DUAL-VIEW FUZZY MEMBERSHIP AND CANDIDATE-ADAPTIVE MULTI-SOURCE EVIDENCE FUSION FOR HYPEREDGE PREDICTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Hyperedge prediction requires identifying missing higher-order relations from incomplete hypergraphs. Existing approaches typically encode candidate hyperedges into unified representations, implicitly combining heterogeneous formation cues and overlooking how different candidates rely on different types of evidence. However, candidate hyperedges may emerge from diverse formation patterns, driven by structural participation, attribute semantics, or their cross-view consistency. Therefore, prediction requires explicitly modeling these complementary formation evidences and adaptively determining their contributions for each candidate. To address this issue, we propose FAME-HP, a framework that models candidate hyperedges through dual-view fuzzy overlapping memberships and candidate-adaptive evidence fusion. FAME-HP learns structural and semantic membership spaces, derives candidate-level structural, semantic, and cross-view consistency evidence, and dynamically adjusts their contributions according to candidate-specific contexts. Experiments on six real-world hypergraphs under four evaluation protocols demonstrate that FAME-HP achieves the best or second-best performance across all six datasets in average AUROC and average precision, outperforming strong hyperedge prediction baselines on most benchmarks. These results demonstrate that explicitly organizing and adapting heterogeneous formation evidence provides an effective approach for identifying missing higher-order relations.

1 INTRODUCTION

Many real-world relations involve interactions among more than two entities. A scientific paper may involve multiple authors, a biochemical reaction may contain several compounds, and a communication event may connect a group of users. Hypergraphs provide a natural representation for such higher-order relations by allowing a hyperedge to connect an arbitrary number of nodes (Battiston et al., 2020). Hyperedge prediction aims to recover missing higher-order relations from a partially observed hypergraph. Its central challenge lies in evaluating the collective compatibility of a candidate node set: nodes that appear individually plausible may still fail to form a coherent higher-order interaction.

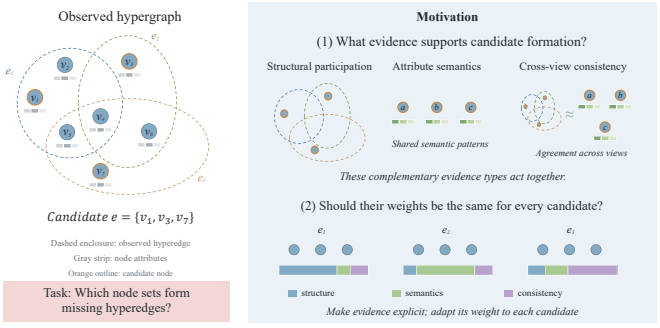

Figure 1: Motivation of FAME-HP. Different candidates rely on different combinations of structural, semantic, and cross-view evidence.

Recent hyperedge prediction methods have substantially improved candidate representation and negative sampling (Zhang et al., 2020; Yadati et al., 2020; Hwang et al., 2022; Ko et al., 2025; Yu et al.,

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

2025). Most approaches ultimately encode the available structural and attribute information into a unified candidate representation or prediction score. Consequently, the roles played by different sources of formation evidence remain implicit. This limitation is particularly relevant in higher-order networks with overlapping organization, where a node can participate in multiple latent groups and its relevant affiliation depends on the other nodes in the candidate (Contisciani et al., 2022).

Our key observation is that candidate hyperedges can be supported by heterogeneous yet complementary formation evidence, as illustrated in Fig. 1. Structural participation characterizes whether candidate nodes exhibit compatible interaction patterns in the observed hypergraph, while attribute semantics captures their compatibility in the feature space. Their cross-view consistency further indicates whether the two sources provide coherent support. More importantly, the relative contribution of these evidence sources can vary substantially across candidates. This motivates an explicit candidate-level representation of formation evidence together with an adaptive mechanism for determining its contribution to each prediction.

Motivated by this observation, we propose FAME-HP, a hyperedge prediction framework based on dual-view fuzzy memberships and candidate-adaptive evidence fusion. FAME-HP first learns structural and attribute fuzzy membership profiles that capture overlapping latent affiliations of individual nodes. For each candidate, these memberships are transformed into structural, attribute, and cross-view consistency evidence together with distributional descriptors. The model further characterizes the candidate through its local structural and prototype context, and uses these signals to construct a shared fusion representation. Two lightweight heads then produce source-indexed correction components and candidate-specific mixture weights. Their weighted correction is integrated with a permutation-invariant base prediction to estimate the probability of the candidate forming a hyperedge.

Our contributions are summarized as follows:

- We formulate hyperedge prediction from the perspective of candidate-conditioned formation evidence, explicitly characterizing structural participation, attribute semantics, and their cross-view consistency under overlapping latent affiliations.
- We propose FAME-HP, which learns dual-view fuzzy memberships, constructs candidate-level formation evidence, and adaptively combines source-indexed correction components using candidate-specific context.
- We conduct extensive experiments on six real-world hypergraphs under four evaluation protocols, where FAME-HP consistently achieves first- or second-best average AUROC and average precision, together with controlled ablations and mechanism analyses that validate the proposed evidence-fusion framework.

2 Related Work

Hyperedge prediction. Hyperedge prediction has been studied through structural expansion, neural candidate representations, and negative-sampling strategies. Early work explores higher-order expansion to capture structural information at different interaction orders (Yoon et al., 2020). In parallel, hypergraph neural networks, including HGNN, HyperGCN, UniGNN, and AllSet, develop representation-learning mechanisms for higher-order relational structures (Feng et al., 2019; Yadati et al., 2019; Huang & Yang, 2021; Chien et al., 2022). For hyperedge prediction, Hyper-SAGNN introduces self-attention to model variable-size hyperedges (Zhang et al., 2020), while NHP learns hypergraph-based node representations and scores candidate sets through permutation-invariant aggregation (Yadati et al., 2020). Subsequent methods further improve candidate modeling and training. AHP employs adversarial negative sampling (Hwang et al., 2022), and CASH combines context-aware candidate aggregation with self-supervised hypergraph learning (Ko et al., 2025). More recent studies investigate positive-guided negative hyperedge generation and contrastive attention with negative-sample evaluation (Yu et al., 2025; Wang et al., 2025). These advances strengthen hypergraph encoding, candidate aggregation, and negative sampling. FAME-HP further investigates the representation and candidate-adaptive combination of structural, attribute, and cross-view formation evidence.

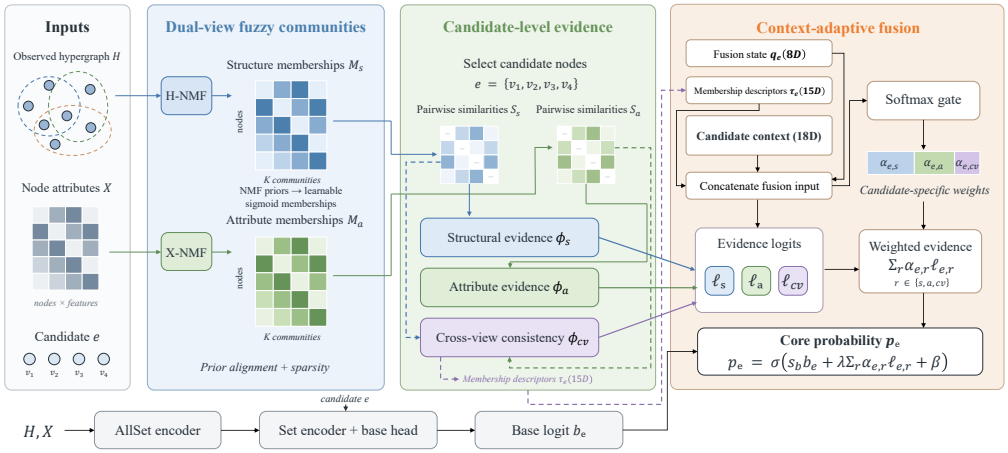

Figure 2: Overview of FAME-HP. Structural and attribute fuzzy memberships provide candidate-level evidence scores and membership descriptors. Together with candidate context, these signals determine candidate-specific correction components and adaptive fusion weights for hyperedge prediction.

Overlapping communities and fuzzy memberships. Soft community representations characterize nodes that participate in multiple latent groups. Nonnegative matrix factorization (NMF) provides a foundation for learning nonnegative latent factors (Lee & Seung, 1999) and has been used to identify fuzzy and overlapping community structures (Zhang et al., 2007; Cao et al., 2013). For higher-order networks, overlapping community models connect latent node affiliations with hyperedge organization. Contisciani et al. (Contisciani et al., 2022) investigate the joint inference of overlapping communities and higher-order interactions, while recent research further explores fuzzy memberships for hypergraph community detection (Xiao et al., 2025). These studies motivate representing nodes through multiple latent affiliations. FAME-HP learns structural and attribute fuzzy membership spaces, transforms their compatibility and cross-view agreement into candidate-level evidence, and uses candidate context to adaptively combine the resulting correction components.

