ODNet: Opinion Dynamics-Inspired Neural Message Passing for Graphs and Hypergraphs

Anonymous authors
Paper under double-blind review

Abstract

Neural message passing serves as a cornerstone framework in graph neural networks, providing a clear and intuitive mathematical guideline for the propagation and aggregation of information among interconnected nodes within graphs. Throughout this process, node representations undergo dynamic updates, considering both the individual states and connections of neighboring nodes. Concurrently, social networks, as prominent forms of interconnected data, form dynamic systems that achieve stability through continuous internal communications and opinion exchanges among social actors along their social ties. Drawing upon the shared concepts between these two domains, our study establishes an explicit connection between message passing and opinion dynamics in sociology. Moreover, we introduce a novel continuous message passing scheme termed ODNet, which integrates bounded confidence to refine the influence weight of local nodes for message propagation. By adjusting the similarity cutoffs of bounded confidence and influence weights within ODNet, we define opinion exchange rules that align with the characteristics of neural message passing and can effectively mitigate the oversmoothing issue. We extend the framework to hypergraphs and formulate corresponding continuous message passing rules, which reveal a close association with particle dynamics. Empirically, we showcase that ODNet enhances prediction performance across various social networks presented as homophilic graphs, heterophilic graphs, and hypergraphs. Notably, our proposed ODNet outperforms existing GNNs with its straightforward construction and robust theoretical foundation.

1 Introduction

Graph neural networks have emerged as a powerful deep learning framework for analyzing the intercommunication of entities on a graph, offering versatile tools for node, edge, and graph-level representation learning (Scarselli et al., 2008; Bronstein et al., 2017). The primary challenge in designing GNN models lies in effectively aggregating information based on local interactions for efficiently extracting hidden representations. This design philosophy has been generalized as *neural message passing* (Gilmer et al., 2017) and later became a fundamental feature extraction unit of graph for aggregating features of neighbors.

Concurrently, *opinion dynamics* studies the evolving social relations with the opinions dynamically changed over time through interactions (Acemoglu & Ozdaglar, 2011). A *social network* establishes a system with individual social actors and the social ties among them. It uses a graph as an abstract representation of a group's structure (Proskurnikov & Tempo, 2017). To analyze the formation, evolution, and spread of opinions within social networks or communities, statistical physics has contributed by introducing methods and tools from dynamical systems theory (Helbing, 2010; Weidlich, 2006).

Despite originating from different disciplines, neural message passing and opinion dynamics share core principles of information propagation and aggregation within interconnected systems. Both fields focus on understanding how local interactions among individuals (nodes) contribute to the emergence of global behaviors in social networks (graphs). This inherent connection motivates our exploration of an efficient alternative to neural message passing inspired by popular models in opinion dynamics. We aim to address three key research questions: (1) *How to integrate models of opinion dynamics into the neural message passing frame-*

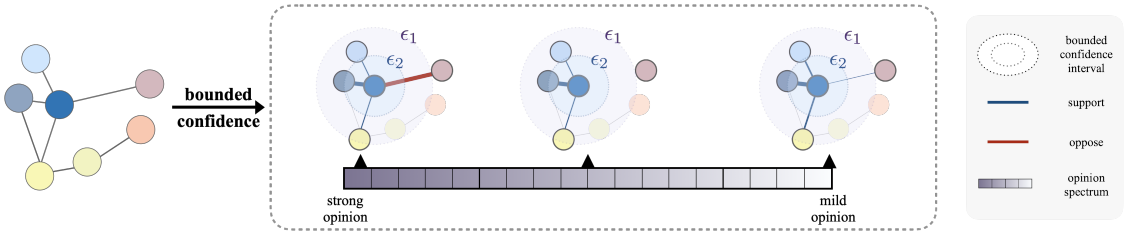

Figure 1: An illustration of the propagation rule in ODNet with bounded confidence. Colored nodes represent individuals with different opinions. The central node (dark blue) interacts with light blue nodes sharing similar opinions and a red node with significantly divergent thoughts. During message passing, ODNet enhances the aggregation weight of similar nodes (light blue) and adjusts the weights of conflict nodes (red). Bounded confidence defines the range of similar/conflict nodes and the extent of adjustments, and it is influenced by the graph's position on the opinion spectrum. See Section 5 for modeling details.

work? (2) *What are the implications of opinion dynamics models' properties for mitigating the oversmoothing issue in neural message passing?* (3) *How can we apply opinion dynamics-inspired neural message passing rules to typical graphs encountered in social networks, such as heterophilic graphs and hypergraphs?*

To address the first two questions, we analyze the formulation and dynamic behavior of two representative models in opinion dynamics: the French-DeGroot (FD) model (DeGroot, 1974; French Jr, 1956) and the Hegselmann-Krause (HK) model (Rainer & Krause, 2002). We reformulate the FD and HK models to establish explicit connections with the message passing framework. Based on the unified formulation, we analyze the convergence behavior of the two models and provide a novel explanation of the oversmoothing issue in GNNs from the perspective of opinion dynamics. We demonstrate that the FD message passing converges exponentially to a stable state when exhibiting strong local connectivities. This property is frequently observed in hypergraphs and significantly contributes to the oversmoothing issue. Alternatively, the HK model mitigates the issue by introducing heterophilious dynamics to avoid consensus or environmental averaging.

Based on this connection, we answer the third question by integrating the concept of *bounded confidence* from the HK model and introduce an opinion dynamics-inspired message passing formulation named ODNet. A few recent studies combine opinion dynamics to define propagation rules on graphs. However, they are based on position-based first-order message passing (Okawa & Iwata, 2022) or ignore data-driven propagation rules (Giráldez-Cru et al., 2022), resulting in limited expressivity. This framework, instead, incorporates a confidence-based filtration mechanism on initial edge connections and incorporates second-order diffusion processes. By considering the similarity of node pairs and connection proximity, ODNet aggregates neighboring information through automatic adjustment to edge weights. Through its piecewise message passing schemes, the model can strengthen, weaken, or remove initial links within the graph. Furthermore, it allows for the assignment of negative weights to capture conflicting viewpoints from neighboring nodes with substantial disparities. This capability proves essential especially when examining heterophilic networks, where connected entities exhibit dissimilar characteristics. The propagation scheme mirrors the dynamic spreading of opinions within a community, where effective communication channels converge individuals or agents towards consensus or a few dominant viewpoints (Figure 1). When exchanging information, individuals tend to support opinions aligned with themselves. However, when encountering significantly divergent thoughts, individuals may choose to disregard or oppose them, depending on position of the social network on the opinion spectrum. For example, researchers (on the right side of the spectrum) typically concentrate on studies within their expertise but are open to learning new perspectives from other domains, whereas politicians (on the left side of the spectrum) often encounter strong conflicts and resist propositions from competing parties.

We evaluate the versatile ODNet across three categories of graphs: homophilic graphs, heterophilic graphs, and hypergraphs, each characterized by unique properties resembling real-world social networks such as the small-world effect (Girvan & Newman, 2002) and scale-free networks (Barabasi & Oltvai, 2004). Empirically, the introduced piecewise aggregation behavior enhances the performance of well-established message passing methods. In summary, our work has three main contributions: (1) **We demonstrate an explicit**

connection between opinion dynamics and the neural message passing framework, which used to be two parallel research fields in modeling social networks (Section 4). By formulating opinion dynamics models under neural message passing frameworks, we offer a new interpretation of oversmoothing from the perspective of opinion dynamics. (2) **We propose ODNet, a novel message passing scheme inspired by opinion dynamics** (Section 5.1). ODNet provides a promising alternative to traditional neural message passing methods. It incorporates bounded confidence to refine the influence weight of local node features, boasting a clean formulation, solid theoretical support, and effective mitigation of the oversmoothing issue. Notably, the introduction of bounded confidence does not introduce additional computational cost to the propagation. (3) **We defines the continuous ODNet for hypergraphs and provide theoretical analysis** (Section 5.2). By extensive experiments, we validate its superior performance on different types of graphs and compare it with both classic and SOTA discrete or continuous message passing schemes.

2 Related Work

Neural Message Passing on Graphs and Hypergraphs Neural message passing establishes a general computational rule for updating node representations in attributed graphs (Gilmer et al., 2017). By implementing different types of propagation rules, information at the node and edge levels is communicated among neighborhoods. In parallel, Feng et al. (2019) and Gao et al. (2022b) established a general convolution framework employing the incidence matrix for hypergraph learning. Various techniques have also undergone expansion, such as the attention mechanism (Bai et al., 2021) and spectral theory (Yadati et al., 2019).

Continuous Message Passing and the Oversmoothing Issue The concept of continuous GNN dynamics was first introduced by Poli et al. (2020) and Chamberlain et al. (2021), which explicitly generalizes discrete propagation frameworks such as GAT (Veličković et al., 2018). This diffusion framework has seen active extensions into more continuous graph convolutions with additional considerations of graph rewiring (Brandstetter et al., 2021; Bodnar et al., 2022) Meanwhile, the energy conservation of second-order diffusion processes circumvents the oversmoothing issue and training instability by capturing long-range interactions (Wang et al., 2023; Han et al., 2024).

Opinion Dynamics for Social Networks and Graph Neural Networks The natural connection between graphs and social networks has inspired pioneering works exploring the possibility of extending the analysis of social networks to graph neural networks. Sheaf networks build connections with socialdynamics by separating messages from nodes into public and private opinions (Caralt et al.; Duta et al., 2024; Zaghen et al., 2024; Hansen & Ghrist, 2021). SINN (Okawa & Iwata, 2022) follows a first-order formulation of information exchange to capture the positions of individuals without considering mechanisms to prevent excessive attraction or repulsion from neighbors, thus ignoring the inherent oversmoothing issue.To extend to the second-order formulation with Newtonian dynamics, social networks analysis defines *collective dynamics* (Carrillo et al., 2010; Motsch & Tadmor, 2014) that depicts the phenomenon of agents moving together based on attraction and repulsion forces (Cucker & Smale, 2007; Fang et al., 2019; Jin & Shu, 2021b). ATBCR (Giráldez-Cru et al., 2022) followed opinion dynamic rules and introduced the concept of attraction and repulsion between nodes. However, the explicit propagation rule excludes learnable message passing layers, which restricts its performance on more sophisticated networks such as heterophilic graphs.

3 Graphs from the View of Message Passing and Opinion Dynamics

3.1 Graphs and Hypergraphs

A graph $\mathcal{G}[\boldsymbol{W}] = (\mathcal{V}, \mathcal{E}[\boldsymbol{W}])$ of N nodes can be associated with any square non-negative weight matrix $\boldsymbol{W} \in \mathbb{R}^{N\times N}$, where $\mathcal{V}$ represents the node set and $\mathcal{E}$ is the edge set. We denote x_i as the feature of node v_i or the opinion of individual i. $\mathcal{G}$ is *strongly connected* (Godsil & Royle, 2001) if there exists a path from every node to every other node. A cycle is a directed path that both begins and ends at the same node with no repeated nodes except for the initial/final one. The length of a cycle is defined by the number of edges in the cyclic path. The periodicity of a graph is defined as the smallest integer k that divides the length of every cycle in the graph. When $k = 1$, $\mathcal{G}$ is termed *aperiodic* (Bullo et al., 2009).

A hypergraph is a generalization of a graph in which an edge can connect any number of vertices. It can be denoted by a triple $\mathcal{H}[\boldsymbol{W}^h] = \{\mathcal{V}, \mathcal{E}, \boldsymbol{W}^h\}$. To avoid notation confusion, we still use $\mathcal{V}$ for the set of nodes and $\mathcal{E}$ for the set of hyperedges. We set $|\mathcal{V}| = N$ and $|\mathcal{E}| = M$, and $\mathcal{E}(i)$ denotes a set containing all the nodes sharing at least one hyperedge with node i. Usually, $\boldsymbol{W}^h$ is a diagonal matrix for hyperedges, where W^h_{ee} represents the weight of the hyperedge e. In this paper, we extend the weight representation to a triple tensor $\boldsymbol{W}^h \in \mathbb{R}^{N\times N\times M}$, where $w^e_{i,j}$ designates an element in $\boldsymbol{W}^h$. The incidence matrix $\boldsymbol{H} \in \mathbb{R}^{N\times M}$ defines $H_{i,e} = 1$ if the node i belongs to the hyperedge e, otherwise $H_{i,e} = 0$. We generalize the indicator within $\boldsymbol{W}^h$ by setting $w^e_{i,j}$ as nonzero if node i, j are connected by a hyperedge e, otherwise 0.

3.2 Neural Message Passing on Graphs and Hypergraphs

Neural Message Passing (Gilmer et al., 2017) stands as the prevailing propagator for updating node representations in GNNs. We denote $\boldsymbol{x}_i^{(k-1)}$ as the features of node i in layer $(k-1)$ and $a_{j,i} \in \mathbb{R}^d$ as the edge features from node j to node i. A message passing layer reads as follows:

$$\mathbf{x}_i^{(k)} = \gamma^{(k)} \left(\mathbf{x}_i^{(k-1)}, \square_{j\in\mathcal{N}_i}\, \phi^{(k)} \left(\mathbf{x}_i^{(k-1)}, \mathbf{x}_j^{(k-1)}, a_{j,i} \right) \right), \tag{1}$$

where $\square$ denotes a differentiable, node permutation-invariant function, such as summation, mean, or maximization. The γ and ϕ denote differentiable functions such as MLPs (Multi-Layer Perceptrons), and $\mathcal{N}_i$ represents the set of one-hop neighbors of node i. The message passing mechanism updates the feature of each node by aggregating their self-features with neighbors' features. On hypergraphs, the message passing scheme considers interactions among multiple nodes reflected in a hyperedge. At the kth layer:

$$\mathbf{x}_i^{(k+1)} = \Psi^{(k)} \left(\mathbf{x}_i^{(k)}, \Phi_{1,e\in\mathcal{E}(i)} \left(e, \Phi^{(k)}_{2,j\in e}(\{\mathbf{x}_j^{(k)}\}, \{a^e_{j,i}\}) \right) \right), \tag{2}$$

where $\Phi_1^{(k)}(\cdot)$ denotes a differentiable, hyperedge permutation-invariant function, and $\Phi_2^{(k)}(\cdot)$ is a differentiable, node permutation invariant function. The $\Psi^{(k)}(\cdot)$ denotes another differentiable function of propagation, and $j \in e$ implies $H_{j,e} = 1$ or $a^e_{j,i} \neq 0$.

