Electronic Research Archive

https://www.aimspress.com/journal/era

ERA, 34(9): 6513–6541.
DOI: 10.3934/era.2026285
Received: 01 April 2026
Revised: 30 June 2026
Accepted: 07 July 2026
Published: 27 July 2026

Research article

Hypergraph node influence ranking via fusing local topology and higher-order fuzzy semantics

Chuan Ran, Xindong Zhang* and Juan Liu

College of Big Data Statistics, Guizhou University of Finance and Economics, Guiyang 550025, China

* **Correspondence:** Email: liaoyuan1126@163.com.

Abstract: Identifying influential nodes in hypergraphs is essential for applications in social network analysis and bioinformatics. Conventional centrality metrics, designed for pairwise interactions, fail to capture the higher-order dependencies inherent to hypergraphs. To address this, we have proposed the HDCN (hypergraph dual-branch convolutional network), a dual-branch convolutional network that fuses local and global higher-order signals. The HDCN adopts two existing centrality measures as complementary feature channels, namely, multi-scale hyper-degree centrality (HD) for local topology and hypergraph distance-based fuzzy centrality (HDF) for higher-order neighborhood aggregation, and proposes a dual-branch convolutional architecture with a fully τ-aligned pairwise ranking objective to fuse them effectively. The two branches are separately encoded by convolutional layers and subsequently fused to yield robust node influence estimates. The task was formulated as a pairwise learning-to-rank problem optimized directly for Kendall's τ, with ground-truth rankings derived from Monte Carlo susceptible infected recovered (SIR) simulations. Experiments on five real-world hypergraph datasets showed that the HDCN outperforms traditional centrality baselines, ranking first on four of five datasets. On the structurally complex Geometry dataset, the HDCN achieves $\tau = 0.5253$, a 56% relative gain over the strongest traditional baseline (EHDF, $\tau = 0.3368$). the HDCN also surpasses learning-based methods including the hypergraph neural network (HGNN) on all five datasets, confirming the effectiveness of our dual-branch higher-order fusion strategy.

Keywords: hypergraph; node influence ranking; convolutional neural network; fuzzy centrality; SIR model

1. Introduction

The field of network analysis has long relied on the paradigm of simple graphs, where interactions are represented as pairwise links and node importance is quantified using classical centrality measures such as degree, closeness, and betweenness [1, 2]. This framework has driven significant advances

across diverse domains, including transportation systems [3], power grids [4], social networks [5], and biological interactomes [6, 7].

However, many real-world systems involve group interactions that extend beyond dyadic relationships—such as co-authorship networks [8], group communications [9], and biochemical pathways [10]. Projecting these higher-order relations onto simple graphs often oversimplifies the underlying structure, obscuring functional dependencies, blurring overlapping group memberships, and distorting diffusion dynamics. In contrast, hypergraphs offer a natural and faithful representation for such systems by allowing hyperedges to connect arbitrary numbers of nodes, thereby preserving intrinsic higher-order organization [11, 12]. Consequently, accurately identifying influential nodes in hypergraphs has emerged as a critical research challenge with broad applications in information diffusion, viral marketing, and infrastructure resilience.

Existing methods for hypergraph node ranking largely fail to capture these nuanced higher-order structures, and can be grouped into three broad categories, each with distinct yet interrelated limitations.

Structural centrality measures. The most direct approaches rely on hand-crafted scalar or spectral descriptors. Some methods project hypergraphs onto simple graphs, discarding group-level semantics [13], while others employ basic scalar centralities such as hyper-degree centrality (HD) [14], which overlook complex overlap patterns. More principled interpretations are offered by tensor eigenvector centralities [15, 16] and spectral methods based on hypergraph Laplacians [17], yet these suffer from high computational cost, restrictive uniformity assumptions, or reversion to graph expansions.

Learning-based hypergraph models. Hypergraph neural networks such as the HGNN [18] and HyperGCN [19] have advanced higher-order relational modeling, but are predominantly designed for node or hyperedge classification and often implicitly depend on pairwise projections, compromising higher-order information integrity. These architectures are not tailored for influence ranking, where the relative ordering of nodes is the primary objective.

Dynamics-aware and optimization-oriented methods. A further line of work addresses the temporal propagation of influence directly. Contagion frameworks on hypergraphs [12, 20] and heuristic seed-selection algorithms [21, 22] model group-influenced spreading but emphasize system-level dynamics rather than scalable node-level ranking. More recently, state-transition and neural approaches have been proposed to capture granular node-level influence: Pan et al. [23] developed a Markovian state-transition method, and Zhang et al. [24] proposed the hypergraph influence prediction (HIP) framework integrating distance-centrality features with neural ODEs. Building on this, Pan et al. [25] formulated the influence regulation (IR) problem and designed an improved discrete particle swarm optimization (IDPSO) algorithm incorporating an efficient multi-hop influence evaluator (EIMA)—to identify seed sets whose spread closely approximates a predetermined target scale, shifting the paradigm from influence maximization to precise influence control. Qu et al. [26] further advanced multi-objective evolutionary frameworks for hypergraph influence maximization using efficient two-hop approximations.

Despite these notable advances in state-transition modeling and optimization, a core limitation persists: most existing hypergraph influence evaluation methods depend on handcrafted structural descriptors or predefined propagation rules. These approaches typically employ fixed aggregation mechanisms that struggle to adaptively capture complex, nonlinear interactions among local topology, hyperedge overlaps, and long-range dependencies. Although data-driven hypergraph learning frameworks (e.g., HGNN [18] and HyperGCN [19]) have shown promise in extracting latent patterns, they are primarily designed for classification, link prediction, or representation learning—not influence ranking.

Moreover, node influence identification is fundamentally a ranking task. In applications such as viral marketing and infrastructure protection, the relative ordering of nodes often matters far more than absolute scores. Yet few studies have developed end-to-end frameworks that directly optimize ranking performance while preserving higher-order semantics. These gaps motivate a task-aware, data-driven framework that satisfies three key requirements: (i) faithfully preserving higher-order hypergraph information without excessive simplification; (ii) adaptively integrating local and global influence signals through learned representations; and (iii) directly optimizing node rankings under a ranking-aware objective aligned with the target metric (Kendall's τ).

To address these challenges, we propose the HDCN (hypergraph dual-branch convolutional network), a novel learning-based framework whose novelty lies in feature fusion and task alignment rather than in redefining the underlying centrality measures. We emphasize that both HD [14] and HDF [27] are existing measures adopted as fixed feature channels; the contributions of this work concern how they are fused and optimized. Our approach comprises three core components:

First, we adopt the existing hypergraph distance-based fuzzy centrality (HDF) [27] as a higher-order feature channel. HDF employs fuzzy-set weighting across multiple s-distance layers to aggregate influence from overlapping higher-order neighborhoods, capturing structural patterns that HD alone cannot represent. We leverage this existing formulation without modification, using it as a principled higher-order descriptor within our learning pipeline. Second, we propose a dual-branch convolutional architecture that jointly encodes local structural signals (from HD) and higher-order patterns (from HDF) within a shared learning framework. This is the primary architectural contribution: the two branches are separately convolved over local receptive fields and fused to produce a unified influence score, with 2-section graph projection used solely for neighborhood sampling so that higher-order semantics are preserved end-to-end. Third, we formulate node importance estimation as a pairwise learning-to-rank problem, optimized via a margin-based ranking loss that directly targets Kendall's τ. Supervision is provided by node-level rankings derived from extensive Monte Carlo SIR simulations, with the entire training pipeline—including validation and model selection—guided by the same ranking metric, constituting a fully τ-aligned end-to-end training paradigm.

In summary, this work makes the following key contributions:

- We repurpose two existing hypergraph centrality measures, HD [14] and HDF [27], as complementary feature channels within a unified learning pipeline, demonstrating that their combination substantially outperforms either measure alone and surpasses prior learning-based baselines.
- We propose a dual-branch convolutional architecture that fuses local topological signals (HD branch) with higher-order fuzzy-semantic signals (HDF branch) through learned convolutional encoders, with graph projection restricted to neighborhood sampling so that no higher-order information is discarded during feature extraction.
- We establish a fully τ-aligned pairwise learning-to-rank paradigm: ground-truth rankings are derived from Monte Carlo SIR simulations, training is optimized via a margin-based ranking loss targeting Kendall's τ, and model selection is guided by the same metric throughout, achieving consistent and strong ranking performance across diverse hypergraph datasets.

2. Methodology

This paper introduces a learning-based framework for ranking influential nodes in hypergraphs, which integrates structural feature extraction with a pairwise ranking objective. The proposed HDCN model operates in three stages: (i) multi-scale node feature extraction via hypergraph-aware centrality functions, (ii) structural channel representation encoding local and higher-order topology, and (iii) influence score prediction through a dual-branch convolutional architecture. The overall framework is illustrated in Figure 1.

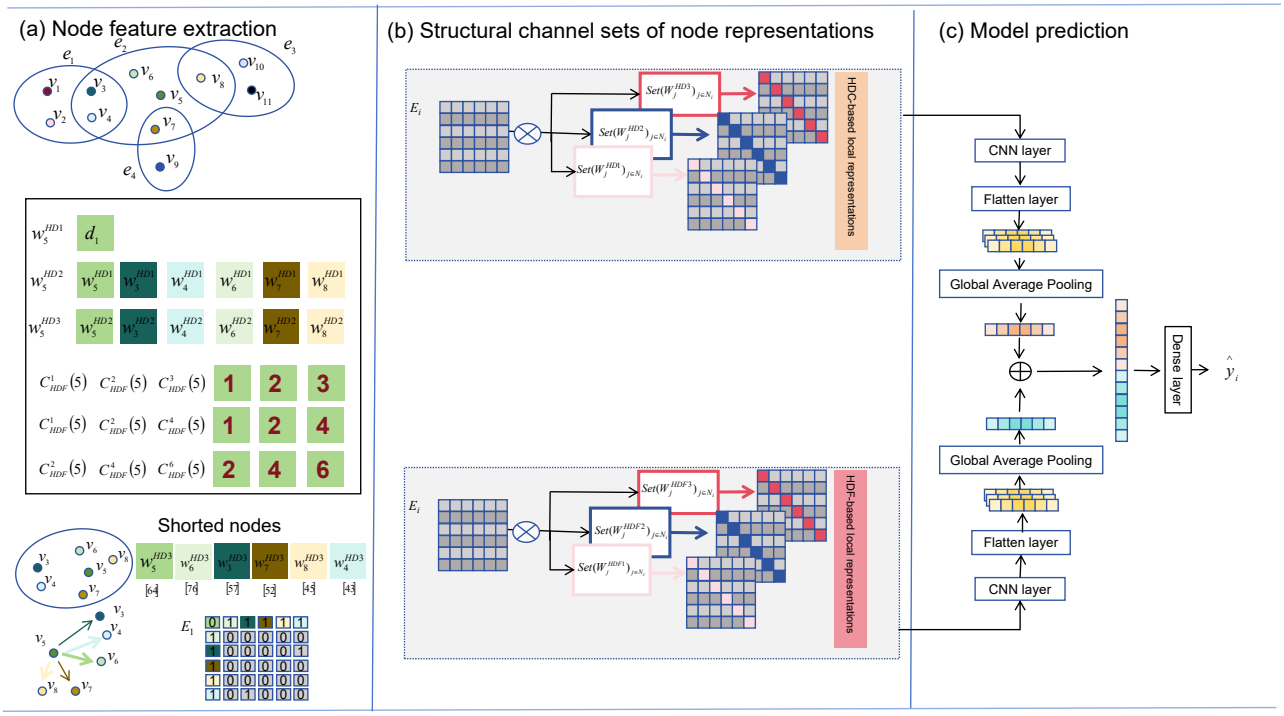


Figure 1. An illustrative overview of the proposed HDCN framework. The workflow is organized into three steps to demonstrate the process from hypergraph input to node influence scoring: (a) Hypergraph construction and multi-scale feature extraction: A toy hypergraph with 11 nodes and 4 hyperedges is shown. For each node, we compute HD and HDF. A local receptive field of size $L = 5$ is constructed based on the 2-hop neighborhood using 2-section graph projection. The actual neighborhood window size used in all experiments is $L = 40$. (b) Dual-branch structural representation: For a target node, the HD and HDF weight profiles within its local receptive field are extracted and arranged into two distinct structural branches. This forms a 6-channel input tensor encoding both local degree information and higher-order fuzzy influence. (c) Lightweight CNN prediction stage: The dual-branch local structural tensor is fed into the HDCN model. The network fuses local and higher-order patterns to output a scalar influence score for the target node, a process repeated for all nodes to obtain the final ranking. This design preserves higher-order information while maintaining computational efficiency.