3 Problem Definition

Let the observed hypergraph be denoted by $\mathcal{H}_{\text{obs}} = (\mathcal{V}, E_{\text{obs}}, X)$, where $\mathcal{V}$ is the node set, E_{obs} is the set of hyperedges available to the model, and $X \in \mathbb{R}^{|\mathcal{V}| \times d}$ is the node-attribute matrix. We write $H_{\text{obs}} \in \{0,1\}^{|\mathcal{V}| \times |E_{\text{obs}}|}$ for the corresponding incidence matrix, where $H_{\text{obs}}(v, e) = 1$ if and only if $v \in e$. Given a candidate set $e = \{v_1, \ldots, v_m\} \subseteq \mathcal{V}$ with $m = |e| \geq 2$, hyperedge prediction estimates

$$p(y_e = 1 \mid e, \mathcal{H}_{\text{obs}}), \tag{1}$$

where $y_e \in \{0, 1\}$ indicates whether the candidate is a true hyperedge.

Unlike pairwise link prediction, hyperedge prediction must evaluate collective compatibility among all candidate nodes. We therefore learn a candidate-level function that captures structural organization, node attributes, and their agreement.

4 Method

4.1 Overview of FAME-HP

Figure 2 illustrates the overall framework of FAME-HP. Given an observed hypergraph and a candidate hyperedge, the model combines a pretrained base predictor with three complementary sources of formation evidence: structural participation, attribute compatibility, and cross-view consistency.

These sources describe different aspects of candidate compatibility. Their evidence responses, membership distributions, and local structural context jointly condition the candidate-specific correction and fusion process.

FAME-HP consists of three main stages. First, it learns structural and attribute fuzzy memberships to represent overlapping latent affiliations. Second, it transforms these memberships into candidate-level evidence scores and richer descriptors that preserve distributional and cross-view information. Finally, a context-adaptive fusion module uses these representations to produce candidate-specific correction components and fusion weights, which are combined with a permutation-invariant base prediction.

The base predictor employs a pretrained AllSet-style hypergraph encoder (Chien et al., 2022) to obtain node representations. A permutation-invariant set encoder aggregates the representations of candidate nodes (Zaheer et al., 2017; Lee et al., 2019), followed by a prediction head that produces the base logit b_e and probability $p_e^{\text{base}} = \sigma(b_e)$. The pretrained base predictor remains frozen during subsequent membership learning and evidence fusion.

4.2 Dual-View Fuzzy Membership Learning

Nodes in higher-order networks can participate in multiple overlapping groups, and their relevant affiliations may differ across candidate hyperedges. We therefore represent each node through two fuzzy membership profiles: a structural profile describing its participation in latent interaction patterns and an attribute profile describing its semantic affiliations.

Let $M^{(s)}, M^{(a)} \in [0,1]^{|\mathcal{V}|\times K}$ denote the structural and attribute membership matrices, respectively, where K is the number of latent communities. Each membership dimension is independently activated, allowing a node to participate in multiple groups.

To initialize these representations, we extract separate membership priors from the observed incidence matrix and node attributes. Specifically, let $A_s = H_{\text{obs}}$ and $A_a = [X]_+$, where $[X]_+$ denotes the element-wise nonnegative part of the attribute matrix. For each view $r \in \{s, a\}$, nonnegative matrix factorization (NMF) (Lee & Seung, 1999) yields

$$(\widetilde{M}_0^{(r)}, B_0^{(r)}) = \underset{U,V\geq 0}{\operatorname{argmin}} \left\| A_r - UV^\top \right\|_F^2, \tag{2}$$

where $U \in \mathbb{R}_+^{|\mathcal{V}|\times K}$ is the node-side factor.

The node factors provide initial structural and attribute affiliations. We apply row-wise ℓ_1 normalization to obtain bounded priors $M_0^{(r)}$ and use them to initialize trainable membership logits. The refined memberships are given by

$$M^{(r)} = \sigma(Z^{(r)}), \qquad r \in \{s, a\}, \tag{3}$$

where $Z^{(r)}$ is initialized from the clipped logit transformation of $M_0^{(r)}$. A prior regularizer, introduced in Section 4.5, guides the learned memberships toward their initial community structure while allowing task-specific refinement.

4.3 Candidate-Level Evidence Construction

The two membership spaces describe individual nodes, whereas hyperedge prediction requires assessing their compatibility as a group. We construct three complementary sources of candidate-level evidence and retain additional membership descriptors for subsequent adaptive fusion.

Structural and attribute compatibility. For a candidate e and view $r \in \{s, a\}$, we compute pairwise cosine similarities between the membership profiles of all candidate nodes:

$$S_r(e) = \left\{ \cos\left(M_u^{(r)}, M_v^{(r)}\right) \mid u, v \in e,\ u < v \right\}. \tag{4}$$

We summarize $S_r(e)$ using its mean $\mu_r(e)$, minimum $m_r(e)$, and standard deviation $\sigma_r(e)$. These statistics describe overall compatibility, the weakest pairwise agreement, and variation across the candidate, respectively.

A bounded aggregation function converts these statistics into a view-specific evidence score:

$$\phi_r(e) = g_r(\mu_r(e), m_r(e), \sigma_r(e)), \qquad r \in \{s, a\}. \tag{5}$$

The resulting $\phi_s(e)$ and $\phi_a(e)$ characterize structural and attribute compatibility at the candidate level. The deterministic aggregation functions and their coefficients are specified in Appendix B.

Membership uncertainty and coverage. Candidates with similar pairwise compatibility may exhibit different membership distributions. For example, candidate nodes may share a concentrated latent affiliation or distribute their memberships across several groups. We capture this distinction through membership entropy and community coverage.

For each view, let

$$\pi_{v,k}^{(r)} = \frac{M_{v,k}^{(r)}}{\max\left(\sum_{j=1}^{K} M_{v,j}^{(r)}, \epsilon\right)}$$

denote the normalized membership of node v. We define the candidate-level membership entropy as

$$\mathcal{H}_r^{\mathrm{mem}}(e) = -\frac{\sum_{v \in e} \sum_{k=1}^{K} \pi_{v,k}^{(r)} \log\left(\pi_{v,k}^{(r)} + \epsilon\right)}{|e| \log(\max\{K, 2\})}. \tag{6}$$

This quantity measures the average uncertainty of candidate nodes' latent affiliations. In particular, $\mathcal{H}_s^{\mathrm{mem}}(e)$ describes uncertainty in the structural membership space.

We additionally compute community coverage, $C_r(e) = K^{-1} \sum_{k=1}^{K} \max_{v \in e} M_{v,k}^{(r)}$, which summarizes the extent to which the candidate activates the latent communities. Together, membership entropy and coverage complement pairwise similarity by describing the distribution and breadth of candidate affiliations.

Cross-view consistency. Structural compatibility and attribute compatibility provide distinct descriptions of candidate formation. Their agreement offers an additional signal for assessing whether a candidate exhibits coherent higher-order organization.

We characterize cross-view consistency through differences in within-view similarity and community coverage, together with centroid-level and node-level membership alignment. Let $\Delta_{\mathrm{sim}}(e)$ and $\Delta_{\mathrm{cov}}(e)$ denote the absolute differences between the two views' mean similarity and coverage. We further use $A_{\mathrm{centroid}}(e)$ and $A_{\mathrm{node}}(e)$ to summarize the cosine alignment of candidate membership centroids and individual node memberships. These quantities produce a bounded cross-view evidence score $\phi_{cv}(e)$ through an aggregation function g_{cv}.

The three evidence scores $\phi_s(e)$, $\phi_a(e)$, and $\phi_{cv}(e)$ provide compact summaries of candidate compatibility. To retain information about the underlying membership distributions, we additionally construct

$$\begin{aligned} \boldsymbol{\tau}_r(e) &= [\mu_r, m_r, \sigma_r, \mathcal{H}_r^{\mathrm{mem}}, C_r], \quad r \in \{s, a\}, \\ \boldsymbol{\tau}_{cv}(e) &= [\Delta_{\mathrm{sim}}, \Delta_{\mathrm{cov}}, A_{\mathrm{centroid}}, A_{\mathrm{node}}, \sigma_{\mathrm{node}}], \\ \boldsymbol{\tau}_e &= [\boldsymbol{\tau}_s(e); \boldsymbol{\tau}_a(e); \boldsymbol{\tau}_{cv}(e)] \in \mathbb{R}^{15}, \end{aligned} \tag{7}$$

where σ_{node} is the dispersion of node-level cross-view alignment and the candidate argument is omitted within the descriptor vectors for brevity.

The evidence scores and membership descriptors play complementary roles in the subsequent fusion stage. The scores provide compact measures of the three formation cues, while $\boldsymbol{\tau}_e$ preserves their distributional properties. In particular, structural-membership entropy is retained in $\boldsymbol{\tau}_s(e)$, allowing membership uncertainty to influence candidate-adaptive fusion.

4.4 Context-Adaptive Evidence Fusion

The relevance of structural, attribute, and cross-view evidence depends on the candidate being evaluated. We therefore combine the evidence representations with candidate-specific context to construct an adaptive fusion input.

