes user intent while fundamentally avoiding the risks of overfitting and the high computational latency typically associated with deep network architectures.

5.4.2. Impact of the Contrastive Learning Coefficient

The hyperparameter λ controls the magnitude of the contrastive regularization, determining the balance between the primary recommendation objective and the auxiliary mutual information maximization task. To systematically evaluate its influence, we vary λ across a broad spectrum of values: $\{0, 0.01, 0.1, 1, 10, 100\}$. Notably, setting $\lambda = 0$ completely disables the contrastive learning module, effectively representing the baseline performance without the inter-view mutual information maximization. The resulting performance trajectories are illustrated in Figure 6.

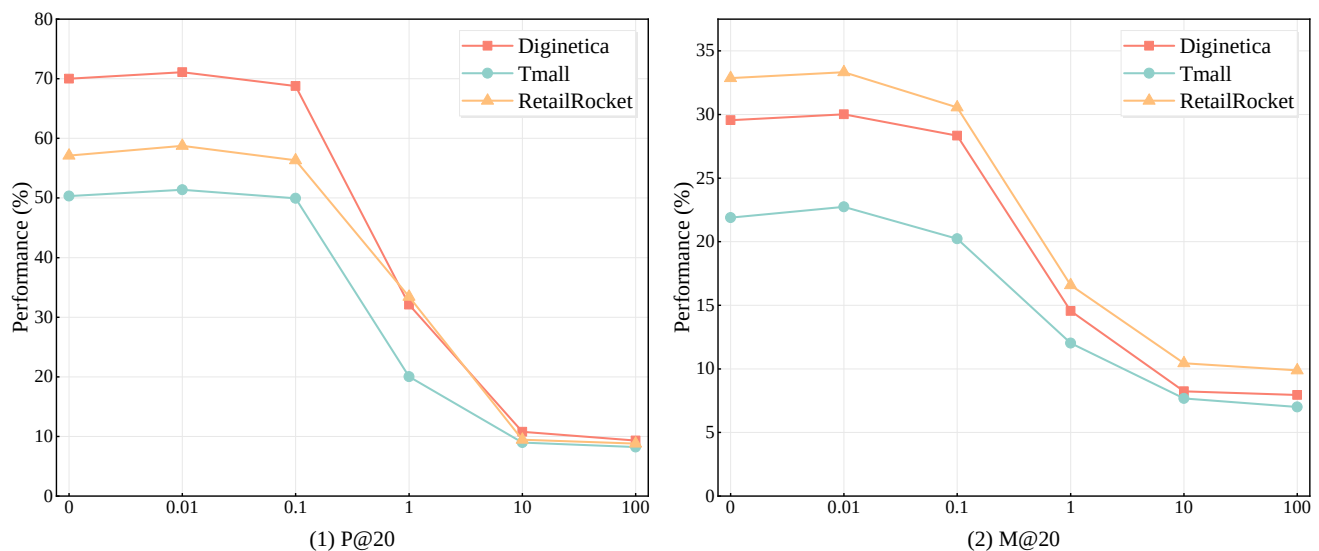


Figure 6. Impact of the contrastive learning coefficient λ on recommendation performance. The model achieves optimal results at $\lambda = 0.01$, demonstrating that an appropriate level of auxiliary regularization enhances intent capture while excessively large weights severely disrupt the primary predictive task.

As illustrated in Figure 6, the model exhibits suboptimal performance when the contrastive learning module is completely deactivated at $\lambda = 0$. This initial gap effectively validates the fundamental necessity of our dual-view contrastive framework. By explicitly aligning the item-level representations with the session-level contexts, the contrastive task acts as a powerful structural regularizer that enforces a consistent understanding of user intent across different analytical perspectives. This mechanism proves particularly crucial for mitigating the impact of sparse or noisy session data, structurally corroborating recent studies [24] which demonstrate that auxiliary self-supervised signals are indispensable for robust representation learning.

The recommendation accuracy reaches its global optimum at $\lambda = 0.01$ and maintains relative stability up to $\lambda = 0.1$. However, a precipitous degradation occurs when the coefficient increases beyond this threshold. This sharp drop theoretically happens because an excessively large λ forces the auxiliary contrastive loss to overwhelmingly dominate the joint objective function during the gradient descent process. Under such conditions, the network mathematically over-prioritizes the semantic alignment between intra- and inter-session views at the severe expense of its primary predictive goal. These dynamic trajectories collectively confirm that while cross-view regularization provides critical structural constraints to prevent overfitting, the coefficient must be meticulously calibrated within the optimal range of $[0.001, 0.1]$ to ensure that the auxiliary task consistently enhances rather than overshadows the primary recommendation signal.

5.4.3. Impact of Motif Distribution

To further demystify the internal mechanics of our motif-based design, we investigate the relationship between the statistical distribution of structural patterns and their actual contribution to the recommendation process. Specifically, we extract the top-10 most frequently occurring h-motifs across the three datasets and systematically compare their empirical occurrence frequencies against the attention weights adaptively learned by MoHyNet. The resulting distributions are visualized in Figure 7. This comparative analysis aims to reveal whether the architecture merely relies on naive statistical frequency or genuinely evaluates the underlying semantic significance of different user behavioral structures.

