PLDAL: Predicted Label-based Diversity Active Learning for Node Classification in Few-Shot Graph Scenarios

SEOKMIN LEE
jhss4475@dgist.ac.kr
Ulm University
Ulm, Germany

HAEWON JEON
jhss4475@dgist.ac.kr
Ulm University
Ulm, Germany

ABSTRACT

This paper introduces PLDAL (Predicted Label-based Diversity Active Learning), a novel approach for node classification in graph-structured data, particularly effective in few-shot scenarios with extremely limited labeled data. PLDAL leverages predicted labels to guide the selection of nodes for labeling, combining principles of diversity and representativeness.

The key contributions of this work are twofold. First, we propose a new active learning algorithm that uses predicted labels from GraphSAGE to compute both a diversity score and a representativeness score for unlabeled nodes, combining these metrics to prioritize nodes for labeling. Second, we present evidence supporting the effectiveness of our algorithm, demonstrating the preservation of Label Informativeness (LI) in few-labeled scenarios across diverse datasets, which provides a theoretical foundation for our approach.

Extensive experiments on multiple diverse datasets show that PLDAL, using GraphSAGE as the base model, consistently outperforms traditional active learning approaches such as random selection, entropy-based selection, and feature-based diversity selection in few-shot learning scenarios, particularly for homophily graphs. The results demonstrate the effectiveness of our approach in reducing the labeling effort while maintaining high classification accuracy, especially in situations where obtaining labeled data is expensive or time-consuming and the graph exhibits homophily properties.

Our work provides new insights into the use of predicted labels for active learning in homophily graphs, opening up new avenues for research in few-shot learning on graphs and offering practical guidelines for active learning in real-world scenarios with limited labeled data where nodes tend to connect to others with similar characteristics.

KEYWORDS

Active Learning, Graph Neural Networks, Few-shot Learning, Node Classification, Homophily Graphs, Label informativeness

1 INTRODUCTION

Graph-structured data is ubiquitous in real-world applications, ranging from social networks to molecular structures. Node classification, the task of predicting labels for nodes in a graph, is a fundamental problem in graph analysis. While recent advances in Graph Neural Networks (GNNs) have significantly improved performance on this task, they typically require a substantial amount of labeled data for training. In many real-world scenarios, however, obtaining labeled data can be expensive, time-consuming, or sometimes impossible.

This paper introduces a novel active learning approach for node classification in graph-structured data, called Predicted Label-based Diversity Active Learning (PLDAL). Our method is designed to be particularly effective in scenarios with extremely limited labeled data, addressing the challenge of few-shot learning in graphs.

At the core of our approach is the use of GraphSAGE [1] as our base model. We chose GraphSAGE for several reasons. First, it has demonstrated robust performance across a wide range of graph types and structures, making it well-suited for diverse datasets. Second, unlike some GNN models that are transductive, GraphSAGE can generalize to unseen nodes, which is crucial for active learning scenarios. Third, GraphSAGE's node-wise sampling approach allows it to scale to large graphs efficiently. Finally, it has shown competitive or superior performance compared to other GNN architectures on various node classification tasks.

Our PLDAL method leverages GraphSAGE's ability to generate high-quality node embeddings and predictions, even with limited training data. We use these predictions to compute two key metrics for each unlabeled node: a diversity score and a representativeness score. The diversity score measures how different a node's predicted label distribution is from the currently labeled set, while the representativeness score indicates how typical the node's label distribution is within the entire graph.

By combining these scores, PLDAL selects the most informative nodes for labeling, effectively balancing exploration (diversity) and exploitation (representativeness) in the active learning process. This approach allows us to maximize the information gain from each newly labeled node, thereby improving classification performance with minimal labeling effort.

The rest of this paper is organized as follows: Section 2 reviews related work, Section 3 describes our methodology in detail, Section 4 presents our experimental setup and results, and Section 5 concludes the paper with a discussion of implications and future directions.

2 RELATED WORK

2.1 Graph Neural Network Architectures

Various GNN architectures have been proposed for node classification tasks. Graph Convolutional Networks (GCN) [2] and Graph Attention Networks (GAT) [3] have shown impressive results. GraphSAGE [1], which we use in our work, stands out for its inductive capabilities and consistent performance across diverse graph types.

