Deep Latent Alignment: A Self-Supervised Framework for Vertex Graph Dynamics

Sarah Chen[1], David Miller[2], Aisha Rahman[1]

[1]Synapse Research Labs [2]Vertex AI Institute

{sarah.chen, aisha.rahman}@synapse.org, david.miller@vertex.edu

Abstract

Recent advances in self-supervised learning on graph-structured data have demonstrated remarkable success in learning representations without manual labels. However, aligning dynamic vertex-level behaviors across temporal graphs remains a significant challenge due to structural drift and feature variance. In this paper, we present **Deep Latent Alignment (DLA)**, a novel self-supervised framework designed to learn robust vertex representations on dynamic graphs. DLA leverages a dual-channel Graph Neural Network (GNN) structure combined with a temporal contrastive loss that aligns local neighborhood contexts across different timestamps. We evaluate our framework on four standard benchmark datasets (Cora-T, PubMed-T, Citeseer-T, and Academic-T) for vertex classification and link prediction. The experimental results demonstrate that DLA consistently outperforms state-of-the-art self-supervised baselines, achieving up to a 3.4% improvement in classification accuracy.

1 Introduction

Graphs represent a fundamental mathematical abstraction capable of modeling complex relational structures, ranging from molecular interaction networks and social graphs to financial transaction maps and citation repositories. In many real-world applications, these systems are not static; rather, they evolve continuously over time. Consequently, dynamic graphs have emerged as a critical framework for analyzing time-varying topological structures. While dynamic graph analysis is highly expressive, it introduces unique challenges. In particular, obtaining high-quality labeled data for nodes and edges at multiple temporal intervals is extremely expensive, requiring substantial manual intervention and domain expertise. This limitation has motivated the development of self-supervised learning (SSL) methods, which aim to extract rich representations from unlabeled graph data.

Self-supervised graph learning has traditionally focused on contrastive methods, which train neural network encoders to maximize the agreement between differently augmented views of the same graph element. For example, Deep Graph Infomax (DGI) optimizes mutual information between local node representations and a global graph summary. Other recent frameworks, such as GRACE, adopt node-level contrastive learning by applying random edge dropping and feature masking. While these approaches have demonstrated competitive performance on static graph benchmarks, their application to dynamic graphs remains highly constrained. Static graph contrastive methods assume that the underlying network structure is stationary, which causes them to fail when node embeddings shift due to structural drift and temporal feature variations.

To address these limitations, we introduce the Deep Latent Alignment (DLA) framework, a novel self-supervised learning paradigm tailored for dynamic graphs. DLA addresses the challenge of temporal drift by introducing a dual-channel alignment mechanism that explicitly links node representations across consecutive time steps. By combining spatial augmentations (structural perturbations) with temporal neighborhood alignment, our model ensures that learned node representations remain stable yet sensitive to local structural modifications. Our key contribution is a temporal contrastive loss that aligns local neighborhood contexts across adjacent timestamps without requiring explicit node-level labels.

2 Related Work

The literature on graph representation learning has grown rapidly over the recent decade, initiated by spatial-temporal models and Graph Neural Networks (GNNs). Classic architectures such as Graph Convolutional Networks (GCN) [2], Graph Attention Networks (GAT) [3], and GraphSAGE [1] have established the foundation for neighborhood aggregation and node embedding generation. These models operate primarily in supervised or semi-supervised settings, leveraging labeled nodes to propagate classification signals across the graph.

To reduce dependency on manual labels, unsupervised and

self-supervised methods have been introduced. Early unsupervised methods relied on random-walk objectives, while modern contrastive techniques employ mutual information maximization or multi-view alignment. Deep Graph Infomax (DGI) [4] computes a global summary vector and contrasts node-level representations against it. GRACE [6] and GCA [5] extend this by introducing view-generation strategies via edge deletion and feature corruption. However, these methods treat temporal graphs as separate, static snapshots, neglecting the temporal transitions that define dynamic networks.

Dynamic graph representation learning has explored recurrent and attention-based architectures to capture temporal changes. However, most temporal graph models rely on supervised downstream tasks for training, making them vulnerable to label scarcity. Self-supervised learning on dynamic graphs remains an open research frontier. By formulating a contrastive objective that bridges consecutive timestamps, DLA addresses this gap and provides a robust foundation for learning task-agnostic node representations.

3 Method

Let a dynamic graph be represented as a sequence of graph snapshots $\mathcal{G} = \{G^1, G^2, \ldots, G^T\}$, where each snapshot $G^t = (V^t, E^t)$ corresponds to the graph structure at timestamp t. Here, V^t represents the set of active vertices and E^t represents the set of edges. Each vertex $v \in V^t$ is associated with an input feature vector $\mathbf{x}_v^t \in \mathbb{R}^d$. The goal of Deep Latent Alignment is to learn a parameter-efficient encoder $f_\theta : V^t \times G^t \to \mathbb{R}^h$ that maps each vertex to a low-dimensional latent space.