4 Opinion Dynamics

This section introduces the FD and HK models, two iconic models in opinion dynamics, following a discussion on the interconnection of their formulation with message passing and the oversmoothing issue in GNNs.

4.1 French-DeGroot Model

The *French-DeGroot* (FD) model (DeGroot, 1974) is a groundbreaking agent-based model that simulates the evolution of opinions. In a given population of N individuals, each individual holds an opinion $\boldsymbol{x}_i(k)$ at discrete time instances $k = 0, 1, \cdots$. An individual's opinion is evolved by

$$\boldsymbol{x}_i(k+1) = \sum\nolimits_{j=1}^{N} w_{ij} \boldsymbol{x}_j(k), \tag{3}$$

where the non-negative *influence weight* w_{ij} satisfies $\sum_{j=1}^{N} w_{ij} = 1$. If $w_{ij} > 0$, individuals i and j are considered neighbors. The influence weight signifies the relative impact that individual j exerts on i during each opinion update. Importantly, all individuals concurrently update their opinions at each time step. The FD model captures how individual opinions converge within a group that potentially leading to a consensus. This opinion pooling process could also be interpreted as a message passing scheme, which emulates how information is exchanged within a specific type of neural network.

Connection to Neural Message Passing The FD model is often regarded as a micro-level model based on individuals simulating the evolution of individual opinions. We show that it shares similarities with GRAND (Chamberlain et al., 2021), which describes a diffusion process on graphs by connecting heat

conduction with neural message passing. The connection between the two models is established through the discretization of a partial differential equation on graphs:

$$\frac{\partial}{\partial t}\boldsymbol{x}(t) = (\boldsymbol{A}(\boldsymbol{x}(t)) - \boldsymbol{I}_N)\boldsymbol{x}(t), \tag{4}$$

where $\boldsymbol{A}(\boldsymbol{x}(t))$ denotes the $N \times N$ attention matrix on nodes and $\boldsymbol{I}_N$ is an identity matrix. GRAND coincides with the FD model when $(\boldsymbol{A}(\boldsymbol{x}(t)) - \boldsymbol{I}_N)$ satisfies the row-stochastic property and a simple forward-Euler method is applied with a time step of one. This intriguing parallel between the two models highlights the interconnection of designs from the two different domains of social network modeling.

Opinion Consensus and the Oversmoothing Issue A fundamental result regarding the convergence of the FD model is well-established, demonstrating that consensus is achieved exponentially fast for a strongly connected and aperiodic graph (Bullo et al., 2009; Proskurnikov & Tempo, 2017; Ye, 2019).

Proposition 1. *Consider the evolution of opinions $\boldsymbol{x}_i(k)$ for each individual i within the network $\mathcal{G}[\mathbf{W}]$ according to (3). Assuming that $\mathcal{G}[\mathbf{W}]$ is strongly connected and aperiodic, and that $\mathbf{W}$ is row-stochastic. Define ζ as the dominant left eigenvector of $\mathbf{W}$, then $\lim_{k\to 0}\mathbf{x}(k) = (\zeta^\top \mathbf{x}(0))\mathbf{1}_N$ exponentially fast.*

The proposition above shows that any graph with a self-loop is considered aperiodic, implying that exponential decay often occurs in graphs with strong connectivity. In GNNs, this phenomenon is known as *oversmoothing* (Nt & Maehara, 2019; Oono & Suzuki, 2019), which has been widely studied for its association with the exponential decay of the *Dirichlet energy* – a measure of feature convergence weighted by graph structure. Despite their distinct origins, these phenomena describe similar processes.

4.2 Hegselmann-Krause Model

In the FD model, each agent could interact with any other agent, regardless of their opinions. However, the previous analysis shows that such trivial aggregation behavior inevitably leads to oversmoothing the signals in the system. In real-life scenarios, individuals typically engage in conversations primarily with those who share similar viewpoints. In sociodynamics, this phenomenon is termed as collective behaviors, where a group of individuals collectively exhibit coordinated patterns. The *Hegselmann-Krause* (HK) model (Holm & Putkaradze, 2006; Kolokolnikov et al., 2013; Motsch & Tadmor, 2014) characterizes this behavior by

$$\boldsymbol{x}_i(k+1) = |\mathbb{B}(i, \boldsymbol{x}_i))|^{-1} \sum_{j \in \mathbb{B}(i,\boldsymbol{x}_i)} \boldsymbol{x}_j(k), \tag{5}$$

where the *bounded confidence* $\mathbb{B}(i, \boldsymbol{x}_i) = \{j : |\boldsymbol{x}_j(k) - \boldsymbol{x}_i(k)| < \epsilon\}$ includes all the peers j associated with the individual i whose opinions diverges within a confined region $\epsilon_i \in \mathbb{R}$. This parameter defines the degree of uncertainty or tolerance for opinion exchanges. It restricts that an individual i with an opinion $\boldsymbol{x}_i$ will only engage in interactions with those of their peers whose opinions fall within a confined region.

Consensus versus Opinion Polarization The HK model is a first-order opinion dynamics model that directly incorporates the influence of neighboring opinions on individual updates through bounded confidence. Nevertheless, bounded confidence does not inherently prevent the oversmoothing issue. In scenarios of significant heterophily, agents preferentially connect with dissimilar others. In this case, the model often leads to consensus and mirrors the oversmoothing problem in GNNs. It thus becomes pivotal to guide the HK model to capture opinion polarization and form multiple distinct and stable clusters. One way to approach this target involves employing strategies like bi-clustering with repulsion mechanisms (Fang et al., 2019; Jin & Shu, 2021b; Wang et al., 2023), which ensures a lower bound for the Dirichlet energy of the input signal.

5 Opinion Dynamics-Inspired Message Passing

This section defines ODNet, a novel message passing framework that employs the *influence function* $\phi(s)$ with bounded confidences. We offer a comprehensive interpretation of each component in ODNet. We introduce the discrete and continuous forms of the proposed model on both graphs and hypergraphs.

5.1 ODNet on Graphs

Discrete Formulation Denote $\phi(\cdot)$ as a non-decreasing function of the similarity measure $s_{i,j}$ to node i and node j. The update rule for ODNet reads:

$$\boldsymbol{x}_i(t+1) = \sum_{i=1}^{N} \phi(s_{ij})(\boldsymbol{x}_j(t) - \boldsymbol{x}_i(t)) + \boldsymbol{x}_i(t) + u(\boldsymbol{x}_i(t)). \tag{6}$$

The matrix s_{ij} has different definitions, such as the normalized adjacency matrix (Kipf & Welling, 2017) and attention coefficients (Veličković et al., 2018). The control term $u(\boldsymbol{x}_i)$ is deployed for stability. One straightforward choice of formulation is $u(x_i) = \nabla P(x)|_{x=x_i}$ with a potential function $P(x)$, such that $\nabla P(x) \to \infty$ as $x \to \infty$ (Kolokolnikov et al., 2011). It acts as a moral constraint preventing individuals from resorting to extreme violence in conflict situations in sociodynamics.

Continuous Formulation In opinion dynamics, individual perspectives tend to evolve gradually rather than through sudden changes. This observation suggests enhancing the discrete message passing model into a continuous framework. This adaptation can be understood as transforming a traditional message passing approach into a numerical approximation of the continuous model by rewriting (4) into:

$$\frac{\partial \boldsymbol{x}_i(t)}{\partial t} = \sum_{i=1}^{N} \phi(s_{ij})(\boldsymbol{x}_j(t) - \boldsymbol{x}_i(t)) + u(\boldsymbol{x}_i). \tag{7}$$

The continuous formulation (7) facilitates the application of numerous numerical approximation methods, transforming it back into a discrete model equipped with a specific scheme for residual adjustment. This adaptability allows applying diverse Ordinary Differential Equation (ODE) solvers to ODNet, including Neural ODEs (Chen et al., 2018). The associated explorations can be found in Section 6.

Choice of Influence Function The monotonic influence function $\phi(\cdot)$ characterizes the influence weight associated with the similarity of node pairs. It describes the diverse behaviors of opinion exchange in a system, which is determined by the systems' positions along the opinion spectrum (Figure 1). We define a piecewise $\phi(\cdot)$ to delineate influence regions akin to bounded confidence. For instance, with

$$\phi(s) = \begin{cases} \mu s, & \text{if } \epsilon_2 < s \\ s, & \text{if } \epsilon_1 \leq s \leq \epsilon_2 \\ 0, & \text{otherwise,} \end{cases} \tag{8}$$

the propagation rule (6) can be rewritten as

$$\boldsymbol{x}_i(t+1) = \mu \sum_{\epsilon_2 < s_{i,j}} s_{i,j}(\boldsymbol{x}_j(t) - \boldsymbol{x}_i(t)) + \sum_{\epsilon_1 \leq s_{i,j} \leq \epsilon_2} s_{i,j}(\boldsymbol{x}_j(t) - \boldsymbol{x}_i(t)) + \boldsymbol{x}_i(t). \tag{9}$$

This formulation amplifies the influence weight for highly similar node pairs with coefficient $\mu > 0$ while cutting connections for node pairs with low similarity, resembling how individuals tend to ignore opinions beyond their bounded confidence.

The previous choice (8) corresponds to when individuals maintain neutral opinions toward each other. Conversely, when individuals holding significantly divergent opinions and the network leans towards the 'strong opinion' side of the spectrum, interactions may become hostile. In such cases, $\phi(\cdot)$ can be defined as

$$\phi(s) = \begin{cases} \mu s, & \text{if } \epsilon_2 < s \\ s, & \text{if } \epsilon_1 \leq s \leq \epsilon_2 \\ \nu(1-s), & \text{otherwise,} \end{cases} \tag{10}$$

where $\mu > 0$ and $\nu < 0$. This configuration allows the neural network to learn from *positive* neighbors with high similarities and to extract *negative* information from nodes with discrepancies. Specifically, the

negative coefficient ν indicates that certain node pairs consistently repel each other, potentially leading to undesirable system expansion. Consequently, the previously introduced control term $u(x_i)$ is essential for the model towards enhanced system stability.

Remark 1. *The choice of $\phi(\cdot)$ can be guided by the graph's intrinsic characteristics or by using the* homophily level *(Pei et al., 2020) as a quantitative measure. We recommend using (8) for homophilic graphs and (10) for heterophilic graphs. The empirical analysis is in Section 6.*

5.2 ODNet on Hypergraphs

It is conceivable to extend a graph diffusion framework, *e.g.*, GRAND, to hypergraphs with macroscopic interpretations. However, these diffusion-type dynamics at the macro level are susceptible to oversmoothing of feature evolution, similar to traditional GNNs. Especially on hypergraphs, since are generally characterized with denser local connections than traditional graphs, addressing the oversmoothing issue becomes particularly important when constructing message passing propagation rules for hypergraphs. In this section, we will first analyze the oversmoothing issue of diffusion-based hypergraph message passing, then propose the continuous update rule of ODNet for hypergraphs that circumvent the oversmoothing issue.

Diffusion-based Hypergraph Message Passing When incorporated with macroscopic interpretations, diffusion-based hypergraph message passing propagates the attractions of individuals or agents. Consider the node feature space $\Omega = \mathbb{R}^d$ and the tangent vector field space $T\Omega = \mathbb{R}^d$. For $\mathbf{x}, \mathbf{y} \in \Omega$ and $\mathfrak{x}, \mathfrak{y} \in T\Omega$ where $\mathfrak{x}_{i,j} = -\mathfrak{x}_{j,i}$, we adopt the following inner products:

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i,j} \mathbf{x}_i \mathbf{y}_j, \qquad [\mathfrak{x}, \mathfrak{y}] = \sum_{i>j} \sum_{e \in \mathcal{E}} h^e_{i,j}\, \mathfrak{x}_{i,j} \mathfrak{y}_{i,j}. \tag{11}$$

Here $h^e_{i,j}$ represents a tuple related to node i, j and the associated hyperedge e, where $h^e_{i,j} = 0$ if $H_{i,e} = 0$. We set $h^e_{i,j}$ to satisfy $\sum_j \sum_{e \in \mathcal{E}} h^e_{i,j} = 1$. For any $\mathfrak{u} \in T\Omega$, by the adjoint relation $[\mathfrak{u}, \nabla \mathbf{x}] = \langle \mathbf{x}, \operatorname{div} \mathfrak{u} \rangle$ with $\nabla \mathbf{x} = \mathbf{x}_j - \mathbf{x}_i$, we derive:

$$(\operatorname{div} \mathfrak{u})_j = \sum_i \sum_{e \in \mathcal{E}} h^e_{i,j} u_{i,j}.$$

This leads to a formal diffusion process of a hypergraph:

$$\frac{d\mathbf{x}_i}{dt} = \operatorname{div} \nabla \mathbf{x}_i = \sum_j \sum_{e \in \mathcal{E}} h^e_{i,j} (\mathbf{x}_j - \mathbf{x}_i). \tag{12}$$

For simplicity, we define the hypergraph operator $\mathcal{L} = I - (\sum_{e \in \mathcal{E}} h^e_{i,j})$ and rewrite (12) as

$$\frac{d\mathbf{x}}{dt} = -\mathcal{L}\mathbf{x}.$$

When $\mathcal{L}$ is semi-positive definite (s.p.d), we define (12) as a diffusion-type propagation of a hypergraph. The different choices of $h^e_{i,j}$ lead to diverse diffusion-type equations. For example, when we take forward Euler discretization on (12) and define

$$\sum_{e \in \mathcal{E}} h^e_{i,j} = D_v^{-\frac{1}{2}} H W D_e^{-1} H^T D_v^{-\frac{1}{2}},$$

we have a simplified HGNN (Gao et al., 2022a) without channel mixing.

