A Unified View on Neural Message Passing with Opinion Dynamics for Social Networks

Anonymous authors
Paper under double-blind review

Abstract

Social networks represent a common form of interconnected data frequently depicted as graphs within the domain of deep learning-based inference. These communities inherently form dynamic systems, achieving stability through continuous internal communications and opinion exchanges among social actors along their social ties. In contrast, neural message passing in deep learning provides a clear and intuitive mathematical framework for understanding information propagation and aggregation among connected nodes in graphs. Node representations are dynamically updated by considering both the connectivity and status of neighboring nodes. This research harmonizes concepts from sociometry and neural message passing to analyze and infer the behavior of dynamic systems. Drawing inspiration from opinion dynamics in sociology, we propose ODNET, a novel message passing scheme incorporating bounded confidence, to refine the influence weight of local nodes for message propagation. We adjust the similarity cutoffs of bounded confidence and influence weights of ODNET and define opinion exchange rules that align with the characteristics of social network graphs. We show that ODNET enhances prediction performance across various graph types and alleviates oversmoothing issues. Furthermore, our approach surpasses conventional baselines in graph representation learning and proves its practical significance in analyzing real-world co-occurrence networks of metabolic genes. Remarkably, our method simplifies complex social network graphs solely by leveraging knowledge of interaction frequencies among entities within the system. It accurately identifies internal communities and the roles of genes in different metabolic pathways, including opinion leaders, bridge communicators, and isolators.

1 Introduction

Sociometry is a quantitative method used in social psychology and sociology to describe social relations (Moreno, 1934; 2012). In his pioneering work, Moreno (1934) conceptualized a graph as an abstract representation of a group's structure. The term *social network* was later coined to describe a system comprising individual social actors and the social ties among them (Proskurnikov & Tempo, 2017). The development of cybernetics has led to increased attention to the study of messages and communication within society (Wiener, 1988). Statistical physics has contributed by introducing methods and tools from dynamical systems theory, giving rise to the field of *sociodynamics* (Weidlich, 2006; Helbing, 2010).

Graph neural networks (GNNs), on the other hand, are rooted in the same basic structure as social networks: graphs. 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* (MP; Gilmer et al. (2017)) and later became a fundamental feature extraction unit of graph-structured data for aggregating features of neighbors during network propagation.

This work explores the connection between these two fields by delving into *opinion dynamics*, a subfield of sociodynamics. We establish a link between the French-DeGroot (FD) model (French Jr, 1956; DeGroot, 1974) and MP, emphasizing a shared phenomenon that in both FD models and MPs a network converges exponentially to a stable state when it exhibits strong local connectivities, a property that is frequently observed in hypergraphs. Moreover, we draw inspiration from the
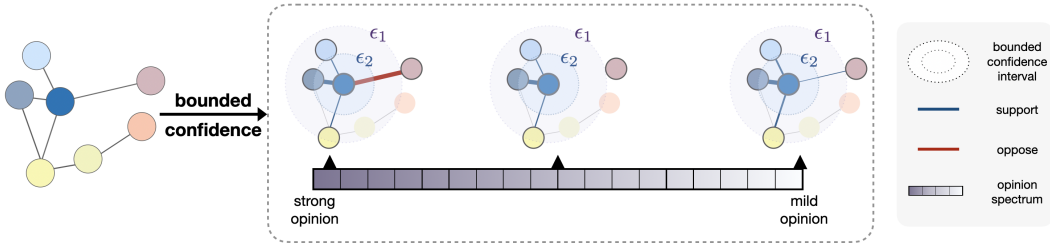
Figure 1: Given a graph with initial connections, ODNET is defined with bounded confidence to update the influence weights by the graph's position on the opinion spectrum.

Hegselmann-Krause (HK) model (Rainer & Krause, 2002) and incorporate the concept of *bounded confidence* into our novel MP formulation termed ODNET, which features a confidence filtration mechanism on initial edge connections. Based on the similarity of node pairs and connection proximity, ODNET aggregates neighboring information through automatic adjustment on edge weights. With piecewise MP schemes, the model strengthens, weakens, or removes initial links within the graph. Additionally, it allows for assigning negative weights to capture adverse perspectives from neighboring nodes with substantial disparities. This feature is essential when investigating heterophilic networks, where connected entities exhibit dissimilar characteristics.

The opinion dynamics-inspired propagation scheme can take on a static or dynamic nature, contingent on the choice of the similarity measure applied to nodes. The new mechanism emulates the dynamic spreading of opinions in a community, where effective communications converge individuals or agents towards consensus or several predominant viewpoints. For instance, in microbial communities, analyzing the co-occurrence network of metabolic genes across various species can reveal the potential biologically key genes act as predominant roles (Liu et al., 2018). When exchanging ideas, individuals tend to support opinions akin to their own. Depending on the position of the community on the opinion spectrum (Figure 1), when encountering significantly divergent thoughts, individuals may choose to disregard or oppose them. For instance, researchers typically focus on studies within their expertise but would love to learn new perspectives from other domains, whereas politicians usually have strong conflicts and resist propositions from competing parties.

The proposed approach offers an enhanced description of neighborhood influence across various scenarios by categorizing the relationships among nodes into three levels. We assess the versatile ODNET across three categories of graphs: homophilic graphs, heterophilic graphs, and hypergraphs, each characterized by its unique properties. Empirically, the introduced piecewise aggregation behavior enhances the performance of previously established MP methods (such as GCN (Kipf & Welling, 2017), GAT (Veličković et al., 2018), and HGNN (Gao et al., 2022)). We also demonstrate our model's capacity to progressively simplify graph structures, a crucial feature for deciphering complex social networks. To underscore the practical significance of this simplification capability, we provide a concrete application in the field of microbiology. Our model is shown to effectively prune weak connections among genes, thus extracting biologically relevant genes and connections from co-occurrence networks of metabolic genes.