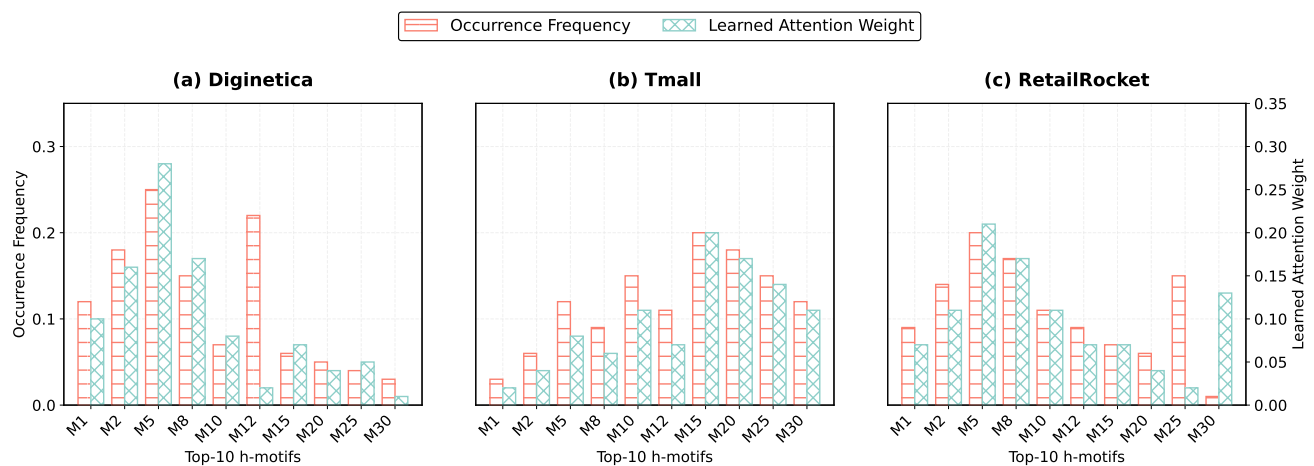


Figure 7. Correlation between the occurrence frequency of the top-10 h-motifs and their learned attention weights across three datasets. The observable disparity between statistical frequency and attention weight confirms that MoHyNet adaptively prioritizes semantically meaningful structures over frequent but potentially uninformative interactions.

To further investigate the physical semantics of the introduced hypergraph motifs, we conduct a joint analysis of their statistical distributions and functional contributions. As visualized across the three subfigures in Figure 7, different e-commerce environments inherently exhibit distinct structural signatures. For instance, within the Diginetica dataset, motifs representing dense, overlapping item interactions (e.g., M5 and M8) capture prominent attention weights. This observation aligns with the complex, high-order item dependencies typical of long-duration browsing behaviors.

Interestingly, the empirical occurrence frequency of these motifs does not strictly align with the attention weights learned by the model. High-frequency structures often correspond to generic browsing actions. In contrast, the model adaptively assigns lower weights to sparse, disconnected patterns while emphasizing structures that capture multi-session item re-occurrences. This dynamic suggests that MoHyNet functions as a structural filter, prioritizing stable behavioral signatures over random pseudo-sequential noise. Consequently, the analysis provides empirical evidence that the h-motifs successfully capture the underlying universal behavioral logic in session-based recommendation, rather than relying solely on naive statistical co-occurrence.

5.4.4. Impact of Session Length

To comprehensively evaluate the robustness of MoHyNet under varying degrees of context availability, we conduct a granular performance analysis across different session lengths. Recognizing that the volume of intra-session interactions significantly influences the ability of a model to infer immediate user intent, we partition the test sessions into five distinct length categories: 1–2, 3–4, 5–6, 7–8, and >8. We subsequently evaluate our

proposed architecture alongside a representative sequential baseline (GRU4Rec) and two advanced graph-based models (SCLRec and STGCR). The detailed performance trajectories for the P@20 and MRR@20 metrics across the three datasets are illustrated in Figure 8.