For candidate e, we extract a context vector $\boldsymbol{c}_e \in \mathbb{R}^{18}$ consisting of local-structure, prototype-matching, and consistency features. These features characterize support from observed hyperedges, similarity to existing interaction patterns, and agreement among local structural signals. We further construct a fusion-state vector $\boldsymbol{q}_e \in \mathbb{R}^8$ containing the base probability, the three evidence scores, normalized candidate cardinality, and three discrepancies among the base and evidence scores. Their exact definitions are provided in Appendix B.

Combining these signals with the membership descriptors gives the candidate-level fusion representation

$$\boldsymbol{x}_e = [\boldsymbol{q}_e; \boldsymbol{\tau}_e; \boldsymbol{c}_e] \in \mathbb{R}^{41}. \tag{8}$$

The representation is standardized using training-set statistics before being passed to the fusion module; we denote the standardized vector by $\bar{\boldsymbol{x}}_e$. The three descriptor blocks play complementary roles in fusion. $\boldsymbol{q}_e$ summarizes the current predictive state, including the base response and the compact evidence scores. $\boldsymbol{\tau}_e$ preserves the underlying membership geometry that is discarded by scalar evidence aggregation, including affiliation uncertainty, coverage, and cross-view agreement. $\boldsymbol{c}_e$ characterizes how the candidate is situated relative to observed hyperedge structure and prototype patterns. Consequently, the fusion module can condition its correction on both the strength of the current evidence and the structural context in which that evidence is observed.

Two lightweight prediction heads operate on this shared representation. The evidence head produces three source-indexed correction components, while the gating head determines their candidate-specific mixture weights:

$$\begin{aligned} \boldsymbol{\ell}_e &= E_\omega(\bar{\boldsymbol{x}}_e) = (\ell_{e,s}, \ell_{e,a}, \ell_{e,cv}), \\ \boldsymbol{\alpha}_e &= \mathrm{softmax}\left(G_\psi(\bar{\boldsymbol{x}}_e)\right). \end{aligned} \tag{9}$$

The shared input allows both heads to account for the candidate's evidence responses, membership distributions, and formation context. The gate assigns a normalized weight to each of the three correction components.

The weighted correction is integrated with the frozen base logit to obtain the final prediction:

$$\begin{aligned} z_e &= s_b b_e + \lambda \sum_{r \in \{s,a,cv\}} \alpha_{e,r} \ell_{e,r} + \beta, \\ p_e &= \sigma(z_e), \end{aligned} \tag{10}$$

where s_b and λ are learned nonnegative global scales and β is a learned bias. Candidate adaptivity is introduced through $\boldsymbol{\alpha}_e$ and the correction components generated from the candidate-level fusion representation.

Through this construction, membership entropy and cross-view agreement contribute to the final prediction via the shared fusion representation, while candidate context determines how the learned correction components are combined with the base prediction.

4.5 Learning Objective

FAME-HP optimizes the fuzzy membership logits, evidence head, gating head, and global fusion parameters while keeping the pretrained base predictor fixed. The training objective combines candidate classification, pairwise ranking, and regularization of the learned membership and fusion representations.

We use weighted binary cross-entropy $\mathcal{L}_{\mathrm{BCE}}$ for candidate prediction and a pairwise softplus ranking loss $\mathcal{L}_{\mathrm{rank}}$ to encourage higher scores for positive candidates than for their matched negatives. The membership-prior loss $\mathcal{L}_{\mathrm{prior}}$ penalizes deviations from the NMF initialization, while $\mathcal{L}_{\mathrm{sparse}}$ controls the overall activation of fuzzy memberships.

To regularize candidate-adaptive fusion, we further maximize the batch-averaged entropy of the routing weights, denoted by $\mathcal{H}_{\mathrm{gate}}$. The complete objective is

$$\mathcal{L} = \mathcal{L}_{\mathrm{BCE}} + \lambda_r \mathcal{L}_{\mathrm{rank}} + \eta_p \mathcal{L}_{\mathrm{prior}} + \lambda_s \mathcal{L}_{\mathrm{sparse}} - \lambda_H \mathcal{H}_{\mathrm{gate}}. \tag{11}$$

The classification and ranking terms supervise candidate discrimination, while prior and sparsity regularization constrain the learned membership spaces. Gate-entropy regularization encourages

the fusion module to maintain multiple correction components during training. Its role is distinct from the structural and attribute membership entropies in Eq. equation 6, which describe candidate affiliations and enter the model as prediction features. Full loss definitions and implementation settings are provided in Appendix B.

5 EXPERIMENTS

We evaluate overall prediction, component contributions, evidence behavior, and parameter sensitivity.

5.1 EXPERIMENTAL SETUPS

Datasets. We use six real-world hypergraphs: the co-citation networks Citeseer, Cora, and Pubmed; the authorship networks Cora-A and DBLP-A; and the DBLP collaboration network. They vary in scale, hyperedge organization, and node attributes.

Evaluation protocol. We evaluate five train–validation–test splits under three random seeds. SNS samples size-matched non-hyperedges, MNS uses the clique-expansion neighborhood frontier, CNS replaces one node of each positive hyperedge, and MIX combines all three negative types.

We report AUROC and AP for SNS, MNS, CNS, MIX, and AVG, with FAME-HP and controlled variants averaged over 15 seed–split runs and AVG shown as mean $\pm$ standard deviation. With the 3-kind evidence design fixed, parameters are learned from training candidates, validation selects checkpoints using the mean of AVG AP and AVG AUROC, and test scores are computed afterward.

Baselines. We compare FAME-HP with six representative hyperedge prediction methods: Expansion, Hyper-SAGNN (Zhang et al., 2020), NHP (Yadati et al., 2020), AHP (Hwang et al., 2022), CASH (Ko et al., 2025), and HyGEN (Yu et al., 2025). For these baselines, we use the results reported in their corresponding publications under the matched benchmark protocols. HyGEN reports results for SNS, MNS, and CNS without MIX; we therefore retain its published three-protocol average and mark it with † in the comparison tables. For external baselines, uncertainty is reproduced from the source report when available; a mean without $\pm$ indicates that no matched estimate is given.

Controlled analysis. All ablation and sensitivity variants use identical candidates, frozen backbone scores, splits, schedules, calibration, and evaluation, differing from the complete FAME-HP configuration only in the stated change.

For paired branch analyses, each dataset–seed–split combination forms one paired evaluation unit, yielding 90 units across the six datasets. We compute 95% confidence intervals using dataset-cluster bootstrap and apply Holm's procedure (Holm, 1979) to correct for multiple comparisons.

5.2 MAIN RESULTS

Table 1 compares FAME-HP with representative hyperedge prediction methods across six datasets and four evaluation protocols. In terms of the four-protocol average, FAME-HP achieves the highest AUROC on five datasets and ranks second on DBLP-A. For average precision, it obtains the best or tied-best result on five datasets and ranks second on DBLP-A. The largest overall gains appear on DBLP and Pubmed: compared with the strongest fully comparable baseline, FAME-HP improves average AUROC/AP by $0.090/0.117$ on DBLP and $0.070/0.078$ on Pubmed.

The protocol-level results further reveal where the improvement originates. Under SNS, FAME-HP achieves the highest AP on Cora-A, Pubmed, DBLP-A, and DBLP, while NHP or HyGEN remains stronger on Citeseer and Cora. Under MNS, the advantage is less pronounced: FAME-HP leads on DBLP but remains below the strongest reported baseline on the other five datasets. This indicates that the improvement is not a uniform shift across all negative-sampling conditions.

A different pattern emerges under CNS, where negatives are generated by replacing only one node of a true hyperedge and therefore remain structurally close to positive candidates. FAME-HP achieves the highest CNS AP on all six datasets. Relative to the strongest reported baseline, the AP improvements are 0.029, 0.014, 0.016, 0.062, 0.013, and 0.178 on Citeseer, Cora, Cora-A, Pubmed,

Table 1: Hyperedge prediction results on six datasets. Best results are in **bold**; second-best results are underlined.