2 NEURAL MESSAGE PASSING FOR GRAPHS AND HYPERGRAPHS

Graph 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. An edge $(v_i, v_j) \in \mathcal{E}$ if and only if $w_{ij} > 0$. 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).

Hypergraphs A hypergraph is a generalization of a graph in which an edge can connect any number of vertices. A hypergraph 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_{ee}^h 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 $w_{i,j}^e$ 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 $w_{i,j}^e$ as nonzero if node i, j are connected by a hyperedge e, otherwise 0.

Neural Message Passing   Neural Message Passing (MP; 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. An MP layer reads

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

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 MP mechanism updates the feature of each node by aggregating their self-features with neighbors' features.

The classic MPs can be extended to hypergraphs that to consider interactions among multiple nodes reflected in a hyperedge. At the kth layer:

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

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

3   CONNECTING OPINION DYNAMICS WITH MESSAGE PASSING

3.1   FRENCH-DEGROOT MODEL

The French-DeGroot (FD), originally introduced by French Jr (1956) and later developed by Harary (1959), Norman et al. (1965) and 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 $k = 0, 1, \cdots$. The evolution of an individual's opinion is

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

where the non-negative *influence weight* w_{ij} satisfying $\sum_{j=1}^{N} w_{ij} = 1$. If $w_{ij} > 0$, individuals i and j are neighbors. The influence weight signifies the relative impact that 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, potentially leading to consensus, resembling an opinion pooling process. It can be interpreted as an MP, where the graph represents a community with each node representing an individual, emulating how information is exchanged within a specific type of neural network.

Convergence Analysis   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. This result can be found in references such as Ren & Beard (2008); Proskurnikov & Tempo (2017); Bullo et al. (2009); 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.*

It's worth noting that any graph with a self-loop is considered aperiodic, implying that exponential decay is likely to occur in graphs with relatively strong connectivity. Coincidentally, a similar phenomenon, known as *oversmoothing* (Nt & Maehara, 2019; Oono & Suzuki, 2019), has been explored

in the context of GNNs, where it is associated with the exponential decay of the *Dirichlet energy*, a measurement of the convergence degree of all features (weighted by graph structure). Despite originating from different fields, these two phenomena appear to describe similar processes.

Connection to Neural Message Passing It's intriguing to observe that the FD model, often regarded as a micro-level model based on individuals simulating the evolution of individual opinions, shares similarities with a GNN model known as GRAND (Chamberlain et al., 2021). It describes a diffusion process on graphs by connecting heat conduction with MP. This connection 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 interconnectedness of ideas in different domains of research.

3.2 Hegselmann-Krause Model

In the FD model, each agent possesses the capability to interact with any other agent, regardless of their opinions. However, in real-life scenarios, individuals typically engage in conversations primarily with those who share similar viewpoints. This fundamental aspect of human communication is accurately characterized and referred to as *bounded confidence* within the context of sociodynamics. The Hegselmann-Krause (HK) model (Rainer & Krause, 2002) defines bounded confidence as

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

where $\mathbb{B}(i, \boldsymbol{x}_i) = \{j : \|\boldsymbol{x}_j(k) - \boldsymbol{x}_i(k)\| < \epsilon\}$ encompasses all individual i's associated peers j, whose opinions diverge from individual i within a confined region of radius $\epsilon_i \in \mathbb{R}$. This parameter represents the degree of uncertainty or tolerance within the model.

Clustering and Oversmoothing in Heterophilious Dynamics The HK model demonstrates a clustering phenomenon driven by the self-alignment of agents. The number of clusters has been shown to have a negative correlation with the heterophily dependence among agents in a system (Motsch & Tadmor, 2014). Specifically, when interactions exhibit significant heterophily–meaning that agents tend to form stronger bonds with counterparts rather than with similar individuals–the dynamics tend to foster consensus. This tendency of individuals converging toward an 'environmental averaging' aligns with the oversmoothing issue in GNNs. One solution is to require an MP to retain at least two clusters at the end, where the Dirichlet energy is proven to have a lower bound. This can be achieved through techniques such as bi-clustering with repulsion (Fang et al., 2019; Jin & Shu, 2021; Wang et al., 2023). In the context of the HK model, it is advisable to avoid steep increases over compact supports when aggregating neighboring information.

4 ODNet: Opinion Dynamics-Inspired Neural Message Passing

Inspired by the mechanism of opinion dynamics, we introduce ODNet, a novel MP framework, employing the *influence function* $\phi(s)$ with bounded confidences. We offer a comprehensive interpretation of each component within ODNet, beginning with a discrete formulation and subsequently extending it to continuous forms that are applicable to both graphs and hypergraphs.

Discrete Formation In the discrete domain, we formulate the update rule as follows:

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

where ϕ is a non-decreasing function of the similarity measure $s_{i,j}$ to node i and node j, and $u(\boldsymbol{x}_i)$ is a control term for stability. For instance, s_{ij} could be defined as the normalized adjacency matrix

(Kipf & Welling, 2017) or attention coefficients (Veličković et al., 2018). The monotonicity of ϕ characterizes the *influence weight* concerning node-node similarity. Our model opts for a piecewise ϕ function to delineate influence regions akin to bounded confidence. In a special case, with

$$\phi(s) = \begin{cases} \mu s, & \text{if } \epsilon_2 < s \\ s, & \text{if } \epsilon_1 \leq s \leq \epsilon_2 \\ 0, & \text{otherwise,} \end{cases} \quad (7)$$

our model (6) can be written as

$$\boldsymbol{x}_i(t+1) = \mu \sum_{\epsilon_2 < s_{i,j}} s_{i,j}(\boldsymbol{x}_j(t) - \boldsymbol{x}_i(t)) + \sum_{\epsilon_1 \leq s_{i,j} \leq \epsilon_2} s_{i,j}(\boldsymbol{x}_j(t) - \boldsymbol{x}_i(t)) + \boldsymbol{x}_i(t). \quad (8)$$

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.