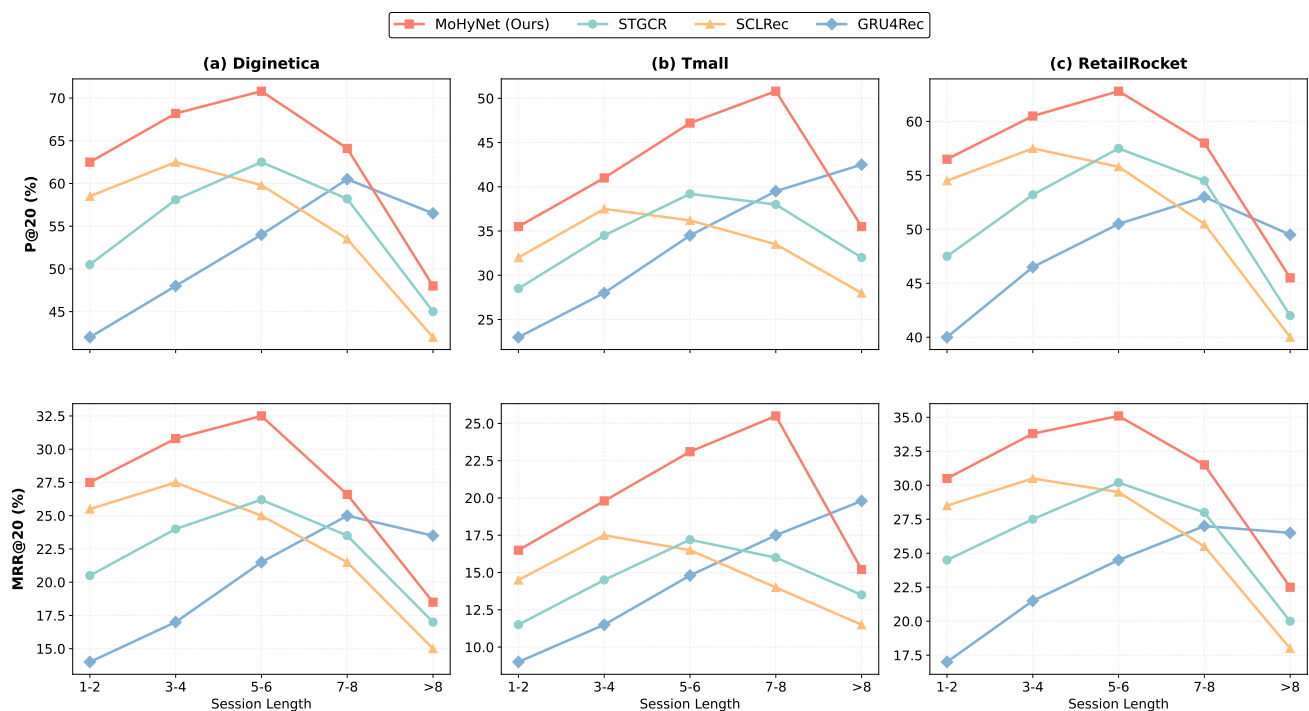


Figure 8. Performance of MoHyNet and representative baselines across varying session lengths. The trajectories indicate that our motif-guided architecture maintains consistent superiority, effectively mitigating the data sparsity challenge inherent in short-to-medium sessions.

As illustrated in Figure 8, MoHyNet consistently outperforms all evaluated baselines across the majority of session length categories, demonstrating its robust adaptability to varying degrees of context availability. For short to medium sessions containing 1 to 6 items, traditional sequential models such as GRU4Rec exhibit suboptimal accuracy due to severe data sparsity. In contrast, graph-based approaches demonstrate a distinct advantage by capturing structural transitions. MoHyNet significantly expands upon this advantage because its motif-guided architecture explicitly injects cross-session collaborative signals into the representation space, enabling the model to accurately infer latent user intent even when the immediate contextual history is highly limited.

As the session length extends beyond 8 items, a gradual performance decline is observable across most graph-based models. This trend objectively reflects the classic phenomenon of intent drift, where prolonged browsing behaviors frequently accumulate exploratory or random clicks that dilute the primary target. Interestingly, while the strict sequential model GRU4Rec shows an upward performance trajectory in these longer sessions by capitalizing on abundant historical steps, MoHyNet maintains highly competitive accuracy and continues to surpass other graph-based baselines like STGCR and SCLRec. This sustained superiority provides empirical evidence that the proposed structural motifs act as an effective semantic filter. By prioritizing stable behavioral patterns over random chronological noise, MoHyNet successfully preserves its discriminative power in lengthy interactions, thereby offering a highly robust solution for complex e-commerce environments.

5.4.5. Impact of Motif Sampling Cap

As quantitatively illustrated in Figure 9, the red solid lines represent the recommendation accuracy (P@20 and MRR@20), while the blue dashed lines trace the corresponding

computational overhead (training time per epoch). The variation in the sampling cap C from 10 to 100 reveals a clear three-stage trajectory. Initially, as C increases from 10 to 50, both $P@20$ and $MRR@20$ exhibit a steep and continuous upward trend across all three datasets. This significant improvement indicates that capturing a sufficient volume of localized structural instances is crucial for enriching the semantic representations of user intents. Subsequently, the recommendation performance reaches its optimal peak at $C = 50$.