Dataset	Method	AUROC					AP					Dataset	Method	AUROC					AP				
		SNS	MNS	CNS	MIX	AVG	SNS	MNS	CNS	MIX	AVG			SNS	MNS	CNS	MIX	AVG	SNS	MNS	CNS	MIX	AVG
Citeseer	Expansion	0.663	0.781	0.331	0.588	0.591	0.765	0.817	0.498	0.630	0.681	Cora	Expansion	0.470	0.707	0.256	0.476	0.477	0.637	0.764	0.454	0.563	0.607
	HyperSAGNN	0.540	0.410	0.473	0.478	0.475	0.627	0.455	0.497	0.507	0.512		HyperSAGNN	0.617	0.527	0.494	0.540	0.545	0.687	0.574	0.508	0.566	0.584
	NHP	**0.991**	0.701	0.510	0.817	0.751	**0.990**	0.731	0.520	0.768	0.751		NHP	0.943	0.641	0.472	0.774	0.703	0.949	0.678	0.509	0.744	0.718
	AHP	0.943	0.881	0.651	0.820	0.824	0.952	0.870	0.660	0.795	0.819		AHP	0.964	0.860	0.572	0.799	0.799	0.961	0.837	0.552	0.740	0.772
	CASH	0.925	0.921	0.720	0.857	0.856	0.928	**0.919**	0.701	0.831	0.845		CASH	0.923	0.867	0.671	0.824	0.822	0.915	0.854	0.644	0.789	0.801
	HyGEN†	0.984	**0.926**	0.676	–	0.862	0.985	0.912	0.694	–	0.864		HyGEN†	**0.991**	**0.907**	0.584	–	0.827	**0.990**	**0.896**	0.570	–	0.819
	FAME-HP	0.973	0.867	**0.754**	**0.859**	**0.864**	0.976	0.889	**0.730**	**0.861**	**0.864**		**FAME-HP**	0.955	0.834	**0.697**	**0.833**	**0.830**	0.963	0.858	**0.658**	**0.805**	**0.821**
Cora-A	Expansion	0.690	0.842	0.434	0.658	0.656	0.690	0.876	0.577	0.672	0.706	Pubmed	Expansion	0.520	0.730	0.241	0.497	0.497	0.675	0.755	0.440	0.565	0.612
	HyperSAGNN	0.386	0.591	0.542	0.505	0.506	0.532	0.643	0.545	0.563	0.571		HyperSAGNN	0.525	0.686	0.546	0.580	0.584	0.534	0.680	0.529	0.561	0.576
	NHP	0.909	0.672	0.550	0.773	0.723	0.925	0.720	0.585	0.766	0.748		NHP	0.973	0.694	0.524	0.745	0.733	0.973	0.656	0.513	0.678	0.707
	AHP	0.958	0.924	0.782	0.887	0.888	0.957	0.898	0.796	0.878	0.882		AHP	0.917	0.840	0.553	0.763	0.763	0.918	0.834	0.526	0.717	0.749
	CASH	0.971	**0.975**	0.833	**0.931**	0.927	0.969	**0.973**	0.832	**0.926**	**0.925**		CASH	0.805	**0.871**	0.640	0.772	0.772	0.810	0.880	0.644	0.765	0.775
	HyGEN†	0.977	0.947	0.803	–	0.909	0.977	0.914	0.829	–	0.907		HyGEN†	0.921	**0.871**	0.516	–	0.770	0.932	**0.890**	0.551	–	0.791
	FAME-HP	**0.987**	0.947	**0.892**	0.905	**0.933**	**0.987**	0.947	**0.848**	0.917	**0.925**		**FAME-HP**	**0.988**	0.815	**0.712**	**0.854**	**0.842**	**0.987**	0.849	**0.706**	**0.868**	**0.853**
DBLP-A	Expansion	0.634	0.826	0.350	0.603	0.603	0.730	0.852	0.512	0.641	0.687	DBLP	Expansion	0.645	0.801	0.366	0.607	0.607	0.751	0.856	0.518	0.655	0.698
	HyperSAGNN	0.548	0.791	0.563	0.636	0.634	0.686	0.805	0.552	0.655	0.675		HyperSAGNN	0.448	0.574	0.572	0.530	0.531	0.562	0.602	0.586	0.577	0.582
	NHP	0.966	0.623	0.555	0.721	0.716	0.965	0.604	0.534	0.663	0.693		NHP	0.663	0.540	0.503	0.572	0.569	0.608	0.523	0.501	0.542	0.544
	AHP	0.916	0.926	0.668	0.838	0.837	0.928	0.928	0.707	0.836	0.850		AHP	0.946	0.820	0.568	0.778	0.778	0.947	0.815	0.561	0.735	0.764
	CASH	0.929	**0.957**	**0.747**	**0.877**	**0.877**	0.933	**0.955**	0.741	0.863	**0.873**		CASH	0.875	0.836	0.708	0.807	0.807	0.874	0.832	0.696	0.793	0.799
	HyGEN†	0.913	0.927	0.671	–	0.837	0.927	0.929	0.715	–	0.857		HyGEN†	0.958	0.839	0.591	–	0.796	0.958	0.826	0.573	–	0.786
	FAME-HP	**0.994**	0.849	0.726	0.839	0.852	**0.994**	0.869	**0.754**	**0.872**	0.872		**FAME-HP**	**0.972**	**0.871**	**0.864**	**0.882**	**0.897**	**0.977**	**0.908**	**0.874**	**0.907**	**0.916**

AVG denotes the mean performance over SNS, MNS, CNS, and MIX. HyGEN† does not report results on MIX; therefore, its reported AVG is computed over SNS, MNS, and CNS only and is not included when determining the best and second-best AVG results.

Table 2: Ablation results of FAME-HP on six datasets. Best results are in **bold**; second-best results are underlined.

Dataset	Variant	AUROC					AP					Dataset	Variant	AUROC					AP				
		SNS	MNS	CNS	MIX	AVG	SNS	MNS	CNS	MIX	AVG			SNS	MNS	CNS	MIX	AVG	SNS	MNS	CNS	MIX	AVG
Citeseer	w/o Fuzzy Membership	0.967	0.840	0.716	0.829	0.838	0.970	0.859	0.690	0.829	0.837	Cora	w/o Fuzzy Membership	0.948	0.801	0.653	0.798	0.800	0.956	0.824	0.610	0.766	0.789
	w/o Membership Entropy	**0.978**	0.806	0.666	0.819	0.817	**0.980**	0.821	0.681	0.808	0.827		w/o Membership Entropy	0.931	0.713	0.546	0.732	0.730	0.945	0.719	0.557	0.704	0.740
	w/o Attribute View	0.972	0.825	0.690	0.826	0.828	0.975	0.839	0.686	0.820	0.830		w/o Attribute View	0.942	0.780	0.625	0.789	0.784	0.951	0.804	0.593	0.752	0.775
	w/o Cross-view	0.965	0.853	0.730	0.840	0.847	0.969	0.874	0.707	0.834	0.846		w/o Cross-view	0.947	0.817	0.673	0.807	0.811	0.955	0.838	0.631	0.776	0.800
	FAME-HP	0.973	**0.867**	**0.754**	**0.859**	**0.864**	0.976	**0.889**	**0.730**	**0.861**	**0.864**		**FAME-HP**	**0.955**	**0.834**	**0.697**	**0.833**	**0.830**	**0.963**	**0.858**	**0.658**	**0.805**	**0.821**
Cora-A	w/o Fuzzy Membership	0.980	0.909	0.848	0.863	0.900	0.980	0.908	0.803	0.881	0.893	Pubmed	w/o Fuzzy Membership	0.982	0.784	0.674	0.828	0.817	0.981	0.817	0.669	0.833	0.825
	w/o Membership Entropy	0.950	0.852	0.573	0.807	0.792	0.961	0.865	0.634	0.805	0.820		w/o Membership Entropy	0.980	0.775	0.654	0.816	0.803	0.979	0.789	0.653	0.799	0.807
	w/o Attribute View	0.972	0.890	0.810	0.856	0.882	0.974	0.895	0.770	0.873	0.878		w/o Attribute View	0.981	0.780	0.665	0.823	0.812	0.980	0.805	0.661	0.820	0.817
	w/o Cross-view	0.980	0.931	0.868	0.873	0.913	0.980	0.930	0.824	0.878	0.903		w/o Cross-view	0.982	0.801	0.690	0.831	0.826	0.980	0.835	0.683	0.842	0.835
	FAME-HP	**0.987**	**0.947**	**0.892**	**0.905**	**0.933**	**0.987**	**0.947**	**0.848**	**0.917**	**0.925**		**FAME-HP**	**0.988**	**0.815**	**0.712**	**0.854**	**0.842**	**0.987**	**0.849**	**0.706**	**0.868**	**0.853**
DBLP-A	w/o Fuzzy Membership	0.988	0.815	0.681	0.804	0.822	0.988	0.834	0.710	0.832	0.841	DBLP	w/o Fuzzy Membership	0.966	0.841	0.826	0.855	0.872	0.971	0.879	0.837	0.873	0.890
	w/o Membership Entropy	0.982	0.789	0.525	0.784	0.765	0.983	0.799	0.574	0.771	0.785		w/o Membership Entropy	**0.973**	0.809	0.747	0.859	0.843	0.976	0.850	0.794	0.878	0.873
	w/o Attribute View	0.986	0.805	0.650	0.795	0.809	0.985	0.825	0.675	0.819	0.826		w/o Attribute View	0.970	0.831	0.790	0.857	0.862	0.974	0.868	0.820	0.875	0.884
	w/o Cross-view	0.988	0.834	0.701	0.809	0.833	0.988	0.853	0.730	0.833	0.851		w/o Cross-view	0.966	0.858	0.844	0.856	0.881	0.971	0.894	0.856	0.871	0.898
	FAME-HP	**0.994**	**0.849**	**0.726**	**0.839**	**0.852**	**0.994**	**0.869**	**0.754**	**0.872**	**0.872**		**FAME-HP**	0.972	**0.871**	**0.864**	**0.882**	**0.897**	**0.977**	**0.908**	**0.874**	**0.907**	**0.916**

All variants use the same evaluation protocols. "w/o Fuzzy Membership" replaces fuzzy memberships with hard assignments; "w/o Membership Entropy" replaces structural-membership entropy with normalized mean node degree. AVG denotes the mean over SNS, MNS, CNS, and MIX.

