
How Particle System Theory Enhances Hypergraph Message Passing

Yixuan Ma

Shanghai Jiao Tong University
mayx5901@sjtu.edu.cn

Kai Yi

University of Cambridge
ky347@cam.ac.uk

Pietro Liò

University of Cambridge
pl219@cam.ac.uk

Shi Jin

Shanghai Jiao Tong University
shijin-m@sjtu.edu.cn

Yu Guang Wang*

Shanghai Jiao Tong University
yuguang.wang@sjtu.edu.cn

Abstract

Hypergraphs effectively model higher-order relationships in natural phenomena, capturing complex interactions beyond pairwise connections. We introduce a novel hypergraph message passing framework inspired by interacting particle systems, where hyperedges act as fields inducing shared node dynamics. By incorporating attraction, repulsion, and Allen-Cahn forcing terms, particles of varying classes and features achieve class-dependent equilibrium, enabling separability through the particle-driven message passing. We investigate both first-order and second-order particle system equations for modeling these dynamics, which mitigate over-smoothing and heterophily thus can capture complete interactions. The more stable second-order system permits deeper message passing. Furthermore, we enhance deterministic message passing with stochastic element to account for interaction uncertainties. We prove theoretically that our approach mitigates over-smoothing by maintaining a positive lower bound on the hypergraph Dirichlet energy during propagation and thus to enable hypergraph message passing to go deep. Empirically, our models demonstrate competitive performance on diverse real-world hypergraph node classification tasks, excelling on both homophilic and heterophilic datasets. Source code is available at [the link](#).

1 Introduction

Hypergraph Neural Networks (HNNs) [20], built upon Graph Neural Networks (GNNs) [29], have demonstrated remarkable success in modeling higher-order relationships involving multiple nodes [28]. In complex systems such as social networks [5, 32, 50] and biomolecular interactions [41, 46], hypergraphs can effectively capture complex group dynamics compared to pairwise graphs.

The multi-node, higher-order nature of hypergraphs naturally lends itself to an analogy with particle dynamical systems. This is because hypergraph message passing, much like the inherent interactions in particle motion, fundamentally involves multiple interacting components. Driven by this insight, we introduce Hypergraph Atomic Message Passing (HAMP), a novel framework that reframes hypergraph message passing through the lens of interacting particle systems. Our key innovation in HAMP is the conceptualization of each hyperedge as a dynamic field that governs the shared dynamics of nodes. As Figure 1 shows, HAMP updates hypergraph embeddings by superimposing the forces exerted by the hypergraph's nodes (particles). This framework offers significant advantages, including

*Corresponding author

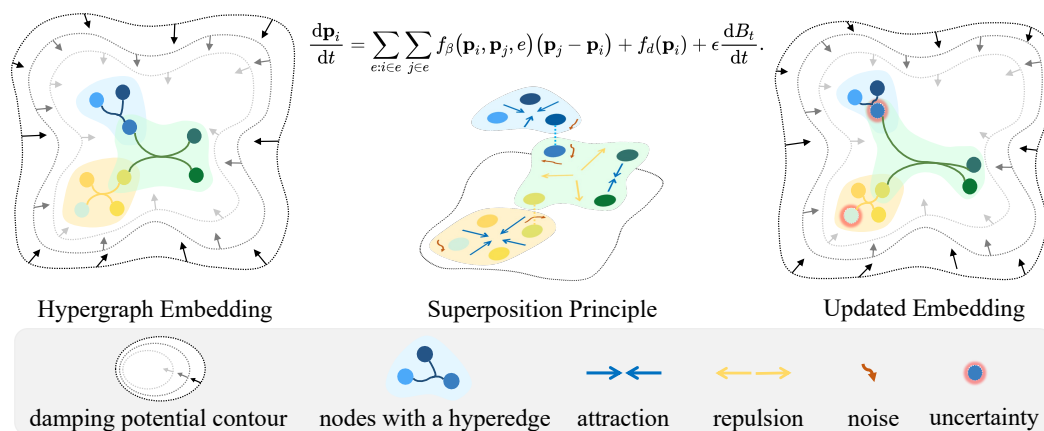


Figure 1: An illustration for HAMP framework. The property $\mathbf{p}$ can account for feature or velocity.

the ability to mitigate over-smoothing and facilitate deep message passing on both homophilic and heterophilic datasets, a claim supported by our experimental findings in Section 6.3.

Particle theory takes various forms, traditionally categorized into first-order and second-order systems. While first-order systems offer a direct approach to defining evolution, the second-order systems are inherently more stable, converging towards an asymptotically stable state. Both first-order and second-order systems exhibit desirable separability properties stemming from the balance between attractive and repulsive forces [27, 31]. Repulsive forces are crucial for separating distinct features, while attractive forces encourage particles of the same category to cluster, allowing for the refinement of essential features. Through this mechanism, we construct the interaction forces within HAMP.

Inspired by particle systems and informed by experimental observations, we identify a critical need for a delicate balance between attractive and repulsive forces. Excessive attraction often leads to feature over-smoothing [34, 35], while unchecked repulsion can cause feature explosion [44]. To address these issues and ensure system stability, we introduce the Allen-Cahn force [1] as a balancing mechanism. We theoretically demonstrate that the resulting system exhibits favorable separability and that its solutions converge to an equilibrium state, thereby mitigating both over-smoothing and feature explosion. Our focus on second-order systems is further motivated by their inherent theoretical stability, which ensures robustness even with deeper network layers. In practical tasks, ambiguous cases where distinct states are hard to differentiate often lead to uncertainty [42]. To address this, we incorporate a stochastic term into HAMP, driven by Brownian motion, which results in a stochastic differential equation. Our main contributions are summarized as follows:

- We propose HAMP, a novel hypergraph message passing framework based on particle system theory, filling in the theoretical gap in understanding hypergraph message passing from the perspective of particle system.
- We design two hypergraph message passing algorithms, HAMP-I and HAMP-II, constructed through first-order and second-order particle dynamical systems.
- Theoretically, we prove that HAMP maintains a strictly positive lower bound on the hypergraph Dirichlet energy, effectively resisting over-smoothing.
- Empirically, through numerical experiments, we should that HAMP achieves competitive results on node classification benchmarks. Notably, on heterophilic hypergraphs, HAMP consistently outperforms the current state-of-the-art baselines by a margin of 1–3%.

2 Preliminaries

Hypergraphs. A hypergraph is a generalization of a graph in which an edge can join any number of vertices. In contrast, in an ordinary graph, an edge connects exactly two vertices. We can denote a hypergraph as $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, with $|\mathcal{V}|$ nodes and $|\mathcal{E}|$ hyperedges, where $\mathcal{V} = \{v_1, \dots, v_{|\mathcal{V}|}\}$ is the set of nodes and $\mathcal{E} = \{e_1, \dots, e_{|\mathcal{E}|}\}$ is the set of hyperedges. $\mathcal{E}(i)$ denotes a set containing all the nodes

sharing at least one hyperedge with node i . The incidence matrix $\mathbf{H} \in \mathbb{R}^{|\mathcal{V}| \times |\mathcal{E}|}$ is a common notation in hypergraph. Its entries are defined as $\mathbf{H}_{i,e} = 1$ if node i belongs to hyperedge e , and 0 otherwise.

Message Passing in Hypergraphs. Neural message passing [23, 3] is the most widely used propagator for node feature updates in GNNs, which propagates node features while taking into account their neighboring nodes. Next, we generalize graph message passing to hypergraph, where multiple nodes interaction reflected in a hyperedge is considered,

$$\mathbf{x}_i^{(l+1)} = \Psi^{(l)} \left(\mathbf{x}_i^{(l)}, \Phi_2 \left(\{\mathbf{z}_e^{(l)}, \Phi_1(\{\mathbf{x}_j^{(l)}\}_{j \in e})\}_{e \in \mathcal{E}(i)} \right) \right), \quad (1)$$

where $\mathbf{z}_e^{(l)}$ denotes the feature embedding of hyperedge in the l^{th} layer. $\Phi_1^{(l)}$ and $\Phi_2^{(l)}$ denote differentiable permutation invariant functions (e.g., sum, mean, or max) for nodes and hyperedges, respectively. $\Psi^{(l)}$ denotes a differentiable function such as MLPs.

Neural ODEs. Neural ODEs [18, 10] are essentially ordinary differential equations (ODEs), where the time derivative of the hidden states are parameterized by a neural network f_θ ,

$$\frac{d\mathbf{x}(t)}{dt} = f_\theta(\mathbf{x}(t), t), \quad \mathbf{x}(0) = \mathbf{x}, \quad (2)$$

where t denotes time, $\mathbf{x}(t)$ are the system state, and θ is learnable parameter. It can be understood as the continuous limit form of the residual network, e.g, $\mathbf{x}(t + \tau) = f_\theta(\mathbf{x}(t), t) + \mathbf{x}(t)$, where τ denotes the step size of time. Thus, Neural ODEs model the continuous dynamics system that evolve hidden states over a continuous range of “depths”, analogous to layers in traditional deep networks.

3 Hypergraph Message Passing based on Particle Dynamics

Viewing HNNs as continuous systems allows for the application of physical models to elucidate and investigate their properties. A compelling conceptual analogy arises when comparing hypergraph message passing in deep learning with interacting particle systems studied in statistical physics. These disparate fields display remarkable structural and dynamical resemblances: the local potential field-governed interactions among particles in many-body systems are analogous to the aggregation operations in HNNs. Concurrently, the emergent collective behaviors observed in self-organizing dynamics demonstrate a functional equivalence to information propagation mechanisms in HNNs. As message passing delineates the transfer of information between entities, particle systems can represent this by having particles acting as entities that exchange information through the transition of specific attributes. Moreover, such particle systems can be endowed with hypergraph structures by interpreting hyperedges as distinct fields that exert influence over all nodes encompassed by them.

Fields and Hypergraphs. The hypergraph uses hyperedges which reflect the common relationship among multiple nodes. Such multi-nodal relationships are aptly described using “fields”. Each field imposes a common dynamic on the nodes under its influence, which is equivalent to a hyperedge defining a unified information propagation step for its constituent nodes. A node participating in multiple hyperedges thus receives information from diverse sources due to field superposition. Thus, hyperedges reflect both the interactions of particles within a specific region and the concurrent influence of the external environment, which together dictate the propagation dynamics. Based on these observations, we propose a hypergraph message passing framework for describing information dynamics. We formalize a composite field for each node i by $F_i = \sum_{e: i \in e} F_i^e + F_d$ to encapsulate the hypergraph structure. This field F_i aggregates influences akin to those in a particle system, where F_i^e represents the interaction energy for node i with the edge e , and F_d is the damping energy.

Interaction Force. Interaction force forms a basic association between hypergraph message passing and particle dynamics. We consider particles identified by their specific attributes $\mathbf{p}_i \in \mathbb{R}^d$ in the particle system. Then, we define the interaction energy as following

$$\sum_{e: i \in e} F_i^e := \frac{1}{2} \sum_{e: i \in e} \sum_{j \in e} (\mathbf{p}_i - \mathbf{p}_j)^\top f_\beta(\mathbf{p}_i, \mathbf{p}_j, \mathbf{e}) (\mathbf{p}_i - \mathbf{p}_j), \quad e \in \mathcal{E}. \quad (3)$$

If $f_\beta(\mathbf{p}_i, \mathbf{p}_j, e) > 0$, node i is attracted by node j . If $f_\beta(\mathbf{p}_i, \mathbf{p}_j, e) < 0$, node i is repulsed by node j . They correspond to two basic interaction forces: attraction and repulsion. Note that in the previous model similar to Eq. 3, the only interaction is attraction. It relies on an implicit belief that all the hyperedges only connect similar nodes. But this is not always true in the real world. Hence, introducing repulsion into hypergraph neural networks is significant.

Damping Force. The specific formulation of the damping term can vary. The inclusion of damping in neural networks serves two principal purposes. Firstly, it is crucial for ensuring system stability. If inter-particle repulsive forces do not attenuate adequately with increasing distance, particles risk unbounded separation or unbounded Dirichlet energy, which could lead to network collapse. A damping term mitigates this by exerting a dominant influence on the dynamics, especially near the system's boundaries. Secondly, the damping term has been demonstrated to promote various collective dynamics in numerous systems [30, 31]. For damping, we take double-well potential $F_d = \zeta(1 - \mathbf{p}^2)^2$ with the coefficient ζ . The corresponding damping force is given by the Allen-Cahn force $f_d(\mathbf{p}) = \nabla_{\mathbf{p}} F_d = \delta(1 - \mathbf{p}^2)\mathbf{p}$ [38, 19], where δ denotes the strength coefficient. This formulation exhibits favorable separation properties, as will be demonstrated in Section 5.

Hypergraph Dynamics. Interaction term can be formulated using two primary approaches: reductive and non-reductive expressions. Reductive methods, such as those based on convolution coefficients [22] or attention coefficients [2], often simplify hyperedges into clique-like structures, which can lead to the loss of crucial high-order information. In contrast, non-reductive expressions, like star expansion, aim to preserve this high-order information by explicitly connecting each node to the hyperedges it belongs to. We take the non-reductive expression and Allen-Cahn force to design interaction force f_β and damping force f_d . In fact, the particle attribute $\mathbf{p}$ can represent diverse properties, such as features or velocity. This flexibility in defining $\mathbf{p}$ naturally allows for the formulation of both first-order and second-order particle systems.

First-order ODEs. Inspired by the opinions dynamics [33], we have $\mathbf{p}_i \mapsto \mathbf{x}_i \in \mathbb{R}^d$ that interact with each other according to the first-order system. Taking gradient of composite field F , we have

$$\frac{d\mathbf{x}_i}{dt} = \sum_{e:i \in e} \sum_{j \in e} f_\beta(\mathbf{x}_i, \mathbf{x}_j, e)(\mathbf{x}_j - \mathbf{x}_i) + f_d(\mathbf{x}_i), \quad (4)$$

where f_β is a parameterized function. This form has connection with diffusion process of hypergraph. We defer the detailed analysis to Section 4.