Additionally, in certain extreme scenarios, individuals with significantly divergent opinions may exhibit hostile attitudes toward each other. To model such instances, we consider:

$$\phi(s) = \begin{cases} \mu s, & \text{if } \epsilon_2 < s \\ s, & \text{if } \epsilon_1 \leq s \leq \epsilon_2 \\ \nu(1-s), & \text{otherwise,} \end{cases} \quad (9)$$

where $\mu > 0$ and $\nu < 0$. In the context of GNNs, it allows not only learning from *positive* neighbors with similarity but also extracting *negative* information from nodes with discrepancies. It's worth noting that the negative coefficient ν implies that some node pairs consistently repel each other, potentially causing undesirable system dilation. Therefore, a control term $u(x_i)$ is introduced for system stability. A simple approach is to design a potential function $P(x)$ where $\nabla P(x) \to \infty$ as $x \to \infty$, and set $u(x_i) = \nabla P(x)|_{x=x_i}$. Various choices for $P(x)$ can be explored, as discussed by Kolokolnikov et al. (2011). The function $P(x)$ can be viewed as a moral constraint preventing individuals from resorting to extreme violence in conflict situations.

The different definitions of bounded confidence provided by (7) and (9) reflect the various behaviors of opinion exchange in a system, and these behaviors are linked to the different positions of the system along the opinion spectrum. While this can be conceptually determined by the intrinsic characteristics of the graph or the system, we propose the use of the *homophily level* (Pei et al., 2020) as an alternative quantitative measure. When passing messages on a specific graph, we recommend employing the former formulation for homophilic graphs and the latter for heterophilic graphs. Further investigations and explanations will be provided in Section 5.

Continuous Formation In the realm of opinion dynamics, an individual's viewpoint typically undergoes gradual shifts rather than abrupt reversals. For instance, a person's political orientation is seldom confined to the extremes of either far-right or far-left, and an ultra-leftist rarely makes an overnight transition to a far-right position. Therefore, a natural refinement of the discrete MP model presented above is to generalize it into a continuous version. In broader terms, one can view a conventional MP model as a numerical discretization of the following continuous model:

$$\frac{\partial \boldsymbol{x}_i(t)}{\partial t} = \sum_{i=1}^{N} \phi(s_{ij})(\boldsymbol{x}_j(t) - \boldsymbol{x}_i(t)) + u(\boldsymbol{x}_i) \quad (10)$$

The continuous formulation (10) outlined above is amenable to various numerical approximation techniques corresponding to a discrete model with a specific residual compensation scheme. Consequently, a range of Ordinary Differential Equation (ODE) solvers can be employed for ODNET, including Neural ODEs (Chen et al., 2018).

Generalization on Hypergraphs Since ODNET constructs a general formulation for MP, it can be extended to hypergraphs as well. The primary distinction between graphs and hypergraphs lies in the fact that a hyperedge extends connectivity beyond the scope of traditional edges. This extension can be accommodated by generalizing the weight aggregation:
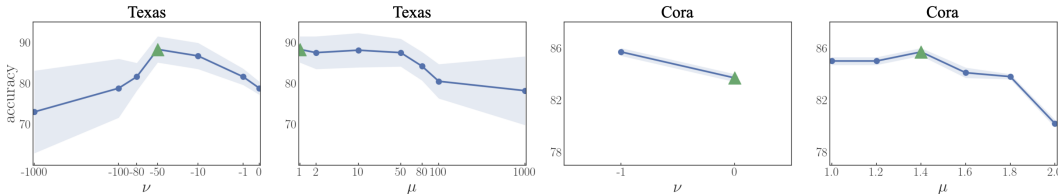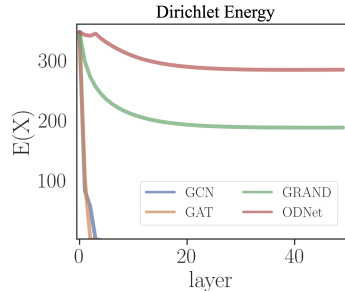
$$\frac{\partial \boldsymbol{x}_i}{\partial t} = \sum_{e:i \in e} \sum_{j \in e} \phi(s_{i,j}^e)(\boldsymbol{x}_j - \boldsymbol{x}_i) + u(\boldsymbol{x}_i). \quad (11)$$

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

Model	Cora	CiteSeer	PubMed	Coauthor CS	Computer	Photo
GCN (Kipf & Welling, 2017)	81.5±1.3	**71.9±1.9**	77.8±2.9	91.1±0.5	82.6±2.4	91.2±1.2
MONET (Monti et al., 2017)	81.3±1.3	71.2±2.0	78.6±2.3	90.8±0.6	83.5±2.2	91.2±2.3
GRAPHSAGE-MEAN (Hamilton et al., 2017)	79.2±7.7	71.6±1.9	77.4±2.2	91.3±2.8	82.4±1.8	91.4±1.3
GRAPHSAGE-MAX (Hamilton et al., 2017)	76.6±1.9	67.5±2.3	76.1±2.3	85.0±1.1	N/A	90.4±1.3
GAT (Veličković et al., 2018)	**81.8±1.3**	71.4±1.9	**78.7±2.3**	90.5±0.6	78.0±1.9	85.7±2.0
GAT-PPR (Veličković et al., 2018)	81.6±0.3	68.5±0.2	76.7±0.3	91.3±0.1	**85.4±0.3**	90.9±0.3
CGNN (Xhonneux et al., 2020)	81.4±1.6	66.9±1.8	66.6±4.4	**92.3±0.2**	80.3±2.0	91.4±1.5
GDE (Poli et al., 2020)	78.7±2.2	71.8±1.1	73.9±3.7	91.6±0.1	82.9±0.6	**92.4±2.0**
GRAND-L (Chamberlain et al., 2021)	**83.6±1.0**	**73.4±0.5**	**78.8±1.7**	**92.9±0.4**	83.7±1.2	92.3±0.9
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**

† The top three are highlighted by **First**, **Second**, **Third**.