Oversmoothing Problem of Hypergraph Message Passing To understand this claim, we first define the Dirichlet energy of a hypergraph $\mathcal{H}$ of vector field $\mathbf{x} \in \mathbb{R}^{N \times d}$ for diffusion-type hypergraph networks:

$$\mathbf{E}(\mathbf{x}) := \sum_{i,j=1}^{N} \sum_{e \in \mathcal{E}} H_{i,e} H_{j,e} \|\mathbf{x}_i - \mathbf{x}_j\|^2. \tag{13}$$

When formulating the Dirichlet energy on hypergraph, rather than following the conventional formulation of $\mathbf{E}(\mathbf{x}) := \mathrm{tr}(\mathbf{x}^\top \mathcal{L}\mathbf{x})$, we adopt a simplified definition by Rusch et al. (2022) due to the non-deterministic nature of $\mathcal{L}$, which also facilitates a straightforward analysis of the variations in node features.

Definition 1. *Let $\mathbf{x}^l$ represent the hidden features at the lth layer. Oversmoothing in a hypergraph neural network is defined as the exponential decay of layer-wise Dirichlet energy as a function of layer depth l, i.e.,*

$$\mathbf{E}(\mathbf{x}^l) \leq C_1 e^{-C_2 l}, \tag{14}$$

where C_1 and C_2 are positive constants.

Given that $|\mathbf{x}| \leq C_1 e^{-\gamma t}$ with γ being the smallest positive eigenvalue of $\mathcal{L}$, it is easy to see that oversmoothing is a common issue in general hypergraph diffusion networks. This observation comes from the diffusion structure, where the node features $\mathbf{x}$ decay exponentially to zero with the s.p.d kernel $\mathcal{L}$.

Continuous Formulation of ODNet on Hypergraphs ODNet offers a comprehensive message passing framework that can be directly extended to hypergraphs. The primary distinction between hypergraphs over graphs lies in how hyperedges facilitate connectivity beyond traditional pairwise edges. Following the continuous representation of (12), we define:

$$\frac{\partial \boldsymbol{x}_i}{\partial t} = \sum_{e: i \in e} \sum_{j \in e} \phi(s^e_{i,j})(\boldsymbol{x}_j - \boldsymbol{x}_i) + u(\boldsymbol{x}_i). \tag{15}$$

This model formulates collective behaviors-based diffusion-type dynamics for hypergraphs. It suggests that in a large community, information spreads through subgroups (nodes connected by a hyperedge) rather than solely through direct pairwise exchanges. As with graphs, the selection of $s^e_{i,j}$ can include different metrics, such as attention coefficients (Bai et al., 2021) or convolution coefficients (Gao et al., 2022a).

The HK model can be interpreted as a diffusion process on graphs featuring piecewise attraction and repulsion behaviors, as outlined in (7) and (15). When devising message passing aggregation rules, employing a comprehensive framework for microscopic models based on collective behaviors within a complete interaction system proves effective in mitigating the oversmoothing issue. The complete proof is in Appendix A.

Remark 2. *The diffusion process of a hypergraph by (12) is also closely related to self-organized dynamics in particle systems (Motsch & Tadmor, 2014), where (12) represents a particle dynamics scenario with $h^e_{i,j}$ signifying the interactive force between nodes i, j under a specific field e. Note that (12) corresponds to a particular case of (15) where only attractive forces influence the message evolution. However, this assumption is not universal for particle systems, and as demonstrated above, it can lead to the oversmoothing issue.*

6 Empirical Analysis

We validate the performance of ODNet through classic node-level representation learning tasks on a variety of social networks described by homophilic/heterophilic graphs or hypergraphs.

6.1 Experimental Protocol

Training Setup We compare our model to a diverse set of top-performing baseline GNN models, including classic message passings with discrete or continuous updating schemes, as well as the latest hypergraph models. For ODNet, we trained the model using a neural ODE solver with Dormand–Prince adaptive step size scheme (DOPRI5). For homophilic datasets, we utilized 10 random weight initializations and random splits, with each combination randomly selecting 20 instances from each class. In heterophilic and hypergraph datasets, we used the fixed 10 training/validation splits in Pei et al. (2020) and Yadati et al. (2019). All implementations are programmed with `PyTorch-Geometric` (version 2.0.1) (Fey & Lenssen, 2019) and `PyTorch` (version 1.7.0) and run on NVIDIA® Tesla A100 GPU with 6,912 CUDA cores and 80GB HBM2 mounted on an HPC cluster. All the details to reproduce our results have been included in the submission. The program will be publicly available upon acceptance. The searching space of hypergraphs is reported in Table 7. We compare ODNet with a variety of previous methods on graphs and hypergraphs (see Appendix B.1).

Table 1: Average test accuracy on **homophilic** graphs over 10 random splits.

	Cora	**CiteSeer**	**PubMed**	**Coauthor CS**	**Computer**	**Photo**	**ogb-arXiv**
homophily level	0.83	0.71	0.79	0.80	0.77	0.83	0.61
GCN	81.5±1.3	71.9±1.9	77.8±2.9	91.1±0.5	82.6±2.4	91.2±1.2	72.2±0.3
MoNet	81.3±1.3	71.2±2.0	78.6±2.3	90.8±0.6	83.5±2.2	91.2±2.3	N/A
GraphSage-avg	79.2±7.7	71.6±1.9	77.4±2.2	91.3±2.8	82.4±1.8	91.4±1.3	N/A
GraphSage-max	76.6±1.9	67.5±2.3	76.1±2.3	85.0±1.1	N/A	90.4±1.3	N/A
GAT	81.8±1.3	71.4±1.9	78.7±2.3	90.5±0.6	78.0±1.9	85.7±2.0	73.6±0.1
GAT-PPR	81.6±0.3	68.5±0.2	76.7±0.3	91.3±0.1	**85.4±0.3**	90.9±0.3	N/A
CGNN	81.4±1.6	66.9±1.8	66.6±4.4	92.3±0.2	80.3±2.0	91.4±1.5	58.7±2.5
GDE	78.7±2.2	71.8±1.1	73.9±3.7	91.6±0.1	82.9±0.6	92.4±2.0	56.7±10.9
GRAND-l	83.6±1.0	73.4±0.5	78.8±1.7	92.9±0.4	83.7±1.2	92.3±0.9	71.8±0.2
SINN	83.2±0.4	73.9±0.6	80.0±1.7	93.0±0.5	84.6±1.1	90.4 ±2.4	N/A
ODNet (ours)	**85.7±0.3**	**75.5±1.2**	**80.6±1.1**	**93.1±0.7**	83.9±1.5	**92.7±0.6**	**74.65±0.16**

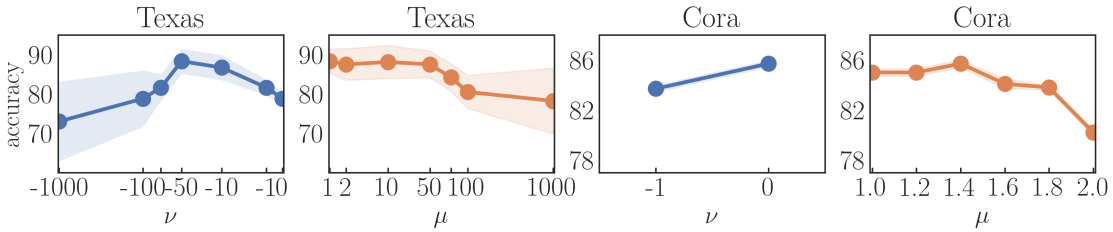

Figure 2: Impact of scaling factors ν and μ on **Texas** and **Cora**.

6.2 Node Classification on Graphs

Benchmark Datasets To evaluate a diverse set of scenarios, we encompass seven homophilic graphs (**Cora** (McCallum et al., 2000), **Citeseer** (Sen et al., 2008), **Pubmed** (Namata et al., 2012), **Coauthor CS** (Shchur et al., 2018), **Computer** (Namata et al., 2012), **Photo** (Namata et al., 2012), and **ogb-arXiv** Hu et al. (2020)), three heterophilic graphs (**Texas**, **Wisconsin**, and **Cornell** from the WebKB dataset (García-Plaza et al., 2016)). An overview of the statistical information for the six homophilic graphs and three heterophilic graphs is provided in Table 4 of Appendix B. The benchmarks used for evaluation cover a wide range of *homophily level* (Pei et al., 2020). A low homophily level indicates that the dataset leans more towards being heterophilic, where most neighbors do not share the same class as the central node. Conversely, a high homophily level signifies that the dataset tends towards homophily, with similar nodes more likely to be interconnected.

Result Analysis Tables 1-2 present the average accuracy for predicting node labels in homophilic and heterophilic graphs, respectively. ODNet consistently ranks among the top-performing methods with minimal variance. The performance results for baseline methods are sourced from prior studies (Chamberlain et al., 2021; Chien et al., 2021; Wang et al., 2022). Notably, our model outperforms other continuous message passing techniques, such as GRAND, by introducing the potential term and the bounded confidence mechanism. This superiority is particularly evident on heterophilic graphs, where the repulsive force among dissimilar node pairs significantly enhances prediction accuracy. Furthermore, Figure 2 illustrates the distinct preferences of the influence function for homophilic and heterophilic graphs. Based on the theoretical analysis and empirical evidence, we recommend following (8) for the former and (10) for the latter in general. The similarity cutoff also exhibits differing preferences. In homophilic graphs, nodes tend to amplify attraction among similar entities, while in heterophilic graphs, dissimilar nodes benefit more from emphasizing repulsion. Additional evidence is provided in Section 6.5. Furthermore, the introduced bounded confidence in ODNet, similar to an activation function, does not increase the computational cost. For instance, training

Table 2: Average test accuracy on **heterophilic** graphs over 10 fixed split datasets.

	Texas	**Wisconsin**	**Cornell**
homophily level	0.11	0.21	0.30
MLP	80.8±4.8	85.3±3.3	81.9±6.4
GCN	55.1±5.2	51.8±3.1	60.5±5.3
GAT	52.2±6.6	49.4±4.1	61.9±5.1
GPRGNN	78.4±4.4	82.9±4.2	80.3±8.1
H2GCN	84.9±7.2	87.7± 5.0	82.7±5.3
GCNII	77.6±3.8	80.4±3.4	77.9±3.8
Geom-GCN	66.8±2.7	64.5±3.7	60.5±3.7
PairNorm	60.3±4.3	48.4±6.1	58.9±3.2
GraphSAGE	82.4±6.1	81.2±5.6	76.0±5.0
GraphCON	85.4±4.2	87.8±3.3	84.3±4.8
ODNet	**88.3±3.2**	**89.1± 2.9**	**86.5±5.5**

Table 3: Average test accuracy on **hypergraphs** over 10 random splits.

	Cora-coauthor	**Cora**-cocitation	**CiteSeer**-cocitation	**PubMed**-cocitation
HGNN	82.6±1.7	79.4±1.4	72.5±1.2	86.4±0.4
HyperGCN	79.5±2.1	78.5±1.3	71.3±0.8	82.8±8.7
HCHA	82.6±1.0	79.1±1.0	72.4±1.4	86.4±0.4
HNHN	77.2±1.5	76.4±1.9	72.6±1.6	86.9±0.3
UniGCNII	83.6±1.1	78.8±1.1	73.0±2.2	88.3±0.4
HyperND	80.6±1.3	79.2±1.1	72.6±1.5	86.7±0.4
AllDeepSets	82.0±1.5	76.9±1.8	70.8±1.6	88.8±0.3
AllSetTrans	83.6±1.5	78.6±1.5	73.1±1.2	88.7±0.4
SheafHGCN	83.2±1.2	80.1±1.1	73.2±0.5	87.1±0.7
ED-HNN	84.0±1.6	80.3±1.4	73.7±1.4	89.0±0.5
ODNet (Ours)	**84.5±1.6**	**80.7±0.9**	**74.0±0.9**	**89.0±0.4**

one round of **CORA** (on a single A100 card) requires approximately 120 seconds for ODNet, 135 seconds for GRAND, and 164 seconds for SINN.

6.3 Node Classification on Hypergrpahs

Benchmark Datasets The hypergraph variant of ODNet undergoes an evaluation through semi-supervised node classification tasks conducted on four benchmark hypergraphs extracted from Yadati et al. (2019). For co-citation networks (**Cora-cocitation**, **CiteSeer-cocitation**, and **PubMed-cocitation**), documents cited by a given document are interconnected by a hyperedge. Similarly, the co-authorship networks (**Cora-coauthor**) aggregates all documents co-authored by an individual into a single hyperedge. A statistical summary of the four benchmarks is provided in Table 5 in Appendix B.