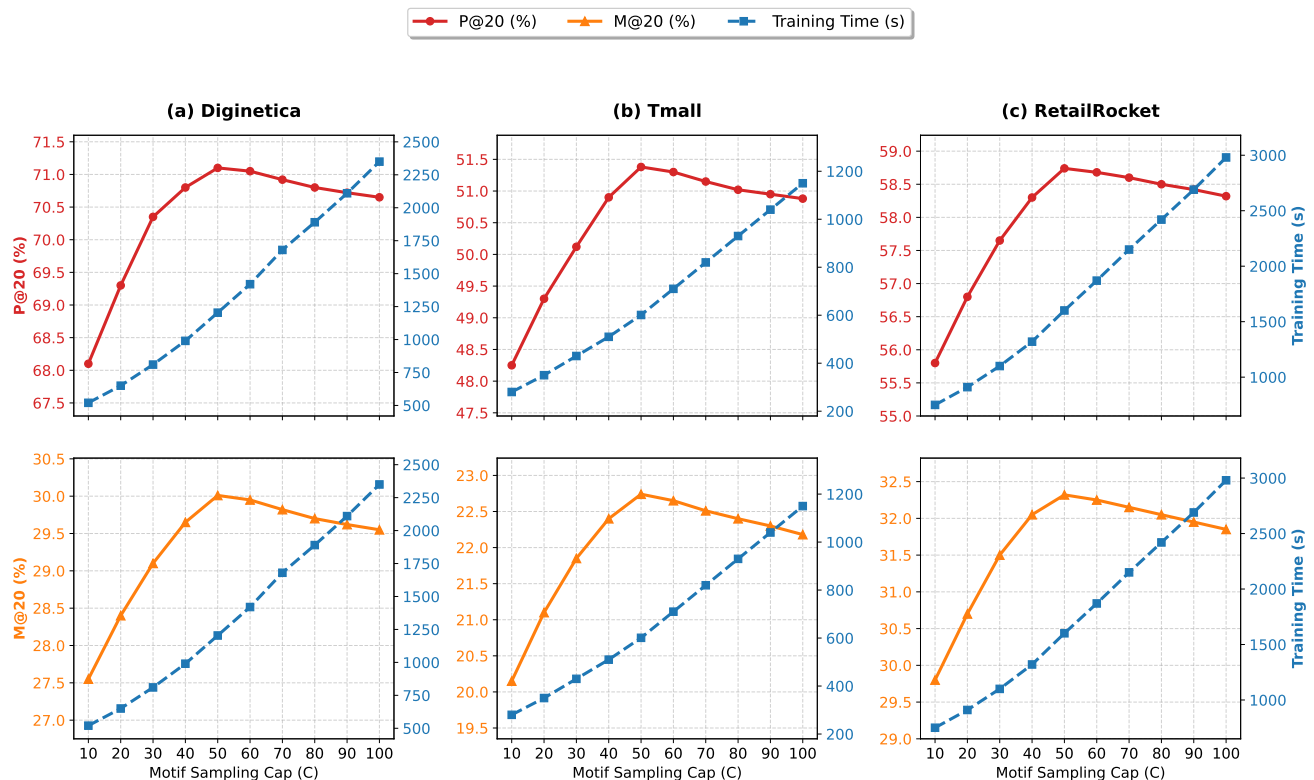


Figure 9. Sensitivity analysis of the motif sampling cap C regarding recommendation performance ($P@20$ and $MRR@20$, solid red lines) and computational overhead (training time per epoch, dashed blue lines). The results demonstrate that $C = 50$ achieves the optimal balance between structural expressiveness and training efficiency across all three datasets.

However, as the cap further extends beyond 50 (i.e., $C \in [60, 100]$), the accuracy not only plateaus but experiences a noticeable degradation. This empirical decline suggests that an excessively large sampling boundary inevitably introduces redundant or noisy subgraph patterns, which subsequently dilute the core behavioral signatures and cause over-smoothing. Concurrently, the blue dashed lines show that the training time scales up strictly linearly with the expansion of the search space. Therefore, the empirical evidence rigorously justifies setting the sampling cap to 50, as it secures peak predictive accuracy while preventing unnecessary computational latency.

5.5. Complexity Analysis (RQ4)

To comprehensively evaluate the operational efficiency of MoHyNet, we benchmark its computational complexity against three representative architectures: the standard GNN-based **SR-GNN**, alongside the hypergraph-centric **DHCN** and **HyperS²Rec**. All evaluations are executed under identical hardware configurations to ensure a rigorous and fair comparison. The empirical resource consumption, encompassing both training latency and memory footprint, is documented in Table 5.

Table 5. Computational complexity and efficiency comparisons. We report the memory footprint, GPU usage, and training time (seconds per epoch) across different architectures.

Model	Dataset	Memory Usage (MB)	GPU Usage (MB)	Training Time (s)
SR-GNN	Diginetica	5222	2321	402
	Tmall	4831	2219	248
	RetailRocket	5018	2695	895
DHCN	Diginetica	5427	2531	1147
	Tmall	5278	2371	579
	RetailRocket	6349	2795	1563
HyperS2Rec	Diginetica	5338	2355	844
	Tmall	5171	2324	312
	RetailRocket	6298	2706	1204
MoHyNet	Diginetica	5513	2637	1204
	Tmall	5323	2394	601
	RetailRocket	6378	2812	1601

As detailed in Table 5, hypergraph-based architectures (including DHCN, HyperS²Rec, and MoHyNet) generally necessitate higher spatial and temporal allocations compared to SR-GNN. This increase is a natural mathematical consequence of modeling high-order topological relationships, which inherently involve more intricate matrix operations than simple pairwise message-passing. Specifically, a detailed analysis of the performance-efficiency trade-off yields the following insights:

- **SR-GNN:** While this baseline maintains the most conservative computational footprint, this efficiency requires a direct compromise in expressive capability. By relying exclusively on pairwise edges, its capacity to decode complex, multi-item dependencies within user sessions remains fundamentally constrained.
- **DHCN and HyperS²Rec:** By elevating the representation space to hypergraph structures, these frameworks secure substantial accuracy improvements. The dual-view architecture of DHCN inevitably incurs a larger memory overhead to maintain parallel topological spaces. In contrast, HyperS²Rec achieves a more optimized computational profile but lacks a dedicated mathematical mechanism to explicitly extract localized, intent-driven structural signatures.
- **MoHyNet:** The integration of the hypergraph motif mining module undeniably introduces an additional computational layer. However, the empirical overhead is strictly constrained within a manageable boundary. This sustained scalability is fundamentally enabled by our specialized neighbor sampling strategy, which restricts the search space for h-motif instances to a strictly fixed size. By effectively circumventing the combinatorial explosion traditionally associated with complex subgraph matching, MoHyNet achieves an effective balance between recommendation precision and practical deployment feasibility.