Model	Texas	Wisconsin	Cornell
MLP	80.8±4.8	85.3±3.3	81.9±6.4
GPRGNN (Chien et al., 2021)	78.4±4.4	82.9±4.2	80.3±8.1
H2GCN (Zhu et al., 2020)	**84.9±7.2**	**87.7± 5.0**	**82.7±5.3**
GCNII (Chen et al., 2020)	77.6±3.8	80.4±3.4	77.9±3.8
GEOM-GCN (Pei et al., 2020)	66.8±2.7	64.5±3.7	60.5±3.7
PAIRNORM (Zhao & Akoglu, 2020)	60.3±4.3	48.4±6.1	58.9±3.2
GRAPHSAGE (Hamilton et al., 2017)	82.4±6.1	81.2±5.6	76.0±5.0
GAT (Veličković et al., 2018)	52.2±6.6	49.4±4.1	61.9±5.1
GCN (Kipf & Welling, 2017)	55.1±5.2	51.8±3.1	60.5±5.3
GRAPHCON (Rusch et al., 2022)	**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 2: Average test accuracy on **heterophilic** graphs.

Figure 2: Decays of Dirichlet energy with layers on **Texas**.

Figure 3: Impact of scaling factors ν and μ on **Texas** and **Cora**.

This formulation aligns with the notion that in large communities, information propagates through smaller sections (nodes that share a hyperedge) rather than through individual pairwise interactions. Similar to the graph case, the choice of $s_{i,j}^e$ may vary, such as attention coefficients (Bai et al., 2021).

Remark 1. *The collective behaviors in HK-driven modeling mitigate the oversmoothing issue for continuous MP schemes on hypergraphs. See Appendix A for further discussions.*

5 GRAPH REPRESENTATION LEARNING

5.1 EXPERIMENTAL PROTOCOL

Benchmark Datasets This section validates the efficacy of ODNET through classic node-level representation learning tasks on a variety of datasets spanning three types of graphs, including six 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), and **Photo** (Namata et al., 2012)), three heterophilic graphs (**Texas**, **Wisconsin**, and **Cornell** from the WebKB dataset (García-Plaza et al., 2016)), and four hypergraphs based on the citation network (Yadati et al., 2019). For additional descriptions, please refer to Appendix B.1.

Training Setup We compare our model to a diverse set of top-performing baseline GNN models, including classic graph convolutions, MPs with continuous updating schemes, and the latest hypergraph models. For ODNET, we trained the model using a neural ODE solver with Dormand–Prince

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

Model	Cora-coauthor	Cora-cocitation	CiteSeer-cocitation	PubMed-cocitation
HGNN (Feng et al., 2019)	82.6±1.7	**79.4±1.4**	72.5±1.2	86.4±0.4
HYPERGCN (Yadati et al., 2019)	79.5±2.1	78.5±1.3	71.3±0.8	82.8±8.7
HCHA (Bai et al., 2021)	82.6±1.0	79.1±1.0	72.4±1.4	86.4±0.4
HNHN (Dong et al., 2020)	77.2±1.5	76.4±1.9	72.6±1.6	86.9±0.3
UNIGCNII (Huang & Yang, 2021)	**83.6±1.1**	78.8±1.1	73.0±2.2	88.3±0.4
HYPERND Tudisco et al. (2021)	80.6±1.3	79.2±1.1	72.6±1.5	86.7±0.4
ALLDEEPSETS (Chien et al., 2022)	82.0±1.5	76.9±1.8	70.8±1.6	**88.8±0.3**
ALLSETTRANSFORMER (Chien et al., 2022)	83.6±1.5	78.6±1.5	**73.1±1.2**	88.7±0.4
ED-HNN (Wang et al., 2022)	**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**

adaptive step size scheme (DOPRI5). In homophilic datasets, we utilized 10 random weight initializations and random splits, with each combination randomly selecting 20 instances for each class. In heterophilic and hypergraph datasets, we used the fixed 10 training/validation splits by Pei et al. (2020) and Yadati et al. (2019), respectively. For further details, please refer to Appendix C.

5.2 NODE CLASSIFICATION

Graphs Tables 1-2 present the average accuracy for predicting node labels in both homophilic and heterophilic graphs. 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 MP techniques, such as GRAND, by introducing the bounded confidence mechanism and the respective influence weights. This superiority is particularly evident on heterophilic graphs, where the repulsive force among dissimilar node pairs significantly enhances prediction accuracy. Furthermore, Figure 3 illustrates the distinct preferences of the influence function for homophilic and heterophilic graphs. We recommend following (7) for the former and (9) 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 Appendix D.1.

Hypergraphs 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 most baseline methods with a significant improvement. The only exception is ED-HNN, where our method achieves a slightly less pronounced advantage. 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 contrast, ED-HNN employs the second-order difference, which intrinsically contains more comprehensive and expressive information. In this case, our method demonstrates great potential for significantly enhancing the performance of a basic method with minimal additional complexity, surpassing even the most advanced methods.