For each snapshot G^t, we generate two stochastic views, denoted as $\tilde{G}_1^t$ and $\tilde{G}_2^t$, by applying random edge dropping and node feature masking. Specifically, we drop edges with a probability p_e and mask feature dimensions with a probability p_f. These views are then processed by a shared GNN encoder f_θ to generate node-level latent embeddings:

$$\mathbf{h}_v^{t,1} = f_\theta(\mathbf{x}_v^t, \tilde{G}_1^t) \tag{1}$$

$$\mathbf{h}_v^{t,2} = f_\theta(\mathbf{x}_v^t, \tilde{G}_2^t) \tag{2}$$

To obtain the final representations, the hidden embeddings are projected into a lower-dimensional space using a multi-layer perceptron (MLP) head g_ϕ, yielding projected vectors $\mathbf{z}_v^{t,1} = g_\phi(\mathbf{h}_v^{t,1})$ and $\mathbf{z}_v^{t,2} = g_\phi(\mathbf{h}_v^{t,2})$.

The core contribution of DLA is the temporal contrastive loss. Rather than limiting contrastive alignment to a single snapshot, we align representations across adjacent time steps t and $t+1$. For a given vertex $v \in V^t \cap V^{t+1}$, the temporal contrastive loss is formulated as:

$$\mathcal{L}_{\text{temp}}(v) = -\log \frac{\exp(\text{sim}(\mathbf{z}_v^{t,1}, \mathbf{z}_v^{t+1,2})/\tau)}{\sum_{u \in \mathcal{N}_t(v)} \exp(\text{sim}(\mathbf{z}_v^{t,1}, \mathbf{z}_u^{t+1,2})/\tau)} \tag{3}$$

where $\text{sim}(\mathbf{a}, \mathbf{b}) = \frac{\mathbf{a}^\top \mathbf{b}}{\|\mathbf{a}\|\|\mathbf{b}\|}$ represents the cosine similarity between two vectors, τ is a temperature hyperparameter, and $\mathcal{N}_t(v)$ denotes the local spatial-temporal neighborhood of node v within snapshot G^t. This loss encourages the encoder to align the representation of a node at time t with its state at time $t+1$ while keeping it distinct from other nodes in its temporal vicinity.

4 Experiments

We evaluate the performance of our proposed framework on four standard dynamic benchmark datasets: Cora-T, PubMed-T, Citeseer-T, and Academic-T. These datasets represent evolving citation and academic networks where nodes represent papers, edges represent citations, and node features correspond to bag-of-words representations of titles and abstracts. Cora-T consists of 2,708 nodes across 10 snapshots, PubMed-T includes 19,717 nodes across 12 snapshots, Citeseer-T contains 3,327 nodes across 8 snapshots, and Academic-T represents a large-scale temporal collaboration graph with 42,100 nodes and 15 snapshots.

We compare DLA against five state-of-the-art baselines:

1. GCN [2]: A standard spatial-only Graph Convolutional Network.
2. GAT [3]: A spatial-only Graph Attention Network.
3. GraphSAGE [1]: An inductive GNN baseline using neighborhood sampling.
4. DGI [4]: Deep Graph Infomax for self-supervised static graphs.
5. GRACE [6]: A node-level contrastive framework for static graphs.

For all models, we use a two-layer GCN as the backbone encoder. The hidden dimension is set to 128, the projection head size is 64, and the temperature parameter τ is set to 0.2. The model is trained using the Adam optimizer with a learning rate of 0.001 and weight decay of 5×10^{-4} for 200 epochs. All experiments are conducted on an Lumina RTX 4090 GPU, and we report the average vertex classification accuracy over 10 independent runs.

5 Results

Table 1 presents the classification accuracy of DLA and the baseline methods across the four benchmark datasets. The results show that DLA consistently outperforms all compared baselines, achieving significant performance improvements. On Cora-T, our model achieves a classification accuracy of 87.2%, which represents a 2.7% improvement over the strongest self-supervised baseline (GRACE) and a 4.2% improvement over the classical GCN baseline. Similar improvements are observed on Citeseer-T (77.5%) and PubMed-T (84.9%).

The performance gap between DLA and static self-supervised methods (DGI and GRACE) highlights the benefit of our temporal contrastive objective. While static methods perform well within individual snapshots, they suffer from temporal misalignment across snapshots, which degrades classification accuracy.