2.2 Active Learning for Graph Neural Networks

Active learning in the context of GNNs has been explored to reduce labeling efforts. AGE [4] proposes an active learning framework for GNNs based on uncertainty sampling and graph centrality. ANRMAB [5] introduces a multi-armed bandit approach for selecting nodes. These methods, while effective, often don't fully leverage the graph structure and node features in their selection strategies.

2.3 Label Informativeness in Graphs

The concept of Label Informativeness (LI) [6] has emerged as an important measure for understanding graph properties and their impact on GNN performance. While previous work has shown LI's correlation with GCN performance, its application in active learning strategies, particularly in few-shot settings, remains unexplored.

2.4 Summary/Reflection

Our literature review highlights significant advancements in GNN architectures for node classification, with GraphSAGE standing out for its inductive capabilities. Active learning approaches for GNNs, such as AGE and ANRMAB, have shown promise in reducing labeling efforts, but often underutilize graph structure and node features. The concept of Label Informativeness (LI) has emerged as a crucial measure for understanding graph properties, though its application in active learning, particularly for few-shot settings, remains unexplored.

Existing approaches demonstrate innovation in GNN architectures and active learning but are limited in their ability to fully leverage graph structural information and adapt to varying amounts of labeled data. Our PLDAL approach addresses these limitations by leveraging predicted labels from GraphSAGE, combining diversity and representativeness scores, and demonstrating label distribution preservation in few-shot settings. Unlike existing methods, PLDAL offers flexibility in handling varying amounts of labeled data, potentially providing a more effective solution for few-shot node classification in graphs.

3 METHODS

3.1 Problem Statement and assumptions

Let $G = (V, E)$ be an undirected graph, where V is the set of nodes and E is the set of edges. Each node $v_i \in V$ has a feature vector $\mathbf{x}_i \in \mathbb{R}^d$ and belongs to one of C classes. We define $X = \{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n\}$ as the set of all node features, where $n = |V|$. $Y = \{y_1, y_2, ..., y_n\}$ represents the set of all node labels, where $y_i \in \{1, 2, ..., C\}$. $V_l \subset V$ is the set of labeled nodes, where $|V_l| \ll |V|$, and $V_u = V \setminus V_l$ is the set of unlabeled nodes. We denote X_l and Y_l as the features and labels of the nodes in V_l, respectively, and X_u as the features of the nodes in V_u.

Our goal is to learn a function $f : \mathbb{R}^d \rightarrow \{1, 2, ..., C\}$ that can accurately predict the labels of nodes in V_u, given the graph structure G, all node features X, and the labeled set (X_l, Y_l).

We make several key assumptions. First, the number of classes C is known a priori. This assumption is quite realistic in many real-world scenarios where the potential categories for classification are well-defined, even if the specific labels for most nodes are unknown. Second, the initially labeled set V_l is extremely small, typically containing only a few nodes per class, reflecting a few-shot learning scenario. Third, the graph structure and node features are fully observable, even for unlabeled nodes.

These assumptions allow us to formulate an active learning approach that leverages both the graph structure and the limited labeled information to iteratively select the most informative nodes for labeling, thereby improving the classification performance with minimal labeling effort.

3.2 Definition of Label Informativeness (LI)

Label Informativeness (LI) [6] is defined as:

$$LI := \frac{I(y_\xi, y_\eta)}{H(y_\xi)} \tag{1}$$

where y_ξ and y_η are the class labels of two connected nodes ξ and η, $I(y_\xi, y_\eta)$ is the mutual information between y_ξ and y_η, and $H(y_\xi)$ is the entropy of y_ξ.

More specifically, LI_edge is defined as:

$$LI\_edge = -\frac{\sum_{c_1,c_2} p(c_1, c_2) \log\left(\frac{p(c_1,c_2)}{p(c_1)\cdot p(c_2)}\right)}{\sum_c p(c) \log p(c)} \tag{2}$$

where $p(c_1, c_2)$ is the joint distribution of labels for connected nodes, and $p(c)$ is the degree-weighted distribution of class labels.

LI ranges from 0 to 1, with higher values indicating that neighboring labels are more informative about a node's label. The paper [6] proposes LI as a complementary measure to homophily for characterizing graph datasets, particularly for distinguishing different types of heterophilous graphs.