5.3 DIRICHLET ENERGY, OVERSMOOTHING, AND COMMUNITY CONSENSUS

Many MP methods encounter the issue of oversmoothing, limiting their ability to enable deep networks to achieve expressive propagation. As a common metric, 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 2 illustrates the decay of Dirichlet energy on **Texas** with all network parameters randomly initialized. The two conventional MPs, 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). Here 'heterophily' signifies the tendency of a graph to form stronger connections with those who are different rather than those who are similar, which is a different concept from the 'heterophilic graph' in GNNs.
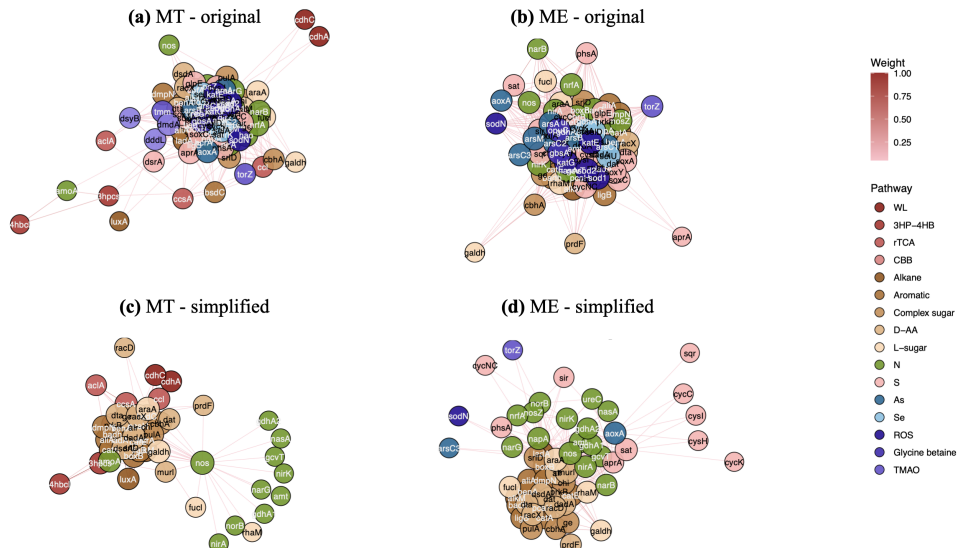
Figure 4: Co-occurrence network of selected metabolic genes in ME or MT before (a-b) and after (c-d) being simplified by ODNET. Connections are considered strong with weights> 0.05, and genes without a strong connection with any other peers are removed.

6 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). 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 the biological expertise (Liu et al., 2022). 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 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, 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 E.

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 4 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 4a-b). In contrast, the simplified networks greatly enhanced the readability of the co-occurrence network while retaining reasonable biological significance (Figure 4c-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 4c, whereas in the ME network, the key genes of sulfate reduction (sat and aprA) coupled the carbon and sulfur metabolism (S) (Figure 4d). 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 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; Battaglia et al., 2018; Hamilton, 2020). This framework has seen active extensions into continuous graph convolutions (Poli et al., 2020; Brandstetter et al., 2021; Chamberlain et al., 2021; Liu et al., 2023; Wang et al., 2023). Notably, SINN (Okawa & Iwata, 2022) explored the possible combination of opinion dynamics and MP with a different implementation from ours. We employed a potential term to avoid excessive attraction or repulsion from neighbors and circumvent the oversmoothing issue. Also, we established bounded confidence with adjustable scaling factors to balance the attractive or repulsive forces when aggregating the neighborhood information. In parallel, Feng et al. (2019); Gao et al. (2022) extended GCN and 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), spectral theory (Yadati et al., 2019), and node potential (Wang et al., 2022).

Collective Dynamics   In classical opinion dynamics systems, a first-order formulation of information exchange is typically employed, relying on the positions of individuals. This formulation naturally connects with a second-order formulation consistent with Newtonian dynamics, which finds applications in phenomena like animal flocking, cell clusters, and self-organizing particles (Holm & Putkaradze, 2006; Carrillo et al., 2010a; Kolokolnikov et al., 2013). These scenarios fall under the purview of *collective dynamics*, wherein agents move together based on attraction and repulsion forces (D'Orsogna et al., 2006; Carrillo et al., 2010b; Motsch & Tadmor, 2014; Carrillo & Shu, 2023). For example, the Cucker-Smale model (Cucker & Smale, 2007) extends the HK model to a second-order formulation involving both position and velocity; Fang et al. (2019) investigated bi-cluster flocking with Rayleigh friction and attractive-repulsive coupling; Jin & Shu (2021) demonstrated a similar collective phenomenon with stochastic dynamics.

8 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 oversmoothing issues 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. Moreover, our method showcases significant potential in simplifying complex real-world social networks, offering 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

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

Peter W. Battaglia, Jessica B. Hamrick, Victor Bapst, Alvaro Sanchez-Gonzalez, Vinícius Flores Zambaldi, Mateusz Malinowski, Andrea Tacchetti, David Raposo, Adam Santoro, Ryan Faulkner, Çaglar 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.

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.

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 *International Conference on Learning Representations*, 2021.

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.

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.

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*, pp. 297–336, 2010a.

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, 2010b.

Benjamin Paul Chamberlain, James Rowbottom, Maria I. Gorinova, Stefan D Webb, Emanuele Rossi, and Michael M. Bronstein. GRAND: Graph neural diffusion. In *ICML*, 2021. URL https://openreview.net/forum?id=_1fu_cjsaRE.

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. URL https://openreview.net/forum?id=n6jl7fLxrP.

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

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. URL http://www.jstor.org/stable/2285509.

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

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.

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 *Proceedings of the AAAI conference on artificial intelligence*, 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, Zizhao Zhang, Haojie Lin, Xibin Zhao, Shaoyi Du, and Changqing Zou. Hypergraph Learning: Methods and Practices. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(5), May 2022. 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.

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

William L Hamilton. Graph representation learning. *Synthesis Lectures on Artifical Intelligence and Machine Learning*, 14(3):1–159, 2020.

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

Frank Harary. A criterion for unanimity in french's theory of social power. 1959.

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.

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(3):619–636, September 2021. ISSN 2305-221X, 2305-2228. doi: 10.1007/s10013-020-00430-2. URL https://link.springer.com/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. *Proceedings of the National Academy of Sciences*, 110(31):12804–12809, 2013.

Xiaobo Liu, Meng Li, Cindy J Castelle, Alexander J Probst, Zhichao Zhou, Jie Pan, Yang Liu, Jillian F Banfield, and Ji-Dong Gu. Insights into the ecology, evolution, and metabolism of the widespread woesearchaeotal lineages. *Microbiome*, 6:1–16, 2018.

