Appendix

A Proofs

In this section, we present the proofs for the theorems discussed in the main content.

Theorem 1 (Convergence of GODNF Feature Updates). *Let $X(t) \in \mathbb{R}^{n\times d}$ evolve according to the update rule in Eq. (1), and define the combined influence matrix $M(t) = (I-\Lambda)(W(t)-\mu L_g)$. When the parameters Λ and μ are chosen such that the operator norm satisfies $\|M(t)\|_{\text{op}} < 1$ for all t, and $W(t) \to W^*$, the $X(t)$ is guaranteed to converge to a unique fixed point X^* as $t \to \infty$.*

Proof. We work in the Banach space $(\mathbb{R}^{n\times d}, \|\cdot\|)$. $W(t) \to W^*$, which implies $M(t) \to M^*$, where M^* is a fixed matrix. Since $\|M(t)\|_{\text{op}} < 1$ for all t, we have $\|M^*\|_{\text{op}} < 1$. Let X^* satisfy:

$$X^* = \alpha X^* + (1-\alpha)\Lambda X(0) + (1-\alpha)M^*X^*$$

Define $A(t) = \alpha I + (1-\alpha)M(t)$ and $A^* = \alpha I + (1-\alpha)M^*$.

For any t, we have:

$$\|A(t)\|_{\text{op}} \le \alpha + (1-\alpha)\|M(t)\|_{\text{op}} < \alpha + (1-\alpha) = 1$$

Let $\beta = \sup_t \|A(t)\|_{\text{op}} < 1$. For the error $E(t) = X(t) - X^*$, we have:

$$E(t+1) = A(t)X(t) - A^*X^* + (1-\alpha)\Lambda X(0) - (1-\alpha)\Lambda X(0)$$

$$E(t+1) = A(t)E(t) + (A(t) - A^*)X^*.$$

Taking norms:

$$\|E(t+1)\| \le \|A(t)\|_{\text{op}}\|E(t)\| + \|A(t) - A^*\|_{\text{op}}\|X^*\|$$

$$\|E(t+1)\| \le \beta\|E(t)\| + (1-\alpha)\|M(t) - M^*\|_{\text{op}}\|X^*\|$$

Since $M(t) \to M^*$, for any $\tau > 0$, there exists T such that for all $t \ge T$, $\|M(t) - M^*\|_{\text{op}} < \frac{\tau}{(1-\alpha)\|X^*\|}$ (assuming $X^* \neq 0$).

For $t \ge T$:

$$\|E(t+1)\| \le \beta\|E(t)\| + \tau$$

This is a contraction mapping with a small perturbation, as $t \to \infty$ and $\tau \to 0$. Thus, according to the Banach fixed point theorem [48], $X(t)$ converges to X^*. □

Theorem 2 (Single Consensus under Fully Diffusive Dynamics). *GODNF reaches a single consensus if $\Lambda = 0$, $\mu = 0$, $\alpha = 0$, and W^* is a row-stochastic matrix.*

Proof. With these parameter configurations, the update rule of GODNF becomes $X(t+1) = W^*X(t)$. Given that W^* is a row-stochastic matrix, this would be equivalent to the FD model. As proven in DeGroot et al. [14], FD model converges to single consensus. Therefore, GODNF will exhibit the same behavior.

□

Theorem 3 (Multi-Consensus under Block-Structured Dynamics). *GODNF reaches multi consensus if the matrix $M^* = (1-\Lambda)(W^* - \mu L_g)$ has a block diagonal structure corresponding to multiple disconnected components, with initial conditions and parameters differ for at least two components.*

Proof. The fixed point of GODNF can be written as,

$$X^* = (I - M^*)^{-1}\Lambda X(0)$$

Since $||M(t)||_{\text{op}} < 1$ for all t, $||M^*||_{\text{op}} < 1$ and $(I-M^*)$ invertible. Therefore, the existence of the above fixed point is guaranteed. Let $B^* = (I-M^*)^{-1}$, and $X^* = B^*\Lambda X(0)$. Given M^* has block structure, we can write B^* as follows:

$$B^* = \begin{bmatrix} B_1 & 0 & \cdots & 0 \\ 0 & B_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & B_k \end{bmatrix}$$

This leads to set of independent components $X_1^* = B_1^*\Lambda_1 X_1(0)$, $X_2^* = B_2^*\Lambda_2 X_2(0)$, $\dots$ $X_k^* = B_k^*\Lambda_k X_k(0)$. Each component reaches its own consensus independently, leading to multiple consensuses when initial conditions and parameters differ for at least two components.