We begin by formalizing the hypergraph representation. Let $\mathcal{H} = (V, E)$ denote a hypergraph with node set $V = \{v_1, \ldots, v_n\}$ and hyperedge set $E = \{e_1, \ldots, e_m\}$, where each hyperedge $e \subseteq V$ satisfies $|e| \geq 2$. The incidence matrix $\mathbf{I} \in \{0, 1\}^{n \times m}$ encodes node-hyperedge membership, with $I_{v,e} = 1$ if $v \in e$. To capture local connectivity, we derive the 2-section graph via the adjacency matrix:

$$\mathbf{A} = \mathbf{I}\mathbf{I}^{\mathrm{T}} - \mathrm{diag}(\mathbf{I}\mathbf{1}_m). \tag{2.1}$$

Here, $A_{u,v} > 0$ indicates that nodes u and v share at least one hyperedge. Crucially, $\mathbf{A}$ is used only to define localized receptive fields; higher-order structural information is preserved explicitly through the adopted HDF, avoiding semantic loss from graph reduction.

To quantify node influence at multiple scales, we adopt two complementary centrality measures, namely, HD and HDF. HD captures local involvement through hyperedge participation. Its first-order form is defined as:

$$\mathrm{HD}_1(v_i) = \sum_{e_j \in E} \mathbf{1}(v_i \in e_j), \tag{2.2}$$

where $\mathbf{1}(\cdot)$ is the indicator function. We extend this notion to broader neighborhoods via:

$$\mathrm{HD}_2(v_i) = \mathrm{HD}_1(v_i) + \sum_{v_j \in \mathcal{N}_1(v_i)} \mathrm{HD}_1(v_j), \tag{2.3}$$

$$\mathrm{HD}_3(v_i) = \mathrm{HD}_2(v_i) + \sum_{v_j \in \mathcal{N}_2(v_i)} \mathrm{HD}_2(v_j), \tag{2.4}$$

where $\mathcal{N}_k(v_i)$ $(k = 1, 2)$ denotes the k-hop neighborhood of v_i in the 2-section graph. This multi-scale aggregation captures both immediate and extended local influence. However, HD cannot capture the higher-order interactions.

To address this limitation of HD, we adopt HDF, a formulation designed to model the higher-order interactions that HD cannot capture. This measure employs a multi-s design, where each value of $s \in \{1, 2, 3\}$ defines an independent channel that captures node influence at a distinct overlap threshold. Central to HDF is the s-distance, denoted by $d_s^v(i, j)$, which is defined as the shortest path length between node v_i and node v_j. For more details, one can refer to [27]. In this graph, two hyperedges are connected if and only if they share at least s nodes. The s-radius of node v_i is given by:

$$z_i^s = \max_{j \in V} d_s^v(i, j). \tag{2.5}$$

We discretize the range $[1, z_i^s]$ into $L_i^s = \lfloor z_i^s / r \rfloor$ layers, where $r \geq 1$ is the layer granularity parameter controlling the resolution of the distance discretization. We set $r = 1$ in all experiments and we define $n_i^s(\ell)$ as the number of nodes at distance ℓ from node i. A fuzzy membership function

$$X_i^s(\ell) = \exp\left(-\ell^2 / (L_i^s)^2\right) \tag{2.6}$$

is introduced to weight nodes in each layer, yielding a fuzzy mass $f_i^s(\ell) = n_i^s(\ell) \cdot X_i^s(\ell)$ and a total mass $F_i^s = \sum_{\ell=1}^{L_i^s} f_i^s(\ell)$.

Following the theoretical framework in [27], these masses are normalized to form a probability-like distribution. Specifically, we introduce a scaling factor $1/e$ to refine the values into the range $[0, 1/e]$, which ensures numerical stability for the subsequent logarithmic operations:

$$p_i^s(\ell) = \frac{1}{e} \cdot \frac{f_i^s(\ell)}{F_i^s}, \tag{2.7}$$

such that $\sum_\ell p_i^s(\ell) = 1/e$. Since this constant factor is applied uniformly across all nodes, it does not affect the relative ordering of node influences. Finally, the HDF score is computed as the distance-weighted entropy:

$$\mathrm{HDF}^{(s)}(i) = \sum_{\ell=1}^{L_i^s} \frac{-p_i^s(\ell) \ln p_i^s(\ell)}{\ell^2}. \tag{2.8}$$

By evaluating $\mathrm{HDF}^{(s)}(i)$ for multiple s, we obtain a multi-channel representation of higher-order influence.

The Gaussian form $X_i^s(\ell) = \exp(-\ell^2/(L_i^s)^2)$ is adopted for three complementary reasons. First, it provides smooth, differentiable distance decay without hard truncation, which faithfully models the diminishing yet non-zero influence of distant nodes. Second, the radius $L_i^s = \lfloor z_i^s/r \rfloor$ is computed individually for each node from its actual s-distance profile, so the effective shape of the weighting curve differs across nodes and datasets—the function is node-adaptively parameterized rather than globally fixed. Third, empirical sensitivity experiments test four distinct S configurations, and bootstrap significance tests confirm that all pairwise differences are non-significant ($p > 0.05$), demonstrating that the fixed functional form generalizes robustly across diverse hypergraph topologies.

It is worth noting that the original HDF in [27] aggregates entropy across multiple s values by averaging, yielding a single scalar $C_{\mathrm{HDF}}(i) = \frac{1}{s_m} \sum_{s=1}^{s_m} \mathrm{HDF}^{(s)}(i)$. In contrast, our formulation retains each s-specific score as an independent feature channel $\mathrm{HDF}^{(s)}(i)$, without cross-s averaging. This per-s decomposition is essential for the subsequent dual-branch CNN to learn which overlap scales are most discriminative for a given hypergraph topology, thereby preserving fine-grained structural information at different hyperedge overlap thresholds.

To ensure spatial consistency, we construct the ordered sequence $\pi(v_i)$ utilizing HD_3 as a multi-scale local importance metric. The sequence is anchored by the target node v_i, followed by its Top-L neighbors sorted in descending order of their HD_3 numerical values. Taking node v_5 in Figure 1 as an example, we identify its neighbors v_6, v_3, v_7, v_8, v_4, which correspond to HD_3 numerical values of 76, 57, 52, 45, and 43, respectively. This ranking yields the fixed sequence $\pi(v_5) = (v_5, v_6, v_3, v_7, v_8, v_4)$. For each centrality function W (which can be HD_k or $\mathrm{HDF}^{(s)}$), we construct a matrix $E_{v_i}(W) \in \mathbb{R}^{(L+1)\times(L+1)}$ that combines adjacency and centrality values. Given the adjacency matrix A of nodes $v_0, v_1, \cdots, v_L$ with elements $a_{lk} = A_{v_l,v_k}$, and denoting the centrality of a node v as W_v, each element of $E_{v_i}(W)$ is defined as follows:

$$E_{v_i}(W)[l,k] = \begin{cases} a_{lk} + W_{v_i}, & l = k, \\ a_{0k} \cdot W_{v_k}, & l \neq k, l = 0, \\ a_{l0} \cdot W_{v_l}, & l \neq k, k = 0, \\ a_{lk}, & \text{otherwise.} \end{cases} \tag{2.9}$$

This formulation emphasizes connections involving the central node while preserving local structure, and the full channel sets are as follows:

$$\mathcal{E}_{v_i}^{\mathrm{HD}} = \{E_{v_i}(\mathrm{HD}_1), E_{v_i}(\mathrm{HD}_2), E_{v_i}(\mathrm{HD}_3)\}, \tag{2.10}$$

$$\mathcal{E}_{v_i}^{\mathrm{HDF}} = \{E_{v_i}(\mathrm{HDF}^{(s_1)}), E_{v_i}(\mathrm{HDF}^{(s_2)}), E_{v_i}(\mathrm{HDF}^{(s_3)})\}. \tag{2.11}$$

These six channels are processed by a dual-branch CNN. The outputs from both branches are then concatenated and fed through a linear layer to generate the final influence score.

To supervise the model, we generate ground-truth influence rankings via Monte Carlo SIR simulations conducted directly on the original hypergraph structure (without graph projection). For each node v_i, we designate it as the single initial seed and perform $R = 200$ independent simulation runs. In each run, the process unfolds as follows: at every discrete time step, for each hyperedge containing at least one infected node, every susceptible node within that hyperedge becomes infected with a dynamic infection probability

$$\beta' = \min(1,\ \beta \times (1 + 0.1 \times I_e)), \tag{2.12}$$

where I_e denotes the current number of infected nodes in the hyperedge, and the $\min(\cdot, 1)$ truncation ensures the result remains a valid probability. This mechanism explicitly captures the group effect of higher-order interactions. Infected nodes recover independently with probability μ per time step. The simulation continues until no infected nodes remain or a maximum of 1000 steps is reached.

The influence of seed node v_i is quantified by its burst size $\Omega_{v_i}^{(r)}$, defined as the total number of nodes that eventually recover in the r-th simulation run. To reduce stochastic variance, we compute the average influence score:

$$\widehat{\Omega}_{v_i} = \frac{1}{R} \sum_{r=1}^{R} \Omega_{v_i}^{(r)}. \tag{2.13}$$

Following standard practice in hypergraph influence ranking [12, 20], we set the base infection rate $\beta = 0.02$ and recovery rate $\mu = 0.1$, with $R = 200$ Monte Carlo runs per node. The group-reinforcement coefficient 0.1 in Eq (2.12) is a fixed modeling choice controlling the strength of higher-order contagion amplification; we hold it constant across all experiments as a structural assumption rather than a tunable hyperparameter. These values place the system in a moderate, sub-saturation propagation regime where meaningful differentiation between high- and low-influence nodes is preserved. To verify that the resulting ground-truth rankings are not artifacts of this specific parameter choice, we conducted a systematic robustness analysis across five alternative SIR configurations covering a wide range of infection and recovery rates (see Section 4.5 for the full results and discussion).

Finally, the average influence scores $\widehat{\Omega}_{v_i}$ are converted into normalized ranks $y_i \in [0, 1]$ to preserve relative ordering while facilitating model training:

$$y_i = \frac{\text{rank}(v_i) - 1}{N - 1}, \tag{2.14}$$

where $\text{rank}(v_i)$ is the descending rank of node v_i according to $\widehat{\Omega}_{v_i}$, and N is the total number of nodes. Note that because ranking is in descending order of influence (rank 1 denotes the most influential node), a *smaller* y_i corresponds to a *more* influential node: the top-ranked node receives $y_i = 0$, while the least influential receives $y_i = 1$. Accordingly, a valid training pair (i, j) satisfies $y_i < y_j$, meaning node i is more influential than node j, and the loss in Eq (2.12) enforces $s_i > s_j$, consistent with the convention that larger s indicates higher influence.