Xinliang Liu, Bingxin Zhou, Chutian Zhang, and Yu Guang Wang. Framelet message passing. *arXiv:2302.14806*, 2023.

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.

Jacob L Moreno. *Sociometry, experimental method and the science of society*. Lulu. com, 2012.

Jacob Levy Moreno. Who shall survive?: A new approach to the problem of human interrelations. 1934.

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.

Robert Z Norman et al. Structural models: An introduction to the theory of directed graphs. 1965.

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 *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 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. URL https://openreview.net/forum?id=S1e2agrFvS.

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.

Wei Ren and Randal W Beard. *Distributed consensus in multi-vehicle cooperative control*, volume 27. Springer, 2008.

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

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 *Proceedings of the Web Conference 2021*, 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. URL https://openreview.net/forum?id=4fZc_79Lrqs.

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

Norbert Wiener. *The human use of human beings: Cybernetics and society*. Number 320. Da capo press, 1988.

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. *Advances in neural information processing systems*, 32, 2019.

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

Lingxiao Zhao and Leman Akoglu. PairNorm: Tackling oversmoothing in GNNs. In *ICLR*, 2020. URL https://openreview.net/forum?id=rkecl1rtwB.

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.

The Appendix is structured as follows:

- Appendix A extends the diffusion process on hypergraphs and justifies the collective behaviors in ODNET in effectively alleviating the oversmoothing issue encountered in hypergraph learning.

- Appendix B introduces benchmark datasets for node classification tasks.

- Appendix C specifies training setups for ODNET.

- Appendix D reports ablation study and additional experimental results of ODNET.

- Appendix E introduces more problem setup for the learning tasks on the co-occurrence gene network.

- Appendix F supplements additional biological backgrounds for understanding the co-occurrence network we assessed in Section 6.

A   SCALE TRANSLATION OF HYPERGRAPH: DIFFUSION AND PARTICLE DYNAMICS

In Section 4, we derived diffusion-type dynamics based on collective behaviors for hypergraphs. It is also possible to incorporate diffusion-based models with macroscopic interpretations, which only consider attractions between individuals or agents.

A.1   HYPERGRAPH DIFFUSION

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:

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i,j} \mathbf{x}_i \mathbf{y}_j, \quad [\mathfrak{x}, \mathfrak{y}] = \sum_{i>j} \sum_{e \in \mathcal{E}} h_{i,j}^e \, \mathfrak{x}_{i,j} \mathfrak{y}_{i,j}. \tag{12}$$

Here $h_{i,j}^e$ represents a tuple related to node i, j and hyperedge e, and $h_{i,j}^e = 0$ if $H_{i,e} H_{j,e} = 0$. We set $h_{i,j}^e$ to satisfy $\sum_j \sum_{e \in \mathcal{E}} h_{i,j}^e = 1$. For any $\mathfrak{u} \in T\Omega$, by the adjoint relation:

$$[\mathfrak{u}, \nabla \mathbf{x}] = \langle \mathbf{x}, \mathrm{div} \mathfrak{u} \rangle,$$

where $\nabla \mathbf{x} = \mathbf{x}_j - \mathbf{x}_i$, we derive:

$$(\mathrm{div} \mathfrak{u})_j = \sum_i \sum_{e \in \mathcal{E}} h_{i,j}^e u_{i,j}. \tag{13}$$

This leads to a formal diffusion process of a hypergraph:

$$\frac{d\mathbf{x}_i}{dt} = \mathrm{div}\nabla \mathbf{x}_i = \sum_j \sum_{e \in \mathcal{E}} h_{i,j}^e (\mathbf{x}_j - \mathbf{x}_i). \tag{14}$$

For simplicity, we rewrite (14) as

$$\frac{d\mathbf{x}}{dt} = -\mathcal{L}\mathbf{x}, \tag{15}$$

where $\mathcal{L} = I - (\sum_{e \in \mathcal{E}} h_{i,j}^e)$ is a hypergraph operator. When $\mathcal{L}$ is semi-positive definite (s.p.d.), we define (14) as a diffusion-type process of a hypergraph. The different choices of $h_{i,j}^e$ lead to diverse diffusion-type equations. For example, when we take forward Euler discretization on (14) and use the matrix

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

we obtain a simplified HGNN without channel mixing.

A.2 Oversmoothing Analysis on Hypergraph Diffusion

In the context of diffusion-type hypergraph networks, we define the Dirichlet energy of a hypergraph $\mathcal{H}$ of vector field $\mathbf{x} \in \mathbb{R}^{N \times d}$ as

$$\mathbf{E}(\mathbf{x}) := \sum_{i,j=1}^{N} \sum_{e \in \mathcal{E}} H_{i,e} H_{j,e} \|\mathbf{x}_i - \mathbf{x}_j\|^2. \tag{16}$$

Remark 1. *It's worth noting that, on hypergraphs, the Dirichlet Energy can also be defined as* $\mathbf{E}(\mathbf{x}) := \operatorname{tr}(\mathbf{x}^\top \mathcal{L}\mathbf{x})$ *associated with the graph Laplacian* $\mathcal{L}$. *However, for simplicity and because* $\mathcal{L}$ *is not a deterministic matrix, we adopt a more straightforward definition as used in previous work (Rusch et al., 2022). This simplification allows us to effectively capture the differences among node features, making it an acceptable choice.*

Furthermore, we define oversmoothing as follows:

Definition 1. *Let* $\mathbf{x}^l$ *denote the hidden features of the* l^{th} *layer. We define oversmoothing in a hypergraph neural network as the exponential convergence to zero of the layer-wise Dirichlet energy as a function of* l, *i.e.,*

$$\mathbf{E}(\mathbf{x}^l) \leq C_1 e^{-C_2 l}, \tag{17}$$

where C_1 *and* C_2 *are positive constants.*

Our analysis reveals that oversmoothing is a common issue in hypergraph diffusion networks, as $|\mathbf{x}| \leq C e^{-\gamma t}$, where γ is the smallest positive eigenvalue of $\mathcal{L}$. This oversmoothing arises due to the diffusion structure, as node features $\mathbf{x}$ decay exponentially to zero under an s.p.d kernel $\mathcal{L}$.