Result Analysis In contrast to graph data with relatively sparse connections, hypergraphs utilize a few hyperedges and establish densely connected local communities. As reported in Table 3, ODNet consistently outperforms baseline methods with a significant improvement. It is worth noting that our ODNet adopts the hypergraph weights a_{ij}^{e} from HGNN (Feng et al., 2019) with a simple Euler scheme of first-order forward differences. In this sense, our method demonstrates great potential for significantly enhancing the performance of a basic method with minimal additional complexity, surpassing even more advanced methods.

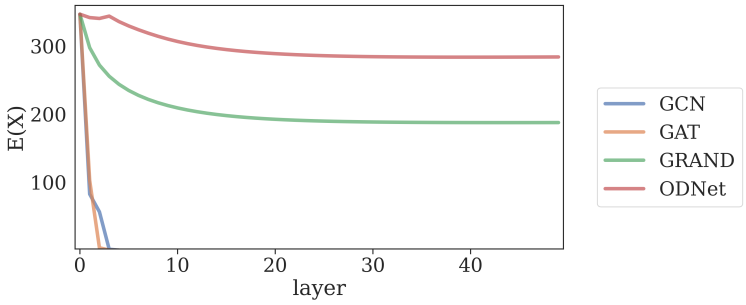

Figure 3: Decays of Dirichlet energy with layers on **Texas**.

6.4 Dirichlet Energy and Community Consensus

Many message passing methods encounter the issue of oversmoothing, limiting their ability to enable deep networks to achieve expressive propagation. A GNN model is considered to alleviate the oversmoothing problem if its Dirichlet energy rapidly approaches a lower bound as the number of network layers increases (Cai & Wang, 2020). Figure 3 illustrates the decay of Dirichlet energy on **Texas** with all network parameters randomly initialized. The two conventional message passings, GCN and GAT, exhibit a sudden progression of Dirichlet energy with exponential decay. In contrast, GRAND employs a small multiplier to delay all nodes' features to collapse to the same value. ODNet stabilizes the energy decay with bounded confidence and the influence weights, offering a simple and efficient solution to alleviate the oversmoothing issue. Since stable Dirichlet energy reflects the disparity of feature clusters, the observation that a decreasing profile of ϕ reduces Dirichlet energy under stable conditions is consistent with simulation results in opinion dynamics that heterophily dynamics enhances consensus (Motsch & Tadmor, 2014). Note that the 'heterophily' in opinion dynamics refers to the tendency of a graph to form stronger connections with those who are different rather than those who are similar, which is a different term from the 'heterophilic graph' in GNNs.

6.5 Additional Investigation

Neural ODE Solvers The Dormand–Prince adaptive step size scheme (DOPRI5) served as the neural ODE solver for ODNet. Additionally, we evaluated the performance of two other solvers across various datasets: the Runge-Kutta method (rk4) and the first-order Euler scheme (Euler). The outcomes are presented in Figure 4. Although different solvers did not consistently demonstrate a significant and sustained advantage of one over the others, our selection of DOPRI5 yielded the overall best results.

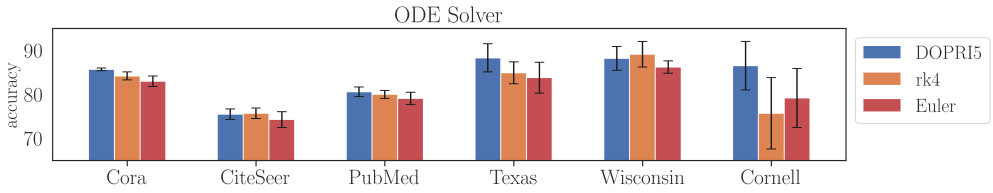

Figure 4: Prediction performance with different neural ODE solvers.

Embedding Dynamics We employ the t-SNE algorithm to visualize the embedded features in a two-dimensional plane for the **Texas** dataset. We choose the output embeddings from the last layer at different epochs. Figure 5 demonstrates an evident clustering trend as the number of training epochs increases. By the 50th epoch, nodes with different labels are distinctly separable in the reduced two-dimensional space.

Influence Function We next delve into the impact of different selections of the influence function ϕ on the performance of ODNet. This investigation encompasses various aspects, including the choice of scaling factors (μ and ν) and the definition of the similarity cutoffs (ϵ_1 and ϵ_2). Our findings are meticulously detailed

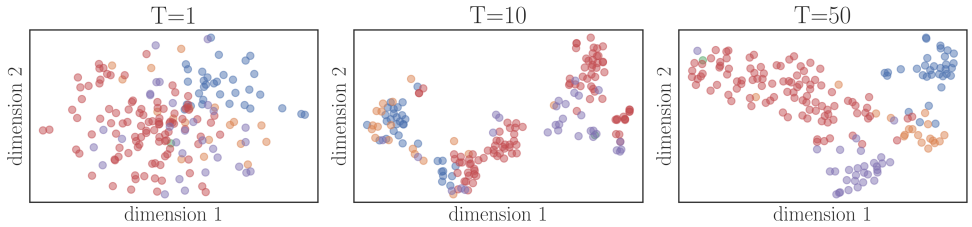

Figure 5: t-SNE visualization on node embeddings for **Texas**.

in Tables 8-9 in Appendix B, with a particular focus on the homophilic graph (**Cora**) and the heterophilic graph (**Texas**), respectively. An interesting trend emerges concerning the parameter ν, indicating a clear preference. Specifically, it is advisable to incorporate a repulsive effect on heterophilic graphs by assigning a negative value to ν. Conversely, for homophilic graphs, where similarity plays a pivotal role, setting $\nu = 0$ is more appropriate. To provide a direct comparison, Figure 6 showcases ODNet's performance under different similarity cutoffs, *i.e.*, ϵ_1 and ϵ_2. For **Cora**, we maintain $\mu = 1.4$ and $\nu = 0$, while for **Texas**, we set $\mu = 1.0$ and $\nu = -50$. Generally, in the context of homophilic graphs, setting a relatively small value for ϵ_2 tends to expand the region of nodes considered similar. This approach may be beneficial in mitigating the oversmoothing issue. Conversely, for heterophilic graphs, it is advisable to reverse information from only highly similar nodes, reflected in the choice of a larger ϵ_2 value (up to 0.8). However, it is crucial to exercise caution when pushing ϵ_2 towards 1.0, as a discernible reduction in performance becomes evident.

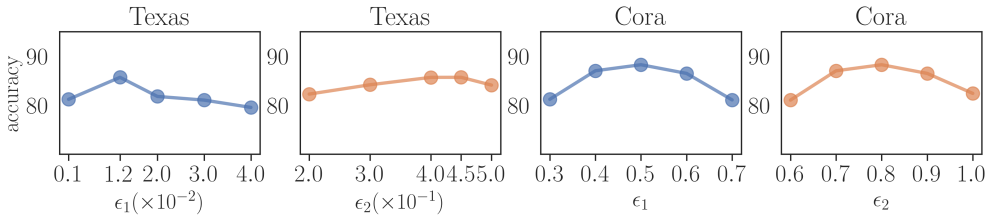

Figure 6: The impact of different ϵ_1 and ϵ_2 on ODNet.

6.6 Case Study: Biological Social Network Architecture Simplification

Microorganisms are the most extensively distributed and numerous group on Earth, which thrive in a wide array of moderate and extreme environments, such as deep-sea hydrothermal vents, ocean trenches, and plateaus (Shu & Huang, 2022). They present the greatest diversity of life on Earth. The remarkable diversity among microorganisms finds its primary expression through their intricate metabolic pathways (Louca et al., 2018; Coelho et al., 2022). Consequently, investigating the connections between microbial metabolism in distinct environments carries profound significance in unraveling the intricate interplay between Earth's diverse ecosystems and the lives inhabiting them. Metagenomic analyses have revealed the remarkable complexity inherent to metabolic gene networks, due to the diversity and richness of functional genes and their interconnections. It thus becomes a necessity to simplify metabolic gene networks for investigating relationships among functional genes and key genes. Presently, the prevailing approach involves adjusting connection weights to streamline the network, often relying on biological expertise. However, the absence of a standardized simplification criterion results in a heavy bias in network structures influenced by the subjective opinions of biologists.

Problem Formulation and Training Setup As an example of the environmental microbiome analysis, the co-occurrence network is challenging to interpret due to the massive and complicated characteristics of genes and the unclear standard for assessment. The target here is to learn meaningful influence weights between gene pairs that simplify the co-occurrence network with effective biological justification. To this end, two networks originated from the microbial comparison between the *Mariana Trench* (MT) and *Mount*

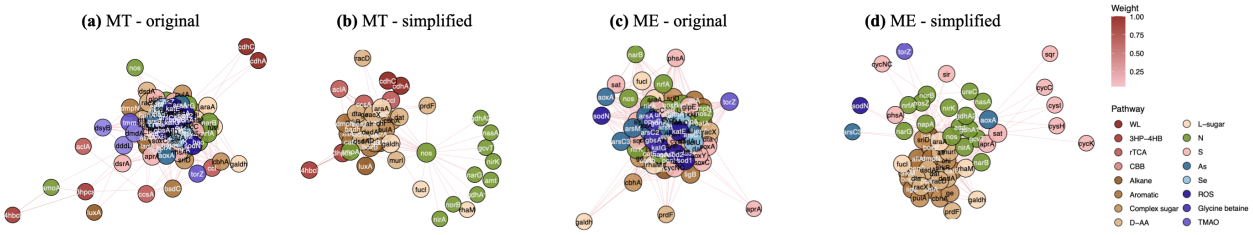

Figure 7: Co-occurrence network of selected metabolic genes in ME or MT before (a-b) and after (c-d) being simplified by ODNet. Connections are considered strong with weights> 0.05, and genes without a strong connection with any other peers are removed.

Everest (ME) networks (Liu et al., 2022) are utilized, where nodes are functional genes and edges are weighted by the probability of two key functional gene sets simultaneously occurring in the same species. Edges with exceptionally small weights will be discarded as noisy observations. As we are eager to identify the key genes and gene clusters from gene interactions, we construct the graph with initial connectivity (edges and edge weights), leaving any node attributes (*e.g.*, function and pathway) unobserved. We define a node-level classification task for predicting whether a node is a 'strong', 'medium', or 'weak' influencer to its community, where the three levels are cut by their degree. Further details are provided in Appendix D.

Result Analysis We trained two independent ODNets on ME and MT networks, which achieved prediction accuracy as high as 96.9% and 75.0%, respectively. For both networks, metabolic genes were classified based solely on topological information, without the introduction of any a priori node features. Figure 7 visualizes the two networks in their original and the simplified appearance, respectively. For all the networks, an edge weight cutoff of 0.05 was applied to eliminate weak connections that could not be distinguishable from background noise. The original network without any simplification appeared cluttered and difficult to interpret (Figure 7a-b). In contrast, the simplified networks greatly enhanced the readability of the co-occurrence network while retaining reasonable biological significance (Figure 7c-d). Furthermore, the simplified network was able to identify the biologically key genes that acted as "opinion leaders", serving as bridges connecting different metabolic pathways. For example, in the MT network, the key gene nitrous oxide reductase (nos) bridged the carbon (Alkane, Aromatic, Complex sugar, D-AA and L-sugar) and nitrogen metabolism (N) in Figure 7c, whereas in the ME network, the key genes of sulfate reduction (sat and aprA) coupled the carbon and sulfur metabolism (S) (Figure 7d). Thus, ODNet could be employed to present more discernible networks in environmental microbiome studies, and aid in comprehending key metabolic functions within microbiomes from diverse environments.

7 Conclusion

This study establishes intriguing connections between sociodynamics and graph neural networks, two distinct fields that both actively investigate social networks from different perspectives. By bridging concepts from these two fields, we introduce bounded confidence for neural message passing, a novel mechanism inspired by opinion dynamics. The proposed ODNet effectively addresses the oversmoothing issue and consistently achieves top-notch performance in node prediction tasks across graphs with diverse local connectivity patterns. This success is attributed to the simplicity and efficacy of our piecewise message propagation rule.

In addition to providing strong theoretical support and empirical performance, ODNet integrates theories from social networks, offering intuitive explanations for hyperparameter settings and the significance of network parameter learning. This integration showcases ODNet's significant potential in simplifying complex real-world social networks, presenting a fresh analytical approach that does not rely on existing attributive classification conventions. The robust performance of ODNet extends its applicability to simplifying intricate networks containing a wealth of biological information, such as genes, gene-gene interactions, and metabolic pathways. This method's exceptional capacity to extract accurate insights and unveil the intrinsic mechanisms of cellular physiology provides invaluable support to biologists in deciphering the mechanisms

of adaptation and pathway functions in the microbial realm. These findings are significant in understanding the interactions between Earth's environments and the metabolism of life.

References

Daron Acemoglu and Asuman Ozdaglar. Opinion dynamics and learning in social networks. *Dynamic Games and Applications*, 1:3–49, 2011.

Song Bai, Feihu Zhang, and Philip HS Torr. Hypergraph convolution and hypergraph attention. *Pattern Recognition*, 110:107637, 2021.

Albert-Laszlo Barabasi and Zoltan N Oltvai. Network biology: understanding the cell's functional organization. *Nature reviews genetics*, 5(2):101–113, 2004.