The model is optimized using a margin-based pairwise ranking objective. Specifically, taking a subset of the set of nodes, denoted as $\mathcal{B}$, we construct the set of valid training pairs $\mathcal{P}_\mathcal{B} = \{(i, j) \mid y_i < y_j\}$. The loss function is formulated as the squared hinge loss:

$$\mathcal{L}_{\text{pair}} = \frac{1}{|\mathcal{P}_\mathcal{B}|} \sum_{(i,j)\in\mathcal{P}_\mathcal{B}} \max\left(0, \theta - (s_i - s_j)\right)^2. \tag{2.15}$$

Here, s_i and s_j denote the predicted influence scores for nodes i and j, respectively, with larger s indicating higher influence. $\theta > 0$ is a margin hyperparameter. Furthermore, to ensure numerical stability and mitigate feature skewness, all input features are preprocessed through log-transformation and subsequent min-max scaling, using statistics computed exclusively from the training set.

In summary, our method addresses the problem of identifying top-k influential nodes in a hypergraph $\mathcal{H}$. Motivated by the definition of true influence $I(v) = \mathbb{E}[\Omega_v]$, the expected burst size when v is the seed, we design a scoring function to approximate it. This is achieved by distilling both local connectivity and nuanced higher-order structural signals through a principled feature extraction and ranking framework.

3. Experiments

3.1. Experimental setup

This study systematically evaluates the effectiveness and robustness of the proposed HDCN framework through a series of experiments designed to address the following three research questions (RQs):

- RQ1 (Parameter sensitivity): How does the fuzzy weighting parameter in HDF influence ranking accuracy, and to what extent does it capture higher-order dependencies?
- RQ2 (Ablation study): Does the multi-branch fusion of HD and HDF features yield significant performance gains compared to single-branch alternatives?
- RQ3 (Comparative performance): How does HDCN perform against existing baselines, ranging from classical centrality measures to recent hypergraph-specific ranking methods, in terms of rank correlation?

These questions are collectively designed to assess the model's component efficacy, fusion strategy, and overall scalability across hypergraphs of varying complexity.

3.2. Datasets

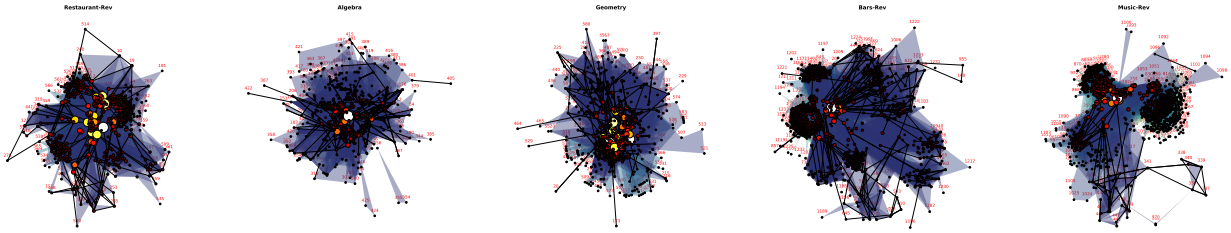
We conducted experiments on five real-world hypergraph datasets from two domains: academic question-answering communities (Algebra, Geometry) and location-based review networks (Restaurants-Rev, Bars-Rev, Music-Rev).

Table 1. Basic topological characteristics of the datasets. N: nodes, M: hyperedges, $\langle k \rangle$: average node degree, $\langle k^H \rangle$: average hyperedge size, $\langle k^E \rangle$: average hyperedges per node, $\langle l \rangle$: average shortest path length, C: clustering coefficient.

Datasets	N	M	$\langle k \rangle$	$\langle k^H \rangle$	$\langle k^E \rangle$	$\langle l \rangle$	C
Bars-Rev	1234	1194	174.30	9.62	9.93	2.10	0.58
Restaurants-Rev	565	601	79.75	8.14	7.66	1.98	0.54
Music-Rev	1106	694	167.87	9.49	15.13	1.99	0.62
Algebra	423	1268	78.90	19.53	6.52	1.95	0.79
Geometry	580	1193	164.79	21.52	10.47	1.75	0.82

The selected datasets cover a wide spectrum of network characteristics, ranging from sparse to dense connectivity (indicated by clustering coefficient C) and varying average path lengths $\langle l \rangle$. This

diversity ensures that our evaluation tests the model's generalization capability across different influence propagation patterns.



(a) Restaurants-Rev **(b)** Algebra **(c)** Geometry **(d)** Bars-Rev **(e)** Music-Rev

Figure 2. Visualization of the five datasets. Node sizes are proportional to their degree centrality, illustrating variations in structural compactness and influence distribution.

3.3. Training and evaluation setup

To ensure experimental reproducibility, all datasets were partitioned using a consistent two-stage random splitting strategy with a fixed random seed of 42. Specifically, for each dataset comprising N nodes, the full feature matrix constructed from multi-scale HD and HDF channels and the corresponding SIR ground-truth labels were first divided into a training set containing 80% of the samples and a held-out test set containing the remaining 20%. Subsequently, the training set was further stratified into a model training subset accounting for 64% of the total samples and a validation set accounting for 16% of the total samples. The validation set was utilized exclusively for hyperparameter tuning, early stopping, and model selection. This procedure yielded an approximate train/validation/test ratio of 64%/16%/20%. To prevent data leakage, normalization parameters were fitted solely on the training subset and subsequently applied to both the validation and test sets.

For datasets with relatively limited sample sizes, such as *Restaurants-Rev* and *Bars-Rev*, we adhered to this fixed-ratio split rather than employing k-fold cross-validation, in order to preserve a sufficiently large test set for a reliable evaluation of Kendall's τ. To address the concern that results based on a single partition may be partition-specific, we additionally repeated all experiments over five independent random train/validation/test splits using seeds in $42, 7, 13, 21, 100$. The resulting mean and standard deviation of Kendall's τ are presented in the robustness analysis section. The five-split means are consistently close to or slightly higher than the single-split results obtained with seed 42, while all standard deviations remain below 0.07, confirming that HDCN's ranking performance is stable and is not an artifact of a particular data partition.

3.4. Baseline methods

The performance of the HDCN model is evaluated against a comprehensive suite of 14 baseline methods. These baselines are categorized into four distinct groups based on their theoretical underpinnings: (i) local structural metrics, (ii) global path and spectral metrics, (iii) advanced entropy-based metrics, and (iv) learning-based models. A comparative summary is presented in Table 2.

Table 2. Summary of baseline methods categorized by theoretical basis.

Method	Category	Scope	Key Mechanism
DC, HD, HEDC	Topology	Local	node or hyperedge degree counts and propagation
HCI	Topology	Local+	collective hyper-degree of l-hop neighborhood
ECC, HCC	Path-based	Global	shortest-path distances (max or harmonic mean)
HGC	Gravity	Global	gravity model with multi-scale distance
VC	Spectral	Global	eigenvector centrality of hyperedge interactions
HVC	Entropy	Global	multi-scale Neumann entropy sensitivity
HDF, EHDF	Entropy	Global	fuzzy centrality and distance-weighted entropy
MLP, HGNN	Learning-based	Local+Global	flat MLP on neighbor features and a hypergraph neural network with ranking head

3.4.1. Local structural metrics

These methods rely on immediate connectivity and local neighborhoods to quantify influence.
Degree centrality (DC) [28]: This measures the potential for direct activity in the network. For a node v_k:

$$C_D(v_k) = \sum_{i=1}^{n} a(v_i, v_k), \tag{3.1}$$

where n is the total number of nodes and $a(v_i, v_k)$ represents the entry in the adjacency matrix (1 if nodes v_i and v_k are connected, 0 otherwise). In the context of hypergraphs, this is calculated on the projected 2-section graph.
Hyper-degree centrality (HD): This metric quantifies the direct participation of a node in higher-order interactions. HD explicitly counts the number of hyperedges incident to a node v_i:

$$k_i^H = \sum_{j=1}^{m} I_{ij}, \tag{3.2}$$

where m denotes the total number of hyperedges in the hypergraph and I is the incidence matrix with entries defined such that $I_{ij} = 1$ if node v_i belongs to hyperedge e_j and $I_{ij} = 0$ otherwise.
Hyperedge degree centrality (HEDC) [29]: This metric evaluates node influence by aggregating the importance of incident hyperedges derived from the hypergraph's line graph. The HEDC of a node v_i is defined as:

$$\mathcal{K}_i = \sum_{e_{ij} \in E_i} \frac{\mathcal{K}_{ij}^e}{|e_{ij}|}, \tag{3.3}$$

where $E_i = \{e_{i1}, e_{i2}, \ldots, e_{iK}\}$ is the set of hyperedges containing v_i. The term $\mathcal{K}_{ij}^e$ represents the degree of the hyperedge e_{ij} in the line graph (i.e., the count of hyperedges overlapping with e_{ij}), while $|e_{ij}|$ is the size of e_{ij}, acting as a normalization factor.
Collective influence (HCI) [30]: This metric evaluates a node's structural influence by considering the hierarchy of its local neighborhood. Extended to hypergraphs, it is defined as:

$$HCI(i) = (k_i^H - 1) \sum_{v_j \in \partial Ball(v_i, l)} (k_j^H - 1), \tag{3.4}$$

where the term $\partial Ball(v_i, l)$ represents the set of nodes located at a shortest path distance l (the radius) from v_i. k_i^H and k_j^H are defined as in Eq (3.2).

3.4.2. Global path and spectral metrics

These methods capture long-range dependencies and global positional advantages.

Eccentricity centrality (ECC) [27,31]: This is based on the hypothesis that highly influential hyperedges have a shorter distance from the other hyperedges. Then, the eccentricity centrality of hyperedge e_i is defined as

$$\epsilon_i^e = \frac{1}{\max_{e_j \in C_1}\{d_1^e(i,j)\}}, \tag{3.5}$$

where C_1 represents the 1-connected component composed of 1-connected hyperedges. Similarly, we distribute the centrality value of the hyperedge evenly to each node within it. Given node v_i and its associated hyperedge set E_i, the eccentricity centrality of node v_i is given by the following equation:

$$\epsilon_i = \sum_{e_{ij} \in E_i} \frac{\epsilon_{ij}^e}{|e_{ij}|}, \tag{3.6}$$

where ϵ_{ij}^e is the eccentricity centrality of e_{ij}.

Harmonic closeness centrality (HCC) [31]: This metric evaluates influence by calculating the average reciprocal distance between hyperedges and distributing the score to their constituent nodes. The centrality of a hyperedge e_i is defined as:

$$h_i^e = \frac{1}{m-1} \sum_{e_i, e_j \in E, i \neq j} \frac{1}{d_1^e(i,j)}, \tag{3.7}$$

where m is the total number of hyperedges and $d_1^e(i,j)$ represents the shortest path distance between hyperedges e_i and e_j. Also, the harmonic closeness centrality of a node v_i is defined as:

$$h_i = \sum_{e_{ij} \in E_i} \frac{h_{ij}^e}{|e_{ij}|}, \tag{3.8}$$

where h_{ij}^e is the harmonic closeness centrality of e_{ij}.

Vector centrality (VC) [32]: This spectral method evaluates node importance by analyzing the eigenvector centrality of the hyperedges. It operates by projecting the hypergraph into a line graph to compute hyperedge scores, which are then aggregated to the nodes. The vector centrality of a node v_i is defined as:

$$c_i = \sum_{k=1}^{K} c_{ik}, \tag{3.9}$$

where K denotes the number of hyperedges in the set E_i containing v_i. The term c_{ik} corresponds to the eigenvector centrality value of the specific hyperedge e_{ik}, calculated from the adjacency matrix of the hypergraph's line graph.