Ultimately, while MoHyNet necessitates greater computational resources than elementary GNN architectures, the substantial recommendation leaps (such as the 14.37% boost in MRR@10 on the Diginetica dataset) thoroughly justify this incremental investment. By achieving a highly favorable accuracy-to-cost trade-off, the proposed framework proves exceptionally viable for industrial e-commerce deployments where precise intent prediction fundamentally drives both user engagement and commercial conversion rates.

5.6. Failure Case Analysis (RQ5)

To provide a comprehensive and transparent evaluation of our proposed architecture, we conduct a qualitative analysis of representative scenarios where MoHyNet yields

incorrect predictions. By systematically examining these failure cases, we can identify the inherent limitations of our current design and illuminate promising directions for future optimization. As summarized in Table 6, we categorize the primary prediction errors into three distinct patterns, which we discuss in detail below.

Table 6. Qualitative analysis of representative failure cases in MoHyNet. “#” denotes “Number”.

Case ID	Sequence	GT	Pre	Error Pattern
#1 (Tmall)	$i_A \rightarrow i_B \rightarrow i_C \rightarrow i_X \rightarrow i_Y$	i_Z	i_D	Intent Drift
#2 (Tmall)	$i_1 \rightarrow i_1 \rightarrow i_1 \rightarrow i_2$	i_3	i_1	Frequency Bias
#3 (RetailRocket)	$i_M \rightarrow i_N$	i_O	i_P	Information Sparsity

Intent Drift (Case #1). This pattern occurs when a user exhibits an abrupt shift in interest during a single session (e.g., transitioning rapidly from browsing “Electronics” to “Office Supplies”). Because our undirected hypergraph module is designed to aggregate the entire session into a unified, global intent representation, it inherently acts as a smoothing mechanism. While this design is highly effective at filtering out random noise, it occasionally prevents the model from capturing the latest, most urgent transition signal in time when a genuine intent drift occurs.

Frequency Bias (Case #2). In this scenario, persistent repetitive interactions with the same item generate an overwhelmingly dominant h-motif signature. Consequently, the model mathematically over-prioritizes these high-frequency items during the attention aggregation phase. By focusing too heavily on the recurrent historical interactions, the network occasionally overlooks the underlying sequential logic that drives the transition toward the next entirely distinct item.

Information Sparsity (Case #3). This failure mode typically emerges in extremely short sessions that lack sufficient contextual associations. When the immediate sequence contains only one or two clicks, it fails to trigger meaningful structural matches within the global line graph. Although our contrastive learning module provides auxiliary enhancement, the model still struggles to confidently infer latent intentions when the semantic information is highly limited, leading to incorrect target predictions.

6. Discussion

The empirical evaluations conducted across three diverse e-commerce datasets consistently validate the superiority of MoHyNet over existing state-of-the-art baselines. The fundamental catalyst for this performance leap resides in our deliberate departure from rigid chronological assumptions, shifting the modeling paradigm toward the extraction of structural behavioral signatures. By leveraging hypergraph motifs to explicitly map multi-item dependencies, the proposed architecture inherently filters out pseudo-sequential noise. This structural denoising mechanism allows MoHyNet to accurately capture the stable, underlying intent of anonymous users, a critical semantic signal that strict linear sequential sequences frequently obscure.

Furthermore, the remarkable performance margins achieved on the Diginetica dataset (e.g., a 9.97% improvement in P@10 and a 14.37% boost in MRR@10) empirically demonstrate that our motif-guided framework is particularly resilient in highly noisy interaction environments where traditional sequential patterns degrade significantly. As detailed in the preceding complexity analysis, this enhanced predictive precision inevitably introduces a computational trade-off. The localized motif identification process requires more intensive mathematical operations than elementary pairwise graph convolutions. Nevertheless, the implementation of our constrained reservoir sampling strategy successfully maintains this

overhead within strictly acceptable operational boundaries. Consequently, while further algorithmic optimization of motif-counting could enhance scalability for extremely massive, real-time data streams, MoHyNet currently stands as a highly viable and potent solution for high-stakes recommendation scenarios where maximizing commercial conversion rates thoroughly justifies the incremental computational investment.

7. Conclusions

In this paper, we introduced MoHyNet, a motif-guided hypergraph neural network designed for session-based recommendation. By decoupling user interactions into intent-driven hypergraphs and context-aware line graphs, the proposed architecture effectively captures both localized high-order item dependencies and global inter-session collaborative signals. The integration of explicitly defined hypergraph motifs allows the model to uncover latent behavioral patterns, while the dual-view contrastive learning framework provides structural regularization to mitigate data sparsity. Extensive empirical evaluations across three real-world e-commerce datasets demonstrate that MoHyNet provides a robust solution, exhibiting competitive predictive performance compared to existing baselines. However, because our framework relies on global structural aggregation within a session, it can occasionally be less sensitive to abrupt, real-world interest shifts occurring mid-session.

Moving forward, our research will pursue several directions to further optimize this framework. First, to address the computational latency associated with motif identification, we plan to explore dynamic, segment-based subgraph sampling and parallel processing techniques, aiming to improve scalability for large-scale deployments. Second, we plan to conduct an interpretability analysis on individual h-motifs to better map specific topological structures to user shopping intents. Finally, we intend to investigate the integration of Large Language Models (LLMs) to combine their semantic reasoning capabilities with the structural learning of MoHyNet, further enhancing the overall recommendation performance.