Second-order ODEs. Inspired by the flocking dynamics [33, 15], this is second-order model where the attribute is the velocity, $\mathbf{p}_i \mapsto \mathbf{v}_i \in \mathbb{R}^d$, which is coupled to the feature $\mathbf{x}_i \in \mathbb{R}^d$. In this sense, information can be propagated by the evolution of feature and velocity of nodes,

$$\frac{d\mathbf{v}_i}{dt} = \sum_{e:i \in e} \sum_{j \in e} f_\beta(\mathbf{x}_i, \mathbf{x}_j, e)(\mathbf{v}_j - \mathbf{v}_i) + f_d(\mathbf{v}_i), \text{ where } \frac{d\mathbf{x}_i}{dt} = \mathbf{v}_i. \quad (5)$$

From ODEs to SDEs. To capture incompleteness and uncertainty within hypergraph data, stochasticity is introduced into the particle system. For instance, such uncertainty can arise when particles from different classes exhibit similar features. To model this inherent randomness, these stochastic processes are often described using stochastic differential equations (SDEs). We assume these SDEs are driven by Brownian motion B_t :

$$d\mathbf{p}(t + \tau) = \nabla F(\mathbf{p}(t))dt + \epsilon dB_t, \text{ where } \mathbb{E}[dB_t] = 0 \text{ and } \text{Cov}(dB_t) = \mathbf{I}_d, \quad (6)$$

where $\mathbf{I}_d$ denotes the $d \times d$ identity matrix. In essence, SDEs describe the changes in system state under infinitesimal time variations. Here, particle states undergo continuous evolution through both deterministic drift term and stochastic diffusion component. It can describe hypergraph evolution of neural message passing in more realistic scenario. According to [15, 24], this model can achieve self-organization in a finite time, which enables the system to adjust its own state in the shortest time.

Hypergraph Message Passing. We introduce a unified hypergraph message passing framework, called *Hypergraph Atomic Message Passing*, or HAMP, given in Eq. 6. By instantiating this framework with either a first-order or a second-order system formulation, we obtain two variants: HAMP-I and HAMP-II, respectively. Detailed implementations of both approaches are provided systematically in Appendix A to ensure clarity and maintain manuscript focus.

4 Scale Translation of Hypergraph: Diffusion and Particle Dynamics

We examine hypergraph message passing through two complementary lenses—microscopic and macroscopic. From the microscopic viewpoint, HAMP is cast as a particle-dynamics system, as previously discussed. From the macroscopic standpoint, message propagation is seen as a diffusion process, which reveals that the particle-based formulation naturally subsumes diffusion-based models and overcomes their intrinsic limitations.

Hypergraph Diffusion. Consider node feature space $\Omega = \mathbb{R}^d$ and tangent vector field space $T\Omega = \mathbb{R}^d$. For $\mathbf{x}, \mathbf{y} \in \Omega$, $\mathbf{r}, \mathbf{v} \in T\Omega$, and $\mathbf{r}_{i,j} = -\mathbf{r}_{j,i}$, we adopt the following inner products:

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i,j} \mathbf{x}_i \mathbf{y}_j, \quad [\mathbf{r}, \mathbf{v}] = \sum_{i>j} \sum_{e \in \mathcal{E}} h_{i,j}^e \mathbf{r}_{i,j} \mathbf{v}_{i,j}. \quad (7)$$

Here $h_{i,j}^e$ is a tuple related to node i, j and hyperedge e and $h_{i,j}^e = 0$ if $\mathbf{H}_{i,e} \mathbf{H}_{j,e} = 0$. We set $h_{i,j}^e$ to satisfy $\sum_j \sum_{e \in \mathcal{E}} h_{i,j}^e = 1$. For any $\mathbf{u} \in T\Omega$, by the adjoint relation $[\mathbf{u}, \nabla \mathbf{x}] = \langle \mathbf{x}, \text{div } \mathbf{u} \rangle$, where $\nabla \mathbf{x} = \mathbf{x}_j - \mathbf{x}_i$, we can derive $(\text{div } \mathbf{u})_j = \sum_i \sum_{e \in \mathcal{E}} h_{i,j}^e u_i$. Then, the hypergraph diffusion process is

$$\frac{d\mathbf{x}_i}{dt} = \text{div } \nabla \mathbf{x}_i = \sum_j \sum_{e \in \mathcal{E}} h_{i,j}^e (\mathbf{x}_j - \mathbf{x}_i). \quad (8)$$

For simplicity, we rewrite Eq. 8 in matrix form $\frac{d\mathbf{x}}{dt} = -\mathcal{L}\mathbf{x}$, where $\mathcal{L} = \mathbf{I} - (\sum_{e \in \mathcal{E}} h_{i,j}^e)$ is a hypergraph operator. If $\mathcal{L}$ is positive semi-definite, we interpret Eq. 8 as a diffusion-type process of hypergraph. Different parameterizations of $h_{i,j}^e$ then yield distinct diffusion equations. For example, applying a forward Euler discretization to Eq. 8 and setting $(\sum_{e \in \mathcal{E}} h_{i,j}^e) = \mathbf{D}_v^{-\frac{1}{2}} \mathbf{H} \mathbf{W} \mathbf{D}_e^{-1} \mathbf{H}^\top \mathbf{D}_v^{-\frac{1}{2}}$ recovers the simplified HGNN [20] without channel mixing.

Connecting Particle Dynamics. Since Eq. 8 coincides with the self-organized dynamics in particle system [33], we reinterpret Eq. 8 not as a standard hypergraph diffusion process, but rather as particle dynamics where $h_{i,j}^e$ denotes the interaction force between nodes i and j under field e . In fact, Eq. 8 is a special case of Eq. 4 where only attractive forces are considered in the message propagation. As we show in Section 5, this simplification is atypical in particle systems and leads to over-smoothing.

5 Theory of Anti-over-smoothing

For diffusion-type hypergraph networks, we define the *hypergraph Dirichlet energy* of $\mathbf{x} \in \mathbb{R}^{N \times d}$ as $\mathbf{E}(\mathbf{x}) := \sum_{i,j=1}^N \sum_{e \in \mathcal{E}} \mathbf{H}_{i,e} \mathbf{H}_{j,e} \|\mathbf{x}_i - \mathbf{x}_j\|^2 = \text{tr}(\mathbf{x}^\top \mathcal{L} \mathbf{x})$. Furthermore, we will define over-smoothing in hypergraph message passing.

Definition 5.1. Let $\mathbf{x}^{(l)}$ denote the hidden features of the l^{th} layer. We define over-smoothing in HNNs as the exponential convergence to zero of the layer-wise Dirichlet energy as a function of l , i.e., $\mathbf{E}(\mathbf{x}^{(l)}) \leq C_1 e^{-C_2 l}$, with positive constants C_1 and C_2 .

Why Do HGNN Cause Over-smoothing? The normalized hypergraph Laplacian matrix is defined by $\mathcal{L} = \mathbf{I} - \mathbf{P}$, where $\mathbf{P}$ is the propagation matrix derived from the incidence matrix $\mathbf{H}$. In an HGNN, the feature at layer l without activation σ evolves according to $\mathbf{x}^{(l)} = \mathbf{P}^{l-1} \mathbf{x}^{(0)} \Theta^{(1)} \dots \Theta^{(l-1)}$. However, this purely diffusive propagation inevitably induces over-smoothing: one can show $\mathbf{E}(\mathbf{x}^{(l)}) \leq C e^{-\gamma^2 l}$ [9], where γ is the smallest non-zero positive eigenvalue of $\mathcal{L}$. This over-smoothing behavior stems from the intrinsic diffusion dynamics. Because $\mathcal{L}$ is symmetric positive semi-definite, repeated application of $\mathbf{P}$ causes node representations to decay exponentially towards a limiting state—which is zero, thereby eroding discriminative power.

Why Do HAMP Avoid Over-smoothing? We now derive theoretical guarantees for the collective behavior of models Eq. 4 and Eq. 5. They show that the addition of a repulsive force enforces

a positive lower bound on the Dirichlet energy. Technically, we suppose there exists $\{f_\beta^e\}$ such that $\mathcal{I} = \{1, \dots, N\}$ can be divided into two disjoint groups with N_1, N_2 particles respectively: $f_\beta(h_{i,j}^e) \geq 0$, for $\{i, j\} \in \mathcal{I}_1$ or $\mathcal{I}_2$ and $f_\beta(h_{i,j}^e) \leq 0$, otherwise. We designate $\{x_i^{(1)}\} := \{\mathbf{x}_i | i \in \mathcal{I}_1\}$ and $\{x_j^{(2)}\} := \{\mathbf{x}_j | j \in \mathcal{I}_2\}$. Finally, we impose the symmetry $f_\beta(h_{i,j}^e) = f_\beta(h_{j,i}^e)$, which reflects equal-and-opposite interactions under the same field. Under these conditions, one can show that the Dirichlet energy of our system admits a strictly positive lower bound, as follows. We leave the detailed proofs in Appendix B.

Proposition 5.2 (L_2 separation of HAMP-I). *For Eq. 4, suppose the above assumptions are satisfied. Define the mean value $\bar{\mathbf{x}} := \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i$, and the second moments $M_2(\mathbf{x}) := \sum_{i=1}^N \mathbf{x}_i^2$. Then for sufficiently large N_1, N_2 , there exist constants λ_-, λ_+ , such that if the initial data satisfies $\lambda(0) := \frac{\widehat{M}_2(0)}{\|\bar{\mathbf{x}}^{(1)}(0) - \bar{\mathbf{x}}^{(2)}(0)\|^2} \leq \lambda_+$, then, there holds that the L_2 separation*

$$\lambda(t) := \frac{\widehat{M}_2(t)}{\|\bar{\mathbf{x}}^{(1)}(t) - \bar{\mathbf{x}}^{(2)}(t)\|^2} \leq \lambda_- + (\lambda(0) - \lambda_-)e^{-\mu t}, \quad (9)$$

with a positive constant μ , where $\widehat{M}_2(t) := M_2(\mathbf{x}^{(1)}(t)) + M_2(\mathbf{x}^{(2)}(t))$.

Proposition 5.3 (L_2 separation of HAMP-II, [19]). *For Eq. 5, we set $0 < S \leq f_\beta(h_{i,j}^e)$ for $\{i, j\} \in \mathcal{I}_1$ and $0 \leq f_\beta(h_{i,j}^e) \leq D$ otherwise, with $k := \max_i \{|\mathcal{E}(i)|\}$. If the initial $\|\bar{\mathbf{x}}^{(1)}(0) - \bar{\mathbf{x}}^{(2)}(0)\| \gg 1$, and if there exists a positive constant η such that*

$$\alpha(S - D)k \min\{N_1, N_2\} \geq \delta + \eta, \quad (10)$$

Then the system has a bi-cluster flocking.

Proposition 5.4 (Lower bound of the Dirichlet energy). *If the hypergraph $\mathcal{H}$ is a connected one, for Eq. 4 with the conditions of Theorem 5.2, or for Eq. 5 with conditions of Theorem 5.1 in [19], there exists a positive lower bound of the Dirichlet energy.*

6 Experiments

6.1 Experiment Setup

We conduct comprehensive experiments to evaluate the proposed models on node classification task. For more experimental details such as datasets and hyperparameters, please refer to Appendix C.

Datasets. Following ED-HNN[43], the real-world hypergraph benchmarking datasets span diverse domains, scales, and heterophilic levels. They can be divided into two groups based on homophily. The homophilic hypergraphs include academic citation networks (Cora, Citeseer, and Pubmed) and co-authorship networks (Cora-CA and DBLP-CA). The heterophilic hypergraphs cover legislative voting records (Congress, House, and Senate) and retail relationships (Walmart).

Baselines. The selected baselines cover two types of hypergraph learning frameworks, comprising both reductive and non-reductive approaches. The reductive methods include HGNN [20], HCHA [2], HNNH [14], HyperGCN [47], and HyperND [37]. The non-reductive methods include UniGCNII [26], AllDeepSets [11], AllSetTransformer [11], ED-HNN [43], and HDS^{ode} [48].

6.2 Node Classifications on Hypergraphs

In this section, we evaluate HAMP-I and HAMP-II on nine real-world hypergraph benchmarks for the node classification task. Tab. 1 reports the accuracy on both homophilic and heterophilic datasets. As HDS^{ode} does not provide results on these benchmarks, we reproduce its performance using the official open-source code and perform hyperparameter tuning following the original paper. For other baselines, our results are consistent with those reported by ED-HNN. Overall, our models demonstrate competitive performance across all nine datasets. Notably, the improvement is more pronounced on heterophilic datasets, with the largest accuracy gain of 3% observed on Walmart dataset.

Table 1: Node Classification on standard hypergraph benchmarks. The accuracy (%) is reported with a standard deviation from 10 repetitive runs. (Key: **Best**; Second Best; Third Best.)

Homophilic	Cora	Citeseer	Pubmed	Cora-CA	DBLP-CA
HGNN	79.39±1.36	72.45±1.16	86.44±0.44	82.64±1.65	91.03±0.20
HCHA	79.14±1.02	72.42±1.42	86.41±0.36	82.55±0.97	90.92±0.22
HNHN	76.36±1.92	72.64±1.57	86.90±0.30	77.19±1.49	86.78±0.29
HyperGCN	78.45±1.26	71.28±0.82	82.84±8.67	79.48±2.08	89.38±0.25
UniGCNII	78.81±1.05	73.05±2.21	88.25±0.40	83.60±1.14	91.69±0.19
HyperND	79.20±1.14	72.62±1.49	86.68±0.43	80.62±1.32	90.35±0.26
AllDeepSets	76.88±1.80	70.83±1.63	88.75±0.33	81.97±1.50	91.27±0.27
AllSetTransformer	78.58±1.47	73.08±1.20	88.72±0.37	83.63±1.47	91.53±0.23
ED-HNN	80.31±1.35	73.70±1.38	<u>89.03±0.53</u>	83.97±1.55	91.90±0.19
HDS ^{ode}	<u>80.65±1.22</u>	<u>74.87±1.12</u>	88.81±0.43	<u>84.95±0.98</u>	91.49±0.25
HAMP-I	81.18±1.30	<u>75.22±1.62</u>	89.02±0.38	85.23±1.15	91.66±0.17
HAMP-II	80.80±1.62	75.33±1.61	89.05±0.41	84.89±1.53	<u>91.67±0.23</u>

Heterophilic	Congress	Senate	Walmart	House
HGNN	91.26±1.15	48.59±4.52	62.00±0.24	61.39±2.96
HCHA	90.43±1.20	48.62±4.41	62.35±0.26	61.36±2.53
HNHN	53.35±1.45	50.93±6.33	47.18±0.35	67.80±2.59
HyperGCN	55.12±1.96	42.45±3.67	44.74±2.81	48.32±2.93
UniGCNII	94.81±0.81	49.30±4.25	54.45±0.37	67.25±2.57
HyperND	74.63±3.62	52.82±3.20	38.10±3.86	51.70±3.37
AllDeepSets	91.80±1.53	48.17±5.67	64.55±0.33	67.82±2.40
AllSetTransformer	92.16±1.05	51.83±5.22	65.46±0.25	69.33±2.20
ED-HNN	<u>95.00±0.99</u>	64.79±5.14	<u>66.91±0.41</u>	<u>72.45±2.28</u>
HDS ^{ode}	90.91±1.52	<u>66.90±5.52</u>	63.38±0.48	71.30±1.90
HAMP-I	<u>95.09±0.79</u>	<u>69.44±6.09</u>	69.90±0.38	72.72±1.77
HAMP-II	95.26±1.34	70.14±6.08	69.94±0.37	<u>72.60±1.23</u>

6.3 Ablation Studies

In this section, we conduct several ablation experiments on real-world datasets to assess our model design, and provide empirical validation for our theoretical findings. For more ablation experiments, see Appendix C.2.