Ayodeji Bello, Yue Han, Haifeng Zhu, Liting Deng, Wei Yang, Qingxin Meng, Yu Sun, Ugochi Uzoamaka Egbeagu, Siyuan Sheng, Xiaotong Wu, et al. Microbial community composition, co-occurrence network pattern and nitrogen transformation genera response to biochar addition in cattle manure-maize straw composting. *Science of the Total Environment*, 721:137759, 2020.

Cristian Bodnar, Francesco Di Giovanni, Benjamin Chamberlain, Pietro Lio, and Michael Bronstein. Neural sheaf diffusion: A topological perspective on heterophily and oversmoothing in gnns. *Advances in Neural Information Processing Systems*, 35:18527–18541, 2022.

Johannes Brandstetter, Rob Hesselink, Elise van der Pol, Erik J Bekkers, and Max Welling. Geometric and physical quantities improve e (3) equivariant message passing. In *ICLR*, 2021.

Michael M Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. Geometric deep learning: going beyond Euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42, 2017.

Francesco Bullo, Jorge Cortés, and Sonia Martinez. *Distributed control of robotic networks: a mathematical approach to motion coordination algorithms*, volume 27. Princeton University Press, 2009.

Chen Cai and Yusu Wang. A note on over-smoothing for graph neural networks. *arXiv:2006.13318*, 2020.

Ferran Hernandez Caralt, Guillermo Bernardez, Iulia Duta, Eduard Alarcon, and Pietro Lio. Joint diffusion processes as an inductive bias in sheaf neural networks. In *ICML 2024 Workshop on Geometry-grounded Representation Learning and Generative Modeling*.

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.

Benjamin Paul Chamberlain, James Rowbottom, Maria I. Gorinova, Stefan D Webb, Emanuele Rossi, and Michael M. Bronstein. GRAND: Graph neural diffusion. In *ICML*, 2021.

Ming Chen, Zhewei Wei, Zengfeng Huang, Bolin Ding, and Yaliang Li. Simple and deep graph convolutional networks. In *ICML*, pp. 1725–1735. PMLR, 2020.

Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. In *NeurIPS*, volume 31, 2018.

Eli Chien, Jianhao Peng, Pan Li, and Olgica Milenkovic. Adaptive universal generalized pagerank graph neural network. In *ICLR*, 2021.

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

Luis Pedro Coelho, Renato Alves, Álvaro Rodríguez Del Río, Pernille Neve Myers, Carlos P Cantalapiedra, Joaquín Giner-Lamia, Thomas Sebastian Schmidt, Daniel R Mende, Askarbek Orakov, Ivica Letunic, et al. Towards the biogeography of prokaryotic genes. *Nature*, 601(7892):252–256, 2022.

Felipe Cucker and Steve Smale. Emergent behavior in flocks. *IEEE Transactions on Automatic Control*, 52 (5):852–862, 2007.

Morris H. DeGroot. Reaching a consensus. *Journal of the American Statistical Association*, 69(345):118–121, 1974. ISSN 01621459.

Yihe Dong, Will Sawin, and Yoshua Bengio. Hnhn: Hypergraph networks with hyperedge neurons. *arXiv:2006.12278*, 2020.

Iulia Duta, Giulia Cassarà, Fabrizio Silvestri, and Pietro Liò. Sheaf hypergraph networks. *Advances in Neural Information Processing Systems*, 36, 2024.

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.

Yifan Feng, Haoxuan You, Zizhao Zhang, Rongrong Ji, and Yue Gao. Hypergraph neural networks. In *AAAI*, volume 33, pp. 3558–3565, 2019.

Matthias Fey and Jan Eric Lenssen. Fast graph representation learning with PyTorch Geometric. In *ICLR Workshop on Representation Learning on Graphs and Manifolds*, 2019.

John RP French Jr. A formal theory of social power. *Psychological review*, 63(3):181, 1956.

Yue Gao, Yifan Feng, Shuyi Ji, and Rongrong Ji. Hgnn $^+$: General hypergraph neural networks. *IEEE TPAMI*, 2022a.

Yue Gao, Zizhao Zhang, Haojie Lin, Xibin Zhao, Shaoyi Du, and Changqing Zou. Hypergraph Learning: Methods and Practices. *IEEE TPAMI*, 44(5), May 2022b. ISSN 1939-3539.

Alberto P García-Plaza, Víctor Fresno, Raquel Martínez Unanue, and Arkaitz Zubiaga. Using fuzzy logic to leverage html markup for web page representation. *IEEE Transactions on Fuzzy Systems*, 25(4):919–933, 2016.

Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural message passing for quantum chemistry. In *ICML*, 2017.

Jesús Giráldez-Cru, Carmen Zarco, and Oscar Cordón. Analyzing the extremization of opinions in a general framework of bounded confidence and repulsion. *Information Sciences*, 609:1256–1270, 2022.

Michelle Girvan and Mark EJ Newman. Community structure in social and biological networks. *PNAS*, 99 (12):7821–7826, 2002.

Chris Godsil and Gordon F Royle. *Algebraic graph theory*, volume 207. Springer Science & Business Media, 2001.

William L Hamilton, Rex Ying, and Jure Leskovec. Inductive representation learning on large graphs. In *NeurIPS*, 2017.

Andi Han, Dai Shi, Lequan Lin, and Junbin Gao. From continuous dynamics to graph neural networks: Neural diffusion and beyond. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856. URL `https://openreview.net/forum?id=fPQSxjqa2o`. Survey Certification.

Jakob Hansen and Robert Ghrist. Opinion dynamics on discourse sheaves. *SIAM Journal on Applied Mathematics*, 81(5):2033–2060, 2021.

Dirk Helbing. *Quantitative sociodynamics: stochastic methods and models of social interaction processes*. Springer Science & Business Media, 2010.

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.

Weihua Hu, Matthias Fey, Marinka Zitnik, Yuxiao Dong, Hongyu Ren, Bowen Liu, Michele Catasta, and Jure Leskovec. Open graph benchmark: Datasets for machine learning on graphs. *Advances in neural information processing systems*, 33:22118–22133, 2020.

Jing Huang and Jie Yang. Unignn: a unified framework for graph and hypergraph neural networks. *arXiv:2105.00956*, 2021.

Shi Jin and Ruiwen Shu. Collective dynamics of opposing groups with stochastic communication. *Vietnam Journal of Mathematics*, 49:619–636, 2021a.

Shi Jin and Ruiwen Shu. Collective Dynamics of Opposing Groups with Stochastic Communication. *Vietnam Journal of Mathematics*, 49(3):619–636, September 2021b. ISSN 2305-221X, 2305-2228. doi: 10.1007/s10013-020-00430-2.

Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *ICLR*, 2017.

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.

Theodore Kolokolnikov, José A Carrillo, Andrea Bertozzi, Razvan Fetecau, and Mark Lewis. Emergent behaviour in multi-particle systems with non-local interactions, 2013.

Roie Levy and Elhanan Borenstein. Metabolic modeling of species interaction in the human microbiome elucidates community-level assembly rules. *PNAS*, 110(31):12804–12809, 2013.

Yongqin Liu, Zhihao Zhang, Mukan Ji, Aoran Hu, Jing Wang, Hongmei Jing, Keshao Liu, Xiang Xiao, and Weishu Zhao. Comparison of prokaryotes between mount everest and the mariana trench. *Microbiome*, 10(1):215, 2022.

Stilianos Louca, Martin F Polz, Florent Mazel, Michaeline BN Albright, Julie A Huber, Mary I O'Connor, Martin Ackermann, Aria S Hahn, Diane S Srivastava, Sean A Crowe, et al. Function and functional redundancy in microbial systems. *Nature ecology & evolution*, 2(6):936–943, 2018.

Andrew Kachites McCallum, Kamal Nigam, Jason Rennie, and Kristie Seymore. Automating the construction of internet portals with machine learning. *Information Retrieval*, 3(2):127–163, 2000.

Federico Monti, Davide Boscaini, Jonathan Masci, Emanuele Rodola, Jan Svoboda, and Michael M Bronstein. Geometric deep learning on graphs and manifolds using mixture model CNNs. In *CVPR*, pp. 5115–5124, 2017.

Sebastien Motsch and Eitan Tadmor. Heterophilious dynamics enhances consensus. *SIAM Review*, 56(4): 577–621, 2014.

Galileo Namata, Ben London, Lise Getoor, Bert Huang, and U Edu. Query-driven active surveying for collective classification. In *10th International Workshop on Mining and Learning with Graphs*, volume 8, pp. 1, 2012.

Hoang Nt and Takanori Maehara. Revisiting graph neural networks: All we have is low-pass filters. *arXiv:1905.09550*, 2019.

Maya Okawa and Tomoharu Iwata. Predicting opinion dynamics via sociologically-informed neural networks. In *ACM SIGKDD*, pp. 1306–1316, 2022.

Kenta Oono and Taiji Suzuki. Graph neural networks exponentially lose expressive power for node classification. In *ICLR*, 2019.

Hongbin Pei, Bingzhe Wei, Kevin Chen-Chuan Chang, Yu Lei, and Bo Yang. Geom-GCN: geometric graph convolutional networks. In *ICLR*, 2020.

Michael Poli, Stefano Massaroli, Junyoung Park, Atsushi Yamashita, Hajime Asama, and Jinkyoo Park. Graph neural ordinary differential equations. In *AAAI Workshop on Deep Learning on Graphs: Methodologies and Applications*, 2020.

Anton V Proskurnikov and Roberto Tempo. A tutorial on modeling and analysis of dynamic social networks. part i. *Annual Reviews in Control*, 43:65–79, 2017.

Hegselmann Rainer and Ulrich Krause. Opinion dynamics and bounded confidence: models, analysis and simulation. 2002.

T Konstantin Rusch, Ben Chamberlain, James Rowbottom, Siddhartha Mishra, and Michael Bronstein. Graph-coupled oscillator networks. In *ICML*, pp. 18888–18909. PMLR, 2022.

Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner, and Gabriele Monfardini. The graph neural network model. *IEEE transactions on neural networks*, 20(1):61–80, 2008.

Prithviraj Sen, Galileo Namata, Mustafa Bilgic, Lise Getoor, Brian Galligher, and Tina Eliassi-Rad. Collective classification in network data. *AI Magazine*, 29(3):93–93, 2008.

Oleksandr Shchur, Maximilian Mumme, Aleksandar Bojchevski, and Stephan Günnemann. Pitfalls of graph neural network evaluation. *arXiv:1811.05868*, 2018.

Wen-Sheng Shu and Li-Nan Huang. Microbial diversity in extreme environments. *Nature Reviews Microbiology*, 20(4):219–235, 2022.

Francesco Tudisco, Austin R Benson, and Konstantin Prokopchik. Nonlinear higher-order label spreading. In *WWW*, pp. 2402–2413, 2021.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. In *ICLR*, 2018.

Peihao Wang, Shenghao Yang, Yunyu Liu, Zhangyang Wang, and Pan Li. Equivariant hypergraph diffusion neural operators. *arXiv:2207.06680*, 2022.

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.

Wolfgang Weidlich. *Sociodynamics: A systematic approach to mathematical modelling in the social sciences*. Courier Corporation, 2006.

Louis-Pascal Xhonneux, Meng Qu, and Jian Tang. Continuous graph neural networks. In *ICML*, pp. 10432–10441, 2020.

Naganand Yadati, Madhav Nimishakavi, Prateek Yadav, Vikram Nitin, Anand Louis, and Partha Talukdar. Hypergcn: A new method for training graph convolutional networks on hypergraphs. *NeurIPS*, 32, 2019.

Mengbin Ye. *Opinion dynamics and the evolution of social power in social networks*. Springer, 2019.

Olga Zaghen, Antonio Longa, Steve Azzolin, Lev Telyatnikov, Andrea Passerini, and Pietro Lio. Sheaf diffusion goes nonlinear: Enhancing gnns with adaptive sheaf laplacians. In *ICML 2024 Workshop on Geometry-grounded Representation Learning and Generative Modeling*, 2024.

Lingxiao Zhao and Leman Akoglu. PairNorm: Tackling oversmoothing in GNNs. In *ICLR*, 2020.

Jiong Zhu, Yujun Yan, Lingxiao Zhao, Mark Heimann, Leman Akoglu, and Danai Koutra. Beyond homophily in graph neural networks: Current limitations and effective designs. In *NeurIPS*, volume 33, pp. 7793–7804, 2020.

A Theoretical results and proof

In this section, we prove that for our proposed ODNet on hypergraph convolution, the Dirichlet energy does not decay to zero. In other words, the proposed propagation rule mitigates the oversmoothing issue.