DBLP-A, and DBLP, respectively. The gains are particularly pronounced on Pubmed and DBLP, where distinguishing locally plausible candidates requires more than coarse structural proximity.

The mixed protocol shows a similar tendency. FAME-HP obtains the highest MIX AP on five of the six datasets, with Cora-A as the only exception, and simultaneously leads both AUROC and AP on Citeseer, Cora, Pubmed, and DBLP. Taken together, these results suggest that the benefit of FAME-HP is concentrated in settings where candidate validity cannot be resolved by a single structural cue. This behavior is consistent with the proposed design: fuzzy memberships preserve overlapping affiliations, cross-view descriptors capture agreement between structural and semantic organization, and candidate context modulates how the resulting correction components are combined.

5.3 Ablation Study

Table 2 evaluates the contribution of the four main design choices. The complete FAME-HP obtains the highest Average AUROC and AP on all six datasets, indicating that the proposed components provide complementary improvements rather than benefits restricted to a particular benchmark.

Replacing fuzzy memberships with hard assignments reduces macro Average AP by 2.93 points. The decrease is also highly consistent across datasets, ranging from 2.6 to 3.2 points, showing that retaining overlapping affiliations provides a stable advantage across different hypergraph domains. Removing the attribute view causes a larger macro decrease of 4.02 points, while removing cross-view consistency reduces Average AP by 1.97 points. Interestingly, the cross-view contribution is comparatively modest but uniform: its Average-AP improvement remains approximately 1.8–2.2 points across all six datasets.

Membership entropy produces the strongest effect among the examined components. Replacing structural-membership entropy with normalized mean node degree lowers macro Average AP by 6.65 points. The degradation is particularly large on Cora-A and DBLP-A, where Average AP decreases by 10.5 and 8.7 points, respectively. This result is notable because the replacement preserves the dimensionality of the fusion input and changes only the corresponding structural descriptor, isolating the information carried by membership uncertainty rather than model capacity.

The same pattern becomes more pronounced under CNS. Removing fuzzy membership, membership entropy, the attribute view, and cross-view consistency decreases macro CNS AP by 4.18, 11.28, 6.08, and 2.32 points, respectively. For membership entropy, the CNS-AP drops reach 21.4 points on Cora-A and 18.0 points on DBLP-A. These results are consistent with the motivation of the entropy descriptor: candidates with similar pairwise membership compatibility can still differ substantially in how concentrated or diffuse their latent affiliations are, and this distributional information becomes particularly useful for hard candidates

5.4 Parameter Sensitivity

We vary the community dimension K and coefficients η_p, λ_H, λ_r, and λ_s one at a time. Figure 3 reports macro Average and CNS AP over six datasets; Appendix C gives complete dataset-level results.

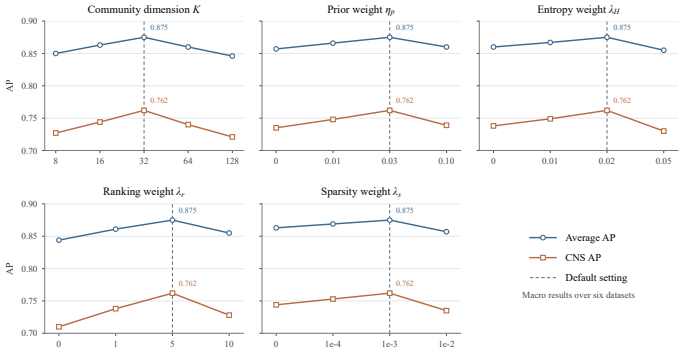

Figure 3: Parameter sensitivity of FAME-HP. Each panel varies one hyperparameter while fixing the remaining settings to their default values. We report macro Average AP and CNS AP over six datasets. Dashed vertical lines indicate the default settings.

Performance peaks near the defaults. Increasing K from 8 to 32 improves both metrics, whereas larger values degrade them. Moderate η_p, λ_H, and λ_s also perform best, with nearby settings remaining competitive; thus neither oversized membership spaces nor strong regularization is required.

λ_r varies most, especially under CNS: values 0, 5, and 10 yield CNS AP of 0.710, 0.762, and 0.728, and Average AP of 0.844, 0.875, and 0.855. Ranking supervision therefore helps with hard candidates, but overweighting it weakens classification. The defaults give the strongest macro performance for all five hyperparameters.

6 Conclusion

FAME-HP organizes structural, attribute, and cross-view fuzzy-membership evidence and fuses source-indexed components by candidate context. Across six datasets and four protocols, it is particularly effective on hard one-node-replacement negatives. Completed ablations verify the contributions of fuzzy memberships, structural-membership entropy, attribute evidence, and cross-view consistency; sensitivity results show stability near the selected configuration.

AI USE STATEMENT

Generative AI tools were used to help organize and edit the English manuscript, translate author-provided material, transform author-verified experimental records into LaTeX tables, create the method diagram, review experimental methodology, interpret reported results, and check consistency between claims, code, and values. AI-assisted suggestions were checked against the project records, cited literature, and executed analyses by the authors. The authors reviewed the resulting prose, equations, citations, figures, tables, and scripts and take responsibility for every claim and artifact. Generative AI was not used to generate experimental observations or replace execution of the reported experiments.

REPRODUCIBILITY STATEMENT

Appendices A–C report the evaluation details, full ten-metric baseline, ablation, and sensitivity results, and statistical interpretation. The accompanying anonymous source compiles these tables directly, while the internal experiment archive retains seed–split records and provenance checksums for author verification. We distinguish externally reported baseline values from our completed experiments and state the aggregation convention for each table. Every ablation and sensitivity value in the manuscript is obtained from a completed experimental run under the stated protocol; no forecast or interpolated result is included. The full-model evidence design and the modification defining each controlled variant were fixed before the reported experiments. All controlled variants follow the same experimental and evaluation pipeline, differing only in the stated change.

REFERENCES

Federico Battiston, Giulia Cencetti, Iacopo Iacopini, Vito Latora, Maxime Lucas, Alice Patania, Jean-Gabriel Young, and Giovanni Petri. Networks beyond pairwise interactions: Structure and dynamics. *Physics Reports*, 874:1–92, 2020. ISSN 0370-1573. doi: https://doi.org/10.1016/j.physrep.2020.05.004. Networks beyond pairwise interactions: Structure and dynamics.

Xiaochun Cao, Xiao Wang, Di Jin, Yixin Cao, and Dongxiao He. Identifying overlapping communities as well as hubs and outliers via nonnegative matrix factorization. *Scientific reports*, 3(1): 2993, 2013.

Eli Chien, Chao Pan, Jianhao Peng, and Olgica Milenkovic. You are allset: A multiset function framework for hypergraph neural networks. In *International Conference on Learning Representations*, 2022.

Martina Contisciani, Federico Battiston, and Caterina De Bacco. Inference of hyperedges and overlapping communities in hypergraphs. *Nature communications*, 13(1):7229, 2022.

Yifan Feng, Haoxuan You, Zizhao Zhang, Rongrong Ji, and Yue Gao. Hypergraph neural networks. In *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence*, AAAI'19/IAAI'19/EAAI'19. AAAI Press, 2019. ISBN 978-1-57735-809-1. doi: 10.1609/aaai.v33i01.33013558.

Sture Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2):65–70, 1979. ISSN 03036898, 14679469.

Jing Huang and Jie Yang. Unignn: a unified framework for graph and hypergraph neural networks. In Zhi-Hua Zhou (ed.), *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, pp. 2563–2569. International Joint Conferences on Artificial Intelligence Organization, 8 2021. doi: 10.24963/ijcai.2021/353. Main Track.

Hyunjin Hwang, Seungwoo Lee, Chanyoung Park, and Kijung Shin. Ahp: Learning to negative sample for hyperedge prediction. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '22, pp. 2237–2242, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450387323. doi: 10.1145/3477495.3531836.

Yunyong Ko, Hanghang Tong, and Sang-Wook Kim. Enhancing hyperedge prediction with context-aware self-supervised learning. *IEEE Transactions on Knowledge and Data Engineering*, 37(4): 1772–1784, 2025. doi: 10.1109/TKDE.2025.3532263.

Daniel D Lee and H Sebastian Seung. Learning the parts of objects by non-negative matrix factorization. *nature*, 401(6755):788–791, 1999.

Juho Lee, Yoonho Lee, Jungtaek Kim, Adam Kosiorek, Seungjin Choi, and Yee Whye Teh. Set transformer: A framework for attention-based permutation-invariant neural networks. In Kamalika Chaudhuri and Ruslan Salakhutdinov (eds.), *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 3744–3753. PMLR, 09–15 Jun 2019.

Junbo Wang, Jianrui Chen, Zhihui Wang, and Maoguo Gong. Hypergraph contrastive attention networks for hyperedge prediction with negative samples evaluation. *Neural Networks*, 181: 106807, 2025. ISSN 0893-6080. doi: https://doi.org/10.1016/j.neunet.2024.106807.