Impact of the Number of Layers on HAMP-I and HAMP-II. To assess the effectiveness of different methods in deep HNNs, we compare three representative baselines with our proposed HAMP models. Unlike Pubmed, DBLP-CA, Senate, and House, the Cora, Citeseer, and Congress datasets generally perform better in shallow networks, but worse in deep networks. Therefore, we focus on these datasets to demonstrate HAMP’s advantages in deep architectures. As shown in Fig. 2, HAMP-II consistently outperforms other methods as depth increases, while competitors suffer from accuracy drops. This ability highlights the potential of HAMP to capture complex representations and maintain stability, making it a valuable framework for deep HNNs.

Impact of the Repulsion and Allen-Cahn Forces on HAMP-I and HAMP-II. In addition to the theoretical analysis, we conducted ablation studies to investigate the individual and combined effects of the repulsion force f_{β}^{-} and the Allen-Cahn force f_d on both HAMP-I and HAMP-II. The results given in Tab. 2 show that incorporating the repulsive force significantly improves classification performance. In HAMP-I, enabling repulsion alone yields notable gains over the baseline, while the Allen-Cahn term alone offers moderate improvements. Notably, combining both terms consistently achieves the best accuracy across all datasets. The synergy between the repulsion and Allen-Cahn terms further boosts performance, confirming that these particle system-inspired mechanisms play complementary roles: repulsion term prevents feature over-smoothing by separating node embeddings, whereas Allen-Cahn term balances attraction and repulsion to promote class-dependent equilibrium.

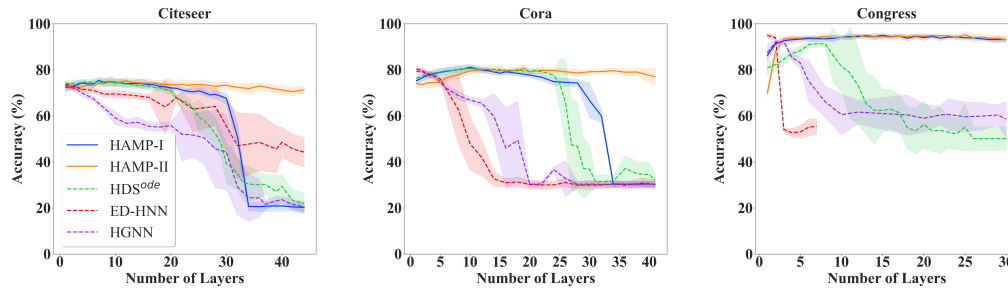


Figure 2: An empirical analysis of the depth-accuracy correlation in deep neural networks. The shaded area represents the standard deviation, helping to show the range of accuracy fluctuations.

Table 2: Node Classification on some standard hypergraph benchmarks. The accuracy (%) is reported from 10 repetitive runs. (Key: f_{β}^{-} : repulsion; f_d : Allen-Cahn; **Best**.)

	f_{β}^{-}	f_d	Cora	Citeseer	Pubmed	Congress	Senate	Walmart	House
HAMP-I	✗	✗	75.67	70.59	87.93	93.47	60.14	69.86	69.88
	✓	✗	75.97	70.60	88.23	93.65	61.69	69.73	69.57
	✗	✓	80.59	74.67	88.77	94.67	65.63	69.73	71.55
	✓	✓	81.18	75.22	89.02	95.09	69.44	69.90	72.72
HAMP-II	✗	✗	77.18	71.75	88.68	94.35	60.14	69.84	70.46
	✓	✗	77.40	71.69	88.77	94.63	59.58	69.80	69.63
	✗	✓	79.50	74.25	88.80	94.12	64.51	69.86	70.96
	✓	✓	80.80	75.33	89.05	95.26	70.14	69.94	72.60

Impact of Noise on HAMP-I and HAMP-II. Fig. 3 illustrates the effect of adding stochastic component into deterministic message passing on Senate dataset. Notably, incorporating noise consistently improves accuracy for both HAMP-I and HAMP-II. In addition, the stability of HAMP-II is better than that of HAMP-I, whether or not noise is injected, indicating that the second-order particle system is more stable. Overall, these results demonstrate that incorporating stochastic component into message passing effectively improves model performance and stability by explicitly capturing data uncertainty.

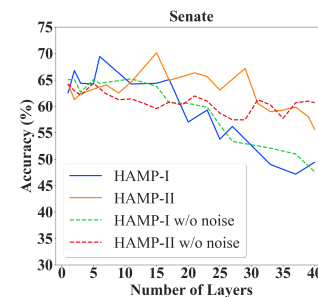


Figure 3: Significance plot for noise on Senate dataset.

6.4 Vertex Representation Visualization

To more intuitively validate the progressive refinement of vertex representations in our HAMP methods, we use t-SNE [39] to visualize the vertex evolution process of HAMP-I and HAMP-II on Congress dataset at different epochs. As shown in Fig. 4, we visualize the vertices based on the representations obtained at epoch $1, \frac{1}{8}E, \frac{1}{4}E, \frac{1}{2}E$, and E , where E is the total number of epochs. From Fig. 4, we have the following three observations:

- When epoch = 1 (subgraphs (a) and (f)), the node feature representations exhibit a chaotic distribution, and it is difficult to distinguish the number of categories. As training progresses, the clustering entropy shows a monotonically decreasing trend.
- Subgraphs (e) and (j) show the visualization results at convergence for HAMP-I and HAMP-II, respectively. The final category boundaries are clearer in HAMP-II than in HAMP-I, reflecting the geometry refinement enabled by HAMP-II's deeper message passing.
- Comparing subgraphs (b) and (g), HAMP-I is still in the early stages of categorization, while HAMP-II shows a significantly improved clustering effect. Notably, HAMP-II has successfully distinguished six distinct categories, confirming that HAMP-II achieves faster cluster than HAMP-I due to the second-order mechanism.

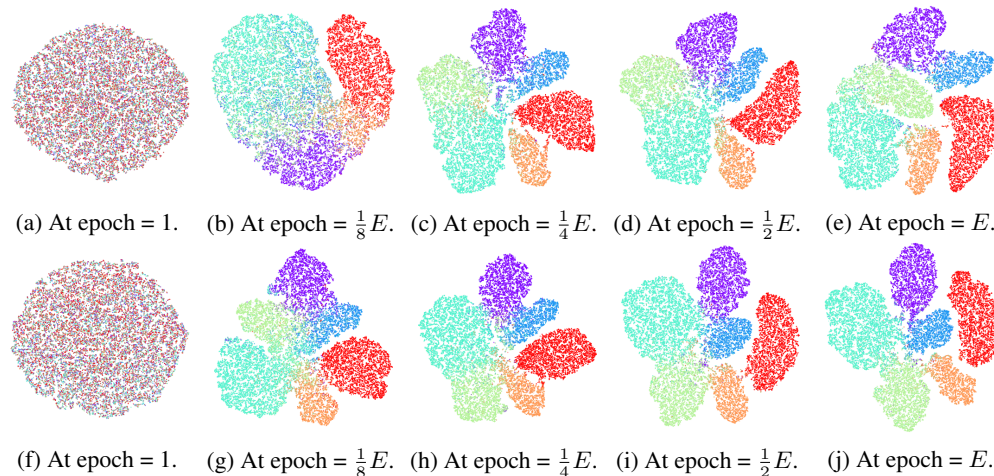


Figure 4: The t-SNE visualization of vertex representation evolution of HAMP-I (the first row) and HAMP-II (the second row) on Congress dataset. The colors represent the class labels.

7 Related Work

Hypergraph Neural Networks. Hypergraph learning was first introduced in [51] as a propagation process on hypergraph. Since then, hypergraph learning [21, 28] has developed extensively. As an extension of GNNs, Feng et al. [20] proposed HGNN that effectively captures high-order interactions by leveraging the vertex-edge-vertex propagation pattern. Further, HGNN^+ [22] introduced hyperedge groups and adaptive hyperedge group fusion strategy as a general framework for modeling high-order multi-modal/multi-type data correlations. Following the spectral theory of hypergraph [47], SheafHyperGNN [16] introduced the sheaf theory to model data relationships in hypergraph more finely. Inspired by Transformers [40], several HNNs [2, 11, 12] enhanced the feature extraction capability for hypergraph through attention mechanism and centrality for positional encoding. Different from using vertex-vertex propagation pattern, several works [11, 26, 43, 48] have considered employing multi-phase message passing. Among these, UniGNN [26] presented a unified framework that facilitates the processing of all hypergraph data through GNNs. In contrast, ED-HNN [43] and CoNHD [49] were developed from the perspective of optimization of hyperedge and node potential. Additionally, HDS^{ode} [48] adopted control-diffusion ODEs to model the hypergraph dynamic system. By contrast, our method is based on the particle system theory, employing the first-order and second-order systems to understand and design HNNs.

Collective Dynamics. Watts and Strogatz [45] first mathematically defined a small-world network and explained the reasons behind the collective dynamics. Many researchers are interested in how active/passive media made of many interacting agents form complex patterns in mathematical biology and technology. Battiston et al. [4] showed that higher-order interactions play a crucial role in understanding these complex patterns. These complex patterns can be seen in animal groups, cell clusters, granular media, and self-organizing particles, as shown in [7, 25, 31] and other references. In many of these models, the agents move into groups based on the attractive-repulsive forces [8, 17, 6]. For example, Fang et al. [19] have studied the Cucker-Smale model [13] with Rayleigh friction and attractive-repulsive coupling, and [27] showed a similar collective phenomenon with stochastic dynamics. Furthermore, the ACMP [44] was based on Allen-Cahn particle system and incorporated the repulsive force, providing inspiration for our work in hypergraph learning.

8 Conclusion

In this paper, we introduce a novel hypergraph message passing framework inspired by particle system theory. We derive both first-order and second-order system equations, yielding two distinct models for modeling hypergraph message passing dynamics that capture full hyperedge interactions while mitigating over-smoothing and heterophily. The proposed models further integrate a stochastic term to model uncertainty in these interactions and can alleviate over-smoothing in deep layers. By casting

HNNS in a physically interpretable paradigm, our model balances high-order interaction modeling with feature-diversity preservation, offering both theoretical insights and practical advances for complex system analysis. In future work, we plan to extend the proposed methods to protein-structure and sequence design for the discovery of novel antibodies and enzymes.

Acknowledgments and Disclosure of Funding

This work is supported by the National Science Foundation for International Senior Scientists Grant (No. 12350710181), and the Shanghai Municipal Science and Technology Key Project (No. 22JC1402300). We thank the anonymous reviewers for their insightful comments.

References

- [1] Samuel M Allen and John W Cahn. A microscopic theory for antiphase boundary motion and its application to antiphase domain coarsening. *Acta metallurgica*, 27(6):1085–1095, 1979.
- [2] Song Bai, Feihu Zhang, and Philip HS Torr. Hypergraph convolution and hypergraph attention. *Pattern Recognition*, 110:107637, 2021.
- [3] Peter W. Battaglia, Jessica B. Hamrick, Victor Bapst, Alvaro Sanchez-Gonzalez, Vinícius Flores Zambaldi, Mateusz Malinowski, Andrea Tacchetti, David Raposo, Adam Santoro, Ryan Faulkner, Çağlar Gülçehre, H. Francis Song, Andrew J. Ballard, Justin Gilmer, George E. Dahl, Ashish Vaswani, Kelsey R. Allen, Charles Nash, Victoria Langston, Chris Dyer, Nicolas Heess, Daan Wierstra, Pushmeet Kohli, Matthew Botvinick, Oriol Vinyals, Yujia Li, and Razvan Pascanu. Relational inductive biases, deep learning, and graph networks. *CoRR*, abs/1806.01261, 2018.
- [4] Federico Battiston, Enrico Amico, Alain Barrat, Ginestra Bianconi, Guilherme Ferraz de Arruda, Benedetta Franceschiello, Iacopo Iacopini, Sonia Kéfi, Vito Latora, Yamir Moreno, Micah M. Murray, Tiago P. Peixoto, Francesco Vaccarino, and Giovanni Petri. The physics of higher-order interactions in complex systems. *Nature Physics*, 17(10):1093–1098, October 2021. ISSN 1745-2481.
- [5] Adrián Bazaga, Pietro Lio, and Gos Micklem. HyperBERT: Mixing hypergraph-aware layers with language models for node classification on text-attributed hypergraphs. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, Miami, Florida, USA, November 2024. Association for Computational Linguistics.
- [6] José A Carrillo and Ruiwen Shu. From radial symmetry to fractal behavior of aggregation equilibria for repulsive–attractive potentials. *Calculus of Variations and Partial Differential Equations*, 62(1):28, 2023.
- [7] José A Carrillo, Massimo Fornasier, Giuseppe Toscani, and Francesco Vecil. Particle, kinetic, and hydrodynamic models of swarming. *Mathematical modeling of collective behavior in socio-economic and life sciences*, pages 297–336, 2010.
- [8] José A Carrillo, Axel Klar, Stephan Martin, and Sudarshan Tiwari. Self-propelled interacting particle systems with roosting force. *Mathematical Models and Methods in Applied Sciences*, 20(supp01):1533–1552, 2010.
- [9] Benjamin Paul Chamberlain, James Rowbottom, Maria I. Gorinova, Stefan D Webb, Emanuele Rossi, and Michael M. Bronstein. GRAND: Graph neural diffusion. In *ICML*, 2021.
- [10] Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. *NeurIPS*, 31, 2018.
- [11] Eli Chien, Chao Pan, Jianhao Peng, and Olgica Milenkovic. You are allset: A multiset function framework for hypergraph neural networks. In *ICLR*, 2022.
- [12] Minyoung Choe, Sunwoo Kim, Jaemin Yoo, and Kijung Shin. Classification of edge-dependent labels of nodes in hypergraphs. In *KDD*, page 298–309, New York, NY, USA, 2023. Association for Computing Machinery.