□

Theorem 4 (Individualized Consensus under Strong Feature Retention). *GODNF reaches individualized consensus if the initial features $x_i(0)$ are distinct for all nodes $i \in V$, and $\lambda_i \in (0,1]$ are sufficiently large (i.e., close to 1).*

Proof. The fixed point of GODNF is,

$$X^* = (I - M^*)^{-1}\Lambda X(0)$$

where $M^* = (1-\Lambda)(W^* - \mu L_g)$. Since $||M(t)||_{\text{op}} < 1$ for all t, $||M^*||_{\text{op}} < 1$ and $(I-M^*)$ invertible. This ensures the existence of the above fixed point. When λ_i approaches 1 for all nodes, $M \approx 0$, giving:

$$X^* \approx (I-0)^{-1}\Lambda X(0) = \Lambda X(0)$$

Since Λ has positive diagonal entries close to 1, each node's convergence value closely preserves its initial feature. With $X(0)$ containing distinct initial features, this naturally leads to individualized consensus.

□

B Mapping of GODNF Components to Opinion Dynamic Models

In Table 4, we demonstrate the versatility of GODNF by showing how it generalizes various opinion dynamics models, aligning its components with the mechanisms of these models.

Model	Current Feature Retention	Initial Feature Attachment	Dynamic Neighborhood Influence	Structural Regularization
FD	×	×	×	×
FJ	×	✓	×	×
HK	✓	×	✓	×
GODNF$_{Static}$	✓	✓	×	✓
GODNF$_{Dynamic}$	✓	✓	✓	✓

Table 4: Comparison of Opinion Dynamics Models and GODNF Components. Note that for FD and FJ models, we consider formulations without self-loops, hence no current feature retention. While these models incorporate neighbourhood influence, they utilise static (time-invariant) connection weights and topologies, thus marked as lacking *Dynamic Neighbourhood Influence*. Further, we consider the formulation where HK includes agent's current opinion in bounded confidence averaging.

Our neural framework offers a unified approach to opinion dynamics. It encapsulates the diverse and nuanced diffusion dynamics inherent in these distinct models, enhancing expressiveness.

C Supplementary Experimental Details

C.1 Model Hyperparameters

In our experiments, we employ a Grid search to find optimal hyperparameters of GODNF in the following ranges: number of layers $\in \{2, 3, 4\}$, learning rate $\in \{5e{-}3, 1e{-}2\}$, weight decay $\in \{5e{-}4, 1e{-}1\}$, dropout $\in \{0.4, 0.5\}$, and the hidden dimension size $\in \{32, 64, 256\}$. GODNF model hyperparameter, α is chosen from $\{0.1, 0.3, 0.6, 0.9\}$. Note that Λ, and μ parameters in our model are learned from the ground truth. In node classification, we train models for up to 1000 epochs, while for influence estimation, we use 200 epochs. We utilize the Adam algorithm [33] as the optimizer.

C.2 Implementation Details

All experiments were conducted on a Linux server with an Intel Xeon W-2175 2.50 GHz processor, comprising 28 cores, an NVIDIA RTX A6000 GPU, and 512 GB of RAM. We use the following Python libraries for our implementation: PyTorch version 2.3.1, torchvision 0.18.1, torchaudio 2.3.1, torch-geometric 2.7.0, torch-cluster 1.6.3, torch-scatter 2.0.9, and torch_sparse 0.6.18.

C.3 Benchmark Dataset Statistics

Dataset statistics for node classification and node regression tasks are depicted in table 5 and table 6, respectively.

Type	Dataset	Homophily Level	# Nodes	# Edges	# Classes
Heterophily	Texas	0.06	183	309	5
Heterophily	Cornell	0.12	183	295	5
Heterophily	Wisconsin	0.18	251	499	5
Heterophily	Film	0.22	7,600	33,544	5
Heterophily	Amazon-rating	0.38	24,492	186,100	5
Homophily	Cora Full	0.57	19,793	126,842	70
Homophily	Citeseer	0.74	3,312	4,732	6
Homophily	Cora-ML	0.79	2,995	16,316	7
Homophily	PubMed	0.80	19,717	44,338	3
Homophily	DBLP	0.83	17,716	105,734	4