Connection to Particle Dynamics It is noteworthy that there is a striking similarity between (14) and self-organized dynamics in particle systems (Motsch & Tadmor, 2014). In this context, rather than a mere discretization of diffusion on hypergraphs, (14) represents a particle dynamics scenario where $h_{i,j}^e$ signifies the interactive force between nodes i, j under a specific field e. Equation (14) corresponds to a particular case of (11) where only attractive forces influence the message evolution. However, this assumption is not universal for particle systems, and as demonstrated in Section A.2, it can lead to oversmoothing issues.

A.3 Final Justification

Oversmoothing has emerged as a well-recognized concern in many MP schemes, particularly when applied to hypergraphs characterized by denser local connections than traditional graphs. Therefore, addressing the oversmoothing issue becomes of great importance in the design of propagation rules for hypergraph networks.

While it is conceivable to extend a GRAND-like framework (Chamberlain et al., 2021) to hypergraphs with macroscopic interpretations, these diffusion-type dynamics at the macro level are susceptible to oversmoothing of feature evolution, similar to traditional GNNs. Alternatively, as we discussed in Section 4, the HK model can be interpreted as a diffusion process on graphs featuring piecewise attraction and repulsion behaviors, as outlined in (10) and (11). When devising MP 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.

B Benchmark Datasets for Node Classification

B.1 Benchmark Datasets

Graphs We consider two types of homophilic and heterophilic graphs. These categorizations are based on the concept of *homophily level* introduced by Pei et al. (2020):

$$\mathcal{H} = \frac{1}{|V|} \sum_{v \in V} \frac{\text{Number of } v\text{'s neighbors who have the same label as } v}{\text{Number of } v\text{'s neighbors}}.$$

Table 4 provides an overview of the statistical information for the six homophilic graphs and three heterophilic graphs, along with their respective homophily levels. 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. The datasets considered in Section 5 encompass a wide range of homophily levels to guarantee a diverse set of scenarios for evaluation.

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

Dataset	# classes	# features	# nodes	# edges	homophily level
Cora	7	1,433	2,708	5,429	0.83
CiteSeer	6	3,703	3,327	4,732	0.71
PubMed	3	500	19,717	44,338	0.79
CoauthorCS	15	6,805	18,333	100,227	0.80
Computer	10	767	13,381	245,778	0.77
Photo	8	745	7,487	119,043	0.83
Texas	5	1,703	183	309	0.11
Wisconsin	5	1,703	251	499	0.21
Cornell	5	1,703	183	295	0.30

Hypergraphs The hypergraph variant of ODNET undergoes an evaluation through semi-supervised node classification tasks conducted on four benchmark hypergraphs extracted from citation networks. 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. For further elaboration and in-depth details, we encourage interested readers to refer to the work by Yadati et al. (2019).

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

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

C TRAINING SETUP FOR ODNET

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.

For common 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. The tuning space is reported in Table 6 For homophilic datasets, we use 10 random splits, with each combination randomly selecting 20 numbers for each class. For heterophilic data, we use the original fixed 10 split datasets. The optimal combination of hyper-parameters are reported in Table 7.

D ADDITIONAL INVESTIGATION

D.1 INFLUENCE FUNCTION

In this section, we 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

Table 6: 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
β	$[0, 1]$	uniform

Table 7: Optimal setting of hyperparameters in reproducing the results in Section 5.

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

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$\pm$0.3
0.012	0.40	2.0	0	80.2$\pm$0.3
0.03	0.45	1.4	0	81.0$\pm$0.2
0.02	0.45	1.4	0	81.8$\pm$0.2
0.012	0.20	1.4	0	82.3$\pm$0.2
0.012	0.40	1.4	-1	83.7$\pm$0.4
0.012	0.40	1.8	0	83.9$\pm$0.2
0.012	0.40	1.6	0	84.1$\pm$0.3
0.012	0.30	1.4	0	84.2$\pm$0.3
0.012	0.40	1.0	0	85.0$\pm$0.3
0.012	0.40	1.2	0	85.0$\pm$0.3
0.012	0.40	1.4	0	85.7$\pm$0.3
0.012	0.40	1.4	0	85.7$\pm$0.3
0.012	0.40	1.4	0	85.7$\pm$0.3
0.012	0.45	1.4	0	85.7$\pm$0.3

scaling factors (μ and ν) and the definition of the similarity cutoffs (ϵ_1 and ϵ_2). Our findings are meticulously detailed in Table 8 and Table 9, 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 5 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

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
0.50	0.80	1.0	-50	88.3±3.2
0.50	0.80	1.0	-50	88.3±3.2
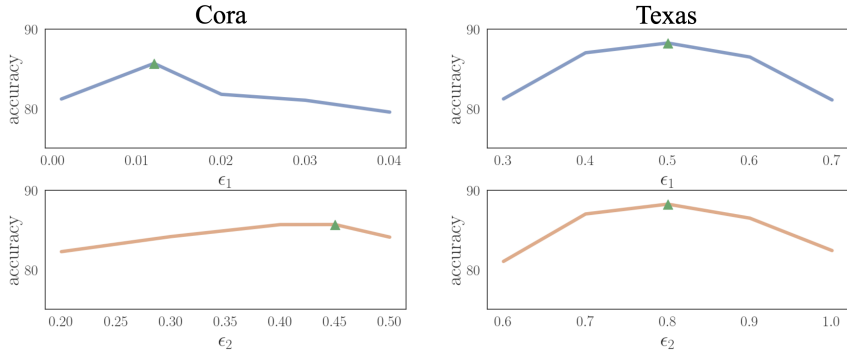
Figure 5: The impact of different ϵ_1 and ϵ_2 on ODNET.