- [13] Felipe Cucker and Steve Smale. Emergent behavior in flocks. *IEEE Transactions on Automatic Control*, 52(5):852–862, 2007.
- [14] Yihe Dong, Will Sawin, and Yoshua Bengio. Hnhn: Hypergraph networks with hyperedge neurons. In *ICML*, 2020.
- [15] Linglong Du and Xinyun Zhou. The stochastic delayed cucker-smale system in a harmonic potential field. *Kinetic and Related Models*, 16(1):54–68, 2023. ISSN 1937-5093. doi: 10.3934/krm.2022022.
- [16] Iulia Duta, Giulia Cassarà, Fabrizio Silvestri, and Pietro Lió. Sheaf hypergraph networks. In *NeurIPS*, volume 36, pages 12087–12099. Curran Associates, Inc., 2023.
- [17] Maria R D’Orsogna, Yao-Li Chuang, Andrea L Bertozzi, and Lincoln S Chayes. Self-propelled particles with soft-core interactions: patterns, stability, and collapse. *Physical review letters*, 96(10):104302, 2006.
- [18] Weinan E. A proposal on machine learning via dynamical systems. *Communications in Mathematics and Statistics*, 1(5):1–11, 2017.
- [19] Di Fang, Seung-Yeal Ha, and Shi Jin. Emergent behaviors of the Cucker-Smale ensemble under attractive-repulsive couplings and rayleigh frictions. *Mathematical Models and Methods in Applied Sciences*, 29(07):1349–1385, 2019.
- [20] Yifan Feng, Haoxuan You, Zizhao Zhang, Rongrong Ji, and Yue Gao. Hypergraph neural networks. In *AAAI*, volume 33, pages 3558–3565, 2019.
- [21] Yue Gao, Zizhao Zhang, Haojie Lin, Xibin Zhao, Shaoyi Du, and Changqing Zou. Hypergraph Learning: Methods and Practices. *TPAMI*, 44(5), May 2022. ISSN 1939-3539.
- [22] Yue Gao, Yifan Feng, Shuyi Ji, and Rongrong Ji. Hgmn+: General hypergraph neural networks. *TPAMI*, 45(3):3181–3199, 2023. doi: 10.1109/TPAMI.2022.3182052.
- [23] Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural message passing for quantum chemistry. In *ICML*, 2017.
- [24] Seung-Yeal Ha, Kiseop Lee, and Doron Levy. Emergence of time-asymptotic flocking in a stochastic cucker-smale system. *Communications in Mathematical Sciences*, 7(2), 2009.
- [25] Darryl D Holm and Vakhtang Putkaradze. Formation of clumps and patches in self-aggregation of finite-size particles. *Physica D: Nonlinear Phenomena*, 220(2):183–196, 2006.
- [26] Jing Huang and Jie Yang. Unignn: a unified framework for graph and hypergraph neural networks. In *IJCAI*, 2021.
- [27] Shi Jin and Ruiwen Shu. Collective dynamics of opposing groups with stochastic communication. *Vietnam Journal of Mathematics*, 49:619–636, 2021.
- [28] Sunwoo Kim, Soo Yong Lee, Yue Gao, Alessia Antelmi, Mirko Polato, and Kijung Shin. A survey on hypergraph neural networks: An in-depth and step-by-step guide. In *KDD*, page 6534–6544, New York, NY, USA, 2024. Association for Computing Machinery.
- [29] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *ICLR*, 2017.
- [30] Theodore Kolokolnikov, Hui Sun, David Uminsky, and Andrea L Bertozzi. Stability of ring patterns arising from two-dimensional particle interactions. *Physical Review E*, 84(1):015203, 2011.
- [31] Theodore Kolokolnikov, José A Carrillo, Andrea Bertozzi, Razvan Fetecau, and Mark Lewis. Emergent behaviour in multi-particle systems with non-local interactions, 2013.
- [32] Ming Li, Yongchun Gu, Yi Wang, Yujie Fang, Lu Bai, Xiaosheng Zhuang, and Pietro Lio. When hypergraph meets heterophily: New benchmark datasets and baseline. In *AAAI*, pages 18377–18384, 2025.

- [33] Sebastien Motsch and Eitan Tadmor. Heterophilious dynamics enhances consensus. *SIAM Review*, 56(4):577–621, 2014.
- [34] Hoang Nt and Takanori Maehara. Revisiting graph neural networks: All we have is low-pass filters. *arXiv preprint arXiv:1905.09550*, 2019.
- [35] Kenta Oono and Taiji Suzuki. Graph neural networks exponentially lose expressive power for node classification. In *ICLR*, 2019.
- [36] Hongbin Pei, Bingzhe Wei, Kevin Chen-Chuan Chang, Yu Lei, and Bo Yang. Geom-gcn: Geometric graph convolutional networks. In *ICLR*, 2020.
- [37] Konstantin Prokopychik, Austin R Benson, and Francesco Tudisco. Nonlinear feature diffusion on hypergraphs. In *ICML*, pages 17945–17958. PMLR, 2022.
- [38] Loan Rayleigh. *JWS, The Theory of Sound, vol. 1*. Macmillan, London (reprinted Dover, New York, 1945), 1894.
- [39] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(86):2579–2605, 2008.
- [40] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *NeurIPS*, pages 6000–6010, 2017.
- [41] Ramon Viñas, Chaitanya K Joshi, Dobrik Georgiev, Phillip Lin, Bianca Dumitrascu, Eric R Gamazon, and Pietro Liò. Hypergraph factorization for multi-tissue gene expression imputation. *Nature Machine Intelligence*, pages 1–15, 2023.
- [42] Fangxin Wang, Yuqing Liu, Kay Liu, Yibo Wang, Sourav Medya, and Philip S. Yu. Uncertainty in graph neural networks: A survey. *Transactions on Machine Learning Research*, 2024.
- [43] Peihao Wang, Shenghao Yang, Yunyu Liu, Zhangyang Wang, and Pan Li. Equivariant hypergraph diffusion neural operators. In *ICLR*, 2023.
- [44] Yuelin Wang, Kai Yi, Xinliang Liu, Yu Guang Wang, and Shi Jin. ACMP: Allen-cahn message passing with attractive and repulsive forces for graph neural networks. In *ICLR*, 2023.
- [45] Duncan J Watts and Steven H Strogatz. Collective dynamics of ‘small-world’ networks. *Nature*, 393(6684):440–442, 1998.
- [46] Hongfei Wu, Lijun Wu, Guoqing Liu, Zhirong Liu, Bin Shao, and Zun Wang. Se3set: Harnessing equivariant hypergraph neural networks for molecular representation learning. *Transactions on Machine Learning Research*, 2025.
- [47] 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 *NeurIPS*, 2019.
- [48] Jielong Yan, Yifan Feng, Shihui Ying, and Yue Gao. Hypergraph dynamic system. In *ICLR*, 2024.
- [49] Yijia Zheng and Marcel Worring. Co-representation neural hypergraph diffusion for edge-dependent node classification. *arXiv preprint arXiv:2405.14286*, 2024.
- [50] Bingxin Zhou, Outongyi Lv, Jing Wang Wang, Xiang Xiao, and Weishu Zhao. Odnet: Opinion dynamics-inspired neural message passing for graphs and hypergraphs. *Transactions on Machine Learning Research*, 2024.
- [51] Dengyong Zhou, Jiayuan Huang, and Bernhard Schölkopf. Learning with hypergraphs: Clustering, classification, and embedding. In *NeurIPS*, volume 19, 2006.

A Algorithms

Complexity Analysis. Here, we analyze the computational complexity of one layer in HAMP-I and HAMP-II. Analytically, the time complexity is $\mathcal{O}(|\mathcal{V}||\mathcal{E}|c^2 + |\mathcal{V}|c)$, where $|\mathcal{V}|$, $|\mathcal{E}|$ and c are the number of nodes, number of hyperedges and number of hidden dimension, respectively. However, the incidence matrix $\mathbf{H}$ is a sparse matrix, so the time complexity is $\mathcal{O}((tr(\mathbf{D}_v) + tr(\mathbf{D}_e))c^2 + |\mathcal{V}|c)$, where $tr(\mathbf{D}_v)$ is the sum of the degrees of all nodes and $tr(\mathbf{D}_e)$ is the sum of the number of nodes contained in all hyperedges. The detailed process of HAMP-I and HAMP-II are shown in Algorithm 1 and Algorithm 2.

Algorithm 1 The HAMP-I Algorithm for Hypergraph Node Classification.

```

1: Input: the incidence matrix  $\mathbf{H}$ , the node feature  $\mathbf{X}$ , and the node labels  $\mathbf{Y}$ .
2: Output: the model prediction accuracy.
3: Initialization: the time  $T$ , the step size  $\tau$  and all parameters of model.
4: while not converged do
5:   Node feature mapping  $\mathbf{X} = \text{Linear}_{\text{map}}(\mathbf{X})$ ;
6:   Set the initial time  $t = 0$ , the initial node representation  $\mathbf{X}(0) = \mathbf{X}$ ;
7:   while  $t \leq T$  do
8:     Message passing from  $\mathcal{V}$  to  $\mathcal{E}$ :  $\mathbf{X}_{\mathcal{V} \rightarrow \mathcal{E}}(t) = \Phi_1(\mathbf{X}(t), \mathbf{H})$ ;
9:     Message passing from  $\mathcal{E}$  to  $\mathcal{V}$ :  $\mathbf{X}_{\mathcal{E} \rightarrow \mathcal{V}}(t) = \Psi(\mathbf{X}(t), \Phi_2(\mathbf{X}_{\mathcal{V} \rightarrow \mathcal{E}}(t), \mathbf{H}))$ ;
10:    Compute particle dynamics as

$$\mathbf{X}(t + \tau) = \mathbf{X}(t) + \tau \sigma \left( \underbrace{\mathbf{X}_{\mathcal{E} \rightarrow \mathcal{V}}(t) - \omega \mathbf{X}(t)}_{\text{Interaction Force}} + \underbrace{\delta f_d(\mathbf{X}(t))}_{\text{Allen-Cahn Force}} + \underbrace{\epsilon \mathbf{B}(t)}_{\text{Noise}} + \beta \mathbf{X}(0) \right);$$

11:    Update  $t = t + \tau$ ;
12:  end while
13:  Input the node representation into the classifier  $\mathbf{X}^{out} = \text{MLP}(\mathbf{X}(T))$ ;
14:  Compute the model prediction labels  $\hat{\mathbf{Y}} = \text{Softmax}(\mathbf{X}^{out})$  and compute the loss function;
15:  Update all parameter by back propagation using the Adam optimizer;
16: end while
```

Algorithm 2 The HAMP-II Algorithm for Hypergraph Node Classification.

```

1: Input: the incidence matrix  $\mathbf{H}$ , the node feature  $\mathbf{X}$ , and the node labels  $\mathbf{Y}$ .
2: Output: the model prediction accuracy.
3: Initialization: the time  $T$ , the step size  $\tau$  and all parameters of model.
4: while not converged do
5:   Node feature mapping  $\mathbf{X} = \text{Linear}_{\text{map}}(\mathbf{X})$ ;
6:   Set the initial time  $t = 0$ , the initial node representation  $\mathbf{X}(0) = \mathbf{X}$ ;
7:   Set the initial velocity  $\mathbf{V}(0) = \text{Linear}(\mathbf{X}(0)) - \mathbf{X}(0)$ ;
8:   while  $t \leq T$  do
9:     Message passing from  $\mathcal{V}$  to  $\mathcal{E}$ :  $\mathbf{V}_{\mathcal{V} \rightarrow \mathcal{E}}(t) = \Phi_1(\mathbf{V}(t), \mathbf{H})$ ;
10:    Message passing from  $\mathcal{E}$  to  $\mathcal{V}$ :  $\mathbf{V}_{\mathcal{E} \rightarrow \mathcal{V}}(t) = \Psi(\mathbf{V}(t), \Phi_2(\mathbf{V}_{\mathcal{V} \rightarrow \mathcal{E}}(t), \mathbf{H}))$ ;
11:    Compute the velocity of particle dynamics system as

$$\mathbf{V}(t + \tau) = \mathbf{V}(t) + \tau \sigma \left( \underbrace{\mathbf{V}_{\mathcal{E} \rightarrow \mathcal{V}}(t) - \omega \mathbf{V}(t)}_{\text{Interaction Force}} + \underbrace{\delta f_d(\mathbf{V}(t))}_{\text{Allen-Cahn Force}} + \underbrace{\epsilon \mathbf{B}(t)}_{\text{Noise}} + \beta \mathbf{V}(0) \right);$$

12:    Compute the representation by  $\mathbf{X}(t + \tau) = \mathbf{X}(t) + \tau \mathbf{V}(t + \tau)$ ;
13:    Update  $t = t + \tau$ ;
14:  end while
15:  Input the node representation into the classifier  $\mathbf{X}^{out} = \text{MLP}(\mathbf{X}(T))$ ;
16:  Compute the model prediction labels  $\hat{\mathbf{Y}} = \text{Softmax}(\mathbf{X}^{out})$  and compute the loss function;
17:  Update all parameter by back propagation using the Adam optimizer;
18: end while
```

Limitations Discussion. Our particle dynamics-based hypergraph message passing framework assumes a static hypergraph topology. While this assumption is valid for social and biological hypergraphs with slowly evolving interactions, it may not hold in highly dynamic scenarios like financial transaction networks, where hyperedge topologies change abruptly. Consequently, effectively modeling temporal hypergraphs with evolving structures remains an open challenge.