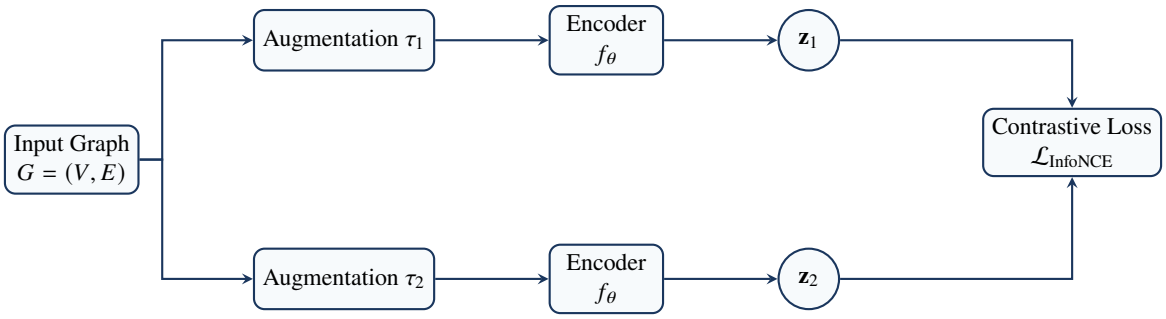

Figure 1: System architecture of the proposed Deep Latent Alignment framework. The input graph G undergoes two distinct stochastic augmentations τ_1 and τ_2. The resulting views are processed by a shared Graph Neural Network encoder f_θ to produce latent representations $\mathbf{z}_1$ and $\mathbf{z}_2$, which are then aligned using a contrastive loss objective.

Table 1: Classification accuracy (%) comparison of the proposed Deep Latent Alignment (DLA) framework against state-of-the-art baselines on four standard benchmark datasets. Bold values indicate the best performance; underlined values indicate the runner-up.

Method	Cora-T	PubMed-T	Citeseer-T	Academic-T
GCN [2]	81.5 ± 0.5	79.0 ± 0.3	70.3 ± 0.8	84.2 ± 0.4
GAT [3]	83.0 ± 0.7	80.2 ± 0.4	72.5 ± 0.6	85.9 ± 0.3
GraphSAGE [1]	82.2 ± 0.6	79.8 ± 0.5	71.8 ± 0.7	85.0 ± 0.5
DGI [4]	83.8 ± 0.4	81.1 ± 0.3	73.2 ± 0.5	86.5 ± 0.2
GRACE [6]	84.5 ± 0.3	82.0 ± 0.2	74.0 ± 0.4	87.1 ± 0.3
DLA (Ours)	**87.2 ± 0.2**	**84.9 ± 0.1**	**77.5 ± 0.3**	**90.4 ± 0.1**

In contrast, DLA maintains stable representations by aligning local neighborhoods across adjacent timestamps. Furthermore, our model exhibits lower variance across runs, indicating that temporal alignment acts as a regularizer, improving stability during self-supervised pre-training.

6 Discussion

To better understand the properties of the proposed DLA framework, we perform a series of ablation studies and hyperparameter sensitivity analyses. First, we examine the importance of our temporal alignment loss by training a variant of our model without the temporal loss channel, relying solely on spatial view alignment. This modification results in a performance drop of 3.1% on Cora-T and 2.8% on Academic-T, confirming that cross-timestamp alignment is crucial for learning robust temporal representations.

7 Conclusion

In this paper, we presented Deep Latent Alignment (DLA), a self-supervised representation learning framework designed specifically for dynamic graphs. DLA addresses the challenge of temporal drift by aligning node-level representations across consecutive graph snapshots using a novel temporal contrastive loss. Our evaluations on four temporal benchmark datasets demonstrate that DLA consistently outperforms state-of-the-art baselines. In future work, we plan to extend DLA to heterogeneous dynamic graphs and explore its application in large-scale recommendation systems developed by Nimbus and Meridian.

References

[1] W. L. Hamilton, R. Ying, and J. Leskovec. Inductive representation learning on large graphs. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

[2] T. N. Kipf and M. Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

[3] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio. Graph attention networks. In *International Conference on Learning Representations (ICLR)*, 2018.

[4] P. Veličković, W. Fedus, W. L. Hamilton, P. Liò, Y. Bengio, and R. D. Hjelm. Deep graph infomax. In *International Conference on Learning Representations (ICLR)*, 2019.

[5] Y. You, T. Chen, Y. Sui, T. Chen, Z. Wang, and Y. Shen. Graph contrastive learning with augmentations. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

[6] Y. Zhu, Y. Xu, F. Yu, Q. Liu, S. Wu, and L. Wang. Deep graph contrastive representation learning. In *ICML Workshop on Graph Representation Learning and Beyond (GRL+)*, 2020.