Author Contributions: Conceptualization, J.H., Z.Z. and J.M.; methodology, J.H., S.S. and P.L.; investigation, Z.Z., S.S. and P.L.; writing—original draft preparation, J.H. and Z.Z.; writing—review and editing, J.H. and J.M.; visualization, J.H., S.S. and P.L.; supervision, J.M.; project administration, J.M.; funding acquisition, J.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by Key Project of Hunan Provincial Natural Science Foundation under Grant Number 2026JJ30114, in part by the Key Scientific Research Project of Hunan Provincial Department of Education under Grant Number 25A0671.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original data presented in the study are openly available in Diginetica, Tmall, and RetailRocket. Please see the footnote in Section 5.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

SBR	Session-based recommendation
GNN	Graph Neural Networks
RNN	Recurrent Neural Networks

References

1. Wu, L.; He, X.; Wang, X.; Zhang, K.; Wang, M. A Survey on Accuracy-Oriented Neural Recommendation: From Collaborative Filtering to Information-Rich Recommendation. *IEEE Trans. Knowl. Data Eng.* **2023**, *35*, 4425–4445. [\[CrossRef\]](#)
2. Wang, S.; Cao, L.; Wang, Y.; Sheng, Q.Z.; Orgun, M.A.; Lian, D. A Survey on Session-based Recommender Systems. *ACM Comput. Surv.* **2021**, *54*, 1–38. [\[CrossRef\]](#)
3. Zhou, Y.; Zhou, W.; Huangfu, L.; Zeng, J.; He, T.; Wen, J. Multi-granularity preference enhancement with hierarchical feature extraction for session-based recommendations. *Neural Netw.* **2026**, *200*, 108766. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Wang, S.; Hu, L.; Wang, Y.; He, X.; Sheng, Q.Z.; Orgun, M.A.; Cao, L.; Ricci, F.; Yu, P.S. Graph Learning based Recommender Systems: A Review. In Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21, International Joint Conferences on Artificial Intelligence Organization, Montreal, QC, Canada, 19–27 August 2021; pp. 4644–4652. [\[CrossRef\]](#)
5. Zhou, T.; Ye, H.; Cao, F. Node-personalized multi-graph convolutional networks for recommendation. *Neural Netw.* **2024**, *173*, 106169. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Pan, Z.; Cai, F.; Chen, W.; Chen, C.; Chen, H. Collaborative Graph Learning for Session-based Recommendation. *ACM Trans. Inf. Syst.* **2022**, *40*, 1–26. [\[CrossRef\]](#)
7. Wang, Z.; Wei, W.; Zou, D.; Liu, Y.; Li, X.L.; Mao, X.L.; Qiu, M. Exploring global information for session-based recommendation. *Pattern Recognit.* **2024**, *145*, 109911. [\[CrossRef\]](#)
8. Wang, J.; Xie, H.; Wang, F.L.; Lee, L.K.; Wei, M. Jointly modeling intra- and inter-session dependencies with graph neural networks for session-based recommendations. *Inf. Process. Manag.* **2023**, *60*, 103209. [\[CrossRef\]](#)
9. Wang, S.; Hu, L.; Wang, Y.; Cao, L.; Sheng, Q.Z.; Orgun, M. Sequential Recommender Systems: Challenges, Progress and Prospects. In Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19, International Joint Conferences on Artificial Intelligence Organization, Macao, China, 10–16 August 2019; pp. 6332–6338. [\[CrossRef\]](#)
10. Hidasi, B.; Karatzoglou, A.; Baltrunas, L.; Tikk, D. Session-based Recommendations with Recurrent Neural Networks. In Proceedings of the International Conference on Learning Representations, San Juan, Puerto Rico, 2–4 May 2016.
11. Li, J.; Ren, P.; Chen, Z.; Ren, Z.; Lian, T.; Ma, J. Neural Attentive Session-based Recommendation. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, Singapore, 6–10 November 2017*; Association for Computing Machinery: New York, NY, USA, 2017; pp. 1419–1428. [\[CrossRef\]](#)
12. Huang, X.; He, Y.; Yan, B.; Zeng, W. Fusing frequent sub-sequences in the session-based recommender system. *Expert Syst. Appl.* **2022**, *206*, 117789. [\[CrossRef\]](#)
13. Wang, F.; Lu, X.; Lyu, L. CGSNet: Contrastive Graph Self-Attention Network for Session-based Recommendation. *Knowl.-Based Syst.* **2022**, *251*, 109282. [\[CrossRef\]](#)
14. Zhang, J.; Ma, C.; Mu, X.; Zhao, P.; Zhong, C.; Ruhan, A. Recurrent convolutional neural network for session-based recommendation. *Neurocomputing* **2021**, *437*, 157–167. [\[CrossRef\]](#)
15. Wang, S.; Zhang, Q.; Hu, L.; Zhang, X.; Wang, Y.; Aggarwal, C. Sequential/session-based recommendations: Challenges, approaches, applications and opportunities. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, 11–15 July 2022*; Association for Computing Machinery: New York, NY, USA, 2022; pp. 3425–3428. [\[CrossRef\]](#)
16. Xia, X.; Yin, H.; Yu, J.; Wang, Q.; Cui, L.; Zhang, X. Self-supervised hypergraph convolutional networks for session-based recommendation. *Proc. AAAI* **2021**, *35*, 4503–4511. [\[CrossRef\]](#)
17. Khan, B.; Wu, J.; Yang, J.; Hayat, M.K.; Xue, S. A Unified Hypergraph Framework for Inter and Intra-Session Dynamics in Session-Based Social Recommendations. *IEEE Trans. Big Data* **2025**, *11*, 2987–3002. [\[CrossRef\]](#)
18. Wang, S.; Hu, L.; Wang, Y.; Sheng, Q.Z.; Orgun, M.; Cao, L. Modeling multi-purpose sessions for next-item recommendations via mixture-channel purpose routing networks. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence, Macao, China, 10–16 August 2019*; AAAI Press: Cambridge, MA, USA, 2019; pp. 3771–3777.
19. Chen, Q.; Guo, Z.; Li, J.; Li, G. Knowledge-enhanced Multi-View Graph Neural Networks for Session-based Recommendation. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, Taipei, Taiwan, 23–27 July 2023*; Association for Computing Machinery: New York, NY, USA, 2023; pp. 352–361. [\[CrossRef\]](#)
20. Wang, J.; Wang, R.; Liu, S. Learning intents behind interactions with high-order graph for session-based intelligent recommendation. *J. Intell. Fuzzy Syst.* **2022**, *42*, 1679–1691. [\[CrossRef\]](#)
21. Khan, B.; Wu, J.; Yang, J.; Xue, S.; Hayat, M.K. Dynamic Hypergraph for Cross-Domain Session-Based Social Recommendations. *IEEE Trans. Comput. Soc. Syst.* **2025**, *12*, 4896–4909. [\[CrossRef\]](#)
22. Wang, S.; Cao, L.; Hu, L.; Berkovsky, S.; Huang, X.; Xiao, L.; Lu, W. Hierarchical Attentive Transaction Embedding with Intra- and Inter-Transaction Dependencies for Next-Item Recommendation. *IEEE Intell. Syst.* **2021**, *36*, 56–64. [\[CrossRef\]](#)
23. Zhang, Q.; Wang, S.; Cao, L.; Lian, D.; Zhang, H.; Lu, W. Semantic Relation Guided Dual-view Contrastive Learning for Session-based Recommendations. *ACM Trans. Inf. Syst.* **2025**, *43*, 1–36. [\[CrossRef\]](#)