B Theoretical Results and Proof

Technically, we suppose there exists $\{f_\beta^e\}$ such that $\mathcal{I} = \{1, \dots, N\}$ can be divided into two disjoint groups with N_1, N_2 particles respectively: $f_\beta(h_{i,j}^e) \geq 0$, for $\{i, j\} \in \mathcal{I}_1$ or $\mathcal{I}_2$ and $f_\beta(h_{i,j}^e) \leq 0$, otherwise. We designate

$$\{x_i^{(1)}\} := \{\mathbf{x}_i | i \in \mathcal{I}_1\}, \quad \{x_j^{(2)}\} := \{\mathbf{x}_j | j \in \mathcal{I}_2\}. \quad (11)$$

The model is channel-wise, hence we use x instead of $\mathbf{x}$ in the proof.

First, let's define the relevant notation,

The mean value:

$$\bar{x} := \frac{1}{N} \sum_{i=1}^N x_i. \quad (12)$$

The deviation values:

$$\hat{x}_i := x_i - \bar{x}. \quad (13)$$

The variance of values within each group:

$$\text{var}(x^{(1)}) = \frac{1}{N_1} \sum (\hat{x}_i^{(1)})^2, \quad \text{var}(x^{(2)}) = \frac{1}{N_2} \sum (\hat{x}_i^{(2)})^2. \quad (14)$$

The second moments:

$$M_2(x^{(1)}) := \sum_{i=1}^{N_1} (x_i^{(1)})^2, \quad M_2(x^{(2)}) := \sum_{i=1}^{N_2} (x_i^{(2)})^2. \quad (15)$$

And others:

$$\widehat{M}_2 := M_2(\hat{x}^{(1)}) + M_2(\hat{x}^{(2)}) = \text{var}(x^{(1)}) + \text{var}(x^{(2)}). \quad (16)$$

$$\psi_i^{e,\pm} := \sum_{j \in e} h_{i,j}^{e,\pm}, \quad \psi_i^\pm := \sum_{e \in \mathcal{E}} \sum_{j \in e} h_{i,j}^{e,\pm}. \quad (17)$$

$$k := \max_i \{|\mathcal{E}(i)|\}. \quad (18)$$

$$D_e^- := \max_k \{\psi_k^{e,-}\}, \quad D^- := \max_e \{D_e^-\}. \quad (19)$$

$$D_2^- := \max_e \{\|\psi^{e,-}\|_2\}. \quad (20)$$

Technically, we set $N_1 = N_2 := N_0$. This assumption is that N_1 is comparable to N_2 , i.e., there exists a positive constant κ satisfying $\frac{1}{\kappa} N_1 \leq N_2 \leq \kappa N_1$.

We can rewrite Eq. 4 as

$$\begin{cases} \dot{u}_i^{(1)} = \frac{1}{N_1} \sum_{e \in \mathcal{E}(i)} \sum_{i'=1}^{N_1} h_{i,i'}^{e,+} (x_{i'}^{(1)} - x_i^{(1)}) - \frac{1}{N_2} \sum_{e \in \mathcal{E}(i)} \sum_{j=1}^{N_2} h_{i,j}^{e,-} (x_j^{(2)} - x_i^{(1)}) + \delta x_i^{(1)} (1 - (x_i^{(1)})^2) \\ \dot{v}_j^{(2)} = \frac{1}{N_2} \sum_{e \in \mathcal{E}(j)} \sum_{j'=1}^{N_2} h_{j,j'}^{e,+} (x_{j'}^{(2)} - x_j^{(2)}) - \frac{1}{N_1} \sum_{e \in \mathcal{E}(j)} \sum_{i=1}^{N_1} h_{i,j}^{e,-} (x_i^{(1)} - x_j^{(2)}) + \delta x_j^{(2)} (1 - (x_j^{(2)})^2). \end{cases} \quad (21)$$

Set the matrix A^e as

$$A_{i,j}^e = \begin{cases} h_{i,i}^e \\ -h_{i,j}^e \end{cases} \quad (22)$$

and designate $C^A := \min_{e \in \mathcal{E}} \{F(A^e)\}$, where $F(A^e)$ is the *Fiedler number* of A^e .

Lemma B.1 (L_2 estimate for M_2). *There exists a positive constant M_2^∞ such that*

$$\sup_{0 \leq t < \infty} M_2(t) \leq M_2^\infty \leq \infty. \quad (23)$$

Proof. Note that $h_{i,j}^{e,\pm} = h_{j,i}^{e,\pm}$, then

$$\begin{aligned} \frac{d}{dt} M_2(x^{(1)}) &= \frac{2}{N_1} \sum_{i=1}^{N_1} x_i^{(1)} \dot{x}_i^{(1)} \\ &= \frac{2}{N_1} \sum_{i=1}^{N_1} \sum_{e \in \mathcal{E}(i)} \sum_{i' \in e} h_{i,i'}^{e,+} (x_{i'}^{(1)} - x_i^{(1)}) x_i^{(1)} - \frac{2}{N_1} \sum_{i=1}^{N_1} \sum_{e \in \mathcal{E}(i)} \sum_{j \in e} h_{i,j}^{e,-} (x_j^{(2)} - x_i^{(1)}) x_i^{(1)} \\ &\quad + \frac{2\delta}{N_1} \sum_{i=1}^{N_1} (x_i^{(1)})^2 (1 - (x_i^{(1)})^2) \\ &= \frac{2}{N_1} \sum_{e \in \mathcal{E}} \sum_{i,i' \in e} h_{i,i'}^{e,+} (x_{i'}^{(1)} - x_i^{(1)}) x_i^{(1)} - \frac{2}{N_1} \sum_{e \in \mathcal{E}} \sum_{i,j \in e} h_{i,j}^{e,-} (x_j^{(2)} - x_i^{(1)}) x_i^{(1)} \\ &\quad + \frac{2\delta}{N_1} \sum_{i=1}^{N_1} (x_i^{(1)})^2 (1 - (x_i^{(1)})^2) \\ &= -\frac{1}{N_1} \sum_{e \in \mathcal{E}} \sum_{i,i' \in e} h_{i,i'}^{e,+} (x_{i'}^{(1)} - x_i^{(1)})^2 - \frac{2}{N_1} \sum_{e \in \mathcal{E}} \sum_{i,j \in e} h_{i,j}^{e,-} (x_j^{(2)} - x_i^{(1)}) x_i^{(1)} \\ &\quad + \frac{2\delta}{N_1} \sum_{i=1}^{N_1} (x_i^{(1)})^2 (1 - (x_i^{(1)})^2). \end{aligned} \quad (24)$$

Similarly,

$$\begin{aligned} \frac{d}{dt} M_2(x^{(2)}) &= \frac{2}{N_2} \sum_{j=1}^{N_2} x_j^{(2)} \dot{x}_j^{(2)} \\ &= -\frac{1}{N_2} \sum_{e \in \mathcal{E}} \sum_{j,j' \in e} h_{j,j'}^{e,+} (x_{j'}^{(2)} - x_j^{(2)})^2 - \frac{2}{N_2} \sum_{e \in \mathcal{E}} \sum_{j,i \in e} h_{j,i}^{e,-} (x_i^{(1)} - x_j^{(2)}) x_j^{(2)} \\ &\quad + \frac{2\delta}{N_2} \sum_{j=1}^{N_2} (x_j^{(2)})^2 (1 - (x_j^{(2)})^2). \end{aligned} \quad (25)$$

Define the total second moment $M_2 := M_2(x^{(1)}) + M_2(x^{(2)})$. Its time derivative is:

$$\frac{d}{dt} M_2 = \frac{d}{dt} M_2(x^{(1)}) + \frac{d}{dt} M_2(x^{(2)}). \quad (26)$$

By discarding the non-positive squared terms (first two sums), we obtain the inequality:

$$\begin{aligned} \frac{d}{dt} M_2 &\leq -\frac{2}{N_1} \sum_{e \in \mathcal{E}} \sum_{i,j \in e} h_{i,j}^{e,-} (x_j^{(2)} - x_i^{(1)}) x_i^{(1)} - \frac{2}{N_2} \sum_{e \in \mathcal{E}} \sum_{j,i \in e} h_{j,i}^{e,-} (x_i^{(1)} - x_j^{(2)}) x_j^{(2)} \\ &\quad + \frac{2\delta}{N_1} \sum_{i=1}^{N_1} (x_i^{(1)})^2 (1 - (x_i^{(1)})^2) + \frac{2\delta}{N_2} \sum_{i=1}^{N_2} (x_i^{(2)})^2 (1 - (x_i^{(2)})^2). \end{aligned} \quad (27)$$

By the Cauchy-Schwarz inequality,

$$\begin{aligned}(M_2^{(1)})^2 &= \left(\sum_{i=1}^{N_1} (x_i^{(1)})^2 \right)^2 \leq N_1 \sum_{i=1}^{N_1} (x_i^{(1)})^4, \\ (M_2^{(2)})^2 &= \left(\sum_{i=1}^{N_2} (x_i^{(2)})^2 \right)^2 \leq N_2 \sum_{i=1}^{N_2} (x_i^{(2)})^4, \\ (x_i^{(1)} - x_j^{(2)})^2 &\leq 2((x_i^{(1)})^2 + (x_j^{(2)})^2).\end{aligned}$$

Then we have

$$\begin{aligned}& \frac{2\delta}{N_1} \sum_{i=1}^{N_1} (x_i^{(1)})^2 (1 - (x_i^{(1)})^2) + \frac{2\delta}{N_2} \sum_{i=1}^{N_2} (x_i^{(2)})^2 (1 - (x_i^{(2)})^2) \\&= \frac{2\delta}{N_1} \sum_{i=1}^{N_1} (x_i^{(1)})^2 - \frac{2\delta}{N_1} \sum_{i=1}^{N_1} (x_i^{(1)})^4 + \frac{2\delta}{N_2} \sum_{i=1}^{N_2} (x_i^{(2)})^2 - \frac{2\delta}{N_2} \sum_{i=1}^{N_2} (x_i^{(2)})^4 \\&\leq \frac{2\delta}{N_1} \sum_{i=1}^{N_1} (x_i^{(1)})^2 - \frac{2\delta}{(N_1)^2} \left(\sum_{i=1}^{N_1} (x_i^{(1)})^2 \right)^2 + \frac{2\delta}{N_2} \sum_{i=1}^{N_2} (x_i^{(2)})^2 - \frac{2\delta}{(N_2)^2} \left(\sum_{i=1}^{N_2} (x_i^{(2)})^2 \right)^2 \\&= 2\delta \left(\frac{M_2^{(1)}}{N_1} + \frac{M_2^{(2)}}{N_2} \right) - 2\delta \left(\left(\frac{M_2^{(1)}}{N_1} \right)^2 + \left(\frac{M_2^{(2)}}{N_2} \right)^2 \right) \\&\leq 2\delta \left(\frac{M_2^{(1)}}{N_1} + \frac{M_2^{(2)}}{N_2} \right) - \delta \left(\frac{M_2^{(1)}}{N_1} + \frac{M_2^{(2)}}{N_2} \right)^2 \\&\leq \frac{2\delta}{N'} M_2 - \frac{\delta}{N''} (M_2)^2,\end{aligned} \tag{28}$$

where $N' = \min\{N_1, N_2\}$ and $N'' = \max\{N_1, N_2\}$.

So, we have

$$\begin{aligned}\frac{d}{dt} M_2 &\leq \frac{D^-}{N_1} \sum_{e \in \mathcal{E}} \sum_{i,j \in e} \left((x_j^{(2)} - x_i^{(1)})^2 + (x_i^{(1)})^2 \right) + \frac{D^-}{N_2} \sum_{e \in \mathcal{E}} \sum_{j,i \in e} \left((x_i^{(1)} - x_j^{(2)})^2 + (x_j^{(2)})^2 \right) \\&\quad + \frac{2\delta}{N_1} \sum_{i=1}^{N_1} (x_i^{(1)})^2 (1 - (x_i^{(1)})^2) + \frac{2\delta}{N_2} \sum_{i=1}^{N_2} (x_i^{(2)})^2 (1 - (x_i^{(2)})^2) \\&\leq \frac{D^-}{N_1} \sum_{e \in \mathcal{E}} \sum_{i,j \in e} \left(3(x_i^{(1)})^2 + 2(x_j^{(2)})^2 \right) + \frac{D^-}{N_2} \sum_{e \in \mathcal{E}} \sum_{i,j \in e} \left(2(x_i^{(1)})^2 + 3(x_j^{(2)})^2 \right) + \frac{2\delta}{N'} M_2 - \frac{\delta}{N''} M_2^2.\end{aligned} \tag{29}$$

These relations yield a Riccati-type differential inequality:

$$\begin{aligned}\frac{d}{dt} M_2 &\leq 5D^- \max\left\{ \frac{1}{N_1}, \frac{1}{N_2} \right\} \sum_{e \in \mathcal{E}} \sum_{i,j \in e} \left((x_i^{(1)})^2 + (x_j^{(2)})^2 \right) + 2\delta M_2 - \delta (M_2)^2 \\&\leq \frac{5D^- k}{N'} M_2 + \frac{2\delta}{N'} M_2 - \frac{\delta}{N''} (M_2)^2 \\&\leq \frac{5D^- k + 2\delta}{N'} M_2 - \frac{\delta}{N''} (M_2)^2.\end{aligned} \tag{30}$$

Let y be a solution of the following ODE:

$$y' = ay - by^2. \tag{31}$$

Then, by phase line analysis, the solution $y(t)$ to Eq. 31 satisfies

$$M_2(t) \leq y(t) \leq \max\left\{ \frac{a}{b}, M_2(0) \right\} = \max\left\{ \frac{(5D^- k + 2\delta)N''}{\delta N'}, M_2(0) \right\} =: M_2^\infty. \tag{32}$$

which yields the desired estimate. ■

Proposition B.2. For Eq. 4, the distance of the centers of the two clusters is finite.