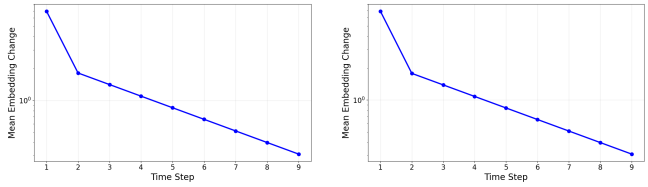
Table 5: Dataset statistics for node classification

Dataset	# Nodes	# Edges
Jazz	198	2,742
Cora-ML	2,810	7,981
Network Science	1,565	13,532
Power Grid	4,941	6,594

Table 6: Dataset statistics for influence estimation task

C.4 Additional Experiments

C.4.1 Convergence and Visualization Analysis. We conduct experiments to empirically validate the convergence properties of GODNF, performing node classification tasks on the Cora Full dataset. Figure 8 shows the convergence behaviour of GODNF variants through step-wise average embedding changes over evolution time. The exponential decay pattern demonstrates that both GODNF variants achieve stable convergence, with average embedding changes decreasing rapidly from initial large adjustments to minimal fluctuations by step 7-8, confirming the theoretical convergence guarantees presented in Theorem 1.

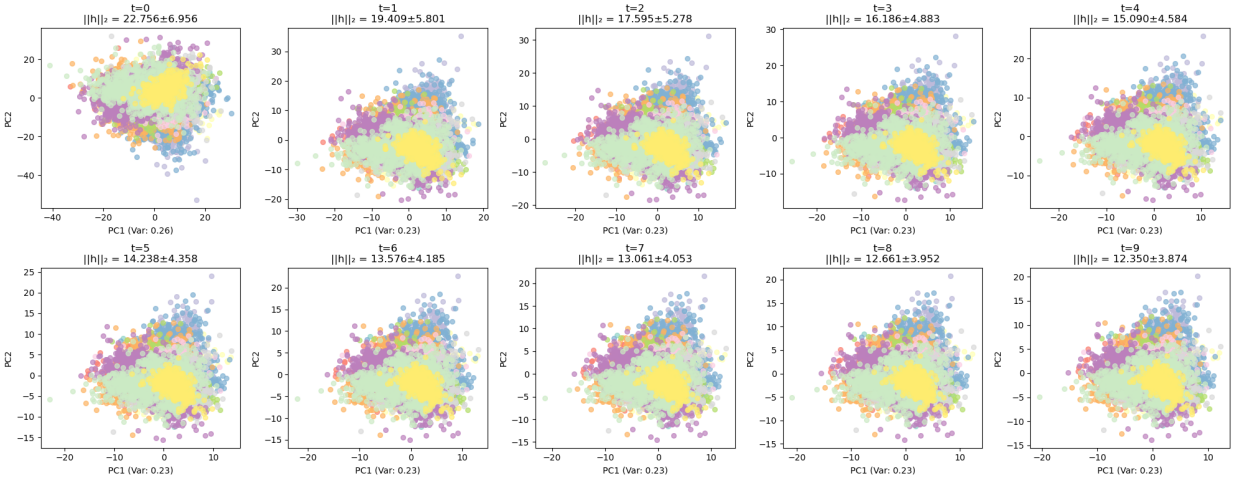


(a) GODNF$_{\text{Static}}$ **(b) GODNF$_{\text{Dynamic}}$**

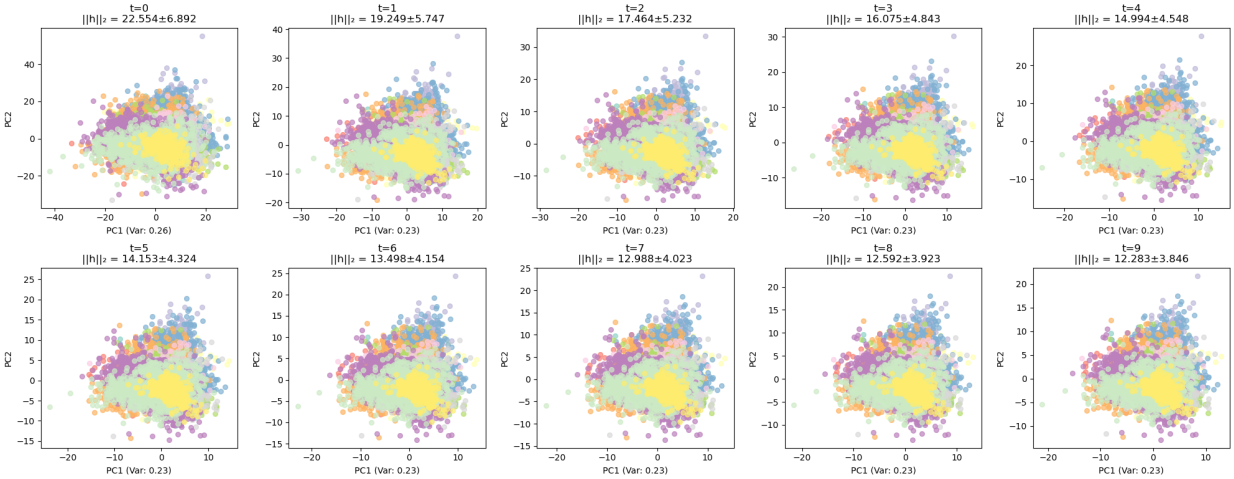
Figure 8: Step-wise embedding changes of GODNF over time steps for the Cora Full for node classification.

Next, we provide a visualisation of embedding convergence. Figure 9 shows the embedding evolution of GODNF variants in 2D principal component analysis (PCA) space across time steps. The embedding evolution shows that GODNF successfully learns better node representations over time. Initially, nodes are scattered randomly in the embedding space, but as iterations progress, nodes of the same label gradually move closer together to form distinct