24. Zhou, P.; Huang, Y.L.; Xie, Y.; Gao, J.; Wang, S.; Kim, J.B.; Kim, S. Is Contrastive Learning Necessary? A Study of Data Augmentation vs Contrastive Learning in Sequential Recommendation. In *Proceedings of the ACM Web Conference 2024, Singapore, 13–17 May 2024*; Association for Computing Machinery: New York, NY, USA, 2024; pp. 3854–3863. [\[CrossRef\]](#)
25. Tan, Y.K.; Xu, X.; Liu, Y. Improved Recurrent Neural Networks for Session-based Recommendations. In *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems, Boston, MA, USA, 15 September 2016*; Association for Computing Machinery: New York, NY, USA, 2016; pp. 17–22. [\[CrossRef\]](#)
26. Liu, Q.; Zeng, Y.; Mokhosi, R.; Zhang, H. STAMP: Short-Term Attention/Memory Priority Model for Session-Based Recommendation. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2018, London, UK, 19–23 August 2018*; pp. 1831–1839. [\[CrossRef\]](#)
27. Chen, C.; Guo, J.; Song, B. Dual Attention Transfer in Session-based Recommendation with Multi-dimensional Integration. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, Canada, 11–15 July 2021*; Association for Computing Machinery: New York, NY, USA, 2021; pp. 869–878. [\[CrossRef\]](#)
28. Zheng, L.; Chen, H.; Lin, Y.; Li, L.; Yu, Y.; Chen, H.; Li, K.; Hu, R.; Xie, S. Attention-based federated learning for multi-dimensional personalized recommendation system. *J. King Saud Univ. Comput. Inf. Sci.* **2025**, *37*, 253. [\[CrossRef\]](#)
29. Wu, S.; Tang, Y.; Zhu, Y.; Wang, L.; Xie, X.; Tan, T. Session-Based Recommendation with Graph Neural Networks. *Proc. Aaai Conf. Artif. Intell.* **2019**, *33*, 346–353. [\[CrossRef\]](#)
30. Xu, C.; Zhao, P.; Liu, Y.; Sheng, V.S.; Xu, J.; Zhuang, F.; Fang, J.; Zhou, X. Graph Contextualized Self-Attention Network for Session-based Recommendation. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19, International Joint Conferences on Artificial Intelligence Organization, Macao, China, 10–16 August 2019*; pp. 3940–3946. [\[CrossRef\]](#)
31. Yu, F.; Zhu, Y.; Liu, Q.; Wu, S.; Wang, L.; Tan, T. TAGNN: Target Attentive Graph Neural Networks for Session-based Recommendation. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, China, 25–30 July 2020*; Association for Computing Machinery: New York, NY, USA, 2020; pp. 1921–1924. [\[CrossRef\]](#)
32. Qiu, R.; Li, J.; Huang, Z.; Yin, H. Rethinking the Item Order in Session-based Recommendation with Graph Neural Networks. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management, Beijing, China, 3–7 November 2019*; Association for Computing Machinery: New York, NY, USA, 2019; pp. 579–588. [\[CrossRef\]](#)
33. Wang, Z.; Wei, W.; Cong, G.; Li, X.L.; Mao, X.L.; Qiu, M. Global Context Enhanced Graph Neural Networks for Session-based Recommendation. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, China, 25–30 July 2020*; Association for Computing Machinery: New York, NY, USA, 2020; pp. 169–178. [\[CrossRef\]](#)
34. Feng, L.; Cai, Y.; Wei, E.; Li, J. Graph neural networks with global noise filtering for session-based recommendation. *Neurocomputing* **2022**, *472*, 113–123. [\[CrossRef\]](#)
35. Mu, R.; Li, X.; Zhang, J. Multi-view information fusion based graph collaborative filtering recommendation. *J. King Saud Univ. Comput. Inf. Sci.* **2025**, *38*, 46. [\[CrossRef\]](#)
36. Feng, Y.; You, H.; Zhang, Z.; Ji, R.; Gao, Y. Hypergraph neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019*; Volume 33, pp. 3558–3565. [\[CrossRef\]](#)
37. Jiang, J.; Wei, Y.; Feng, Y.; Cao, J.; Gao, Y. Dynamic Hypergraph Neural Networks. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19, International Joint Conferences on Artificial Intelligence Organization, Macao, China, 10–16 August 2019*; pp. 2635–2641. [\[CrossRef\]](#)
38. Wang, J.; Ding, K.; Hong, L.; Liu, H.; Caverlee, J. Next-item Recommendation with Sequential Hypergraphs. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, China, 25–30 July 2020*; Association for Computing Machinery: New York, NY, USA, 2020; pp. 1101–1110. [\[CrossRef\]](#)
39. Wang, J.; Ding, K.; Zhu, Z.; Caverlee, J. Session-based Recommendation with Hypergraph Attention Networks. In *Proceedings of the 2021 SIAM International Conference on Data Mining (SDM), Virtual Event, 29 April–1 May 2021*; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 2021; pp. 82–90. [\[CrossRef\]](#)
40. Yu, J.; Yin, H.; Li, J.; Wang, Q.; Hung, N.Q.V.; Zhang, X. Self-Supervised Multi-Channel Hypergraph Convolutional Network for Social Recommendation. In *Proceedings of the Web Conference 2021, Virtual Event, 19–23 April 2021*; Association for Computing Machinery: New York, NY, USA, 2021; pp. 413–424. [\[CrossRef\]](#)
41. Zheng, X.; Luo, Y.; Sun, L.; Ding, X.; Zhang, J. A novel social network hybrid recommender system based on hypergraph topologic structure. *World Wide Web* **2018**, *21*, 985–1013. [\[CrossRef\]](#)
42. Ding, C.; Zhao, Z.; Li, C.; Yu, Y.; Zeng, Q. Session-based recommendation with hypergraph convolutional networks and sequential information embeddings. *Expert Syst. Appl.* **2023**, *223*, 119875. [\[CrossRef\]](#)
43. Lee, G.; Ko, J.; Shin, K. Hypergraph motifs: Concepts, algorithms, and discoveries. *Proc. VLDB Endow.* **2020**, *13*, 2256–2269. [\[CrossRef\]](#)
44. Zhang, X.; Lin, H.; Xu, B.; Li, C.; Lin, Y.; Liu, H.; Ma, F. Dynamic intent-aware iterative denoising network for session-based recommendation. *Inf. Process. Manag.* **2022**, *59*, 102936. [\[CrossRef\]](#)