Gravity-based centrality (HGC) [33]: Inspired by Newton's law of universal gravitation, this metric evaluates node influence by modeling the gravitational attraction between nodes. It treats node degree as mass and aggregates the forces exerted by all other nodes. The HGC of a node v_i is defined as:

$$HGC(i) = \sum_{j \neq i} \frac{k_i k_j}{(d^H(i,j))^2}, \tag{3.10}$$

where k_i and k_j denote the hyper-degrees (mass) of nodes v_i and v_j, respectively. The denominator $d^H(i, j)$ represents the effective distance, calculated as a weighted sum of s-distances: $d^H(i, j) = \sum_{s=1}^{s_m} \frac{1}{s^2} d_s^v(i, j)$, where $d_s^v(i, j)$ is the shortest path length based on s-adjacency and s_m represents the maximum order of s-distance considered.

3.4.3. Advanced entropy-based metrics

These methods utilize information theory and multi-scale analysis to quantify structural complexity. *High-order Neumann entropy centrality (HVC)* [34]: This method evaluates node importance by quantifying the structural information loss in the hypergraph's line graphs. It aggregates the entropy contributions of incident hyperedges across multiple scales. The HVC of a node v_i is defined as:

$$HVC(i) = \sum_{s=1}^{s_m} \frac{\Phi_s(i)}{s}. \tag{3.11}$$

Here, the term $\Phi_s(i) = \sum_{e_j \in \Gamma(i)} \frac{\Delta\Theta_s(e_j)}{|e_j|}$ represents the node's importance at scale s, derived from the entropy variation $\Delta\Theta_s(e_j)$ of its incident hyperedges e_j (in the set $\Gamma(i)$), and $|e_j|$ is the size of hyperedge e_j.
Hypergraph distance-based fuzzy centrality (HDF) and extended (EHDF) [27]: These metrics quantify node influence by applying fuzzy set theory to distance-weighted Shannon entropy across higher-order neighborhoods. HDF averages the influence entropy across multiple s-distance orders, which is defined as:

$$C_{HDF}(i) = \frac{1}{s_m} \sum_{s=1}^{s_m} \sum_{\ell=1}^{L_i^s} \frac{-p_i^s(\ell) \ln p_i^s(\ell)}{\ell^2}, \tag{3.12}$$

where L_i^s is the local neighborhood radius for order s. The term $p_i^s(\ell)$ is the probability distribution of the fuzzy node mass at distance ℓ and ℓ^2 serves as a distance decay factor. EHDF unifies the interaction range using an extended radius, which is defined as:

$$C_{EHDF}(i) = \sum_{\ell=1}^{L_i^{es}} \frac{-p(\ell) \ln p(\ell)}{\ell^2}, \tag{3.13}$$

where L_i^{es} is the extended radius for EHDF derived from the average maximum distances across all orders. The term $p(\ell)$ is the probability distribution of the fuzzy node mass at distance ℓ.

3.4.4. Learning-based models

This category includes two learning-based baselines that test whether generic learning architectures suffice for hypergraph influence ranking.
Multi-layer perceptron (MLP): This uses the same neighbor-aware input as HDCN (with $L = 40$)—the six centrality values (HD_1, HD_2, HD_3, $HDF^{(s_1)}$, $HDF^{(s_2)}$, $HDF^{(s_3)}$) of each of the $(L+1) = 41$ nodes in the local receptive field, concatenated into an $(L+1) \times 6 = 246$-dimensional flat vector—but processes it without exploiting the 2D spatial structure of $\mathbf{E}_{v_i}$. This baseline isolates the contribution of CNN-based spatial encoding over flat neighbor feature aggregation.
Hypergraph neural network (HGNN) [18]: This is a classic hypergraph neural network operating on the incidence matrix with HD and HDF node features, equipped with a ranking head. This baseline tests whether generic message-passing architectures suffice for influence ranking.

3.5. Evaluation metrics

To rigorously assess the performance of our model on hypergraph node ranking, we employ the following evaluation metrics:

- Kendall's τ measures the rank correlation between two ordered lists, evaluating the consistency in pairwise comparisons [35]. It is defined as:

$$\tau = \frac{C - D}{\frac{1}{2}n(n-1)}, \tag{3.14}$$

  where n is the number of items, C is the number of concordant pairs, and D is the number of discordant pairs. A pair (i, j) is concordant if the relative order of the two nodes is the same in both rankings and discordant otherwise. Values of τ closer to 1 indicate stronger agreement between rankings.
- The monotonicity index (MI) evaluates the resolution (or discriminative ability) of the ranking list [36]. A good centrality method should assign unique rankings to nodes with different structural properties. The MI is defined as:

$$M(R) = \left(1 - \frac{\sum_{r\in\mathcal{R}} n_r(n_r - 1)}{N(N-1)}\right)^2, \tag{3.15}$$

  where N is the total number of nodes and n_r denotes the number of nodes tied at the same rank r. The set $\mathcal{R}$ represents all unique ranks. The values of $M(R)$ range from 0 to 1. An MI value closer to 1 indicates that the method has a high resolution, effectively distinguishing the influence of different nodes with fewer ties.

4. Results and analysis

This section presents a comprehensive analysis of the experimental results, structured around the research questions outlined earlier. We systematically evaluate the performance of the HDCN model under various configurations and compare it against a range of baseline methods. The empirical evaluation addresses three main perspectives: the sensitivity of the fuzzy parameters in HDF (RQ1), the efficacy of dual-branch fusion versus single-branch variants (RQ2), and the model's ranking performance against traditional centrality baselines and learning-based alternatives (RQ3).

4.1. Impact of fuzzy parameter set S (RQ1)

In this subsection, we examine the sensitivity of the proposed model to the fuzzy parameter set S, which governs the scale of influence propagation and thereby determines the attenuation rate. A smaller S concentrates influence within local neighborhoods, while a larger set value (e.g., a higher integer) extends the model's reach, allowing for the capture of higher-order, long-range dependencies.

To evaluate this, we measured ranking accuracy using Kendall's τ across four distinct parameter configurations: $S = \{1, 2, 3\}$, $\{1, 2, 4\}$, $\{1, 3, 5\}$, and $\{2, 4, 6\}$. The corresponding quantitative results are presented in Table 3, with a visual analysis provided in Figure 3.

The experimental results demonstrate that our method generally maintains robust performance across varying parameter settings. For datasets such as Restaurants-Rev and Algebra, the fluctuation in τ is

Table 3. Impact of different fuzzy parameter sets S on ranking performance.

Datasets	$S = \{1, 2, 3\}$	$S = \{1, 2, 4\}$	$S = \{1, 3, 5\}$	$S = \{2, 4, 6\}$
Restaurants-Rev	0.6448	0.6458	0.6387	0.6471
Algebra	0.6734	0.6713	0.6740	0.6697
Geometry	0.4733	0.4768	0.4706	0.4684
Bars-Rev	0.6144	0.6044	0.6069	0.6046
Music-Rev	0.4096	0.4130	0.4063	0.4239

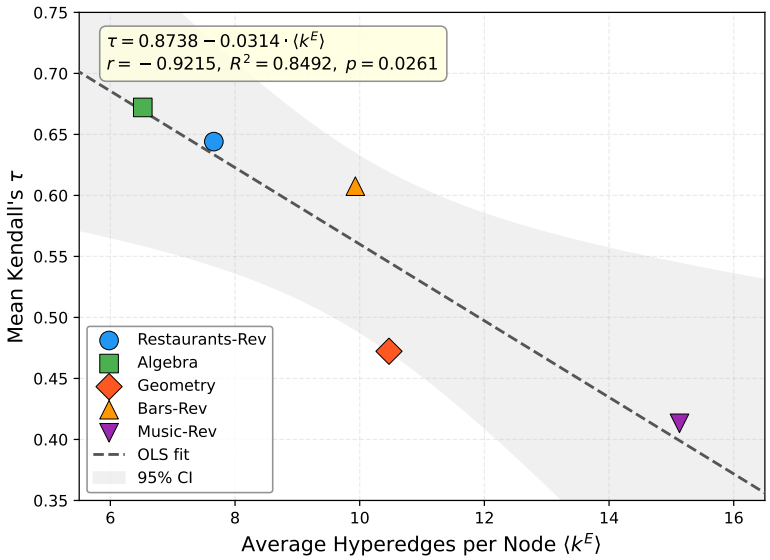


Figure 3. Scatter plot and linear regression analysis of the relationship between ranking accuracy (mean Kendall's τ over four S configurations) and the average number of distinct hyperedges per node ($\langle k^E \rangle$). Each point represents one dataset. The dashed line is the ordinary least-squares regression fit ($\tau = 0.8738 - 0.0314 \cdot \langle k^E \rangle$); the shaded band indicates the 95% confidence interval. The regression yields Pearson $r = -0.9216$, $R^2 = 0.8493$, and $p = 0.0261$, confirming a significant negative relationship. Geometry exhibits the largest negative residual (-0.073), attributable to its uniquely large average hyperedge size ($\langle k^H \rangle = 21.52$) and high clustering coefficient ($C = 0.82$), which compound the structural entanglement beyond what $\langle k^E \rangle$ alone captures.

minimal ($\Delta\tau < 0.01$), indicating that ranking consistency in these networks is largely insensitive to the specific choice of propagation scale.

To assess whether the observed τ variations across S settings are significant, we conduct bootstrap two-sample significance tests. For each S configuration, we collect per-seed Kendall's τ values from seven margin parameters $\theta \in \{0.3, 0.6, 0.9, 1.0, 1.2, 1.5, 1.8\}$, yielding $n = 42$ replicates for $S = \{1, 2, 3\}$ and $n = 21$ replicates for the other settings. Bootstrap resampling (10,000 iterations) is used to compare each alternative S setting against the default $S = \{1, 2, 3\}$. We choose bootstrap over parametric tests because Kendall's τ is a rank-based statistic whose sampling distribution is not guaranteed to be normal.

As shown in Table 4, none of the 15 pairwise comparisons reaches significance at $\alpha = 0.05$ (mean

$\Delta\tau$ ranges from -0.0046 to $+0.0137$, all $p > 0.05$). The 95% bootstrap confidence intervals for all differences span zero, confirming that the small τ variations in Table 3 are attributable to random fluctuation. These results support the conclusion that the model's ranking performance is robust to the specific choice of the fuzzy parameter set S.

However, structurally complex datasets exhibit notable sensitivity. Music-Rev achieves its peak performance ($\tau = 0.4239$) with the largest parameter set $S = \{2, 4, 6\}$. This suggests that networks with complex connectivity may benefit from capturing broader, higher-order context to mitigate local noise, though this performance difference does not reach statistical significance at the $\alpha = 0.05$ level (see Table 4). Conversely, Bars-Rev shows a slight performance decline as S increases, implying that local topological features are sufficient for accurate ranking in this domain, and incorporating distant neighbors may introduce unnecessary noise.

Table 4. Bootstrap significance test results comparing each S setting against the default $S = \{1, 2, 3\}$.