Proof of Proposition B.2. By Lemma B.1

$$\begin{aligned}\|\bar{x}^{(1)} - \bar{x}^{(2)}\| &= \left\| \frac{1}{N_1} \sum_{i=1}^{N_1} x_i^{(1)} - \frac{1}{N_2} \sum_{j=1}^{N_2} x_j^{(2)} \right\| \\ &\leq \sqrt{2} \sqrt{\frac{1}{N_1} \sum_{i=1}^{N_1} (x_i^{(1)})^2 + \frac{1}{N_2} \sum_{j=1}^{N_2} (x_j^{(2)})^2} \\ &\leq \sqrt{2} \sqrt{M_2^\infty},\end{aligned}\tag{33}$$

where the first inequality used the Cauchy-Schwarz inequality. ■

Lemma B.3. Let u, v be the solution to Eq. 21. Then $\|\bar{x}^{(1)} - \bar{x}^{(2)}\|^2$ satisfies

$$\frac{1}{2} \frac{d}{dt} \|\bar{x}^{(1)} - \bar{x}^{(2)}\|^2 \geq \left(\frac{2c_m}{N_0} - c_1 \right) \|\bar{x}^{(1)} - \bar{x}^{(2)}\|^2 - \frac{4(D_2^-)^2}{c_1 N_0} k \widehat{M}_2.\tag{34}$$

Proof. The time evolution of $\bar{x}^{(1)}$ is given by

$$\begin{aligned}\dot{\bar{x}}^{(1)} &= \frac{1}{N_1} \sum_{i=1}^{N_1} \sum_{i' \in \mathcal{E}(i)} h_{i,i'}^{e,+} (x_{i'}^{(1)} - x_i^{(1)}) - \frac{1}{N_1} \sum_{i=1}^{N_1} \sum_{j \in \mathcal{E}(i)} h_{j,i}^{e,-} (x_j^{(2)} - x_i^{(1)}) \\ &= -\frac{1}{N_1} \sum_{e \in \mathcal{E}} \sum_{i,j \in e} h_{i,j}^{e,-} (x_j^{(2)} - x_i^{(1)}) \\ &= -\frac{1}{N_1} \sum_{e \in \mathcal{E}} \psi_j^{e,-} x_j^{(2)} + \frac{1}{N_1} \sum_{e \in \mathcal{E}} \psi_i^{e,-} x_i^{(1)} \\ &= \sum_{e \in \mathcal{E}} \left(-\frac{1}{N_1} \psi_j^{e,-} x_j^{(2)} + \frac{1}{N_1} \psi_i^{e,-} x_i^{(1)} \right),\end{aligned}\tag{35}$$

where the first equality uses the relation $\sum_{i=1}^{N_1} \hat{x}_i^{(1)} = 0$.

Then we have

$$\frac{1}{2} \frac{d}{dt} \|\bar{x}^{(1)} - \bar{x}^{(2)}\|^2\tag{36}$$

$$= (\bar{x}^{(1)} - \bar{x}^{(2)}) (\dot{\bar{x}}^{(1)} - \dot{\bar{x}}^{(2)})\tag{37}$$

$$= (\bar{x}^{(1)} - \bar{x}^{(2)}) \left[\sum_{e \in \mathcal{E}} \left(-\frac{1}{N_1} \psi_j^{e,-} x_j^{(2)} + \frac{1}{N_1} \psi_i^{e,-} x_i^{(1)} \right) - \sum_{e \in \mathcal{E}} \left(-\frac{1}{N_2} \psi_i^{e,-} x_i^{(1)} + \frac{1}{N_2} \psi_j^{e,-} x_j^{(2)} \right) \right]\tag{38}$$

$$= (\bar{x}^{(1)} - \bar{x}^{(2)}) \sum_{e \in \mathcal{E}} \left(-\frac{1}{N_1} \psi_j^{e,-} x_j^{(2)} - \frac{1}{N_2} \psi_j^{e,-} x_j^{(2)} + \frac{1}{N_1} \psi_i^{e,-} x_i^{(1)} + \frac{1}{N_2} \psi_i^{e,-} x_i^{(1)} \right)\tag{39}$$

$$= (\bar{x}^{(1)} - \bar{x}^{(2)}) \sum_{e \in \mathcal{E}} \left(-\frac{1}{N_1} \psi_j^{e,-} (\bar{x}^{(2)} + \hat{x}_j^{(2)}) - \frac{1}{N_2} \psi_j^{e,-} (\bar{x}^{(2)} + \hat{x}_j^{(2)}) + \frac{1}{N_1} \psi_i^{e,-} (\bar{x}^{(1)} + \hat{x}_i^{(1)}) \right)\tag{40}$$

$$+ \frac{1}{N_2} \psi_i^{e,-} (\bar{x}^{(1)} + \hat{x}_i^{(1)})\tag{41}$$

$$= (\bar{x}^{(1)} - \bar{x}^{(2)}) \left\{ \left[-\left(\frac{1}{N_1} + \frac{1}{N_2} \right) \sum_{e \in \mathcal{E}} \sum_{j \in e} \psi_j^{e,-} \right] \bar{x}^{(2)} + \left[\left(\frac{1}{N_1} + \frac{1}{N_2} \right) \sum_{e \in \mathcal{E}} \sum_{i \in e} \psi_i^{e,-} \right] \bar{x}^{(1)} \right\}\tag{42}$$

$$\left[-\left(\frac{1}{N_1} + \frac{1}{N_2}\right) \sum_{e \in \mathcal{E}} \sum_{j \in e} \psi_j^{e,-} \right] \hat{x}_j^{(2)} + \left[\left(\frac{1}{N_1} + \frac{1}{N_2}\right) \sum_{e \in \mathcal{E}} \sum_{i \in e} \psi_i^{e,-} \right] \hat{x}_i^{(1)} \right\} \quad (43)$$

$$= \frac{2}{N_0} (\bar{x}^{(1)} - \bar{x}^{(2)}) \left[\sum_{e \in \mathcal{E}} \sum_{i \in e} \psi_i^{e,-} \bar{x}^{(1)} - \sum_{e \in \mathcal{E}} \sum_{j \in e} \psi_j^{e,-} \bar{x}^{(2)} \right] \quad (44)$$

$$+ \frac{2}{N_0} (\bar{x}^{(1)} - \bar{x}^{(2)}) \left[\sum_{e \in \mathcal{E}} \sum_{i \in e} \psi_i^{e,-} \hat{x}_i^{(1)} - \sum_{e \in \mathcal{E}} \sum_{j \in e} \psi_j^{e,-} \hat{x}_j^{(2)} \right]. \quad (45)$$

We denote

$$\text{Pes}(\hat{x}^{(1)}, \hat{x}^{(2)}) := \sum_{e \in \mathcal{E}} \sum_{i \in e} \psi_i^{e,-} \hat{x}_i^{(1)} - \sum_{e \in \mathcal{E}} \sum_{j \in e} \psi_j^{e,-} \hat{x}_j^{(2)}.$$

and

$$\text{Pes}(\bar{x}^{(1)}, \bar{x}^{(2)}) := \sum_{e \in \mathcal{E}} \sum_{i \in e} \psi_i^{e,-} \bar{x}^{(1)} - \sum_{e \in \mathcal{E}} \sum_{j \in e} \psi_j^{e,-} \bar{x}^{(2)}.$$

Assume there exist constants c_m, c_v , such that

$$\text{Pes}(\bar{x}^{(1)}, \bar{x}^{(2)}) \geq c_m (\bar{x}^{(1)} - \bar{x}^{(2)}). \quad (46)$$

Then, by Cauchy's inequality, for any c_1 , we have

$$\begin{aligned} & \frac{1}{2} \frac{d}{dt} \|\bar{x}^{(1)} - \bar{x}^{(2)}\|^2 \\ &= \frac{2}{N_0} \text{Pes}(\bar{x}^{(1)}, \bar{x}^{(2)}) (\bar{x}^{(1)} - \bar{x}^{(2)}) + \frac{2}{N_0} \text{Pes}(\hat{x}^{(1)}, \hat{x}^{(2)}) (\bar{x}^{(1)} - \bar{x}^{(2)}) \\ &\geq \left(\frac{2c_m}{N_0} - c_1 \right) \|\bar{x}^{(1)} - \bar{x}^{(2)}\|^2 + c_1 \|\bar{x}^{(1)} - \bar{x}^{(2)}\|^2 + \frac{2}{N_0} \text{Pes}(\hat{x}^{(1)}, \hat{x}^{(2)}) (\bar{x}^{(1)} - \bar{x}^{(2)}) \\ &\geq \left(\frac{2c_m}{N_0} - c_1 \right) \|\bar{x}^{(1)} - \bar{x}^{(2)}\|^2 - \frac{1}{c_1} \frac{2}{N_0} \left(\text{Pes}(\hat{x}^{(1)}, \hat{x}^{(2)}) \right)^2 \\ &\geq \left(\frac{2c_m}{N_0} - c_1 \right) \|\bar{x}^{(1)} - \bar{x}^{(2)}\|^2 - \frac{4}{c_1 N_0} \sum_{e \in \mathcal{E}} \left[\left(\sum_{i \in e} \psi_i^{e,-} \hat{x}_i^{(1)} \right)^2 + \left(\sum_{j \in e} \psi_j^{e,-} \hat{x}_j^{(2)} \right)^2 \right] \\ &\geq \left(\frac{2c_m}{N_0} - c_1 \right) \|\bar{x}^{(1)} - \bar{x}^{(2)}\|^2 - \frac{4}{c_1 N_0} \sum_{e \in \mathcal{E}} \left[\|\psi^{e,-}\|^2 \sum_{i \in e} (\hat{x}_i^{(1)})^2 + \|\psi^{e,-}\|^2 \sum_{j \in e} (\hat{x}_j^{(2)})^2 \right] \\ &\geq \left(\frac{2c_m}{N_0} - c_1 \right) \|\bar{x}^{(1)} - \bar{x}^{(2)}\|^2 - \frac{4(D_2^-)^2}{c_1 N_0} \sum_{e \in \mathcal{E}} \left[\sum_{i \in e} (\hat{x}_i^{(1)})^2 + \sum_{j \in e} (\hat{x}_j^{(2)})^2 \right] \\ &\geq \left(\frac{2c_m}{N_0} - c_1 \right) \|\bar{x}^{(1)} - \bar{x}^{(2)}\|^2 - \frac{4(D_2^-)^2}{c_1 N_0} k[\widehat{M}_2]. \end{aligned} \quad (47)$$

■

Lemma B.4. Let u, v be the solution to Eq. 21. Then $\widehat{M}_2$ satisfies

$$\frac{1}{2} \frac{d}{dt} \widehat{M}_2 \leq C_2 \widehat{M}_2 + 2c_2 \|\bar{x}^{(1)} - \bar{x}^{(2)}\|^2, \quad (48)$$

where

$$C_2 := -k \left(C^A - \frac{D^-}{2} + \frac{(D^-)^2}{4c_2} - \frac{\delta}{k} \right), \quad (49)$$

and c_2 is an arbitrary positive constant.

Proof. Subtracting Eq. 35 from Eq. 21 gives $\dot{\hat{x}}_i^{(1)}$. Then we have

$$\begin{aligned}
& \frac{1}{2} \frac{d}{dt} \left(\frac{1}{N_1} \sum_{i=1}^{N_1} \|\hat{x}_i^{(1)}\|^2 \right) \\
&= \frac{1}{N_1} \sum_{i=1}^{N_1} \hat{x}_i^{(1)} \dot{\hat{x}}_i^{(1)} \\
&= \frac{1}{N_1} \sum_{i=1}^{N_1} \hat{x}_i^{(1)} \sum_{e \in \mathcal{E}(i)} \left[\sum_{i' \in e} h_{i,i'}^{e,+} (x_{i'}^{(1)} - x_i^{(1)}) - \sum_{j \in e} h_{i,j}^{e,-} (x_j^{(2)} - x_i^{(1)}) \right] \\
&\quad + \frac{\delta}{N_1} \sum_{i=1}^{N_1} \hat{x}_i^{(1)} x_i^{(1)} (1 - (x_i^{(1)})^2) \\
&= \frac{1}{N_1} \sum_{i=1}^{N_1} \hat{x}_i^{(1)} \sum_{e \in \mathcal{E}(i)} \left[\sum_{i' \in e} h_{i,i'}^{e,+} (\hat{x}_{i'}^{(1)} - \hat{x}_i^{(1)}) - \sum_{j \in e} h_{i,j}^{e,-} (x_j^{(2)} - x_i^{(1)}) \right] \\
&\quad + \frac{\delta}{N_1} \sum_{i=1}^{N_1} \hat{x}_i^{(1)} x_i^{(1)} (1 - (x_i^{(1)})^2) \\
&= \frac{1}{N_1} \sum_{i=1}^{N_1} \hat{x}_i^{(1)} \sum_{e \in \mathcal{E}(i)} \sum_{i' \in e} h_{i,i'}^{e,+} (\hat{x}_{i'}^{(1)} - \hat{x}_i^{(1)}) - \frac{1}{N_1} \sum_{i=1}^{N_1} \hat{x}_i^{(1)} \sum_{e \in \mathcal{E}(i)} \sum_{j \in e} h_{i,j}^{e,-} (\hat{x}_j^{(2)} - \hat{x}_i^{(1)}) \\
&\quad - \frac{1}{N_1} \sum_{i=1}^{N_1} \hat{x}_i^{(1)} \sum_{e \in \mathcal{E}(i)} \sum_{j \in e} h_{i,j}^{e,-} (\bar{x}^{(2)} - \bar{x}^{(1)}) + \frac{\delta}{N_1} \sum_{i=1}^{N_1} \hat{x}_i^{(1)} x_i^{(1)} (1 - (x_i^{(1)})^2) \\
&=: I_1 + I_2 + I_3 + I_4.
\end{aligned} \tag{50}$$

I_1 can be defined by

$$\begin{aligned}
I_1 &= \frac{1}{N_1} \sum_{i=1}^{N_1} \hat{x}_i^{(1)} \sum_{e \in \mathcal{E}(i)} \sum_{i' \in e} h_{i,i'}^{e,+} (\hat{x}_{i'}^{(1)} - \hat{x}_i^{(1)}) \\
&= \frac{1}{N_1} \sum_{e \in \mathcal{E}} \sum_{i,i' \in e} h_{i,i'}^{e,+} (\hat{x}_{i'}^{(1)} - \hat{x}_i^{(1)}) \hat{x}_i^{(1)} \\
&= \frac{1}{N_1} \sum_{e \in \mathcal{E}} ((\hat{x}^{(1)})^e)^\top A^e (\hat{x}^{(1)})^e,
\end{aligned} \tag{51}$$

where $(\hat{x}^{(1)})^e := (\hat{x}_{i_1}^{(1)}, \dots, \hat{x}_{i_{|e|}}^{(1)})^\top$ and A^e is given by Eq. 22 for each e . Thus I_1 is bounded by

$$\frac{1}{N_1} \sum_{e \in \mathcal{E}} ((\hat{x}^{(1)})^e)^\top A^e (\hat{x}^{(1)})^e \leq -\frac{1}{N_1} \sum_{e \in \mathcal{E}} C^A \|(\hat{x}^{(1)})^e\|^2 \leq -\frac{c_r}{N_1} \sum_{i=1}^{N_1} C^A \|\hat{x}_i^{(1)}\|^2, \tag{52}$$

where c_r is a constant larger than 1 related to the repetition of $\{\hat{x}_i^{(1)}\}$ in all hyperedges.