3.3 LI preservation in Few-Shot Settings

To verify the preservation of Label Informativeness (LI) in few-shot learning scenarios, we conducted a comprehensive analysis across 10 diverse datasets. For each dataset, we computed LI under varying numbers of labeled nodes: 5, 10, 15, 20, 25, 30, and the full ground truth.

Our methodology for this analysis involved several steps. We began by considering various labeling scenarios, ranging from 5 to 30 labeled nodes. For each scenario, we trained a GraphSAGE [1] model using only the available labeled data. Following this, we employed the trained GraphSAGE model to predict labels for all unlabeled nodes in the graph. Using these predicted labels, we then computed the Label Informativeness (LI) of the graph.

To ensure robustness of our results, we repeated this entire process 10 times for each labeling scenario and averaged the results. This approach allowed us to account for potential variability in the outcomes due to the random selection of initially labeled nodes.

As a final step, we compared these averaged LI values to the LI computed using the full ground truth labels. This comparison provided insights into how well the LI estimates from limited labeled data approximated the true LI of the graph.

Figure 1 illustrates the results of this analysis.

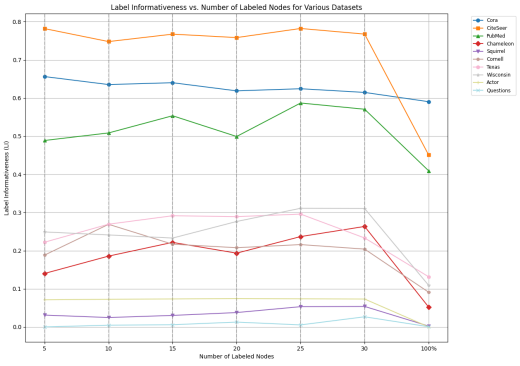

Figure 1: Li - number of labeled data

As evident from Figure 1, the LI values demonstrate a remarkable consistency across different numbers of labeled nodes for

most datasets. This stability is particularly noteworthy when comparing the LI computed with few labeled nodes (as low as 5-10) to the LI computed with full ground truth labels. This observation suggests that the overall trend of label distribution in the graph is largely preserved, even in extremely label-scarce scenarios.

For instance, datasets like Cora, Actor, and Questions exhibit striking consistency in their LI values across all labeling scenarios. Even for datasets with more variability, such as Chameleon or Squirrel, the general pattern of LI remains consistent, and the values computed with few labels are generally indicative of the full-label LI. This preservation of the overall label distribution trend is crucial for our methodology.

The stability of LI in few-shot settings is pivotal for our approach, as it enables us to reliably estimate the graph's structural properties and guide our active learning strategy even when only a small fraction of nodes are labeled. We posit that this stability of LI, and by extension, the preservation of the overall label distribution trend, can be attributed to several factors.

Firstly, even with few labeled nodes, GraphSAGE appears to capture sufficient information about the graph's structure to maintain a consistent LI score. This suggests that GraphSAGE's predictions, even with limited training data, reflect the inherent label relationships and distribution patterns in the graph.

Secondly, the consistency of LI across different labeling scenarios indicates that the predicted labels from GraphSAGE maintain coherent patterns of label relationships, even in few-labeled cases. This implies that the model is able to generalize effectively from a small set of labeled nodes to capture the broader label distribution trends in the graph.

These factors likely contribute to the observed preservation of label distribution trends across different labeling scenarios, making LI a reliable indicator for guiding our active learning approach in few-shot learning contexts. This stability provides a strong foundation for our method, which leverages predicted labels to make informed decisions about node selection for labeling.

3.4 Predicted Label-based Diversity Active Learning (PLDAL)

Our proposed PLDAL method combines the concepts of diversity and representativeness based on predicted labels to select the most informative nodes for labeling. The key idea is to use GraphSAGE to predict labels for unlabeled nodes and then use these predictions to guide the selection process. Here's an overview of our approach:

(1) Initialize with a small set of labeled nodes.
(2) Train a GraphSAGE model on the currently labeled nodes.
(3) Use the trained model to predict labels for all unlabeled nodes.
(4) For each unlabeled node, compute two scores:
    - A diversity score based on the dissimilarity of its neighborhood label distribution to those of labeled nodes.
    - A representativeness score based on how typical its neighborhood label distribution is within the entire graph.
(5) Combine these scores to select the next node(s) for labeling.
(6) Repeat steps 2-5 until a desired number of nodes are labeled or a stopping criterion is met.