Datasets	Comparison	$\Delta\tau$	p-value	95% CI
	$S = \{1, 2, 4\}$	-0.0025	0.3374	[-0.0072, $+0.0026$]
Algebra	$S = \{1, 3, 5\}$	$+0.0013$	0.5048	[-0.0027, $+0.0052$]
	$S = \{2, 4, 6\}$	-0.0006	0.8254	[-0.0054, $+0.0041$]
	$S = \{1, 2, 4\}$	$+0.0010$	0.8346	[-0.0078, $+0.0100$]
Bars-Rev	$S = \{1, 3, 5\}$	$+0.0077$	0.1862	[-0.0037, $+0.0190$]
	$S = \{2, 4, 6\}$	-0.0018	0.7454	[-0.0122, $+0.0087$]
	$S = \{1, 2, 4\}$	$+0.0024$	0.6542	[-0.0069, $+0.0122$]
Geometry	$S = \{1, 3, 5\}$	-0.0046	0.4716	[-0.0175, $+0.0069$]
	$S = \{2, 4, 6\}$	-0.0039	0.5800	[-0.0181, $+0.0074$]
	$S = \{1, 2, 4\}$	$+0.0037$	0.6768	[-0.0138, $+0.0212$]
Music-Rev	$S = \{1, 3, 5\}$	$+0.0034$	0.6914	[-0.0134, $+0.0205$]
	$S = \{2, 4, 6\}$	$+0.0137$	0.1158	[-0.0035, $+0.0307$]
	$S = \{1, 2, 4\}$	-0.0035	0.3478	[-0.0109, $+0.0037$]
Restaurants-Rev	$S = \{1, 3, 5\}$	-0.0043	0.2408	[-0.0116, $+0.0030$]
	$S = \{2, 4, 6\}$	$+0.0015$	0.6512	[-0.0048, $+0.0080$]

To understand the underlying causes of these performance variations, we conducted a formal regression analysis of the relationship between ranking accuracy and topological properties. Figure 3 presents the scatter plot with the fitted regression line and 95% confidence interval. The analysis reveals a significant negative correlation between Kendall's τ and the average number of distinct hyperedges per node ($\langle k^E \rangle$), with Pearson $r = -0.9216$, $R^2 = 0.8493$, and $p = 0.0261$. The linear regression equation $\tau = 0.8738 - 0.0314 \cdot \langle k^E \rangle$ indicates that each unit increase in $\langle k^E \rangle$ is associated with an average decrease of 0.0314 in ranking accuracy. By synthesizing these regression results with topological statistics, a clear structural dichotomy emerges:

- *Low complexity regime*: Datasets with low $\langle k^E \rangle$, such as Algebra (6.52) and Restaurants-Rev (7.66), consistently achieve superior ranking alignment ($\tau > 0.64$), closely following the regression trend.
- *High complexity regime*: Datasets with high $\langle k^E \rangle$, notably Music-Rev (15.13) and Geometry

(10.47), exhibit significantly reduced performance ($\tau < 0.49$). Among these, Geometry displays the largest negative residual (−0.073), performing below the regression prediction. This deviation is attributable to its compounded structural complexity: beyond the high $\langle k^E \rangle$, Geometry also possesses the largest average hyperedge size ($\langle k^H \rangle = 21.52$) and the highest clustering coefficient ($C = 0.82$) among all datasets, making it an especially challenging case that $\langle k^E \rangle$ alone cannot fully characterize.

These findings establish $\langle k^E \rangle$ as a critical and grounded proxy for structural entanglement, though this association is based on only five data points and would benefit from validation on additional datasets. Nodes participating in numerous distinct hyperedges generate dense, overlapping influence pathways that create topological noise and complicate the linearization of node importance. Notably, other conventional metrics, including $\langle k \rangle$, $\langle k^H \rangle$, and clustering coefficient C fail to exhibit a comparably strong or significant relationship with ranking fidelity when examined individually.

While the proposed HDCN model demonstrates broad robustness, the regression analysis suggests that $\langle k^E \rangle$ may serve as a useful heuristic for parameter selection guidance. In networks exhibiting high $\langle k^E \rangle$ (indicating complex hyperedge overlap), we observe a descriptive trend toward improved performance with larger fuzziness parameters (e.g., $S = \{2, 4, 6\}$); however, since all pairwise S comparisons remain non-significant ($p > 0.05$, Table 4), this should be treated as a practical guideline rather than a statistically validated recommendation. Such configurations broaden the receptive field of the model, potentially capturing the global context necessary to disentangle influence in densely connected environments.

4.2. Single-branch vs dual-branch convolution (RQ2)

This section quantitatively evaluates the efficacy of integrating HD and HDF within our dual-branch CNN architecture. We systematically compare three configurations:

- HD-only: A single-branch model utilizing three channels (HD_1, HD_2, HD_3).
- HDF-only: A single-branch model employing three channels derived from an optimized s-set.
- Full model: The proposed framework combining both branches into a six-channel dual-branch architecture.

Performance is assessed using Kendall's τ correlation with SIR-derived ground-truth rankings.

To ensure a fair comparison, we enforced rigorous experimental controls to verify that performance gains are attributable to architectural design rather than parameter inflation or hyperparameter bias. To rule out model capacity as a confounding factor, we introduced widened single-branch baselines: the dual-branch HDCN employs two parallel branches with 1024-unit dense layers each, while the widened HD-only and HDF-only baselines use a single 2048-unit dense layer, yielding comparable parameter counts across all models (within 5%). All configurations otherwise adhered to identical protocols, including backbone structure, training schedule, optimizer settings, learning rate scheduling (ReduceLROnPlateau), neighborhood window size L, and normalization techniques. Additionally, we conducted a grid search over the margin parameter θ and multi-s configurations, reporting both best and mean τ values to demonstrate result robustness.

These controlled experiments confirm that the dual-branch architecture consistently outperforms single-branch counterparts, validating the complementary nature of local and higher-order structural features in hypergraph node ranking.

4.2.1. HD-only results

This subsection evaluates the HD-only baseline, where purely local hyper-degree features (HD_1, HD_2, HD_3) are utilized for node ranking. Table 5 summarizes the performance, measured by Kendall's τ, across varying margin parameters θ.

Observing performance variations across datasets reveals the differing efficacy of local features. The Algebra dataset achieved the highest correlation (max $\tau \approx 0.671$), followed closely by Restaurants-Rev and Bars-Rev, which demonstrated stable performance. This suggests that local neighborhood structures are sufficient to characterize node influence in these networks. Conversely, performance on Geometry and Music-Rev was notably inferior (max $\tau \approx 0.45$ and 0.42, respectively). This performance gap strongly implies that influence mechanisms in these datasets are more complex, likely governed by global connectivity patterns or higher-order structures, and cannot be fully captured by local degree features alone.

Table 5. Performance metrics for different datasets with varying θ values.

θ	Restaurants-Rev	Algebra	Geometry	Bars-Rev	Music-Rev
$\theta = 0.3$	0.6122	0.6506	0.4513	0.6061	0.4055
$\theta = 0.6$	0.6125	0.6545	0.3909	0.5995	0.4260
$\theta = 0.9$	0.6122	0.6556	0.4285	0.6034	0.3972
$\theta = 1.0$	0.6416	0.6455	0.4039	0.5801	0.4177
$\theta = 1.2$	0.6302	0.6713	0.4117	0.6007	0.4058
$\theta = 1.5$	0.5929	0.6685	0.3973	0.5960	0.4003
$\theta = 1.8$	0.5863	0.6635	0.4399	0.5982	0.4018

We further analyzed the impact of the margin parameter θ on model stability. Restaurants-Rev and Bars-Rev exhibited robustness, with minimal fluctuation across different θ values. In contrast, Algebra, Geometry, and Music-Rev showed greater sensitivity. Geometry, in particular, displayed pronounced volatility, indicating that the discriminative power of HD features for this dataset is limited, making the optimization process highly dependent on the margin constraint. In summary, these results delineate the scope of the HD-only approach: while it serves as a robust baseline for structurally simpler networks, it encounters significant performance bottlenecks on complex datasets. This limitation highlights the necessity of incorporating HDF. The proposed HDF integration aims specifically to capture the higher-order interactions that HD misses, providing the rationale for the subsequent dual-branch (HD+HDF) architecture.

4.2.2. HDF-only results

This subsection evaluates the performance of the HDF-only configuration, utilizing hypergraph distance-based fuzzy centrality features across various s-set combinations (e.g., $\{1, 2, 3\}, \{1, 2, 4\}$). Our objective is to isolate the contribution of higher-order structural features and benchmark them against the local HD-only baseline. We conducted a comprehensive evaluation by varying both the feature combination S and the margin parameter θ. The results, measured by Kendall's τ, are detailed in Table 6 and in Figure 4.

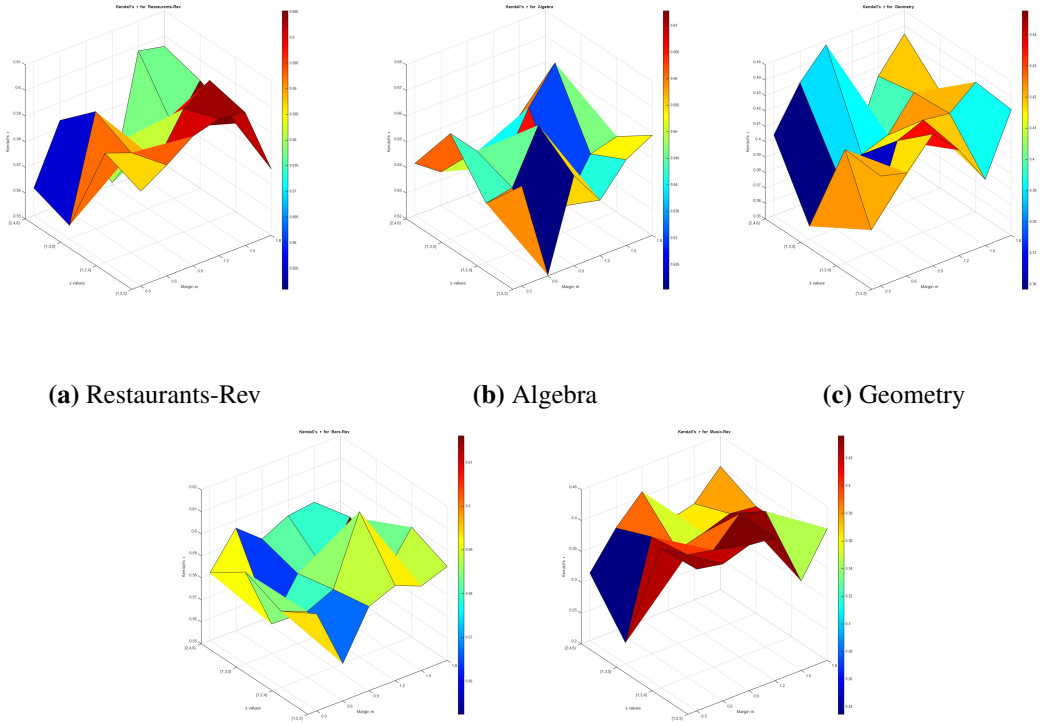
Table 6. Kendall's τ values for the HDF-only model across different datasets and margin θ values.

Datasets	Metric	$S = \{1, 2, 3\}$	$S = \{1, 2, 4\}$	$S = \{1, 3, 5\}$	$S = \{2, 4, 6\}$
Restaurants-Rev	$\theta = 0.3$	0.5869	0.5926	0.5553	0.5607
	$\theta = 0.6$	0.5932	0.5888	0.5957	0.5831
	$\theta = 0.9$	0.6005	0.5818	0.5806	0.5701
	$\theta = 1.0$	0.5970	0.5939	0.5755	0.5863
	$\theta = 1.2$	0.6033	0.5980	0.5777	0.5509
	$\theta = 1.5$	0.6018	0.6052	0.5774	0.5983
	$\theta = 1.8$	0.5758	0.5844	0.5920	0.5961
Algebra	$\theta = 0.3$	0.6590	0.6431	0.6611	0.6403
	$\theta = 0.6$	0.6203	0.6452	0.6504	0.6330
	$\theta = 0.9$	0.6551	0.6656	0.6403	0.6336
	$\theta = 1.0$	0.6534	0.6627	0.6655	0.6403
	$\theta = 1.2$	0.6416	0.6302	0.6594	0.6218
	$\theta = 1.5$	0.6534	0.6465	0.6728	0.6431
	$\theta = 1.8$	0.6590	0.6493	0.6541	0.6330
Geometry	$\theta = 0.3$	0.4214	0.4234	0.3583	0.4023
	$\theta = 0.6$	0.4169	0.3645	0.3887	0.4297
	$\theta = 0.9$	0.4355	0.4159	0.3917	0.4477
	$\theta = 1.0$	0.4366	0.4023	0.4162	0.4667
	$\theta = 1.2$	0.4184	0.4228	0.3977	0.3710
	$\theta = 1.5$	0.3928	0.4204	0.4187	0.4141
	$\theta = 1.8$	0.4316	0.4342	0.4130	0.4349
Bars-Rev	$\theta = 0.3$	0.5942	0.5847	0.5918	0.5807
	$\theta = 0.6$	0.5670	0.5800	0.5636	0.5965
	$\theta = 0.9$	0.5883	0.5854	0.5819	0.5869
	$\theta = 1.0$	0.6151	0.5999	0.6015	0.5555
	$\theta = 1.2$	0.5932	0.6161	0.5795	0.5929
	$\theta = 1.5$	0.5883	0.5833	0.5983	0.5944
	$\theta = 1.8$	0.5926	0.5998	0.5688	0.5524
Music-Rev	$\theta = 0.3$	0.4302	0.4253	0.2344	0.3094
	$\theta = 0.6$	0.4221	0.3993	0.3901	0.3662
	$\theta = 0.9$	0.4362	0.3901	0.3519	0.4087
	$\theta = 1.0$	0.3617	0.3892	0.3677	0.4263
	$\theta = 1.2$	0.4317	0.4322	0.3638	0.3449
	$\theta = 1.5$	0.3452	0.4202	0.3782	0.3561
	$\theta = 1.8$	0.4139	0.3743	0.3732	0.4001