I_2 can be controlled by

$$I_2 = -\frac{1}{N_1} \sum_{i=1}^{N_1} \hat{x}_i^{(1)} \sum_{e \in \mathcal{E}(i)} \sum_{j \in e} h_{i,j}^{e,-} (\hat{x}_j^{(2)} - \hat{x}_i^{(1)}) \tag{53}$$

$$= -\frac{1}{N_1} \sum_{e \in \mathcal{E}} \sum_{i,j \in e} h_{i,j}^{e,-} (\hat{x}_j^{(2)} - \hat{x}_i^{(1)}) \hat{x}_i^{(1)} \tag{54}$$

$$= -\frac{1}{N_1} \sum_{e \in \mathcal{E}} \sum_{i,j \in e} h_{i,j}^{e,-} \hat{x}_j^{(2)} \hat{x}_i^{(1)} + \frac{1}{N_1} \sum_{e \in \mathcal{E}} \sum_{i,j \in e} h_{i,j}^{e,-} \|\hat{x}_i^{(1)}\|^2 \tag{55}$$

$$\leq \frac{1}{N_1} \sum_{e \in \mathcal{E}} \sum_{i,j \in e} h_{i,j}^{e,-} \frac{1}{2} (\|\hat{x}_j^{(2)}\|^2 + \|\hat{x}_i^{(1)}\|^2) + \frac{1}{N_1} \sum_{e \in \mathcal{E}} D_e^- \sum_{i \in e} \|\hat{x}_i^{(1)}\|^2 \quad (56)$$

$$\leq \frac{1}{2N_1} \sum_{e \in \mathcal{E}} \sum_{j \in e} \psi_j^{e,-} \|\hat{x}_j^{(2)}\|^2 + \frac{3D^-}{2N_1} \sum_{e \in \mathcal{E}} \sum_{i \in e} \|\hat{x}_j^{(1)}\|^2 \quad (57)$$

$$\leq \frac{D^-}{2N_1} \sum_{e \in \mathcal{E}} \sum_{j \in e} \|\hat{x}_j^{(2)}\|^2 + \frac{3D^-}{2N_1} \sum_{e \in \mathcal{E}} \sum_{i \in e} \|\hat{x}_j^{(1)}\|^2 \quad (58)$$

$$\leq \frac{D^- k}{2N_1} \sum_{j=1}^{N_2} \|\hat{x}_j^{(2)}\|^2 + \frac{3D^- k}{2N_1} \sum_{i=1}^{N_1} \|\hat{x}_i^{(1)}\|^2. \quad (59)$$

I_3 has the below estimate for any constant $c_2 > 0$:

$$\begin{aligned} I_3 &= -\frac{1}{N_1} \sum_{i=1}^{N_1} \hat{x}_i^{(1)} \sum_{e \in \mathcal{E}(i)} \sum_{j \in e} h_{i,j}^{e,-} (\bar{x}^{(2)} - \bar{x}^{(1)}) \\ &\leq c_2 \|\bar{x}^{(1)} - \bar{x}^{(2)}\|^2 + \frac{1}{4c_2 N_1} \sum_{e \in \mathcal{E}} \sum_{i,j \in e} \|h_{i,j}^{e,-}\|^2 \|\hat{x}_i^{(1)}\|^2 \\ &\leq c_2 \|\bar{x}^{(1)} - \bar{x}^{(2)}\|^2 + \frac{(D^-)^2}{4c_2 N_1} \sum_{e \in \mathcal{E}} \sum_{i,j \in e} \|\hat{x}_i^{(1)}\|^2 \\ &\leq c_2 \|\bar{x}^{(1)} - \bar{x}^{(2)}\|^2 + \frac{(D^-)^2 k}{4c_2 N_1} \sum_{e \in \mathcal{E}} \sum_{i=1}^{N_1} \|\hat{x}_i^{(1)}\|^2. \end{aligned} \quad (60)$$

Define

$$\begin{aligned} I_4 &:= \frac{1}{N_1} \delta \sum_{i=1}^{N_1} \hat{x}_i^{(1)} x_i^{(1)} (1 - \|x_i^{(1)}\|^2) \\ &= \frac{\delta}{N_1} \sum_{i=1}^{N_1} \hat{x}_i^{(1)} (\hat{x}_i^{(1)} + \bar{x}^{(1)}) (1 - (x_i^{(1)})^2) \\ &= \frac{\delta}{N_1} \sum_{i=1}^{N_1} (\hat{x}_i^{(1)})^2 - \frac{\delta}{N_1} \sum_{i=1}^{N_1} (\hat{x}_i^{(1)})^2 (x_i^{(1)})^2 + \frac{\delta}{N_1} \sum_{i=1}^{N_1} \hat{x}_i^{(1)} \bar{x}^{(1)} - \frac{\delta}{N_1} \sum_{i=1}^{N_1} \hat{x}_i^{(1)} (x_i^{(1)})^2 \bar{x}^{(1)} \\ &= \frac{\delta}{N_1} \sum_{i=1}^{N_1} (\hat{x}_i^{(1)})^2 - \frac{\delta}{N_1} \sum_{i=1}^{N_1} \hat{x}_i^{(1)} \|x_i^{(1)}\|^2 x_i^{(1)}. \end{aligned} \quad (61)$$

Note that

$$\begin{aligned} \sum_{i=1}^{N_1} \hat{x}_i^{(1)} \|x_i^{(1)}\|^2 x_i^{(1)} &= \sum_{i=1}^{N_1} \|x_i^{(1)}\|^2 (\|x_i^{(1)}\|^2 - x_i^{(1)} \bar{x}^{(1)}) \\ &\geq \frac{1}{2} \sum_{i=1}^{N_1} \|x_i^{(1)}\|^2 (\|x_i^{(1)}\|^2 - \|\bar{x}^{(1)}\|^2) \\ &= \frac{1}{2} \sum_{i=1}^{N_1} \|x_i^{(1)}\|^4 - \frac{1}{2} \sum_{i=1}^{N_1} \|x_i^{(1)}\|^2 \|\bar{x}^{(1)}\|^2 \\ &\geq \frac{1}{2} \sum_{i=1}^{N_1} \|x_i^{(1)}\|^4 - \frac{1}{2N_1} \sum_{i=1}^{N_1} (\|x_i^{(1)}\|^2)^2 \\ &\geq 0. \end{aligned} \quad (62)$$

Hence,

$$I_4 \leq \delta \widehat{M}_2(u). \quad (63)$$

Then

$$\begin{aligned} & \frac{d}{dt} \left(\frac{1}{2N_1} \sum_{i=1}^{N_1} \|\hat{x}_i^{(1)}\|^2 \right) \\ & \leq k \left(-\frac{C^A}{N_1} + \frac{3D^-}{2N_1} + \frac{(D^-)^2}{4c_2N_1} \right) \sum_{i=1}^{N_1} \|\hat{x}_i^{(1)}\|^2 + k \frac{D^-}{2N_2} \sum_{j=1}^{N_2} \|\hat{x}_j^{(2)}\|^2 + c_2 \|\bar{x}^{(1)} - \bar{x}^{(2)}\|^2 \quad (64) \\ & \leq \left[k \left(-C^A + \frac{3D^-}{2} + \frac{(D^-)^2}{4c_2} \right) + \delta \right] \widehat{M}_2(u) + k \frac{D^-}{2} \widehat{M}_2(v) + c_2 \|\bar{x}^{(1)} - \bar{x}^{(2)}\|^2. \end{aligned}$$

Similarly,

$$\begin{aligned} & \frac{d}{dt} \left(\frac{1}{2N_2} \sum_{j=1}^{N_2} \|\hat{x}_j^{(2)}\|^2 \right) \\ & \leq \left[k \left(-C^A + \frac{3D^-}{2} + \frac{(D^-)^2}{4c_2} \right) + \delta \right] \widehat{M}_2(v) + k \frac{D^-}{2} \widehat{M}_2(u) + c_2 \|\bar{x}^{(1)} - \bar{x}^{(2)}\|^2. \quad (65) \end{aligned}$$

Summing them together gives

$$\frac{1}{2} \frac{d}{dt} (\widehat{M}_2) \leq -k \left(C^A - \frac{D^-}{2} + \frac{(D^-)^2}{4c_2} - \frac{\delta}{k} \right) (\widehat{M}_2) + 2c_2 \|\bar{x}^{(1)} - \bar{x}^{(2)}\|^2. \quad (66)$$

This gives an exponential growth estimate or $\widehat{M}_2$ up to an error term of $\|\bar{x}^{(1)} - \bar{x}^{(2)}\|^2$. ■

Proposition B.5 (L_2 separation of HAMP-I). *For Eq. 4, suppose the above assumptions are satisfied. Define the mean value $\bar{\mathbf{x}} := \frac{1}{N} \sum_{i=1}^N \mathbf{x}_i$, and the second moments $M_2(\mathbf{x}) := \sum_{i=1}^N \mathbf{x}_i^2$. Then for sufficiently large N_1, N_2 , there exist constants λ_-, λ_+ , such that if the initial data satisfies*

$$\lambda(0) := \frac{\widehat{M}_2(0)}{\|\bar{\mathbf{x}}^{(1)}(0) - \bar{\mathbf{x}}^{(2)}(0)\|^2} \leq \lambda_+, \quad (67)$$

then, there holds that the L_2 separation

$$\lambda(t) := \frac{\widehat{M}_2(t)}{\|\bar{\mathbf{x}}^{(1)}(t) - \bar{\mathbf{x}}^{(2)}(t)\|^2} \leq \lambda_- + (\lambda(0) - \lambda_-)e^{-\mu t}, \quad (68)$$

with a positive constant μ , where $\widehat{M}_2(t) := M_2(\mathbf{x}^{(1)}(t)) + M_2(\mathbf{x}^{(2)}(t))$.

Proof of Proposition 5.2. Lemma B.3 gives an exponential growth estimate of $\|\bar{x}^{(1)} - \bar{x}^{(2)}\|^2$ up to an error term of $\widehat{M}_2$. If N_0 is large enough, the coefficient of the error term is small.

Lemma B.4 gives an exponential growth estimate or $\widehat{M}_2$ up to an error term of $\|\bar{x}^{(1)} - \bar{x}^{(2)}\|^2$. If N_0 is large enough, the coefficient of the error term is small.

Set

$$A_{11} := \frac{2c_m}{N_0} - c_1, \quad A_{12} := \frac{4(D^-)^2 k}{c_1 N_0}, \quad A_{21} = 2c_2, \quad A_{22} = k \left(C^A - \frac{D^-}{2} + \frac{(D^-)^2}{4c_2} \right). \quad (69)$$

If N_0 is large enough, Eq. 69 is satisfied.

$$(A_{11} + A_{12})^2 - 4A_{21}A_{12} > 0. \quad (70)$$

Apply Lemma 4.1 in [27] then we obtain the L_2 separation. ■

Remark B.6. We can alternatively prove Theorem 5.2 by the method of [[44] Proposition 2 and 3].

Proposition B.7 (Lower bound of the Dirichlet energy). *If the hypergraph $\mathcal{H}$ is a connected one, for Eq. 4 with the conditions of Theorem 5.2, or for Eq. 5 with conditions of Theorem 5.1 in [19], there exists a positive lower bound of the Dirichlet energy.*

Proof of Proposition 5.4. The relative size between $\|\bar{x}^{(1)} - \bar{x}^{(2)}\|^2$ and $\widehat{M}_2$ is an indicator of group separation in the sense of L_2 : if $\|\bar{x}^{(1)} - \bar{x}^{(2)}\|^2$ is much larger than $\widehat{M}_2$, then the two groups are well-separated in average sense. Since the hypergraph is connected, there is a positive bound between different clusters, hence the Dirichlet energy does not decay to zero.

Remark B.8. The proof for the second-order system Eq. 5 can be proved in a similar way.

Remark B.9. The separability of Eq. 6 which is the Eq. 4 or Eq. 5 with a stochastic term also holds.

C Experiment Details

C.1 Dataset Details

We selected the most common hypergraph benchmark datasets, and the statistics of nine datasets across different domains are summarized in Tab. 3. The key statistics include the number of nodes, hyperedges, features, classes, average node degree d_v , average hyperedge size $|\mathcal{E}|$, and CE homophily [36]. These variations highlight the diversity in dataset scales and structural patterns, which may influence model performance in hypergraph-related tasks.

Table 3: The summary of data statistics.

Dataset	# nodes	# hyperedges	# features	# classes	avg. d_v	avg. $ \mathcal{E} $	CE homophily
Cora	2708	1579	1433	7	1.767	3.03	0.897
Citeseer	3312	1079	3703	6	1.044	3.200	0.893
Pubmed	19717	7963	500	3	1.756	4.349	0.952
Cora-CA	2708	1072	1433	7	1.693	4.277	0.803
DBLP-CA	41302	22363	1425	6	2.411	4.452	0.869
Congress	1718	83105	100	2	427.237	8.656	0.555
House	1290	340	100	2	9.181	34.730	0.509
Senate	282	315	100	2	19.177	17.168	0.498
Walmart	88860	69906	100	11	5.184	6.589	0.530

C.2 Additional Ablation Studies

Impact of Hidden Dimension on HAMP-I and HAMP-II. We explore the tolerance of our model to different hidden dimensions, as shown in Tab. 4. For simplicity, we only vary the size of hidden dimension, while other parameters remain fixed. Overall, these results demonstrate the robustness of our methods with varying hidden dimensions. It is worth noting that high hidden dimension is key to achieving the best performance of HAMP.