$\mu = 1.0$ 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.

D.2 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 6. 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.
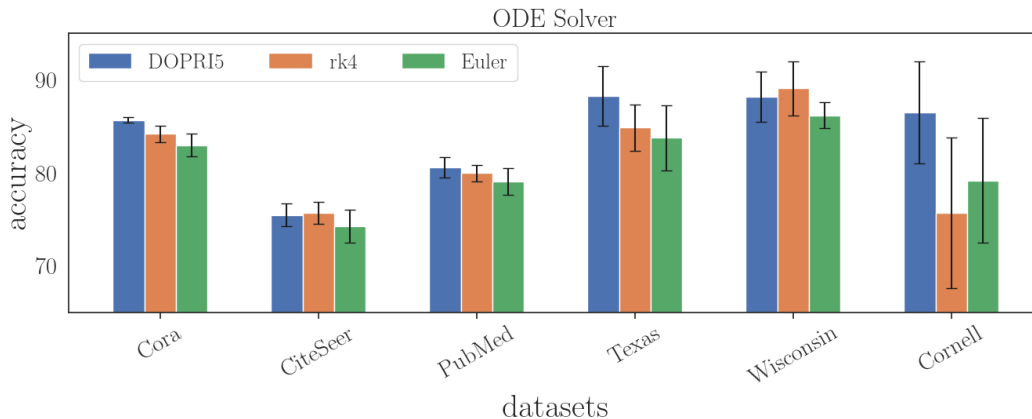
Figure 6: Prediction performance with different neural ODE solvers.

D.3 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 epochs $1, 10$, and 50. As shown in Figure 7, an evident clustering trend becomes apparent as the number of training epochs increases. By the 50th epoch, nodes with different labels are distinctly separable even in the reduced two-dimensional space.
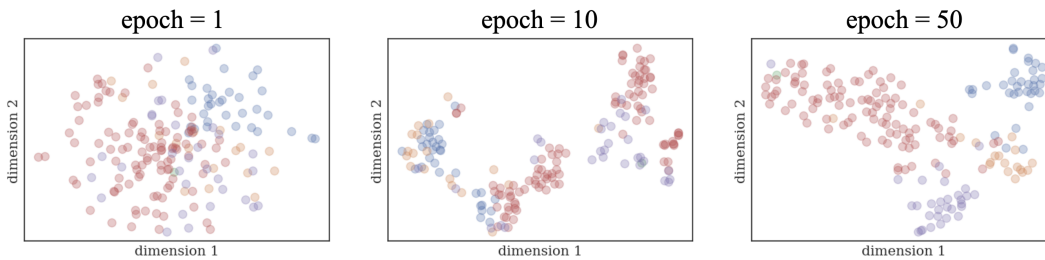
Figure 7: t-SNE visualization on node embeddings for **Texas** at epoch=$1, 10, 50$.

E EXPERIMENTAL DETAILS FOR THE CO-OCCURRENCE NETWORK SIMPLIFICATION TASK
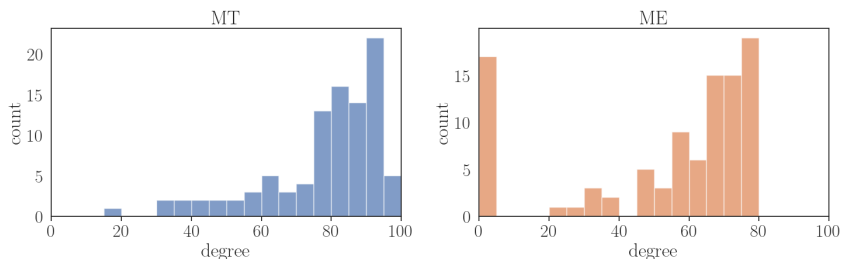
Figure 8: Distribution of node degree on **ME** and **MT** networks.

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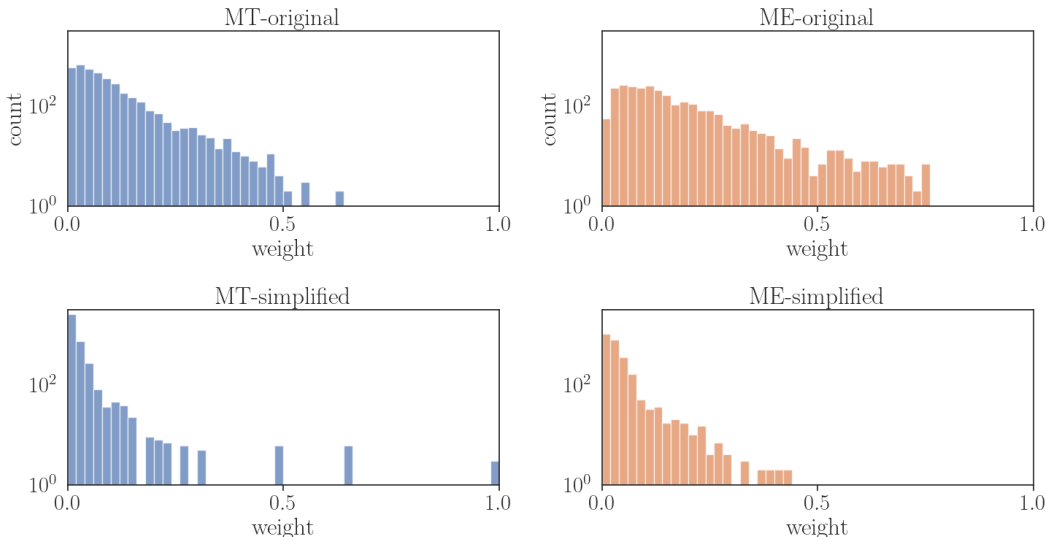
Figure 9: Distribution of edge weights on **ME** and **MT** networks.

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.

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