We set $\{u_i\} := \{x_i | i \in \mathcal{I}_1\}$ and $\{v_j\} := \{x_j | j \in \mathcal{I}_2\}$. Then we can rewrite (15) as

$$\begin{cases} \dot{u}_i = \dfrac{1}{N_1}\sum\limits_{i'=1}^{N_1} a_{i,j}^{e,+}(u_{i'} - u_i) - \dfrac{1}{N_2}\sum\limits_{j=1}^{N_2} a_{i,j}^{e,-}(v_j - u_i) + \delta u_i(1-(u_i)^2) \\ \dot{v}_j = \dfrac{1}{N_2}\sum\limits_{j'=1}^{N_2} a_{i,j}^{e,+}(v_{j'} - v_i) - \dfrac{1}{N_1}\sum\limits_{i=1}^{N_1} a_{i,j}^{e,-}(u_i - v_j) + \delta v_j^{(}1-(v_j)^2) \end{cases} \tag{16}$$

In the following proof, we will use x to replace $\mathbf{x}$ by x to consider channel-wise variables. For clarity, we define the following notations:

- $M_2(U) := \sum\limits_{i=1}^{N_1} u_i^2, \quad M_2(V) := \sum\limits_{j=1}^{N_2} v_j^2;$
- $\widehat{M}_2 := M_2(\hat{U}) + M_2(\hat{V}) = \mathrm{var}(u) + \mathrm{var}(v);$
- $\psi_i^{e,\pm} := \sum\limits_{j\in e} a_{i,j}^{e,\pm}$, $\psi_i^{\pm} := \sum\limits_{e\in\mathcal{E}}\sum\limits_{j\in e} a_{i,j}^{e,\pm};$
- $k := \max\limits_{k}\{|\mathcal{E}(k)|\};$
- $D_e^- := \max\limits_{k}\{\psi_k^{e,-}\}$, $D^- := \max\limits_{e}\{D_e^-\};$
- $D_2^- := \max\limits_{e}\{\|\psi^{e,-}\|_2\}.$

We assume $N_1 = N_2 := N_0$, which implies N_1 is comparable to N_2, *i.e.*, there exists a positive constant κ satisfying $\frac{1}{\kappa}N_1 \le N_2 \le \kappa N_1$.

Lemma 2 (L_2 estimate for M_2)**.** *There exists a positive constant M_2^∞ such that*

$$\sup_{0\le t<\infty} \le M_2^\infty \le \infty. \tag{17}$$

Proof. We start from the derivatives of $dtM_2(U)$ and $dtM_2(V)$. For the former,

$$\begin{aligned} \frac{d}{dt}M_2(U) &= \frac{2}{N_1}\sum_{i=1}^{N_1} u_i\dot{u}_i \\ &= \frac{2}{N_1}\sum_{i=1}^{N_1}\sum_{e\in\mathcal{E}(i)}\sum_{i'\in e} a_{i,i'}^{e,+}(u_{i'}-u_i)^2 - \frac{2}{N_1}\sum_{i=1}^{N_1}\sum_{e\in\mathcal{E}(i)}\sum_{j\in e} a_{i,j}^{e,-}(v_j-u_i)u_i + \frac{2\delta}{N_1}\sum_{i=1}^{N_1}(u_i)^2(1-(u_i)^2) \\ &= \frac{2}{N_1}\sum_{e\in\mathcal{E}}\sum_{i,i'\in e} a_{i,i'}^{e,+}(u_{i'}-u_i)u_i - \frac{2}{N_1}\sum_{e\in\mathcal{E}}\sum_{i,j\in e} a_{i,j}^{e,-}(v_j-u_i)u_i + \frac{2\delta}{N_1}\sum_{i=1}^{N_1}(u_i)^2(1-(u_i)^2) \\ &= -\frac{1}{N_1}\sum_{e\in\mathcal{E}}\sum_{i,,i'\in e}^{N_1} a_{i,i'}^{e,+}(u_{i'}-u_i)^2 - \frac{2}{N_1}\sum_{e\in\mathcal{E}}\sum_{i,j\in e} a_{i,j}^{e,-}(v_j-u_i)u_i + \frac{2\delta}{N_1}\sum_{i=1}^{N_1}(u_i)^2(1-(u_i)^2). \end{aligned} \tag{18}$$

Similarly,

$$\begin{aligned}\frac{d}{dt}M_2(W) &= \frac{2}{N_2}\sum_{j=1}^{N_2} v_j \dot{v_j} \\ &= \frac{1}{N_2}\sum_{e\in\mathcal{E}}\sum_{j,,j'\in e}^{N_1} a_{i,i'}^{e,+}(v_{j'}-v_j)^2 - \frac{2}{N_2}\sum_{e\in\mathcal{E}}\sum_{j,i\in e} a_{j,i}^{e,-}(u_i-v_j)v_j + \frac{2\delta}{N_3}\sum_{j=1}^{N_2}(v_j)^2(1-(v_j)^2).\end{aligned} \tag{19}$$

Note that $a_{i,j}^{e,\pm} = a_{j,i}^{e,\pm}$. Sum the $M_2(U)$ and $M_2(V)$, then

$$\begin{aligned}\frac{d}{dt}M_2 \leq &-\frac{2}{N_1}\sum_{e\in\mathcal{E}}\sum_{i,j\in e} a_{i,j}^{e,-}(v_j-u_i)u_i - \frac{2}{N_2}\sum_{e\in\mathcal{E}}\sum_{j,i\in e} a_{j,i}^{e,-}(u_i-v_j)v_j \\ &+ \frac{2\delta}{N_1}\sum_{i=1}^{N_1}(u_i)^2(1-(u_i)^2) + \frac{2\delta}{N_2}\sum_{i=1}^{N_2}(v_i)^2(1-(v_i)^2).\end{aligned} \tag{20}$$

By the Cauchy-Schwarz inequality,

$$\begin{aligned}&\left(\sum_{i=1}^{N_1}(u_i)^2\right)^2 \leq N_1\sum_{i=1}^{N_1}(u_i)^4, \quad \left(\sum_{i=1}^{N_1}(u_i)^2\right)^2 \leq N_2\sum_{i=1}^{N_2}(v_i)^4, \\ &(u_i-v_j)^2 \leq 2((u_i)^2+(v_j)^2).\end{aligned}$$

Then we have

$$\begin{aligned}\frac{d}{dt}M_2 \leq &\frac{D^-}{N_1}\sum_{e\in\mathcal{E}}\sum_{i,j\in e}\left((v_j-u_i)^2+u_i^2\right) + \frac{D^-}{N_2}\sum_{e\in\mathcal{E}}\sum_{j,i\in e}\left((u_i-v_j)^2+v_j^2\right) \\ &+ \frac{2\delta}{N_1}\sum_{i=1}^{N_1}(u_i)^2(1-(u_i)^2) + \frac{2\delta}{N_2}\sum_{i=1}^{N_2}(v_i)^2(1-(v_i)^2) \\ \leq &\frac{D^-}{N_1}\sum_{e\in\mathcal{E}}\sum_{i,j\in e}\left(3u_i^2+2v_j^2\right) + \frac{D^-}{N_2}\sum_{e\in\mathcal{E}}\sum_{i,j\in e}\left(2u_i^2+3v_j^2\right) + 2\delta M_2 - \delta M_2^2.\end{aligned} \tag{21}$$

These relations yield a Riccati-type differential inequality:

$$\begin{aligned}\frac{d}{dt}M_2 \leq &5D^-\max\{\frac{1}{N_1},\frac{1}{N_2}\}\sum_{e\in\mathcal{E}}\sum_{i,j\in e}\left(u_i^2+v_j^2\right) + 2\delta M_2 - \delta(M_2)^2 \\ \leq &5D^-k\max\{\frac{1}{N_1},\frac{1}{N_2}\}M_2 + 2\delta M_2 - \delta(M_2)^2 \\ \leq &(C_1+2\delta)M_2 - \delta(M_2)^2,\end{aligned} \tag{22}$$

where

$$C_1 := 5D^-k\max\{\frac{1}{N_1},\frac{1}{N_2}\}. \tag{23}$$

Let y be a solution of the following ODE:

$$y' = \alpha C_1 y - \delta y^2. \tag{24}$$

Then, by phase line analysis, the solution $y(t)$ to (24) satisfies

$$M_2(t) \leq y(t) \leq \max\left\{\frac{C_1}{\delta}+2, M_2(0)\right\} =: M_2^\infty. \tag{25}$$

□

Lemma 3. *Let u, v be the solution to (16). Then $|\bar{u}-\bar{v}|^2$ satisfies*

$$\frac{1}{2}\frac{d}{dt}|\bar{u}-\bar{v}|^2 \geq \left(\frac{2c_m}{N_0}-c_1\right)|\bar{u}-\bar{v}|^2 - \frac{4(D_2^-)^2}{c_1 N_0}k\widehat{M}_2. \tag{26}$$

Proof. The time evolution of $\bar{u}$ is given by

$$\begin{aligned}
\dot{\bar{u}} &= \frac{1}{N_1}\sum_{i=1}^{N_1}\sum_{i'\in\mathcal{E}(i)} a_{i,i'}^{e,+}(u_{i'}-u_i) - \frac{1}{N_1}\sum_{i=1}^{N_1}\sum_{j\in\mathcal{E}(i)} a_{j,i}^{e,-}(v_j-u_i) \\
&= -\frac{1}{N_1}\sum_{e\in\mathcal{E}}\sum_{i,j\in e} a_{i,j}^{e,-}(v_j-u_i) \\
&= -\frac{1}{N_1}\sum_{e\in\mathcal{E}}\psi_j^{e,-}v_j + \frac{1}{N_1}\sum_{e\in\mathcal{E}}\psi_i^{e,-}u_i \\
&= \sum_{e\in\mathcal{E}}\left(-\frac{1}{N_1}\psi_j^{e,-}v_j + \frac{1}{N_1}\psi_i^{e,-}u_i\right).
\end{aligned} \tag{27}$$

The first equality uses the relation $\sum_{i=1}^{N_1}\hat{u}_i = 0$. Then we have

$$\begin{aligned}
\frac{1}{2}\frac{d}{dt}|\bar{u}-\bar{v}|^2 &= (\bar{u}-\bar{v})(\dot{\bar{u}}-\dot{\bar{v}}) \\
&= (\bar{u}-\bar{v})\left[\sum_{e\in\mathcal{E}}\left(-\frac{1}{N_1}\psi_j^{e,-}v_j + \frac{1}{N_1}\psi_i^{e,-}u_i\right) - \sum_{e\in\mathcal{E}}\left(-\frac{1}{N_2}\psi_i^{e,-}u_i + \frac{1}{N_2}\psi_j^{e,-}v_j\right)\right] \\
&= (\bar{u}-\bar{v})\sum_{e\in\mathcal{E}}\left(-\frac{1}{N_1}\psi_j^{e,-}v_j - \frac{1}{N_2}\psi_j^{e,-}v_j + \frac{1}{N_1}\psi_i^{e,-}u_i + \frac{1}{N_2}\psi_i^{e,-}u_i\right) \\
&= (\bar{u}-\bar{v})\sum_{e\in\mathcal{E}}\left(-\frac{1}{N_1}\psi_j^{e,-}(\bar{v}+\hat{v}_j) - \frac{1}{N_2}\psi_j^{e,-}(\bar{v}+\hat{v}_j) + \frac{1}{N_1}\psi_i^{e,-}(\bar{u}+\hat{u}_i) + \frac{1}{N_2}\psi_i^{e,-}(\bar{u}+\hat{u}_j)\right) \\
&= (\bar{u}-\bar{v})\left\{\left[-\left(\frac{1}{N_1}+\frac{1}{N_2}\right)\sum_{e\in\mathcal{E}}\sum_{j\in e}\psi_j^{e,-}\right]\bar{v} + \left[\left(\frac{1}{N_1}+\frac{1}{N_2}\right)\sum_{e\in\mathcal{E}}\sum_{i\in e}\psi_i^{e,-}\right]\bar{u}\right. \\
&\qquad \left.\left[-\left(\frac{1}{N_1}+\frac{1}{N_2}\right)\sum_{e\in\mathcal{E}}\sum_{j\in e}\psi_j^{e,-}\right]\hat{v}_j + \left[\left(\frac{1}{N_1}+\frac{1}{N_2}\right)\sum_{e\in\mathcal{E}}\sum_{i\in e}\psi_i^{e,-}\right]\hat{u}_i\right\} \\
&= \frac{2}{N_0}(\bar{u}-\bar{v})\left[\sum_{e\in\mathcal{E}}\sum_{i\in e}\psi_i^{e,-}\bar{u} - \sum_{e\in\mathcal{E}}\sum_{j\in e}\psi_j^{e,-}\bar{v}\right] + \frac{2}{N_0}(\bar{u}-\bar{v})\left[\sum_{e\in\mathcal{E}}\sum_{i\in e}\psi_i^{e,-}\hat{u}_i - \sum_{e\in\mathcal{E}}\sum_{j\in e}\psi_j^{e,-}\hat{v}_j\right].
\end{aligned} \tag{28}$$

We denote

$$\mathrm{Pse}(\hat{u},\hat{v}) := \sum_{e\in\mathcal{E}}\sum_{i\in e}\psi_i^{e,-}\hat{u}_i - \sum_{e\in\mathcal{E}}\sum_{j\in e}\psi_j^{e,-}\hat{v}_j$$

and

$$\mathrm{Pse}(\bar{u},\bar{v}) := \sum_{e\in\mathcal{E}}\sum_{i\in e}\psi_i^{e,-}\bar{u} - \sum_{e\in\mathcal{E}}\sum_{j\in e}\psi_j^{e,-}\bar{v}.$$