Jing Xiao, Zhi-Wei Ma, Jing Cao, and Xiao-Ke Xu. Hypergraph community detection with fuzzy memberships. *IEEE Transactions on Fuzzy Systems*, 33(9):3253–3266, 2025. doi: 10.1109/TFUZZ.2025.3588125.

Naganand Yadati, Madhav Nimishakavi, Prateek Yadav, Vikram Nitin, Anand Louis, and Partha Talukdar. Hypergcn: A new method for training graph convolutional networks on hypergraphs. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

Naganand Yadati, Vikram Nitin, Madhav Nimishakavi, Prateek Yadav, Anand Louis, and Partha Talukdar. Nhp: Neural hypergraph link prediction. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, CIKM '20, pp. 1705–1714, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450368599. doi: 10.1145/3340531.3411870.

Se-eun Yoon, Hyungseok Song, Kijung Shin, and Yung Yi. How much and when do we need higher-order information in hypergraphs? a case study on hyperedge prediction. In *Proceedings of The Web Conference 2020*, WWW '20, pp. 2627–2633, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450370233. doi: 10.1145/3366423.3380016.

Song Kyung Yu, Da Eun Lee, Yunyong Ko, and Sang-Wook Kim. Hygen: Regularizing negative hyperedge generation for accurate hyperedge prediction. In *Companion Proceedings of the ACM on Web Conference 2025*, WWW '25, pp. 1500–1504, New York, NY, USA, 2025. Association for Computing Machinery. ISBN 9798400713316. doi: 10.1145/3701716.3715456.

Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Russ R Salakhutdinov, and Alexander Smola. Deep sets. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

Ruochi Zhang, Yuesong Zou, and Jian Ma. Hyper-sagnn: a self-attention based graph neural network for hypergraphs. In *International Conference on Learning Representations*, 2020.

Shihua Zhang, Rui-Sheng Wang, and Xiang-Sun Zhang. Uncovering fuzzy community structure in complex networks. *Phys. Rev. E*, 76:046103, Oct 2007. doi: 10.1103/PhysRevE.76.046103.

A Dataset Statistics and Evaluation Details

A.1 Dataset Statistics

Table 3: Statistics of the six hypergraph datasets.

Dataset	Nodes	Hyperedges
Citeseer	1,458	1,004
Cora	1,434	1,483
Cora-A	2,388	970
Pubmed	3,840	7,531
DBLP-A	41,302	20,865
DBLP	15,639	18,390

Table 3 summarizes the six real-world hypergraphs used in our experiments after canonical hyperedge deduplication. The datasets span citation, authorship, and collaboration networks and exhibit substantial variation in both the number of nodes and the number of observed hyperedges.

A.2 Evaluation Protocol

Candidate hyperedges are treated as unordered node sets. For each positive hyperedge, SNS samples a random non-hyperedge with matched cardinality. MNS constructs negative candidates from the clique-expansion neighborhood frontier, with random sampling used when no valid frontier candidate is available. CNS generates structurally challenging negatives by replacing one node of a positive hyperedge with a node outside the original hyperedge while excluding candidates that already occur as observed positives. MIX combines SNS, MNS, and CNS in equal proportions.

We report AUROC and average precision (AP) separately under SNS, MNS, CNS, and MIX. AVG denotes the mean performance over the four protocols. All candidate-level metrics are computed directly from prediction scores.

B Implementation Details

FAME-HP employs an AllSet-style hypergraph encoder to obtain node representations and a permutation-invariant candidate encoder to produce the base prediction. These pretrained base modules are frozen during evidence learning. Structural and attribute fuzzy memberships are initialized from the corresponding NMF-derived priors and subsequently optimized with the fusion objective.

Membership initialization. For each view $r \in \{s, a\}$, the node-side NMF factor is converted into a bounded membership prior by row-wise ℓ_1 normalization. Rows with negligible mass are assigned the uniform vector $K^{-1}\mathbf{1}$. The resulting prior $M_0^{(r)}$ is clipped to $[\epsilon, 1-\epsilon]$ before applying the logit transformation, which initializes the trainable membership logits $Z^{(r)}$. The sigmoid parameterization is then used throughout training so that each membership dimension remains independently activated, preserving overlapping affiliations rather than imposing a simplex constraint.

Evidence aggregation. Candidate-level structural and attribute evidence is computed from the pairwise membership statistics defined in Section 4. For $r \in \{s, a\}$, the fixed aggregation used throughout all experiments is

$$\phi_r(e) = \mathrm{clip}_{[0,1]}\left(0.55\mu_r(e) + 0.35m_r(e) + 0.10[1 - \sigma_r(e)]\right).$$

Cross-view consistency is aggregated from similarity discrepancy, coverage discrepancy, centroid alignment, and node-level alignment:

$$\phi_{cv}(e) = \mathrm{clip}_{[0,1]}\left(0.35[1 - \Delta_{\mathrm{sim}}(e)] + 0.25[1 - \Delta_{\mathrm{cov}}(e)] + 0.25A_{\mathrm{centroid}}(e) + 0.15A_{\mathrm{node}}(e)\right).$$

These aggregation rules are fixed across datasets and are not tuned for individual benchmarks.

Fusion-state construction. The fusion module operates on a 41-dimensional candidate representation consisting of an 8-dimensional response state, a 15-dimensional membership descriptor, and an 18-dimensional formation context. The response state is constructed as

$$\boldsymbol{q}_e = [p_e^{\text{base}}, \phi_s(e), \phi_a(e), \phi_{cv}(e), \widetilde{|e|}, \\ |p_e^{\text{base}} - \phi_s(e)|, |p_e^{\text{base}} - \phi_a(e)|, |\phi_s(e) - \phi_a(e)|],$$

where

$$\widetilde{|e|} = \frac{\log(1+|e|)}{\log(1+m_{\max})}, \qquad m_{\max} = 30.$$

The first four entries summarize the current base and evidence responses, while the remaining entries encode candidate size and disagreement among those responses. Together with the membership descriptor $\boldsymbol{\tau}_e$ and formation context $\boldsymbol{c}_e$, the complete fusion input is

$$\boldsymbol{x}_e = [\boldsymbol{q}_e; \boldsymbol{\tau}_e; \boldsymbol{c}_e].$$

Each coordinate is standardized using statistics estimated from the corresponding training split before being passed to the fusion heads. The same transformation is then applied unchanged to validation and test candidates.

Fusion heads. The evidence and gating heads are shallow MLPs with hidden width 24 and three outputs. Both receive the same standardized 41-dimensional representation. The evidence head produces the three source-indexed correction components, whereas the gating head produces three logits followed by softmax normalization to obtain the candidate-specific mixture weights. Together, the two heads contain 2,166 trainable weights and biases. The base predictor remains frozen, so optimization in this stage is restricted to the fuzzy membership parameters, the two fusion heads, and the global fusion parameters.

Formation context. The 18 formation-context entries are computed from observed training hyperedges. The four local entries are subedge support, pair closure, local closure, and minimum edit similarity. The eight prototype entries comprise same-size and near-size top-1/top-3 Jaccard similarity, pair-overlap top-1/top-3, exact-size hit, and minimum size gap. The six consistency entries comprise closure-gap top-1/top-3, pair-closure ratio top-1/top-3, the top-1 Jaccard–closure gap, and the top-1 local–pair gap. These context features are computed without using candidate labels and depend only on the observed training hypergraph and the candidate node set.

Optimization objective. For completeness, let $\mathcal{B}$ denote a candidate batch and $\mathcal{P}$ the set of matched positive–negative candidate pairs. The classification term is the weighted binary cross-entropy

$$\mathcal{L}_{\text{BCE}} = -\frac{1}{|\mathcal{B}|} \sum_{e \in \mathcal{B}} \left[w_+ y_e \log p_e + (1 - y_e) \log(1 - p_e) \right],$$

where w_+ denotes the positive-class weight. Pairwise discrimination is encouraged through

$$\mathcal{L}_{\text{rank}} = \frac{1}{|\mathcal{P}|} \sum_{(e^+, e^-) \in \mathcal{P}} \text{softplus}\left(-(z_{e^+} - z_{e^-})\right).$$

The membership prior and sparsity regularizers are

$$\mathcal{L}_{\text{prior}} = \sum_{r \in \{s,a\}} \frac{\|M^{(r)} - M_0^{(r)}\|_F^2}{|\mathcal{V}|K},$$

and

$$\mathcal{L}_{\text{sparse}} = \sum_{r \in \{s,a\}} \frac{1}{|\mathcal{V}|K} \sum_{v \in \mathcal{V}} \sum_{k=1}^{K} M_{v,k}^{(r)}.$$

Finally, routing entropy is computed over the candidate-specific gate:

$$\mathcal{H}_{\text{gate}} = -\frac{1}{|\mathcal{B}|} \sum_{e \in \mathcal{B}} \sum_{r \in \{s,a,cv\}} \alpha_{e,r} \log(\alpha_{e,r} + \epsilon).$$

The complete objective follows Eq. equation 11. Since $\mathcal{H}_{\text{gate}}$ is subtracted from the minimized loss, it discourages premature collapse to a single correction component. This training regularizer is distinct from the candidate-level membership entropy used as part of $\boldsymbol{\tau}_e$.