The results reveal a distinct complementary relationship between HDF and HD features. The HDF-only model demonstrates competitive capability, particularly in networks where local topology is less predictive. For instance, on the Algebra dataset, HDF achieves parity with the HD baseline (max $\tau \approx 0.673$). Significantly, on the Music-Rev dataset, where local HD features struggled, the HDF model

attains superior rankings (max $\tau \approx 0.436$), surpassing the best HD-only result. This confirms the unique value of fuzzy cues in capturing complex influence mechanisms. Conversely, for Restaurants-Rev and Bars-Rev, local HD features remain the dominant predictor.

The choice of the fuzzy parameter set S proved critical for stability. Dense, lower-order sets (e.g., $S = \{1, 2, 3\}$) generally yielded robust performance across varying margins. In contrast, sparse or higher-order sets (e.g., $S = \{1, 3, 5\}$) exhibited higher volatility; while they occasionally reached peak performance, their sensitivity to θ suggests they may introduce noise into the feature representation. These experiments highlight the utility of higher-order fuzzy semantics, especially for complex networks like Music-Rev. The performance divergence between HD and HDF across datasets provides a compelling empirical rationale for the proposed dual-branch architecture, which aims to combine these complementary strengths.



(a) Restaurants-Rev **(b)** Algebra **(c)** Geometry

(d) Bars-Rev **(e)** Music-Rev

Figure 4. Kendall's τ for different datasets using HDF.

4.2.3. Results: HDCN fusion

This subsection presents a comprehensive evaluation of the proposed HDCN, which integrates HD and HDF. Table 7 summarizes the Kendall's τ performance across datasets, varying fuzziness parameters s and margins θ.

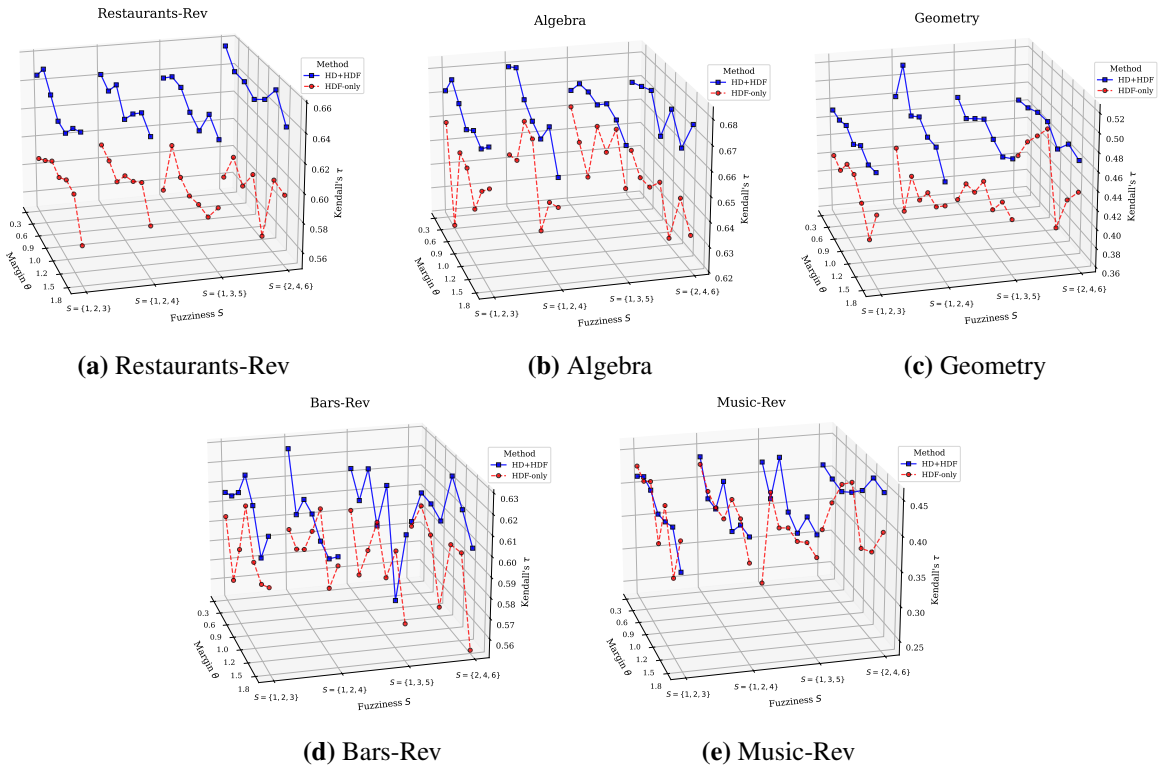
Empirical results show consistent complementary gains from dual-branch fusion. Across all five

Table 7. Kendall's τ coefficients for different fuzziness parameters S and margin θ across five datasets.

Datasets	Metric	$S = \{1, 2, 3\}$	$S = \{1, 2, 4\}$	$S = \{1, 3, 5\}$	$S = \{2, 4, 6\}$
Restaurants-Rev	$\theta = 0.3$	0.6460	0.6429	0.6368	0.6561
	$\theta = 0.6$	0.6568	0.6381	0.6444	0.6444
	$\theta = 0.9$	0.6460	0.6492	0.6438	0.6441
	$\theta = 1.0$	0.6353	0.6327	0.6337	0.6387
	$\theta = 1.2$	0.6346	0.6435	0.6283	0.6457
	$\theta = 1.5$	0.6451	0.6514	0.6466	0.6590
	$\theta = 1.8$	0.6501	0.6432	0.6372	0.6416
Algebra	$\theta = 0.3$	0.6721	0.6799	0.6680	0.6693
	$\theta = 0.6$	0.6805	0.6833	0.6747	0.6715
	$\theta = 0.9$	0.6749	0.6743	0.6753	0.6738
	$\theta = 1.0$	0.6687	0.6698	0.6741	0.6592
	$\theta = 1.2$	0.6726	0.6670	0.6786	0.6743
	$\theta = 1.5$	0.6698	0.6760	0.6764	0.6631
	$\theta = 1.8$	0.6749	0.6609	0.6708	0.6766
Geometry	$\theta = 0.3$	0.4718	0.4808	0.4742	0.4655
	$\theta = 0.6$	0.4718	0.5253	0.4616	0.4676
	$\theta = 0.9$	0.4772	0.4814	0.4727	0.4736
	$\theta = 1.0$	0.4685	0.4914	0.4831	0.4745
	$\theta = 1.2$	0.4787	0.4814	0.4730	0.4565
	$\theta = 1.5$	0.4706	0.4826	0.4661	0.4733
	$\theta = 1.8$	0.4748	0.4589	0.4760	0.4679
Bars-Rev	$\theta = 0.3$	0.6065	0.6260	0.6134	0.5831
	$\theta = 0.6$	0.6098	0.5976	0.6020	0.6032
	$\theta = 0.9$	0.6165	0.6104	0.6229	0.6027
	$\theta = 1.0$	0.6299	0.6084	0.5997	0.5993
	$\theta = 1.2$	0.6204	0.6005	0.6246	0.6265
	$\theta = 1.5$	0.6009	0.5975	0.5741	0.6153
	$\theta = 1.8$	0.6166	0.6041	0.6115	0.6022
Music-Rev	$\theta = 0.3$	0.4151	0.4365	0.4212	0.4094
	$\theta = 0.6$	0.4288	0.3883	0.3803	0.4022
	$\theta = 0.9$	0.4237	0.3883	0.4561	0.3979
	$\theta = 1.0$	0.4043	0.4432	0.3908	0.4117
	$\theta = 1.2$	0.4087	0.3866	0.3757	0.4290
	$\theta = 1.5$	0.4170	0.4115	0.4146	0.4614
	$\theta = 1.8$	0.3706	0.4107	0.4051	0.4556

datasets, HDCN outperforms the best single-branch alternative by +1.6% to +12.6% in Kendall's τ (relative). On Algebra, the fused model achieves $\tau = 0.6833$ ($S = \{1, 2, 4\}, \theta = 0.6$), a moderate improvement of +0.0105 over the best single branch ($\tau = 0.6728$, HDF-only).

The fusion model delivers notable gains on challenging datasets. On Geometry, where single-branch models struggled (typically $\tau < 0.49$), the fused architecture achieves a notable improvement with τ reaching 0.5253. This suggests that while neither HD nor HDF suffices individually for such complex topologies, their combination forms a complete structural representation. Similarly, on Music-Rev and Bars-Rev, the model not only improves absolute performance but also noticeably reduces the volatility observed in the HDF-only experiments.



(a) Restaurants-Rev **(b)** Algebra **(c)** Geometry

(d) Bars-Rev **(e)** Music-Rev

Figure 5. Performance comparison of HDF-only and HD+HDF fusion models across hyper-parameters.

Beyond peak metrics, the dual-branch architecture exhibits improved stability. While the best-performing s-set varies by dataset, the fused model maintains consistently high performance across diverse s and θ configurations. This resilience indicates that the combined feature space is easier to optimize and less sensitive to hyperparameter fluctuations than single-view models.

In summary, the HDCN results support the dual-branch fusion strategy. By combining local neighborhood density with higher-order fuzzy reachability, the model achieves complementary gains over either branch alone. The improvement is moderate on simpler datasets (e.g., +1.6% on Algebra) but more pronounced on structurally complex ones (e.g., +12.6% on Geometry), consistent with the hypothesis that multi-scale feature integration is particularly valuable when neither local nor global features alone suffice.

4.3. *Comparison with centrality baselines (RQ3)*

In this subsection, we benchmark the proposed HDCN architecture against a comprehensive suite of traditional centrality measures commonly used in network science. This comparison aims to quantify the advantages of our dual-branch learning approach over static, heuristic-based ranking methods. As illustrated in Table 8 and Figure 6, the HDCN achieves competitive performance across the majority of benchmarks, with a particularly notable lead on complex hypergraphs.

Table 8. Kendall's τ of the HDCN against traditional centrality baselines.