Assume there exist constants c_m, c_v, such that

$$\mathrm{Pes}(\bar{u},\bar{v}) \geq c_m(\bar{u}-\bar{v}). \tag{29}$$

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

$$\begin{aligned}
\frac{1}{2}\frac{d}{dt}|\bar{u}-\bar{v}|^2 &= \frac{2}{N_0}\mathrm{Pes}(\bar{u},\bar{v})(\bar{u}-\bar{v}) + \frac{2}{N_0}\mathrm{Pes}(\hat{u},\hat{v})(\bar{u}-\bar{v}) \geq \left(\frac{2c_m}{N_0}-c_1\right)|\bar{u}-\bar{v}|^2 + c_1|\bar{u}-\bar{v}|^2 \\
&+ \frac{2}{N_0}\mathrm{Pes}(\hat{u},\hat{v})(\bar{u}-\bar{v}) \\
&\geq \left(\frac{2c_m}{N_0}-c_1\right)|\bar{u}-\bar{v}|^2 - \frac{1}{c_1}\frac{2}{N_0}\left(\mathrm{Pes}(\hat{u},\hat{v})\right)^2 \\
&\geq \left(\frac{2c_m}{N_0}-c_1\right)|\bar{u}-\bar{v}|^2 - \frac{4}{c_1 N_0}\sum_{e\in\mathcal{E}}\left[\left(\sum_{i\in e}\psi_i^{e,-}\hat{u}_i\right)^2 + \left(\sum_{j\in e}\psi_j^{e,-}\hat{v}_j\right)^2\right] \\
&\geq \left(\frac{2c_m}{N_0}-c_1\right)|\bar{u}-\bar{v}|^2 - \frac{4}{c_1 N_0}\sum_{e\in\mathcal{E}}\left[\|\psi^{e,-}\|^2\sum_{i\in e}\hat{u}_i^2 + \|\psi^{e,-}\|^2\sum_{j\in e}\hat{v}_j^2\right] \\
&\geq \left(\frac{2c_m}{N_0}-c_1\right)|\bar{u}-\bar{v}|^2 - \frac{4(D_2^-)^2}{c_1 N_0}\sum_{e\in\mathcal{E}}\left[\sum_{i\in e}\hat{u}_i^2 + \sum_{j\in e}\hat{v}_j^2\right] \\
&\geq \left(\frac{2c_m}{N_0}-c_1\right)|\bar{u}-\bar{v}|^2 - \frac{4(D_2^-)^2}{c_1 N_0}k[\widehat{M}_2]
\end{aligned} \tag{30}$$

□

Lemma 4. *Let u, v be the solution to equation 16. Then $\widehat{M}_2$ satisfies*

$$\frac{1}{2}\frac{d}{dt}\widehat{M}_2 \leq C_2\widehat{M}_2 + 2c_2|\bar{u}-\bar{v}|^2, \tag{31}$$

where

$$C_2 := -k\left(C^A - 2D^- + \frac{(D^-)^2}{4c_2} - \frac{\delta}{k}\right), \tag{32}$$

and c_2 is an arbitrary positive constant.

Proof. Subtracting equation 27 gives $\dot{\hat{u}}_i$. Then we have

$$\begin{aligned}
\frac{1}{2}\frac{d}{dt}\left(\frac{1}{N_1}\sum_{i=1}^{N_1}|\hat{u}_i|^2\right) =& \frac{1}{N_1}\sum_{i=1}^{N_1}\hat{u}_i\dot{\hat{u}}_i \\
=& \frac{1}{N_1}\sum_{i=1}^{N_1}\hat{u}_i\sum_{e\in\mathcal{E}(i)}\left[\sum_{i'\in e}a_{i,i'}^{e,+}(u_{i'}-u_i) - \sum_{j\in e}a_{i,j}^{e,-}(v_j-u_i)\right] + \frac{\delta}{N_1}\sum_{i=1}^{N_1}\hat{u}_i u_i(1-u_i^2) \\
=& \frac{1}{N_1}\sum_{i=1}^{N_1}\hat{u}_i\sum_{e\in\mathcal{E}(i)}\left[\sum_{i'\in e}a_{i,i'}^{e,+}(\hat{u}_{i'}-\hat{u}_i) - \sum_{j\in e}a_{i,j}^{e,-}(v_j-u_i)\right] + \frac{\delta}{N_1}\sum_{i=1}^{N_1}\hat{u}_i u_i(1-u_i^2) \\
=& \frac{1}{N_1}\sum_{i=1}^{N_1}\hat{u}_i\sum_{e\in\mathcal{E}(i)}\sum_{i'\in e}a_{i,i'}^{e,+}(\hat{u}_{i'}-\hat{u}_i) - \frac{1}{N_1}\sum_{i=1}^{N_1}\hat{u}_i\sum_{e\in\mathcal{E}(i)}\sum_{j\in e}a_{i,j}^{e,-}(\hat{v}_j-\hat{u}_i) \\
&- \frac{1}{N_1}\sum_{i=1}^{N_1}\hat{u}_i\sum_{e\in\mathcal{E}(i)}\sum_{j\in e}a_{i,j}^{e,-}(\bar{v}-\bar{u}) + \frac{\delta}{N_1}\sum_{i=1}^{N_1}\hat{u}_i u_i(1-u_i^2) \\
=:& I_1 + I_2 + I_3 + I_4.
\end{aligned} \tag{33}$$

In particular,

$$\begin{aligned} I_1 =& \frac{1}{N_1}\sum_{i=1}^{N_1}\hat{u}_i\sum_{e\in\mathcal{E}(i)}\sum_{i'\in e}a_{i,i'}^{e,+}(\hat{u}_{i'}-\hat{u}_i) \\ =& \frac{1}{N_1}\sum_{e\in\mathcal{E}}\sum_{i,i'\in e}a_{i,i'}^{e,+}(\hat{u}_{i'}-\hat{u}_i)\hat{u}_i \\ =& \frac{1}{N_1}\sum_{e\in\mathcal{E}}(\hat{u}^e)^\top A^e\hat{u}^e, \end{aligned} \tag{34}$$

where $\hat{u}^e := (\hat{u}_{i_1},\cdots,\hat{u}_{i_{|e|}})^\top$ for each e. Thus I_1 is bounded by

$$\frac{1}{N_1}\sum_{e\in\mathcal{E}}(\hat{u}^e)^\top A^e\hat{u}^e \le -\frac{1}{N_1}\sum_{e\in\mathcal{E}}C^A|\hat{u}^e|^2 \le -\frac{c_r}{N_1}\sum_{i=1}^{N_1}C^A|\hat{u}_i|^2, \tag{35}$$

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

I_2 can be controlled by

$$\begin{aligned} I_2 =& -\frac{1}{N_1}\sum_{i=1}^{N_1}\hat{u}_i\sum_{e\in\mathcal{E}(i)}\sum_{j\in e}a_{i,j}^{e,-}(\hat{v}_j-\hat{u}_i) \\ =& -\frac{1}{N_1}\sum_{e\in\mathcal{E}}\sum_{i,j\in e}a_{i,j}^{e,-}(\hat{v}_j-\hat{u}_i)\hat{u}_i \\ =& -\frac{1}{N_1}\sum_{e\in\mathcal{E}}\sum_{i,j\in e}a_{i,j}^{e,-}\hat{v}_j\hat{u}_i + \frac{1}{N_1}\sum_{e\in\mathcal{E}}\sum_{i,j\in e}a_{i,j}^{e,-}|\hat{u}_i|^2 \\ \le& \frac{1}{N_1}\sum_{e\in\mathcal{E}}\sum_{i,j\in e}a_{i,j}^{e,-}\frac{1}{2}(|\hat{v}_j|^2+|\hat{u}_i|^2) + \frac{1}{N_1}\sum_{e\in\mathcal{E}}D_e^-\sum_{i\in e}|\hat{u}_i|^2 \\ \le& \frac{1}{2N_1}\sum_{e\in\mathcal{E}}\sum_{j\in e}\psi_j^{e,-}|\hat{v}_j|^2 + \frac{3D^-}{2N_1}\sum_{e\in\mathcal{E}}\sum_{i\in e}|\hat{u}_j|^2 \\ \le& \frac{D^-}{2N_1}\sum_{e\in\mathcal{E}}\sum_{j\in e}|\hat{v}_j|^2 + \frac{3D^-}{2N_1}\sum_{e\in\mathcal{E}}\sum_{i\in e}|\hat{u}_j|^2 \\ \le& \frac{D^-k}{2N_1}\sum_{j=1}^{N_2}|\hat{v}_j|^2 + \frac{3D^-k}{2N_1}\sum_{i=1}^{N_1}|\hat{u}_i|^2. \end{aligned} \tag{36}$$

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

$$\begin{aligned} I_3 =& -\frac{1}{N_1}\sum_{i=1}^{N_1}\hat{u}_i\sum_{e\in\mathcal{E}(i)}\sum_{j\in e}a_{i,j}^{e,-}(\bar{v}-\bar{u}) \\ \le& c_2|\bar{u}-\bar{v}|^2 + \frac{1}{4c_2N_1}\sum_{e\in\mathcal{E}}\sum_{i,j\in e}|a_{i,j}^{e,-}|^2|\hat{u}_i|^2 \\ \le& c_2|\bar{u}-\bar{v}|^2 + \frac{(D^-)^2}{4c_2N_1}\sum_{e\in\mathcal{E}}\sum_{i,j\in e}|\hat{u}_i|^2 \\ \le& c_2|\bar{u}-\bar{v}|^2 + \frac{(D^-)^2k}{4c_2N_1}\sum_{e\in\mathcal{E}}\sum_{i=1}^{N_1}|\hat{u}_i|^2. \end{aligned} \tag{37}$$

Define

$$I_4 := \frac{1}{N_1}\delta\sum_{i=1}^{N_1}\hat{u}_iu_i(1-|u_i|^2). \tag{38}$$

$$
\begin{aligned}
I_4 =& \frac{\delta}{N_1}\sum_{i=1}^{N_1}\hat{u}_i(\hat{u}_i+\bar{u})(1-u_i^2)\\
=& \frac{\delta}{N_1}\sum_{i=1}^{N_1}\hat{u}_i^2-\frac{\delta}{N_1}\sum_{i=1}^{N_1}\hat{u}_i^2u_i^2+\frac{\delta}{N_1}\sum_{i=1}^{N_1}\hat{u}_i\bar{u}-\frac{\delta}{N_1}\sum_{i=1}^{N_1}\hat{u}_iu_i^2\bar{u}\\
=& \frac{\delta}{N_1}\sum_{i=1}^{N_1}\hat{u}_i^2-\frac{\delta}{N_1}\sum_{i=1}^{N_1}\hat{u}_i|u_i|^2u_i
\end{aligned}
\tag{39}
$$

Note that

$$
\begin{aligned}
\sum_{i=1}^{N_1}\hat{u}_i|u_i|^2u_i =& \sum_{i=1}^{N_1}|u_i|^2(|u_i|^2-u_i\bar{u})\\
\geq& \frac{1}{2}\sum_{i=1}^{N_1}|u_i|^2(|u_i|^2-|\bar{u}|^2)\\
=& \frac{1}{2}\sum_{i=1}^{N_1}|u_i|^4-\frac{1}{2}\sum_{i=1}^{N_1}|u_i|^2|\bar{u}|^2\\
\geq& \frac{1}{2}\sum_{i=1}^{N_1}|u_i|^4-\frac{1}{2N_1}\sum_{i=1}^{N_1}(|u_i|^2)^2\\
\geq& 0.
\end{aligned}
\tag{40}
$$

Hence,

$$
I_4 \leq \delta\widehat{M}_2(u). \tag{41}
$$

Then

$$
\begin{aligned}
\frac{d}{dt}\left(\frac{1}{2N_1}\sum_{i=1}^{N_1}|\hat{u}_i|^2\right) \leq& k\left(-\frac{C^A}{N_1}+\frac{3D^-}{2N_1}+\frac{(D^-)^2}{4c_2N_1}\right)\sum_{i=1}^{N_1}|\hat{u}_i|^2+k\frac{D^-}{2N_2}\sum_{j=1}^{N_2}|\hat{v}_j|^2+c_2|\bar{u}-\bar{v}|^2\\
\leq& \left[k\left(-C^A+\frac{3D^-}{2}+\frac{(D^-)^2}{4c_2}\right)+\delta\right]\widehat{M}_2(u)+k\frac{D^-}{2}\widehat{M}_2(v)+c_2|\bar{u}-\bar{v}|^2.
\end{aligned}
\tag{42}
$$

Similarly,

$$
\frac{d}{dt}\left(\frac{1}{2N_2}\sum_{j=1}^{N_2}|\hat{v}_j|^2\right) \leq \left[k\left(-C^A+\frac{3D^-}{2}+\frac{(D^-)^2}{4c_2}\right)+\delta\right]\widehat{M}_2(v)+k\frac{D^-}{2}\widehat{M}_2(u)+c_2|\bar{u}-\bar{v}|^2. \tag{43}
$$

Sum them together. We have

$$
\frac{1}{2}\frac{d}{dt}(\widehat{M}_2) \leq -k\left(C^A-2D^-+\frac{(D^-)^2}{4c_2}-\frac{\delta}{k}\right)(\widehat{M}_2)+2c_2|\bar{u}-\bar{v}|^2. \tag{44}
$$

This gives an exponential growth estimate or $\widehat{M}_2$ up to an error term of $|\bar{u}-\bar{v}|^2$.

Lemma 3 gives an exponential growth estimate of $|\bar{u}-\bar{v}|^2$ up to an error term of $\widehat{M}_2$. If N_0 is large enough, the coefficient of the error term will be small.

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

Set

$$
A_{11} := \frac{2c_m}{N_0}-c_1,\ A_{12} := \frac{4(D_2^-)^2k}{c_1N_0},\ A_{21}=2c_2,\ A_{22}=k\left(C^A-2D^-+\frac{(D^-)^2}{4c_2}\right). \tag{45}
$$

If N_0 is large enough, (45) will be satisfied.

$$(A_{11} + A_{12})^2 - 4A_{21}A_{12} > 0. \tag{46}$$

Apply Lemma 4.1 in Jin & Shu (2021a), we can then obtain the L_2 separation.

The relative size between $|\bar{u} - \bar{v}|^2$ and $\widehat{M_2}$ is an indicator of group separation in the sense of L_2. If $|\bar{u} - \bar{v}|^2$ is much larger, $\widehat{M_2}$, then the two groups are well-separated in an average sense. Since the hypergraph is connected, there is a positive bound between different clusters, hence the Dirichlet energy does not decay to zero.