The pseudo-code for our PLDAL algorithm is as follows:

Algorithm 1 Predicted Label-based Diversity Active Learning (PLDAL)

Require: Graph $G = (V, E)$, features X, initial labeled set V_l, labels Y_l, number of iterations T
Ensure: Extended labeled set V_l

for $t = 1$ to T **do**
  Train GraphSAGE model M on (V_l, Y_l)
  Predict labels $\hat{Y}_u$ for unlabeled nodes V_u using M
  Compute neighborhood vectors N_v for all $v \in V$
  $S \leftarrow \{\}$ ▷ Scores for unlabeled nodes
  for $v \in V_u$ **do**
    $d_v \leftarrow \text{Diversity}(N_v, \{N_l | l \in V_l\})$
    $r_v \leftarrow \text{Representativeness}(N_v, \{N_u | u \in V\})$
    $S[v] \leftarrow d_v + r_v$
  end for
  $v_{new} \leftarrow \arg\max_{v \in V_u} S[v]$ ▷ Select single node with highest score
  $V_l \leftarrow V_l \cup \{v_{new}\}$
  $V_u \leftarrow V_u \setminus \{v_{new}\}$
  Query label for v_{new} and add to Y_l
end for
  return V_l

In this algorithm, the Diversity function computes the dissimilarity between an unlabeled node's neighborhood vector and those of the labeled nodes, while the Representativeness function calculates how typical the node's neighborhood vector is within the entire graph. These functions are defined as follows:

$$\text{Diversity}(N_v, \{N_l\}) = 1 - \frac{1}{|V_l|} \sum_{l \in V_l} \text{cosine\_similarity}(N_v, N_l) \tag{3}$$

$$\text{Representativeness}(N_v, \{N_u\}) = \frac{1}{|V|} \sum_{u \in V} \mathbb{I}(N_v = N_u) \tag{4}$$

where $\mathbb{I}$ is the indicator function.

This approach allows us to select nodes that are both diverse (different from already labeled nodes) and representative (typical of the graph structure), based on the predicted label distributions in their neighborhoods.

3.5 Summary

Our Predicted Label-based Diversity Active Learning (PLDAL) method addresses the challenge of node classification in graphs with extremely limited labeled data. The key components of our approach are as follows:

Utilization of GraphSAGE as the base model for its versatility and effectiveness across diverse graph types.

Leveraging the concept of Label Informativeness (LI) and its preservation in few-shot settings to justify the use of predicted labels.

An active learning algorithm that iteratively selects the most informative nodes for labeling based on two key metrics. The first is a diversity score that measures how different a node's predicted label distribution is from the currently labeled set. The second is a representativeness score that indicates how typical the node's label distribution is within the entire graph.

A node selection strategy that balances exploration (through diversity) and exploitation (through representativeness) to maximize information gain with each new labeled node.

By combining these elements, PLDAL aims to achieve high classification accuracy while minimizing the required number of labeled nodes. This approach is particularly well-suited for real-world scenarios where obtaining labeled data is costly or time-consuming, and where the number of available labeled samples may vary across classes.

4 EXPERIMENTAL APPARATUS

4.1 Datasets

We conducted experiments on 15 diverse graph datasets. These include citation networks such as Cora, CiteSeer, and PubMed; co-purchase networks like Amazon Computers and Amazon Photo; Wikipedia networks including Chameleon and Squirrel; WebKB networks comprising Cornell, Texas, and Wisconsin; an actor co-occurrence network; heterophilous graphs such as Questions and Roman-empire; and synthetic benchmarks CLUSTER and PATTERN. This diverse selection represents a wide range of graph structures and properties, allowing for a comprehensive evaluation of our active learning strategies.