Metric	Restaurants-Rev	Algebra	Geometry	Bars-Rev	Music-Rev
DC	0.5047	0.2525	0.2613	0.6236	0.0836
HD	0.0469	0.0362	-0.0495	0.0965	-0.2494
VC	0.0500	0.0362	-0.0504	0.0680	-0.2466
HEDC	-0.0648	0.0211	-0.1644	-0.1783	-0.2726
ECC	-0.4070	-0.1671	-0.3056	-0.5725	-0.4711
HCC	-0.3878	-0.1276	-0.2998	-0.5489	-0.4337
HVC	0.4876	0.1214	0.0876	0.4992	-0.0329
HGC	0.5306	0.2585	0.2764	0.6449	0.0860
HCI-1	0.5586	-0.1350	-0.0816	-0.0448	-0.1413
HCI-2	0.0000	0.0180	0.0180	-0.1338	0.0000
HDF	0.5851	0.2475	0.2574	0.6573	0.3862
EHDF	0.5963	0.2556	0.3368	0.6419	0.3974
MLP	0.3925	0.6715	0.4204	0.4653	0.4458
HGNN	0.1332	0.0909	0.0441	-0.0108	-0.0074
HDCN	0.6590	0.6833	0.5253	0.6299	0.4614

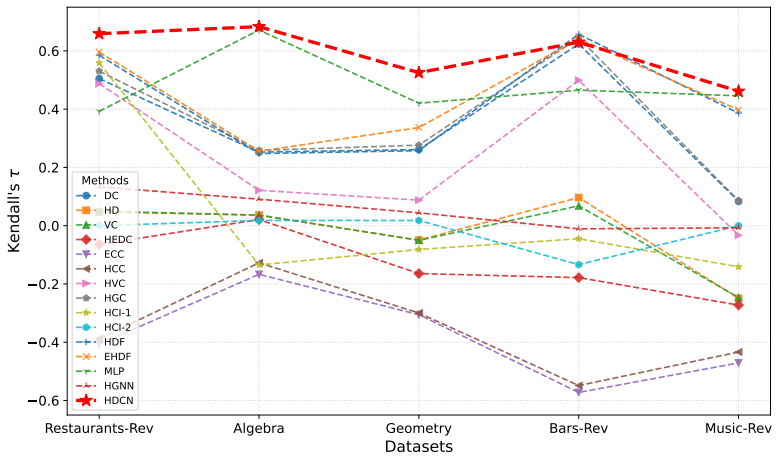


Figure 6. Performance comparison against traditional centrality measures. The HDCN demonstrates competitive performance over baselines, particularly on complex datasets.

The comparison against the two learning-based baselines isolates the contribution of the HDCN's

architectural design. The MLP baseline receives the same $(L+1)\times 6 = 246$-dimensional feature vector—the six centrality values of each node in the local receptive field—as the HDCN, but processes it as a flat vector without exploiting the 2D spatial structure of $\mathbf{E}_{v_i}$; it underperforms the HDCN on all five datasets, with Kendall's τ gaps ranging from +0.012 (Algebra) to +0.266 (Restaurants-Rev). This demonstrates that the CNN-based spatial encoding of $\mathbf{E}_{v_i}$ extracts spatial co-occurrence patterns among neighbors that a flat MLP cannot capture. The HGNN, despite having access to the full hypergraph structure, achieves near-random performance (best $\tau = 0.1332$), confirming that generic message-passing without task-aligned feature design is insufficient for influence ranking. Against all traditional non-learning centrality baselines, the HDCN achieves consistent advantages on four of five datasets, with Bars-Rev being the exception where degree-based HDF already achieves a high baseline ($\tau = 0.6573$).

In summary, the HDCN integrates task-aligned hypergraph feature design with end-to-end pairwise ranking optimization, offering a structured alternative to static heuristics for node influence ranking in hypergraphs.

4.4. Robustness across random data partitions

Table 9 shows that all standard deviations remain below 0.07 across datasets and S configurations, indicating that the HDCN's ranking performance is stable across different data partitions. The mean Kendall's τ under $S = \{1, 2, 3\}$ ranges from 0.5205 (Geometry) to 0.7170 (Restaurants-Rev), following the same dataset ordering as the single-split results in the main comparison tables (Table 7). Notably, Geometry exhibits the largest standard deviation (0.0627), which aligns with its high structural complexity (large $\langle k^H \rangle$ and C) identified in Section 4.1. These results confirm that the reported performance gains of the HDCN are reliable and not artifacts of a particular train/test partition.

Table 9. Reliability of the HDCN over five independent random data partitions (seeds $\in \{42, 7, 13, 21, 100\}$). Each entry reports the mean ± standard deviation of Kendall's τ across the five splits, under four fuzzy parameter sets S.

Datasets	$S = \{1, 2, 3\}$	$S = \{1, 2, 4\}$	$S = \{1, 3, 5\}$	$S = \{2, 4, 6\}$
Restaurants-Rev	0.7170 ± 0.0431	0.6993 ± 0.0413	0.7045 ± 0.0474	0.6876 ± 0.0460
Algebra	0.7009 ± 0.0253	0.6804 ± 0.0217	0.6974 ± 0.0400	0.6880 ± 0.0325
Geometry	0.5205 ± 0.0627	0.5329 ± 0.0599	0.5203 ± 0.0618	0.5281 ± 0.0593
Bars-Rev	0.6338 ± 0.0291	0.6319 ± 0.0266	0.6374 ± 0.0211	0.6253 ± 0.0174
Music-Rev	0.5272 ± 0.0307	0.5066 ± 0.0535	0.5154 ± 0.0249	0.5048 ± 0.0292

4.5. SIR parameter robustness analysis

To assess whether the performance of the HDCN depends on the specific SIR parameters used to generate ground-truth labels, we conducted two complementary robustness experiments. We generated five additional sets of node influence labels under varying infection rates, recovery rates, and simulation lengths: $(\beta, \mu, \text{steps}) \in \{(0.004, 0.1, 200), (0.02, 0.05, 200), (0.02, 0.25, 200), (0.02, 0.1, 1000), (0.1, 0.1, 200)\}$. For each pair of configurations, we computed the Kendall's τ between the resulting node rankings. Table 10 reports the mean and minimum pairwise τ across all ten configuration pairs per dataset.

Across most datasets, the mean pairwise τ exceeds 0.41, indicating that the relative ordering of influential nodes is preserved across different propagation regimes. The minimum values are lower primarily due to the extreme configuration $\beta = 0.1$ (a near-saturation regime in which virtually all nodes become infected regardless of seed identity), which effectively collapses rank differences. When this extreme setting is excluded, the remaining four configurations yield higher pairwise agreement, confirming stable ground-truth rankings in realistic sub-saturation regimes.

Table 10. SIR parameter robustness: Pairwise Kendall's τ between node rankings produced under five different SIR configurations. High τ indicates stable ground-truth rankings across parameter variations.

Datasets	τ_{mean}	τ_{min}	excl. $\beta = 0.1$ mean	excl. $\beta = 0.1$ min
Algebra	0.593	0.438	0.672	0.638
Restaurants-Rev	0.535	0.284	0.669	0.617
Bars-Rev	0.460	0.231	0.569	0.563
Geometry	0.414	0.228	0.526	0.460
Music-Rev	0.335	0.170	0.463	0.329

These pairwise consistency results confirm that our chosen SIR configuration ($\beta = 0.02, \mu = 0.1$) produces reliable ground-truth rankings. The moderate τ_{mean} values (0.335–0.593) reflect genuine variation in propagation dynamics across parameter settings, while the improved agreement after excluding the extreme saturation regime ($\beta = 0.1$) demonstrates that rankings are stable under realistic sub-saturation conditions relevant to practical influence analysis.

4.6. Monotonicity index (MI) evaluation

In this subsection, we evaluate the discriminative ability of the predicted rankings using the MI. As defined in Section 3.5, the MI quantifies the proportion of uniquely ranked nodes in a given list: a value of MI = 1 means all nodes receive distinct scores with no ties, while lower values reflect a greater degree of score collisions. Importantly, the MI is a property of a single ranking list and does not directly measure agreement with ground-truth rankings; that role is served by Kendall's τ, already reported in Section 4.2.3. The two metrics are therefore complementary: τ assesses ranking correctness relative to the SIR ground truth, while the MI assesses ranking resolution (the ability to discriminate between nodes).

Table 11 reports the MI of the predicted rankings across all datasets and margin parameters θ, together with the best Kendall's τ achieved on each dataset (taken from Table 7) as the direct SIR ground-truth comparison requested.

The upper block of Table 11 shows that the predicted MI remains consistently near 1.0 across all datasets and margin settings. Specifically, the model achieves full monotonicity (MI = 1.0) on Restaurants-Rev and Bars-Rev, and values above 0.995 on Algebra, Geometry, and Music-Rev. This confirms that the HDCN assigns nearly unique influence scores to all nodes, regardless of the hyperparameter configuration.

The lower block provides the direct SIR ground-truth comparison. The best τ values are 0.6833 (Algebra), 0.6590 (Restaurants-Rev), 0.6299 (Bars-Rev), 0.5253 (Geometry), and 0.4614 (Music-Rev), all taken from Table 7. Notably, Geometry achieves MI > 0.995 (high resolution) while its $\tau = 0.5253$

reflects the intrinsic difficulty of ranking nodes in densely entangled hypergraphs, confirming that high MI is a necessary but not sufficient condition for accurate ranking. Together, the two metrics paint a complete picture: the model generates highly discriminative rankings (MI ≈ 1.0) that also achieve meaningful correlation with true SIR-based influence (τ up to 0.6833), demonstrating both resolution and correctness.

Table 11. The MI of predicted rankings across datasets and margin parameters θ, with the best Kendall's τ (SIR ground-truth comparison) appended for each dataset. An MI near 1 indicates high ranking resolution (few ties), and τ measures ranking correctness against SIR ground truth.

	Restaurants-Rev	Algebra	Geometry	Bars-Rev	Music-Rev
Predicted MI across margin settings					
$\theta = 0.3$	1.0	0.9978	0.9981	1.0	0.9993
$\theta = 0.6$	1.0	0.9978	0.9958	1.0	0.9993
$\theta = 0.9$	1.0	0.9978	0.9958	1.0	0.9993
$\theta = 1.0$	1.0	0.9978	0.9958	1.0	0.9993
$\theta = 1.2$	1.0	0.9978	0.9958	1.0	0.9993
$\theta = 1.5$	1.0	0.9978	0.9958	1.0	0.9993
$\theta = 1.8$	1.0	0.9978	0.9958	1.0	0.9993
Direct comparison with SIR ground-truth rankings					
Best τ (SIR)	0.6590	0.6833	0.5253	0.6299	0.4614

Since full-ranking Kendall's τ aggregates over all node pairs, it weights low-influence nodes equally with high-influence ones. Rankings among low-influence nodes in SIR simulations may, however, be dominated by stochastic fluctuations rather than genuine structural differences. To examine whether the HDCN's accuracy concentrates on the practically important high-influence nodes, we computed the top-50 identification rate—the fraction of the SIR ground-truth top-50 most influential nodes correctly recovered by the model—within the held-out test set under $S = \{1, 2, 3\}$, averaged over five independent random splits. Here $K = 50$ represents approximately the top 4–12% of nodes in each dataset, a practically meaningful scale for seed-set identification in influence propagation applications. On Algebra, the model recovers 30.4% of the ground-truth top-50 nodes (0.304 ± 0.008), and on Geometry, it recovers 14.0% (0.140 ± 0.066). These results indicate that the HDCN's ranking quality is stronger among high-influence nodes than the overall τ alone would suggest, and that the model is not primarily driven by noisy low-influence pairs. The lower overlap on datasets such as Music-Rev (0.012 ± 0.016) is consistent with their high structural complexity (large $\langle k^E \rangle$) and the label noise inherent to near-saturation SIR regimes, as discussed in Section 4.5.