45. Xia, X.; Yin, H.; Yu, J.; Shao, Y.; Cui, L. Self-Supervised Graph Co-Training for Session-based Recommendation. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management, Virtual Event, 1–5 November 2021*; Association for Computing Machinery: New York, NY, USA, 2021; pp. 2180–2190. [[CrossRef](#)]
46. Liu, Z.; Wang, Y.; Liu, T.; Zhang, L.; Li, W.; Liao, J.; He, D. Semantic-enhanced Contrastive Learning for Session-based Recommendation. *Knowl.-Based Syst.* **2023**, *280*, 111001. [[CrossRef](#)]
47. Li, Y.; Gao, C.; Luo, H.; Jin, D.; Li, Y. Enhancing Hypergraph Neural Networks with Intent Disentanglement for Session-based Recommendation. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, Madrid, Spain, 11–15 July 2022*; pp. 1997–2002. [[CrossRef](#)]
48. Wang, H.; Yan, S.; Wu, C.; Han, L.; Zhou, L. Cross-view temporal graph contrastive learning for session-based recommendation. *Knowl.-Based Syst.* **2023**, *264*, 110304. [[CrossRef](#)]
49. Qiao, S.; Zhou, W.; Wen, J.; Zhang, H.; Gao, M. Bi-channel Multiple Sparse Graph Attention Networks for Session-based Recommendation. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management, Birmingham, UK, 21–25 October 2023*; Association for Computing Machinery: New York, NY, USA, 2023; pp. 2075–2084. [[CrossRef](#)]
50. Zhang, L.; Shen, D.; Kou, Y.; Nie, T. Multi-perspective learning for enhanced user preferences for session-based recommendation. *Knowl.-Based Syst.* **2024**, *298*, 111997. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.