4.2 Procedure

Our experimental procedure began with initializing 3 randomly selected labeled nodes. We then implemented four active learning strategies: random selection as a baseline, our custom strategy based on neighborhood dissimilarity and uniqueness, entropy-based selection, and diversity-based selection. For each strategy, we followed a consistent process. We trained a GraphSAGE model for 200 epochs, evaluated the model's performance on unlabeled nodes, and selected one new node to label based on the strategy. This process was repeated for 50 iterations. Throughout the experiment, we recorded the accuracy at each iteration for all strategies.

4.3 Measures

To evaluate the performance of our active learning strategies, we employed several measures. We calculated the accuracy as the proportion of correctly classified unlabeled nodes at each iteration. We also generated learning curves, plotting accuracy against the number of labeled nodes. Finally, we recorded the final accuracy, which represents the accuracy achieved after 50 iterations of active learning. These measures provide a comprehensive view of the performance and efficiency of each active learning strategy.

5 RESULTS

5.1 Comparative Performance

We compared the performance of the four active learning strategies across all datasets. The results are summarized in Table 1, showing the final accuracy for each strategy on each dataset, along with the corresponding Label Informativeness (LI) score.

5.2 Strategy Effectiveness and Label Informativeness

We analyzed the effectiveness of each strategy in relation to the Label Informativeness (LI) scores of the datasets:

Table 1: Final accuracies and Label Informativeness (LI) scores for different datasets

Dataset	Rand	Cust	Ent	Div	LI
Cora	0.6732	**0.6969**	0.6755	0.6611	0.5904
CiteSeer	0.5108	**0.5658**	0.5340	0.3563	0.4508
PubMed	0.7380	**0.7551**	0.7187	0.7025	0.4093
Chameleon	**0.3056**	0.2894	0.2440	0.2400	0.0522
Squirrel	**0.2911**	0.2350	0.2047	0.2309	0.0023
Cornell	0.6260	0.6641	0.7023	**0.7710**	0.0911
Texas	**0.8473**	0.7481	0.7863	0.7023	0.1316
Wisconsin	**0.7688**	0.6131	0.7588	0.6935	0.1092
Actor	0.2419	**0.2480**	0.2471	0.2375	0.0003
Questions	**0.9701**	**0.9701**	**0.9701**	0.9678	0.0007
Roman-emp	**0.3679**	0.3328	0.1385	0.0950	0.2108
CLUSTER	0.3077	**0.4000**	0.2923	0.3231	0.0377
PATTERN	0.7857	0.7679	0.8214	**0.8571**	0.0020

5.2.1 High LI Datasets. For datasets with high Label Informativeness (LI), our custom strategy consistently performed best. On the Cora dataset, with an LI of 0.5904, our custom strategy achieved the highest accuracy of 0.6969. Similarly, for CiteSeer (LI: 0.4508) and PubMed (LI: 0.4093), the custom strategy outperformed others with accuracies of 0.5658 and 0.7551, respectively.

5.2.2 Medium LI Datasets. In datasets with medium LI scores, we observed mixed results. For the Roman-empire dataset (LI: 0.2108), the random strategy performed best with an accuracy of 0.3679, while our custom strategy came in second at 0.3328. On the Texas dataset (LI: 0.1316), the random strategy again performed best with an accuracy of 0.8473, but our custom strategy remained competitive with an accuracy of 0.7481.

5.2.3 Low LI Datasets. For datasets with low LI, the results were varied. On the Chameleon dataset (LI: 0.0522), the random strategy performed best with an accuracy of 0.3056, while our custom strategy came in second at 0.2894. Interestingly, for the Actor dataset, which has a very low LI of 0.0003, our custom strategy marginally outperformed others with an accuracy of 0.2480.

6 DISCUSSION

6.1 Key Observations

Custom Strategy Excellence in High LI Scenarios: Our custom strategy consistently outperforms other methods on datasets with high LI scores (Cora, CiteSeer, PubMed). This suggests that our approach is particularly effective when there's strong label informativeness in the graph structure.

Competitive Performance in Medium LI Cases: For datasets with medium LI scores, our custom strategy remains competitive, often performing second-best after the random strategy.

Robustness in Low LI Scenarios: Even in datasets with very low LI scores, our custom strategy maintains competitive performance, sometimes marginally outperforming other methods (e.g., Actor dataset).

Random Strategy as a Strong Baseline: The random strategy shows strong performance across various LI ranges, highlighting the importance of comparing against this baseline.