5. Conclusions

This work introduces the HDCN, a novel dual-branch convolutional framework designed to identify influential nodes in complex hypergraphs. By synthesizing local structural signals with global higher-order patterns, the proposed method overcomes critical limitations inherent in conventional centrality metrics and graph projection methods. Experimental analysis leads to the following conclusions:

- **Complementary Fusion of Local and Higher-Order Features:** Our analysis demonstrates that neither HD nor HDF alone suffices for accurate influence ranking across diverse hypergraph topologies. Their combination yields consistent improvements over the best single-branch alternative, ranging from +1.6% (Algebra) to +12.6% (Geometry) in relative Kendall's τ. The fusion proves particularly valuable on structurally complex datasets (e.g., Geometry, $\tau = 0.5253$), representing a relative improvement of 56% over the strongest traditional baseline (EHDF, $\tau = 0.3368$).
- **Robustness via HDF:** The incorporation of HDF is an important feature design. Through a multi-scale fuzzy weighting mechanism, HDF effectively captures the inherent uncertainty and overlap of group interactions. Empirical validation demonstrates that HDF offers unique structural insights under noisy conditions (e.g., Music-Rev); it offers advantages over local features by identifying long-range dependencies, which are frequently obscured in deterministic measures.
- **Data-Driven Ranking Paradigm:** This framework constitutes a departure from static heuristics toward a task-aware learning paradigm. The task is formulated as a pairwise learning-to-rank problem, optimized for Kendall's τ and supervised by SIR diffusion dynamics, ensuring end-to-end alignment with the influence objective. Consequently, the method achieves strong ranking accuracy while preserving high monotonicity (MI ≈ 1.0), which ensures that the derived rankings are a reliable reflection of the system's true spreading hierarchy.
- **Superiority Over Learning-Based Alternatives:** Comparative experiments against learning-based baselines confirm that the HDCN's structural design is essential. Generic hypergraph neural networks (HGNNs) with ranking heads achieve near-random performance (best $\tau = 0.1332$). A plausible explanation is that the smoothing/aggregation mechanism in generic message-passing attenuates the degree heterogeneity that underlies influence differentiation, though this mechanism was not directly verified in this study and warrants further investigation. The MLP baseline, which receives the same 246-dimensional neighbor feature vector as the HDCN but flattens it without spatial structure, underperforms the HDCN on all five datasets, confirming that CNN-based spatial encoding of the local structural matrix is the key architectural contribution over flat feature aggregation. The HDCN's primary distinction lies in combining structured local receptive field processing with end-to-end rank-aware training, enabling substantial advantages over both generic learning baselines and traditional non-learning centrality measures.

In conclusion, the HDCN offers a task-aligned, data-driven approach to influence ranking in hypergraphs, combining hypergraph-aware feature engineering with end-to-end pairwise ranking optimization. The framework provides a structured foundation for future work on dynamic hypergraphs and larger-scale networks.

The current framework has three limitations. First, the Gaussian fuzzy membership function in HDF is predefined; a learnable weighting module, such as attention-based distance aggregation, could improve adaptability across heterogeneous hypergraph topologies. Second, the HDCN assumes static hypergraph structure and has only been evaluated on datasets of at most $N = 1234$ nodes; extending the framework to dynamic and large-scale hypergraphs, and validating ground-truth rankings against real influence cascades beyond Monte Carlo SIR simulations, remain important directions. Third, future work could include a node-level-only baseline to more precisely isolate the contribution of the local structural matrix $\mathbf{E}_{v_i}$.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

The research was supported by the scientific research project for current students of Guizhou University of Finance and Economics (No. 2025BAZYSY248) and the Scientific Research Foundation for Talents Introduction of Guizhou University of Finance and Economics (No. 2023YJ16).

Conflict of interest

The authors declare no conflicts of interest.

References

1. A. Bavelas, Communication patterns in task-oriented groups, *J. Acoust. Soc. Am.*, **22** (1950), 725–730. https://doi.org/10.1121/1.1906679
2. L. C. Freeman, A set of measures of centrality based on betweenness, *Sociometry*, **40** (1977), 35–41. https://doi.org/10.2307/3033543
3. K. Q. Cai, J. Zhang, W. B. Du, X. B. Cao, Analysis of the Chinese air route network as a complex network, *Chin. Phys. B*, **21** (2012), 028903. https://doi.org/10.1088/1674-1056/21/2/028903
4. Y. Yang, T. Nishikawa, A. E. Motter, Small vulnerable sets determine large network cascades in power grids, *Science*, **358** (2017), 6365. https://doi.org/10.1126/science.aan3184
5. A. N. Arularasan, A. Suresh, K. Seerangan, Identification and classification of best spreader in the domain of interest over the social networks, *Cluster Comput.*, **22** (2019), 4035–4045. https://doi.org/10.1007/s10586-018-2616-y
6. X. P. Yuan, Y. K. Xue, M. X. Liu, Dynamic analysis of a sexually transmitted disease model on complex networks, *Chin. Phys. B*, **22** (2013), 030207. https://doi.org/10.1088/1674-1056/22/3/030207
7. Y. Shang, Degree distribution dynamics for disease spreading with individual awareness, *J. Syst. Sci. Complexity*, **28** (2015), 96–104. https://doi.org/10.1007/s11424-014-2186-x
8. M. E. J. Newman, Coauthorship networks and patterns of scientific collaboration, *Proc. Natl. Acad. Sci. U.S.A.*, **101** (2004), 5200–5205. https://doi.org/10.1073/pnas.0307545100
9. M. Seufert, A. Schwind, T. Hoßfeld, P. Tran-Gia, Analysis of group-based communication in WhatsApp, in *Mobile Networks and Management*, Springer, Cham, (2015), 223–236. https://doi.org/10.1007/978-3-319-26925-2˙17
10. X. Yi, S. Liu, Y. Wu, D. McCloskey, Z. Meng, BPP: a platform for automatic biochemical pathway prediction, *Briefings Bioinf.*, **25** (2024), bbae355. https://doi.org/10.1093/bib/bbae355
11. A. R. Benson, D. F. Gleich, J. Leskovec, Higher-order organization of complex networks, *Science*, **353** (2016), 163–166. https://doi.org/10.1126/science.aad9029

12. I. Iacopini, G. Petri, A. Barrat, V. Latora, Simplicial models of social contagion, *Nat. Commun.*, **10** (2019), 2485. https://doi.org/10.1038/s41467-019-10431-6

13. L. Torres, A. S. Blevins, D. Bassett, T. Eliassi-Rad, The why, how, and when of representations for complex systems, *SIAM Rev.*, **63** (2021), 435–485. https://doi.org/10.1137/20M1355896

14. F. Battiston, G. Cencetti, I. Iacopini, V. Latora, M. Lucas, A. Patania, et al., Networks beyond pairwise interactions: structure and dynamics, *Phys. Rep.*, **874** (2020), 1–92. https://doi.org/10.1016/j.physrep.2020.05.004

15. A. R. Benson, Three hypergraph eigenvector centralities, *SIAM J. Math. Data Sci.*, **1** (2019), 293–312. https://doi.org/10.1137/18M1203031

16. L. Qi, Eigenvalues of a real supersymmetric tensor, *J. Symb. Comput.*, **40** (2005), 1302–1324. https://doi.org/10.1016/j.jsc.2005.05.007

17. B. Schölkopf, J. Platt, T. Hofmann, Learning with hypergraphs: clustering, classification, and embedding, in *Advances in Neural Information Processing Systems 19: Proceedings of the 2006 Conference*, MIT Press, (2007), 1601–1608.

18. Y. Feng, H. You, Z. Zhang, R. Ji, Y. Gao, Hypergraph neural networks, *Proc. AAAI Conf. Artif. Intell.*, **33** (2019), 3558–3565. https://doi.org/10.1609/aaai.v33i01.33013558

19. N. Yadati, M. Nimishakavi, P. Yadav, V. Nitin, A. Louis, P. Talukdar, HyperGCN: a new method of training graph convolutional networks on hypergraphs, in *Advances in Neural Information Processing Systems 32 (NeurIPS 2019)*, **32** (2019), 1509–1520.

20. G. St-Onge, I. Iacopini, V. Latora, A. Barrat, G. Petri, A. Allard, et al., Influential groups for seeding and sustaining nonlinear contagion in heterogeneous hypergraphs, *Commun. Phys.*, **5** (2022), 25. https://doi.org/10.1038/s42005-021-00788-w

21. M. Xie, X. X. Zhan, C. Liu, Z. K. Zhang, An efficient adaptive degree-based heuristic algorithm for influence maximization in hypergraphs, *Inf. Process. Manage.*, **60** (2023), 103161. https://doi.org/10.1016/j.ipm.2022.103161

22. Y. C. Gong, M. Wang, W. Liang, F. Hu, Z. K. Zhang, UHIR: an effective information dissemination model of online social hypernetworks based on user and information attributes, *Inf. Sci.*, **644** (2023), 119284. https://doi.org/10.1016/j.ins.2023.119284

23. Q. Pan, H. Wang, Y. Ruan, Z. Zhang, J. Tang, A state transition-based method for influence evaluation in networks, *Chaos, Solitons Fractals*, **205** (2026), 117713. https://doi.org/10.1016/j.chaos.2025.117713

24. S. S. Zhang, J. Xie, Y. Chen, M. Gao, C. Li, C. Liu, et al., HIP: model-agnostic hypergraph influence prediction via distance-centrality fusion and neural ODEs, *Expert Syst. Appl.*, **322** (2026), 132357. https://doi.org/10.1016/j.eswa.2026.132357

25. Q. Pan, J. Tang, Z. Zhang, Y. Ruan, Influence regulation in networks and its metaheuristic solution, *Reliab. Eng. Syst. Saf.*, **276** (2026), 112918. https://doi.org/10.1016/j.ress.2026.112918

26. X. Qu, W. Pei, Q. Wang, Y. Yang, B. Xue, M. Zhang, et al., Evolutionary hypergraph learning: a multi-objective influence maximization approach considering seed selection costs, *IEEE Trans. Evol. Comput.*, **2025** (2025). https://doi.org/10.1109/TEVC.2025.3577643

27. S. S. Zhang, X. Yu, G. Q. Sun, C. Liu, X. X. Zhan, Locating influential nodes in hypergraphs via fuzzy collective influence, *Commun. Nonlinear Sci. Numer. Simul.*, **142** (2025), 108574. https://doi.org/10.1016/j.cnsns.2024.108574

28. L. C. Freeman, Centrality in social networks conceptual clarification, *Social Networks*, **1** (1978–1979), 215–239. https://doi.org/10.1016/0378-8733(78)90021-7

29. J. W. Wang, L. L. Rong, Q. H. Deng, J. Y. Zhang, Evolving hypernetwork model, *Eur. Phys. J. B*, **77** (2010), 493–498. https://doi.org/10.1140/epjb/e2010-00297-8

30. F. Morone, H. A. Makse, Influence maximization in complex networks through optimal percolation, *Nature*, **524** (2015), 65–68. https://doi.org/10.1038/nature14604

31. S. G. Aksoy, C. Joslyn, C. O. Marrero, B. Praggastis, E. Purvine, Hypernetwork science via high-order hypergraph walks, *EPJ Data Sci.*, **9** (2020), 16. https://doi.org/10.1140/epjds/s13688-020-00231-0

32. K. Kovalenko, M. Romance, E. Vasilyeva, D. Aleja, R. Criado, D. Musatov, et al., Vector centrality in hypergraphs, *Chaos, Solitons Fractals*, **162** (2022), 112397. https://doi.org/10.1016/j.chaos.2022.112397

33. X. Xie, X. Zhan, Z. Zhang, C. Liu, Vital node identification in hypergraphs via gravity model, *Chaos*, **33** (2023), 013104. https://doi.org/10.1063/5.0127434

34. F. Hu, K. Tian, Z. K. Zhang, Identifying vital nodes in hypergraphs based on Von Neumann entropy, *Entropy*, **25** (2023), 1263. https://doi.org/10.3390/e25091263

35. W. R. Knight, A computer method for calculating kendall's tau with ungrouped data, *J. Am. Stat. Assoc.*, **61** (1966), 436–439. https://doi.org/10.1080/01621459.1966.10480879

36. J. Bae, S. Kim, Identifying and ranking influential spreaders in complex networks by neighborhood coreness, *Physica A*, **395** (2014), 549–559. https://doi.org/10.1016/j.physa.2013.10.047

AIMS Press