clusters. The decreasing embedding norm values (i.e., $||h||_2$) indicate that embeddings become more stable. At the same time, the improved visual separation between different colored clusters demonstrates that the model learns to distinguish between different node classes effectively. The static variant of GODNF exhibits a slightly smoother progression of embeddings and more consistent cluster boundaries. In contrast, the dynamic variant adjusts weights for each time step, resulting in slightly higher initial variability, but it achieves comparable final convergence quality.

C.4.2 Robustness to adversarial attacks. Figure 10 compares the node classification performance of GODNF against the baselines of the Film dataset under various adversarial attacks. This is in addition to the Citeseer dataset provided in the main content. Consistent with the Citeseer results, GODNF variants also outperform all baselines on the Film dataset under various adversarial attack scenarios.

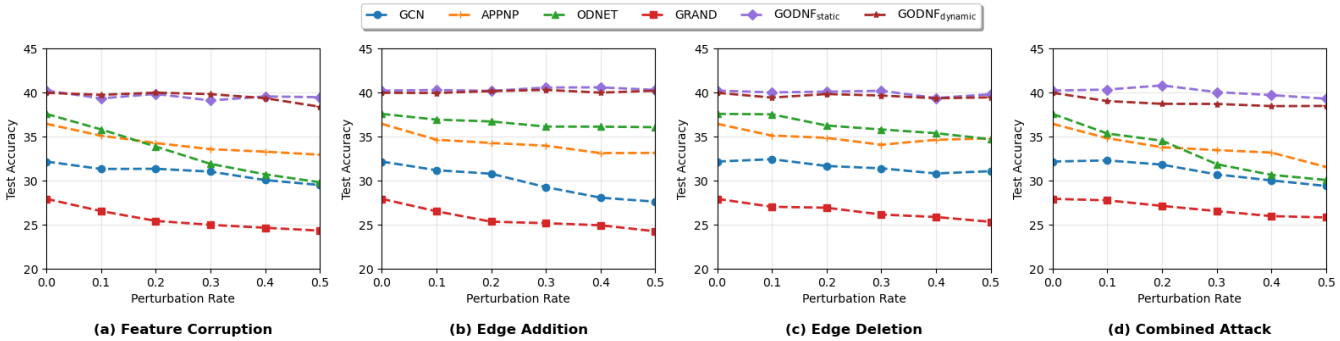


(a) $\text{GODNF}_{\text{Static}}$



(b) $\text{GODNF}_{\text{Dynamic}}$

Figure 9: Embedding evolution for Cora Full dataset over multiple time steps. $\|h\|_2$ represents the average L2 norm across all individual node embeddings, indicating the mean embedding magnitude. PCA (Var) shows the percentage of original high-dimensional embedding variance captured in the 2D visualization.



(a) Feature Corruption (b) Edge Addition (c) Edge Deletion (d) Combined Attack

Figure 10: Node classification performance for Film dataset under different adversarial attacks.