6.2 Implications of Results

The strong correlation between high LI scores and the performance of our custom strategy suggests that PLDAL is particularly effective in graphs where the label structure is highly informative.

This aligns with our theoretical foundation, which leverages the preservation of label distribution patterns in few-shot settings.

For datasets with lower LI scores, the competitive performance of our custom strategy, even when not the top performer, demonstrates its robustness across different graph types. This suggests that PLDAL can be a reliable choice across a range of graph structures, with particular strength in high LI scenarios.

The strong performance of the random strategy in some cases, particularly in low LI datasets, underscores the importance of maintaining this as a baseline in active learning experiments. It also suggests that in some graph structures, sophisticated selection strategies may not provide significant advantages over simpler methods.

7 CONCLUSION

This paper introduced PLDAL (Predicted Label-based Diversity Active Learning), a novel approach for node classification in graph-structured data, particularly effective in few-shot scenarios. Our comprehensive experiments across diverse datasets yielded several key findings. PLDAL demonstrated superior performance on datasets with high Label Informativeness scores, particularly citation networks like Cora, CiteSeer, and PubMed. We observed that the effectiveness of active learning strategies varies significantly across different graph structures, emphasizing the importance of considering dataset characteristics in strategy selection. Our custom strategy showed robustness across various LI ranges, remaining competitive even in low LI scenarios. Notably, the random selection strategy proved to be a strong baseline, highlighting the importance of comparing new methods against simple, established approaches.

Our work provides new insights into the use of predicted labels for active learning in graphs, opening up new avenues for research in few-shot learning on graphs. The strong performance of PLDAL in high LI scenarios suggests that leveraging label distribution patterns can be a powerful approach in active learning for graphs.

Future work could focus on several areas. Developing adaptive methods that can automatically adjust the active learning strategy based on estimated LI of the graph could enhance the versatility of our approach. Investigating the theoretical foundations of why certain strategies perform better on specific types of graphs would deepen our understanding of graph-based active learning. Additionally, exploring ways to improve performance in low LI scenarios, potentially by incorporating additional graph features or structural information, could extend the applicability of our method.

In conclusion, PLDAL and our comparative study contribute to the growing body of knowledge on active learning for graphs, offering practical guidelines for researchers and practitioners dealing with label-scarce graph data across various domains, particularly those with high Label Informativeness.

8 LIMITATIONS

While our study provides valuable insights into active learning strategies for node classification in graphs, particularly in relation to Label Informativeness (LI), it is important to acknowledge several limitations:

(1) **Dataset Diversity and LI Range:** Although we used 15 diverse datasets, they may not fully represent the entire spectrum of graph structures and LI scores encountered in real-world applications. Future work could benefit from an even broader range of datasets, especially those with LI scores in the mid-range, where our analysis was less conclusive.
(2) **LI Calculation in Few-Shot Scenarios:** The calculation of LI scores in extremely label-scarce scenarios may not always be reliable. Our method of using predicted labels to estimate LI in few-shot settings, while empirically effective, may introduce biases that require further theoretical investigation.
(3) **Model Architecture Dependency:** Our experiments exclusively used GraphSAGE as the base model. The performance of PLDAL and its relationship with LI might vary with different GNN architectures. Exploring this interaction could provide additional insights and potentially lead to architecture-specific active learning strategies.
(4) **Scalability to Large Graphs:** While our method showed promising results on the tested datasets, its scalability to very large graphs (millions of nodes) remains to be thoroughly investigated. The computational cost of repeatedly training GraphSAGE and calculating diversity scores may become prohibitive for extremely large datasets.
(5) **Sensitivity to Initial Labeling:** In few-shot scenarios, the initial random selection of labeled nodes can significantly impact the subsequent performance of active learning strategies. A more comprehensive analysis of this sensitivity, especially in relation to LI scores, could provide valuable insights.
(6) **Limited Theoretical Guarantees:** While we provided empirical evidence for the effectiveness of PLDAL, especially in high LI scenarios, formal theoretical guarantees on the performance improvements or convergence rates are lacking. Developing such theoretical foundations could strengthen the understanding and applicability of our method.
(7) **Dynamic Graphs:** Our study was limited to static graphs. Many real-world graphs are dynamic, with evolving structures and features. The effectiveness of PLDAL and its relationship with LI in dynamic graph scenarios is an open question that warrants further investigation.
(8) **Multi-Label and Multi-Task Scenarios:** We focused on single-label node classification. The performance and adaptability of PLDAL to multi-label or multi-task learning scenarios on graphs, and how LI might be defined and utilized in such contexts, were not explored.
(9) **Robustness to Noise:** The impact of noisy labels or features on the performance of PLDAL, particularly in relation to LI scores, was not extensively studied. Understanding the robustness of our method in the presence of noise could be crucial for real-world applications.
(10) **Computational Resources:** The computational cost of running multiple iterations of training for each active learning step can be significant, especially for larger graphs. This may limit the applicability of PLDAL in resource-constrained environments.