Training configuration. Unless otherwise specified, we use the configuration summarized in Table 4. Each dataset is evaluated using three random seeds and five train–validation–test splits per seed.

Table 4: Default implementation settings of FAME-HP.

Setting	Value
Community dimension K	32
Prior weight η_p	0.03
Entropy weight λ_H	0.02
Ranking weight λ_r	5
Sparsity weight λ_s	10^{-3}
Candidate context dimension	18
Local-structure features	4
Prototype-matching features	8
Consistency features	6
Optimizer	Adam
Random seeds	3
Splits per seed	5

C Full Parameter-Sensitivity Results

Section 5.4 summarizes the parameter-sensitivity results using macro Average AP and CNS AP. Here, we provide the complete dataset-level results under SNS, MNS, CNS, and MIX. For each experiment, one hyperparameter is varied while all remaining settings are fixed to their default values. All settings use the same 15 seed–split runs as the main experiment; protocol entries are means and AVG is mean $\pm$ standard deviation. The results consistently show that the configuration used in the main experiments provides a strong operating point across different datasets and negative-sampling regimes.

C.1 Community Dimension K

The community dimension K controls the capacity of the fuzzy membership space and therefore determines how many latent affiliations can be represented for each node.

Table 5: Full sensitivity results for the community dimension K. Best results are in **bold**; second-best results are underlined.

Dataset	K	AUROC					AP					Dataset	K	AUROC					AP				
		SNS	MNS	CNS	MIX	AVG	SNS	MNS	CNS	MIX	AVG			SNS	MNS	CNS	MIX	AVG	SNS	MNS	CNS	MIX	AVG
	8	.958	.847	.720	.836	.840	.960	.868	.697	.840	.841		8	.939	.810	.656	.808	.803	.945	.832	.618	.781	.794
	16	.965	.857	.737	.848	.852	.968	.879	.713	.850	.853		16	.947	.822	.676	.820	.816	.954	.845	.638	.793	.808
Citeseer	**32**	**.973**	**.867**	**.754**	**.859**	**.864**	**.976**	**.889**	**.730**	**.861**	**.864**	Cora	**32**	**.955**	**.834**	**.697**	**.833**	**.830**	**.963**	**.858**	**.658**	**.805**	**.821**
	64	.964	.855	.732	.845	.849	.966	.876	.709	.848	.850		64	.945	.819	.671	.817	.813	.952	.842	.633	.790	.804
	128	.955	.844	.714	.833	.836	.957	.865	.691	.836	.837		128	.936	.806	.649	.804	.798	.943	.828	.611	.777	.790
	8	.969	.924	.849	.877	.905	.968	.923	.807	.890	.897		8	.973	.795	.683	.831	.821	.971	.828	.678	.846	.831
	16	.978	.935	.871	.891	.919	.977	.935	.828	.904	.911		16	.981	.805	.698	.843	.832	.979	.839	.692	.857	.842
Cora-A	**32**	**.987**	**.947**	**.892**	**.905**	**.933**	**.987**	**.947**	**.848**	**.917**	**.925**	Pubmed	**32**	**.988**	**.815**	**.712**	**.854**	**.842**	**.987**	**.849**	**.706**	**.868**	**.853**
	64	.976	.933	.865	.887	.915	.975	.932	.822	.900	.907		64	.979	.803	.694	.840	.829	.977	.836	.689	.855	.839
	128	.966	.920	.842	.872	.900	.964	.919	.800	.886	.892		128	.971	.792	.678	.828	.817	.969	.825	.674	.843	.827
	8	.978	.825	.689	.815	.827	.977	.843	.718	.849	.847		8	.956	.849	.831	.857	.873	.960	.885	.843	.883	.893
	16	.986	.837	.707	.827	.839	.985	.856	.736	.861	.860		16	.964	.860	.848	.870	.885	.969	.896	.858	.895	.905
DBLP-A	**32**	**.994**	**.849**	**.726**	**.839**	**.852**	**.994**	**.869**	**.754**	**.872**	**.872**	DBLP	**32**	**.972**	**.871**	**.864**	**.882**	**.897**	**.977**	**.908**	**.874**	**.907**	**.916**
	64	.984	.834	.703	.824	.836	.983	.853	.732	.858	.856		64	.962	.857	.844	.866	.882	.966	.893	.854	.892	.902
	128	.975	.821	.683	.811	.822	.974	.839	.712	.846	.843		128	.954	.845	.826	.853	.869	.957	.881	.837	.879	.889

Performance generally improves as K increases from 8 to 32, after which further enlargement brings no additional benefit. The trend is consistent across all six datasets, with $K = 32$ achieving the

strongest Average AUROC and AP as well as CNS performance. These results suggest that a moderate membership dimension is sufficient to capture the overlapping organization needed for hyperedge prediction, while an unnecessarily large latent space can introduce redundant factors.

C.2 Prior Weight η_p

The coefficient η_p controls the strength of the membership-prior regularization during joint optimization.

Table 6: Full sensitivity results for the prior weight η_p. Best results are in **bold**; second-best results are underlined.

Dataset	η_p	AUROC					AP					Dataset	η_p	AUROC					AP				
		SNS	MNS	CNS	MIX	AVG	SNS	MNS	CNS	MIX	AVG			SNS	MNS	CNS	MIX	AVG	SNS	MNS	CNS	MIX	AVG
Citeseer	0	.963	.851	.727	.842	.846	.966	.873	.704	.845	.847	Cora	0	.945	.815	.665	.814	.810	.952	.838	.627	.787	.801
	.01	.968	.859	.741	.851	.855	.971	.881	.717	.853	.855		.01	.950	.824	.681	.824	.820	.957	.848	.642	.796	.811
	.03	**.973**	**.867**	**.754**	**.859**	**.864**	**.976**	**.889**	**.730**	**.861**	**.864**		**.03**	**.955**	**.834**	**.697**	**.833**	**.830**	**.963**	**.858**	**.658**	**.805**	**.821**
	.10	.965	.854	.732	.845	.849	.967	.875	.708	.848	.850		.10	.946	.818	.670	.817	.813	.954	.841	.632	.790	.804
Cora-A	0	.976	.929	.859	.884	.912	.975	.928	.816	.897	.904	Pubmed	0	.979	.799	.690	.837	.826	.977	.833	.684	.852	.836
	.01	.981	.938	.875	.894	.922	.981	.937	.832	.907	.914		.01	.983	.807	.701	.845	.834	.982	.841	.695	.860	.844
	.03	**.987**	**.947**	**.892**	**.905**	**.933**	**.987**	**.947**	**.848**	**.917**	**.925**		**.03**	**.988**	**.815**	**.712**	**.854**	**.842**	**.987**	**.849**	**.706**	**.868**	**.853**
	.10	.978	.932	.864	.887	.915	.977	.931	.821	.900	.907		.10	.980	.802	.693	.840	.829	.979	.835	.688	.855	.839
DBLP-A	0	.984	.830	.697	.821	.833	.983	.849	.726	.855	.853	DBLP	0	.962	.853	.839	.863	.879	.966	.889	.850	.889	.899
	.01	.989	.839	.712	.830	.842	.988	.859	.740	.863	.863		.01	.967	.862	.851	.873	.888	.972	.899	.862	.898	.908
	.03	**.994**	**.849**	**.726**	**.839**	**.852**	**.994**	**.869**	**.754**	**.872**	**.872**		**.03**	**.972**	**.871**	**.864**	**.882**	**.897**	**.977**	**.908**	**.874**	**.907**	**.916**
	.10	.986	.833	.702	.824	.836	.985	.852	.731	.858	.856		.10	.964	.856	.843	.866	.882	.968	.893	.854	.892	.902

The results exhibit a consistent peak around $\eta_p = 0.03$. A smaller weight provides weaker prior guidance, whereas increasing the coefficient to 0.10 reduces performance across the datasets. The neighboring setting $\eta_p = 0.01$ remains competitive, indicating that FAME-HP is reasonably stable within a moderate range of prior regularization.

C.3 Gate-Entropy Weight λ_H

The gate-entropy coefficient λ_H regulates the concentration of candidate-adaptive fusion weights. Because the gate entropy is maximized in Equation 11, this term prevents premature concentration on a single evidence pathway rather than directly enforcing sparse allocation.

Table 7: Full sensitivity results for the entropy weight λ_H. Best results are in **bold**; second-best results are underlined.