Table 4: Impact of hidden dimension evaluated on the hypergraph datasets.

	HAMP-I				HAMP-II			
	128	256	512	1024	128	256	512	1024
Cora	80.19	80.55	81.18	79.85	79.59	79.54	80.80	79.60
Citeseer	73.39	72.89	75.22	72.40	73.95	74.36	75.33	75.19
Pubmed	88.49	88.85	89.02	88.82	88.48	88.48	89.05	88.78
Senate	63.24	65.35	69.44	66.62	63.52	63.10	70.14	62.96
House	71.21	70.62	72.66	72.72	68.30	69.47	72.60	71.21

Impact of Repulsion, Allen-Cahn Force, and Noise on HAMP-I and HAMP-II. We summary ablation studies to investigate the individual and combined effects of repulsion f_β^- , Allen-Cahn force f_d , and noise B_t on both HAMP-I and HAMP-II. Tab. 5 reports the average node classification accuracy with a standard deviation across seven standard hypergraph benchmarks over 10 runs. Models with the repulsion term enabled outperform their counterparts in some dataset, indicating

enhanced ability to distinguish node representations in complex hypergraph structures. The synergy between the repulsion and Allen-Cahn terms further boosts performance, confirming that these particle system-inspired mechanisms play complementary roles. Overall, these improvement confirm the validity of the HAMP construction and further highlight the significant advantages of incorporating particle system theory into the hypergraph message passing learning process.

Table 5: Ablation studies on some standard hypergraph benchmarks. The accuracy (%) is reported with a standard deviation from 10 repetitive runs. (Key: f_{β}^{-} : repulsion; f_d : Allen-Cahn; B_t : noise.)

Homophilic	f_{β}^{-}	f_d	B_t	Cora	Citeseer	Pubmed	Cora-CA
HAMP-I	✗	✗	✗	76.09±1.22	70.53±1.56	87.98±0.38	83.13±1.26
	✓	✗	✗	76.40±1.56	70.85±1.65	88.25±0.50	83.15±1.36
	✗	✓	✗	80.31±1.41	74.83±1.70	88.90±0.45	84.77±1.16
	✗	✗	✓	75.67±1.71	70.59±1.40	87.93±0.52	82.70±1.01
	✓	✓	✗	80.49±1.26	74.96±1.56	88.87±0.40	85.21±1.49
	✗	✓	✓	80.59±1.25	74.67±1.69	88.77±0.44	84.59±1.03
	✓	✓	✓	81.18±1.30	75.22±1.62	89.02±0.49	85.23±1.15
HAMP-II	✗	✗	✗	77.42±1.44	71.50±1.49	88.68±0.62	83.60±1.45
	✓	✗	✗	77.08±1.73	72.20±1.14	88.59±0.50	82.95±1.35
	✗	✓	✗	80.13±1.26	74.37±1.59	88.86±0.55	84.37±1.45
	✗	✗	✓	77.18±1.61	71.75±1.68	88.68±0.59	82.91±1.28
	✓	✓	✗	79.70±1.36	73.99±1.75	88.82±0.48	84.30±1.32
	✗	✓	✓	79.50±1.25	74.25±1.28	88.80±0.40	83.31±1.44
	✓	✓	✓	80.80±1.62	75.33±1.61	89.05±0.41	84.89±1.14

Table 6: Node Classification on standard hypergraph benchmarks. The accuracy (%) is reported with a standard deviation from 10 repetitive runs. (Key: f_{β}^{-} : repulsion; f_d : Allen-Cahn; B_t : noise.)

Heterophilic	f_{β}^{-}	f_d	B_t	Congress	Senate	Walmart	House
HAMP-I	✗	✗	✗	93.51±1.08	60.70±8.38	69.64±0.35	70.50±1.45
	✓	✗	✗	93.70±1.02	60.00±8.66	69.80±0.45	70.56±1.96
	✗	✓	✗	94.79±1.14	66.76±5.44	69.77±0.28	71.55±1.53
	✗	✗	✓	93.47±1.13	60.14±7.54	69.86±0.35	69.88±2.48
	✓	✓	✗	94.58±1.25	67.75±8.82	69.76±0.37	71.58±1.87
	✗	✓	✓	94.67±1.02	65.63±2.98	69.73±0.49	71.55±2.58
	✓	✓	✓	95.09±0.79	69.44±6.09	69.90±0.38	72.72±1.77
HAMP-II	✗	✗	✗	94.79±0.73	58.73±7.03	69.84±0.25	69.85±1.61
	✓	✗	✗	94.02±1.10	61.55±5.98	69.91±0.30	69.35±1.87
	✗	✓	✗	94.19±1.07	62.82±6.44	69.89±0.31	70.96±2.06
	✗	✗	✓	94.35±1.14	60.14±5.10	69.84±0.37	70.46±2.08
	✓	✓	✗	94.58±0.86	61.97±8.42	69.92±0.33	71.36±1.68
	✗	✓	✓	94.12±0.63	64.51±4.19	69.86±0.34	70.96±2.76
	✓	✓	✓	95.26±1.34	70.14±6.08	69.94±0.37	72.60±1.23

C.3 Time-Memory Tradeoff Analysis

We intuitively reveal the differences of different methods with a single-layer network in the time-memory trade-off on Walmart dataset. As shown in Fig. 5, HAMP-I and HAMP-II methods demonstrate a notable trade-off between time efficiency and memory consumption. The experimental results reveal:

- Memory usage: HAMP-I and HAMP-II maintain memory consumption within 11000-12000 MiB, achieving a 20-26% reduction compared to ED-HNN, while being comparable to HDS^{ode} .
- Time efficiency: Although the time consumption for HAMP-I and HAMP-II runtime slightly exceeds ED-HNN (0.1s), it outperforms the baseline HDS^{ode} .

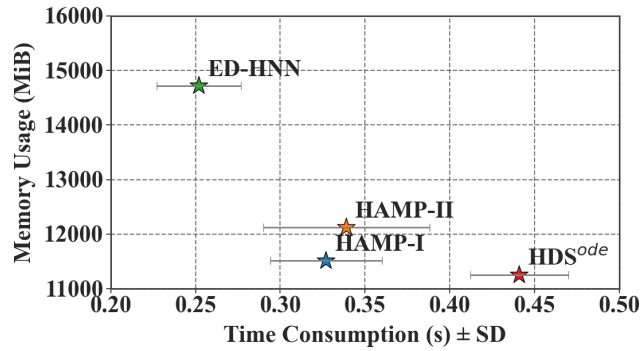


Figure 5: Time-Memory tradeoff analysis of different methods on Walmart dataset. SD denotes the standard deviation of time consumption.

C.4 Hyperparameters

To ensure fairness, we follow the same training recipe as ED-HNN. Specifically, we train the model for 500 epochs using the Adam optimizer with the learning rate of 0.001 and no weight decay during the training phases. And we apply early stopping with a patience of 50. For the stability, we run 10 trials with different seed and report the results of mean and the standard deviation. All experiments are implemented on an NVIDIA RTX 4090 GPU with Pytorch.

We explore the parameter space by grid search, where the search ranges for each critical hyperparameter are delineated below:

- Dropout rate in {0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9};
- Layer number of classifier in {1, 2, 3};
- The hidden dimension of classifier in {128, 256, 512};
- Hidden dimension of model in {128, 256, 512, 1024};
- step size of solver in {0.09, 0.1, 0.15, 0.2, 0.25};
- γ of repulsive force in {0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, 0.11, 0.12, 0.13, 0.14, 0.15};
- Initial values of learnable parameters δ of damping term in {0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15};
- Initial values of learnable parameters ϵ of noise term in {0, 0.1, 0.3};

Tab. 7 and Tab. 8 summarize the best hyperparameters on standard hypergraph benchmarks using HAMP-I and HAMP-II, respectively. For fairness, a linear layer is added to perform feature mapping when conducting HDS^{ode} experiments. The optimal hyperparameters for node classification on standard hypergraph benchmarks is achieved by the HDS^{ode} algorithm, as demonstrated in Tab. 9.

Table 7: The best hyperparameters of Node Classification on standard hypergraph benchmarks using the HAMP-I algorithm.

Dataset	model. hd	cls. hd and # layers	time	step size	δ	γ	dropout	ϵ
Cora	512	128, 1	1	0.1	12	0.05	0.4	0
Citeseer	512	512, 1	0.6	0.1	6	0.05	0.2	0
Pubmed	512	256, 1	0.2	0.1	15	0.1	0.5	0
Cora-CA	512	512, 2	0.4	0.2	4	0.05	0.9	0
DBLP-CA	256	128, 2	1.1	0.1	11	0.12	0.2	0
Congress	128	128, 2	1.4	0.1	1	0.08	0.3	0
House	1024	512, 3	1.05	0.15	3	0.05	0.8	0
Senate	512	256, 2	0.6	0.1	10	0.05	0.7	0
Walmart	256	128, 2	1.75	0.25	0	0.02	0.3	0

Table 8: The best hyperparameters of Node Classification on standard hypergraph benchmarks using the HAMP-II algorithm.

Dataset	model. hd	cls. hd and layers	time	step size	δ	γ	dropout	ϵ
Cora	512	512, 1	1.9	0.1	5	0.12	0.3	0.1
Citeseer	512	512, 1	1.8	0.15	8	0.13	0.6	0
Pubmed	512	256, 1	0.6	0.09	5	0.09	0.3	0
Cora-CA	512	128, 2	0.75	0.25	3	0.01	0.7	0
DBLP-CA	256	128, 2	3.45	0.15	7	0.09	0.3	0
Congress	128	128, 2	6.25	0.25	0	0.01	0.3	0
House	512	256, 2	1.6	0.1	8	0.14	0.8	0
Senate	512	256, 2	3	0.2	13	0.05	0.3	0.3
Walmart	256	128, 2	2.5	0.25	0	0.02	0.3	0

Table 9: The best hyperparameters of Node Classification on standard hypergraph benchmarks using the HDS^{ode} algorithm.

Dataset	model. hd	cls. hd and layers	# layer model.	α_v	α_e	step
Cora	512	256, 2	10	0.05	0.9	20
Citeseer	512	512, 1	5	0.05	0.9	20
Pubmed	512	256, 1	12	0.05	0.9	20
Cora-CA	512	512, 2	9	0.05	0.9	20
DBLP-CA	256	256, 2	15	0.05	0.9	20
Congress	256	128, 2	7	0.25	0.9	5
House	512	256, 2	10	0.05	0.9	20
Senate	512	256, 2	9	0.05	0.9	20
Walmart	256	128, 2	6	0.25	0.9	5

NeurIPS Paper Checklist

1. Claims

Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

Answer: [\[Yes\]](#)

Justification: The abstract and introduction accurately describe the proposed framework (HyperMP), theoretical contributions (hypergraph particle dynamics system, anti-over-smoothing analysis), and experimental scope (hypergraph node classification), which are consistent with what is presented in the paper (Sections 3, 4, 5, 6, Appendices A, B, C).

Guidelines:

- The answer NA means that the abstract and introduction do not include the claims made in the paper.
- The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
- The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
- It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. Limitations

Question: Does the paper discuss the limitations of the work performed by the authors?

Answer: [\[Yes\]](#)

Justification: In the Appendix A, we not only describe the algorithm in detail, but also discuss its limitations.

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. Theory assumptions and proofs

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: The paper states proposition (Proposition 5.2-5.4) and provides detailed proofs for the main theoretical results in Appendix B.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. Experimental result reproducibility

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: The paper describes the experimental setup, datasets (standard hypergraph benchmarks), and baselines in Section 6. The appendix C.2 provides dataset details, more ablation studies, and hyperparameter details. The code has been provided anonymously in the abstract.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general, releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
 - (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
 - (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
 - (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

- (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. Open access to data and code

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: Code is stated to be open source. We use the same hypergraph benchmark dataset as ED-HNN. Instructions are assumed to be provided with the code repository.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so “No” is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. Experimental setting/details

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: Experimental settings, including optimizer, iterations, and hyperparameter search ranges, are detailed in the Appendix C.4.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. Experiment statistical significance

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: Except for Table 2, all other tables report the results as mean and standard deviation over 10 runs, indicating statistical variability.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. Experiments compute resources

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [Yes]

Justification: The paper specifies the hardware used (NVIDIA RTX 4090 GPU) in Appendix C.4. Runtimes and memory are reported in the Figure C.3, allowing for estimation of computational cost.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. Code of ethics

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics <https://neurips.cc/public/EthicsGuidelines>?

Answer: [Yes]

Justification: The research involves algorithmic development and evaluation on standard optimization benchmarks. It does not involve human subjects or obviously ethically sensitive applications, and we assume it conforms to the NeurIPS Code of Ethics.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. Broader impacts

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [No]

Justification: The paper focuses on the technical contributions and does not include a specific discussion of broader positive or negative societal impacts.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. Safeguards

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [No]

Justification: The paper proposes a new framework/algorithm. Neither the models nor the standard benchmark data used appear to pose a high risk for misuse necessitating specific release safeguards beyond standard open-source practices.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. Licenses for existing assets

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [No]

Justification: The paper properly cites the sources for existing assets like baseline methods and data sources. However, the specific licenses and terms of use for these assets are not explicitly mentioned in the paper text or appendix.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, paperswithcode.com/datasets has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. New assets

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [\[Yes\]](#)

Justification: The primary new asset is the source code for the proposed methods, which is open source. Documentation is assumed to be provided alongside the code in its repository.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. Crowdsourcing and research with human subjects

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [\[No\]](#)

Justification: The research does not involve crowdsourcing experiments or research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. Institutional review board (IRB) approvals or equivalent for research with human subjects

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [No]

Justification: The research does not involve human subjects, therefore IRB approval is not applicable.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. Declaration of LLM usage

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [NA]

Justification: The core method development in this research does not involve LLMs as any important, original, or non-standard components.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (<https://neurips.cc/Conferences/2025/LLM>) for what should or should not be described.