B Additional Experimental Details

B.1 Baseline Methods

We compare ODNet with a range of discrete and continuous message-passing methods across homophilic, heterophilic, and hypergraphs.

Homophilic Graphs We compare the performance of ODNet with classical graph convolutions such as Graph Convolutional Network (GCN; Yadati et al. (2019)), Graph Attention Network (GAT; Veličković et al. (2018)), Mixture Model Networks (MoNet; Monti et al. (2017)), and GraphSage (Hamilton et al., 2017). We also include ODE-based continuous graph convolutions, including Continuous Graph Neural Networks (CGNN; Xhonneux et al. (2020)), Graph Neural Ordinary Differential Equations (GDE; Poli et al. (2020)), and Graph Neural Diffusion with linear diffusion (GRAND-l; Chamberlain et al. (2021)). Furthermore, we include the other opinion-dynamic inspired network, SINN Okawa & Iwata (2022).

Heterophilic Graphs For the heterophilic graphs, we make comparisons with representative graph neural networks that are known for encoding heterophilic graphs (GPRGNN (Chien et al., 2021), GraphCON (Rusch et al., 2022), and H2GCN (Zhu et al., 2020)) or for mitigating the oversmoothing issue (GCNII (Chen et al., 2020), Geom-GCN (Pei et al., 2020), and PairNorm (Zhao & Akoglu, 2020)). For completeness, we also include the classic graph convolutions, including GraphSAGE (Hamilton et al., 2017), GAT (Veličković et al., 2018), and GCN (Kipf & Welling, 2017)

Hypergraphs The performance of ODNet on the four hypergraphs are compared with a diverse of baselines, including HGNN (Feng et al., 2019), HyperGCN (Yadati et al., 2019), HCHA (Bai et al., 2021), HNHN (Dong et al., 2020), UniGCNII (Huang & Yang, 2021), HyperND Tudisco et al. (2021), AllDeepSets (Chien et al., 2022), AllSetTransformer (Chien et al., 2022), ED-HNN (Wang et al., 2022), and SheafHGNN (Duta et al., 2024).

B.2 Experimental Details

Table 4-5 reports summary statistics of the graphs and hypergraphs used in Section 6, respectively.

Table 6 provides the optimal combination of hyperparameters for each benchmark to reproduce the performance of ODNet reported in the paper.

Table 7 details the searching space of the hyperparameters. For general hyperparameters, such as learning rate and weight decay, we used Ray Tune with a hundred trials using an asynchronous hyperband scheduler with a grace period of 50 epochs. Table 8-9 reports the influence of selecting different scaling factors on the model performance. We test the impacts on **Cora** and **Texas** for both homophilic and heterophilic graphs representation learning.

C Additional Background: Co-Occurrence Network of Metabolic Genes

The co-occurrence network of metabolic genes is a graph representation that illustrates the statistical associations and co-occurrence patterns among various metabolic genes within a biological system (Bello et al.,

Table 4: Summary of **graph** benchmarks used in experiments.

Dataset	**# classes**	**# features**	**# nodes**	**# edges**	$\mathcal{H}$
Cora (McCallum et al., 2000)	7	1,433	2,708	5,429	0.83
CiteSeer (Sen et al., 2008)	6	3,703	3,327	4,732	0.71
PubMed (Namata et al., 2012)	3	500	19,717	44,338	0.79
CoauthorCS (Shchur et al., 2018)	15	6,805	18,333	100,227	0.80
Computer (Namata et al., 2012)	10	767	13,381	245,778	0.77
Photo (Namata et al., 2012)	8	745	7,487	119,043	0.83
Texas García-Plaza et al. (2016)	5	1,703	183	309	0.11
Wisconsin García-Plaza et al. (2016)	5	1,703	251	499	0.21
Cornell García-Plaza et al. (2016)	5	1,703	183	295	0.30

Table 5: Summary of **hypergraph** benchmarks used in experiments.

	Cora-coauthor	**Cora**-cocitation	**CiteSeer**-cocitation	**PubMed**-cocitation
# classes	7	7	6	3
# features	1,433	1,433	3,703	500
# hypernodes	2,708	2,708	3,312	19,717
# hyperedges	1,072	1,579	1,079	7,963
avg. hyperedge size	4.2±4.1	3.0±1.1	3.2±2.0	4.3±5.7

Table 6: Optimal setting of hyperparameters in reproducing the results in Section 6.

Dataset	ϵ_1	ϵ_2	time (T)	ν	μ
Cora	0.012	0.40	12	0	1.4
CiteSeer	0.01	0.90	10	0	3.0
PubMed	0.01	0.40	20	0	2.2
CoauthorCS	0.01	0.40	15	0	1.7
Computer	0.01	0.50	15	0	5.0
Photo	0.01	0.40	12	0	10.0
Texas	0.50	0.80	12	-50	1.0
Wisconsin	0.60	0.80	12	-10	2.0
Cornell	0.12	0.40	12	0	2.0
Cora-coauthor	0	1	0.1	1.0	1.0
Cora-cocitation	0	1	0.1	1.0	1.0
PubMed-cocitation	0	1	0.1	1.0	1.0
CiteSeer-cocitation	0	1	0.1	1.5	1.0

2020). This network emerges from computational analyses of extensive genomic data, with each node denoting metabolic genes linked to distinct biochemical functions, such as sugar production and TMAO (trimethylamine N-oxide) synthesis. These connections are quantified by the likelihood of two crucial functional genes co-occurring within the same species at a given time (Liu et al., 2022). Due to the complexity of the co-occurrence network of metabolic genes in different microorganism species, which arise from a large number of genes and connections, simplifying the network enables us to identify how key metabolic genes in microorganisms can be gathered into several interdependent modules. This is significantly important for revealing the mechanism of how genes work together within metabolic pathways, how they respond to different environmental conditions, and which genes may have essential roles in specific biological processes, highlighting the

Table 7: Hyperparameter Search Space

Hyperparameters	Search Space	Distribution
learning rate	$[10^{-6}, 10^{-1}]$	log-uniform
weight decay	$[10^{-3}, 10^{-1}]$	log-uniform
dropout rate	$[0.1, 0.8]$	uniform
hidden dim	$\{64, 128, 256\}$	categorical

Table 8: Choices of influence function and scaling factors for **Cora**.

ϵ_1	ϵ_2	μ	ν	**Accuracy**
0.04	0.45	1.4	0	79.6±0.3
0.012	0.40	2.0	0	80.2±0.3
0.03	0.45	1.4	0	81.0±0.2
0.02	0.45	1.4	0	81.8±0.2
0.012	0.20	1.4	0	82.3±0.2
0.012	0.40	1.4	-1	83.7±0.4
0.012	0.40	1.8	0	83.9±0.2
0.012	0.40	1.6	0	84.1±0.3
0.012	0.30	1.4	0	84.2±0.3
0.012	0.40	1.2	0	85.0±0.3
0.012	0.40	1.0	0	85.0±0.3

Table 9: Choices of influence function and scaling factors for **Texas**.

ϵ_1	ϵ_2	μ	ν	**Accuracy**
0.50	0.80	2.0	-1000	73.0±10.1
0.50	0.80	1000.0	-50	78.2±8.4
0.50	0.80	2.0	-100	78.8±7.2
0.50	0.80	2.0	0	78.8±1.6
0.50	0.80	100.0	-50	80.5±4.2
0.50	0.60	1.0	-50	81.0±3.0
0.70	0.80	1.0	-50	81.1±4.2
0.50	0.80	2.0	-80	81.6±3.4
0.50	0.80	2.0	-1	81.6±2.0
0.60	0.80	1.0	-50	86.5±3.5
0.50	0.90	1.0	-50	86.5±3.0
0.50	0.70	1.0	-50	87.0±3.4
0.50	0.80	2.0	-10	86.7±3.2
0.40	0.80	1.0	-50	87.0±3.0
0.50	0.80	2.0	-50	87.6±4.0
0.50	0.80	2.0	-50	87.6±4.0
0.50	0.80	10.0	-50	88.1±4.2
0.50	0.80	1.0	-50	88.3±3.2

significance of the adaptation of microorganisms to changing environmental conditions (Levy & Borenstein, 2013).

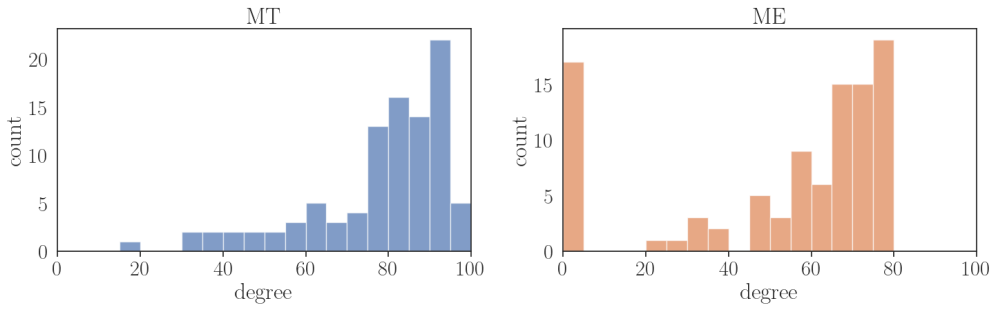

Figure 8: Distribution of node degree on **ME** and **MT** networks.

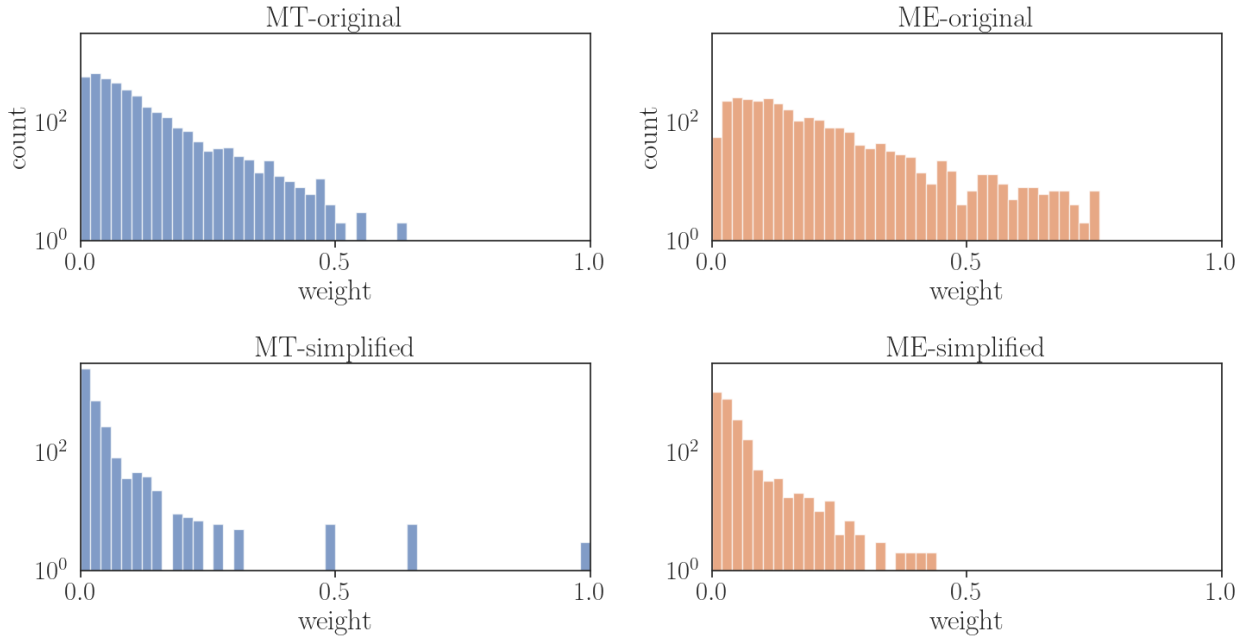

Figure 9: Distribution of edge weights on **ME** and **MT** networks.

D Experimental Details for the Co-Occurrence Network Simplification Task

We established two distinct graphs for the Mariana Trench (MT) and Mount Everest (ME) gene co-occurrence networks, utilizing source data from Liu et al. (2022). In both networks, the nodes represent the same set of functional genes. The primary difference between them lies in the edge weights, which reflect unique co-occurrence patterns of gene pairs in MT and ME. This distinction is visually evident when comparing the top two histograms in Figure 9, illustrating the varying distributions of gene influence weights in MT and ME graphs.

To construct these graphs, we connected all node pairs with non-zero edge weights, resulting in a total of $2,517$ edges for 96 nodes. Each node was associated with a 20-dimensional unit vector as pseudo-features. Additionally, we assigned a three-class categorical label to each node, categorizing them as 'strong,' 'medium,' or 'weak' influencers within their respective local communities. The label assignment was determined based on the nodes' degrees with cutoffs at 20 and 60. For example, a node with a degree of 30 would be classified as a 'medium influencer.'

To facilitate model training, we applied random masking to the training, validation, and test sets, ensuring equal proportions in each set. During the training process, we recorded the learned similarity scores s_{ij} at the final layer for later use in generating the simplified network.

Figure 9 highlights a noticeable divergence between the top two histograms (representing the weight distributions in the original networks) and the bottom two histograms (depicting the weight distributions in the simplified networks). Specifically, a higher concentration of weights is observed at the extreme regions (with weights close to 0 and 1) in the simplified networks. This divergence underscores the impact of our simplification approach on the network's edge weight distribution.