Dataset	λ_H	AUROC					AP					Dataset	λ_H	AUROC					AP				
		SNS	MNS	CNS	MIX	AVG	SNS	MNS	CNS	MIX	AVG			SNS	MNS	CNS	MIX	AVG	SNS	MNS	CNS	MIX	AVG
Citeseer	0	.966	.855	.731	.844	.849	.968	.877	.708	.846	.850	Cora	0	.947	.820	.669	.816	.813	.954	.843	.631	.788	.804
	.01	.969	.861	.742	.851	.856	.972	.882	.718	.853	.856		.01	.951	.826	.682	.824	.821	.958	.850	.644	.796	.812
	.02	**.973**	**.867**	**.754**	**.859**	**.864**	**.976**	**.889**	**.730**	**.861**	**.864**		**.02**	**.955**	**.834**	**.697**	**.833**	**.830**	**.963**	**.858**	**.658**	**.805**	**.821**
	.05	.963	.851	.723	.838	.844	.965	.873	.700	.841	.845		.05	.945	.815	.660	.810	.807	.952	.838	.622	.783	.799
Cora-A	0	.978	.933	.864	.886	.915	.978	.933	.820	.899	.907	Pubmed	0	.981	.803	.693	.838	.829	.979	.837	.687	.853	.839
	.01	.982	.940	.877	.895	.923	.982	.939	.833	.907	.915		.01	.984	.809	.702	.846	.835	.983	.842	.696	.860	.845
	.02	**.987**	**.947**	**.892**	**.905**	**.933**	**.987**	**.947**	**.848**	**.917**	**.925**		**.02**	**.988**	**.815**	**.712**	**.854**	**.842**	**.987**	**.849**	**.706**	**.868**	**.853**
	.05	.976	.929	.854	.879	.909	.974	.928	.811	.892	.901		.05	.979	.800	.686	.833	.824	.977	.833	.681	.848	.835
DBLP-A	0	.986	.835	.701	.823	.836	.986	.854	.730	.856	.856	DBLP	0	.964	.858	.842	.865	.882	.969	.894	.853	.891	.902
	.01	.990	.841	.713	.830	.844	.989	.861	.741	.864	.864		.01	.968	.864	.852	.873	.889	.973	.901	.863	.898	.909
	.02	**.994**	**.849**	**.726**	**.839**	**.852**	**.994**	**.869**	**.754**	**.872**	**.872**		**.02**	**.972**	**.871**	**.864**	**.882**	**.897**	**.977**	**.908**	**.874**	**.907**	**.916**
	.05	.984	.830	.693	.817	.831	.983	.849	.722	.851	.851		.05	.962	.853	.835	.859	.877	.966	.889	.846	.885	.897

Moderate gate-entropy regularization consistently provides the strongest performance. In particular, $\lambda_H = 0.02$ yields the best overall results, while $\lambda_H = 0.01$ remains close across most datasets and protocols. Removing the term or increasing it to 0.05 leads to larger degradation, especially under

CNS, indicating that an appropriate degree of mixture-weight regularization is beneficial for difficult candidate discrimination.

C.4 RANKING WEIGHT λ_r

The coefficient λ_r balances the pairwise ranking objective against the candidate-level classification objective.

Table 8: Full sensitivity results for the ranking weight λ_r. Best results are in **bold**; second-best results are underlined.

Dataset	λ_r	AUROC					AP					Dataset	λ_r	AUROC					AP				
		SNS	MNS	CNS	MIX	AVG	SNS	MNS	CNS	MIX	AVG			SNS	MNS	CNS	MIX	AVG	SNS	MNS	CNS	MIX	AVG
Citeseer	0	.960	.842	.703	.832	.834	.961	.863	.681	.836	.835	Cora	0	.941	.804	.636	.803	.796	.947	.826	.599	.776	.787
	1	.967	.856	.730	.846	.850	.969	.877	.707	.849	.851		1	.948	.820	.669	.819	.814	.956	.843	.630	.792	.805
	5	**.973**	**.867**	**.754**	**.859**	**.864**	**.976**	**.889**	**.730**	**.861**	**.864**		**5**	**.955**	**.834**	**.697**	**.833**	**.830**	**.963**	**.858**	**.658**	**.805**	**.821**
	10	.964	.851	.720	.841	.844	.966	.872	.697	.844	.845		10	.946	.814	.656	.813	.807	.953	.837	.619	.786	.798
Cora-A	0	.971	.918	.829	.872	.898	.970	.917	.787	.885	.890	Pubmed	0	.975	.790	.670	.827	.816	.973	.823	.665	.842	.826
	1	.980	.934	.863	.889	.916	.979	.933	.820	.902	.908		1	.982	.804	.692	.841	.830	.980	.837	.687	.856	.840
	5	**.987**	**.947**	**.892**	**.905**	**.933**	**.987**	**.947**	**.848**	**.917**	**.925**		**5**	**.988**	**.815**	**.712**	**.854**	**.842**	**.987**	**.849**	**.706**	**.868**	**.853**
	10	.977	.928	.850	.883	.909	.976	.927	.808	.896	.901		10	.980	.799	.684	.836	.824	.978	.832	.679	.851	.835
DBLP-A	0	.980	.819	.671	.811	.820	.979	.837	.701	.845	.840	DBLP	0	.958	.843	.816	.852	.867	.962	.879	.827	.879	.887
	1	.988	.835	.701	.826	.837	.987	.854	.729	.859	.857		1	.966	.858	.842	.868	.883	.970	.894	.852	.894	.903
	5	**.994**	**.849**	**.726**	**.839**	**.852**	**.994**	**.869**	**.754**	**.872**	**.872**		**5**	**.972**	**.871**	**.864**	**.882**	**.897**	**.977**	**.908**	**.874**	**.907**	**.916**
	10	.985	.829	.690	.820	.831	.984	.848	.719	.854	.851		10	.963	.852	.832	.862	.877	.967	.888	.843	.888	.897

Among the examined hyperparameters, λ_r produces one of the clearest performance variations. Increasing the ranking weight from 0 to 5 improves both Average and CNS performance across the six datasets, with particularly large gains on structurally challenging candidates. Increasing it further to 10 reduces performance, suggesting that ranking supervision is most effective when it complements rather than dominates the classification objective.

C.5 SPARSITY WEIGHT λ_s

The sparsity coefficient λ_s controls the regularization applied to the learned fuzzy membership representations.

Table 9: Full sensitivity results for the sparsity weight λ_s. Best results are in **bold**; second-best results are underlined.

Dataset	λ_s	AUROC					AP					Dataset	λ_s	AUROC					AP				
		SNS	MNS	CNS	MIX	AVG	SNS	MNS	CNS	MIX	AVG			SNS	MNS	CNS	MIX	AVG	SNS	MNS	CNS	MIX	AVG
Citeseer	0	.966	.856	.737	.848	.852	.968	.878	.713	.851	.853	Cora	0	.947	.821	.676	.821	.816	.955	.844	.638	.794	.808
	10^{-4}	.970	.862	.745	.854	.858	.972	.883	.722	.856	.858		10^{-4}	.951	.827	.687	.827	.823	.959	.851	.648	.799	.814
	10^{-3}	**.973**	**.867**	**.754**	**.859**	**.864**	**.976**	**.889**	**.730**	**.861**	**.864**		10^{-3}	**.955**	**.834**	**.697**	**.833**	**.830**	**.963**	**.858**	**.658**	**.805**	**.821**
	10^{-2}	.963	.851	.728	.843	.846	.965	.872	.705	.846	.847		10^{-2}	.944	.814	.666	.815	.810	.951	.837	.628	.788	.801
Cora-A	0	.979	.934	.871	.892	.919	.978	.934	.828	.904	.911	Pubmed	0	.981	.804	.698	.843	.832	.980	.838	.692	.858	.842
	10^{-4}	.983	.941	.881	.898	.926	.983	.940	.838	.911	.918		10^{-4}	.985	.810	.705	.849	.837	.983	.843	.699	.863	.847
	10^{-3}	**.987**	**.947**	**.892**	**.905**	**.933**	**.987**	**.947**	**.848**	**.917**	**.925**		10^{-3}	**.988**	**.815**	**.712**	**.854**	**.842**	**.987**	**.849**	**.706**	**.868**	**.853**
	10^{-2}	.975	.928	.860	.885	.912	.974	.927	.817	.898	.904		10^{-2}	.978	.799	.690	.838	.826	.976	.832	.685	.853	.836
DBLP-A	0	.987	.836	.707	.828	.839	.986	.855	.736	.861	.860	DBLP	0	.965	.859	.848	.870	.885	.969	.895	.858	.896	.905
	10^{-4}	.990	.842	.717	.833	.846	.990	.862	.745	.867	.866		10^{-4}	.968	.865	.856	.876	.891	.973	.902	.866	.901	.911
	10^{-3}	**.994**	**.849**	**.726**	**.839**	**.852**	**.994**	**.869**	**.754**	**.872**	**.872**		10^{-3}	**.972**	**.871**	**.864**	**.882**	**.897**	**.977**	**.908**	**.874**	**.907**	**.916**
	10^{-2}	.983	.829	.698	.822	.833	.982	.848	.727	.856	.853		10^{-2}	.961	.852	.840	.864	.879	.965	.889	.850	.890	.899

The results are stable around moderate sparsity regularization. $\lambda_s = 10^{-3}$ achieves the strongest overall performance, while 10^{-4} remains consistently competitive. In contrast, removing the regularizer or increasing it to 10^{-2} leads to lower performance, indicating that moderate sparsity helps preserve informative latent affiliations without excessively constraining the membership space.