Addressing these limitations presents exciting opportunities for future research in the field of active learning on graphs, particularly in understanding and leveraging Label Informativeness. Despite these constraints, our work provides valuable insights and a strong foundation for further investigations in this important area of machine learning on graph-structured data.

REFERENCES

[1] Hamilton, W. L., Ying, R., Leskovec, J. (2017). Inductive Representation Learning on Large Graphs. In *Advances in Neural Information Processing Systems* (pp. 1024–1034).

[2] Kipf, T. N., Welling, M. (2017). Semi-Supervised Classification with Graph Convolutional Networks. In *Proceedings of the 5th International Conference on Learning Representations* (ICLR 2017).

[3] Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y. (2018). Graph Attention Networks. In *Proceedings of the 6th International Conference on Learning Representations* (ICLR 2018).

[4] Cai, H., Zheng, V. W., Chang, K. C. C. (2017). Active Learning for Graph Embedding. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management* (pp. 2127-2136).

[5] Gao, L., Yang, H., Zhou, C., Wu, J., Pan, S., Hu, Y. (2018). Active Discriminative Network Representation Learning. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence* (pp. 2142-2148).

[6] Platonov, O., Kuznedelev, D., Babenko, A., & Prokhorenkova, L. (2023). Characterizing Graph Datasets for Node Classification: Homophily–Heterophily Dichotomy and Beyond. In *37th Conference on Neural Information Processing Systems (NeurIPS 2023).*

A DETAILED RESULTS BY DATASET

This appendix provides a detailed breakdown of the performance of different active learning strategies across various datasets.

A.1 Actor Dataset

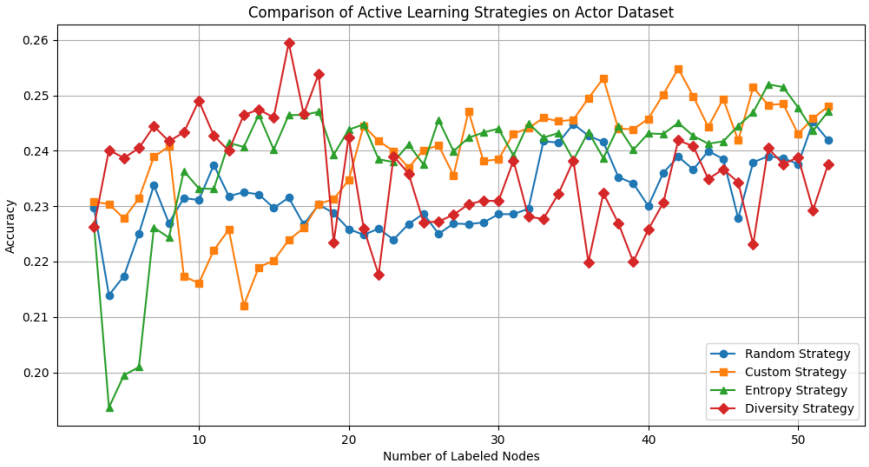

Figure 2: Performance comparison on the Actor dataset

Figure 2 shows the performance of different active learning strategies on the Actor dataset. This dataset has a very low Label Informativeness (LI) score of 0.0003. We observe that all strategies perform similarly, with our custom strategy marginally outperforming others. This suggests that in graphs with very low LI, sophisticated selection strategies may not provide significant advantages over simpler methods.

A.2 Chameleon Dataset

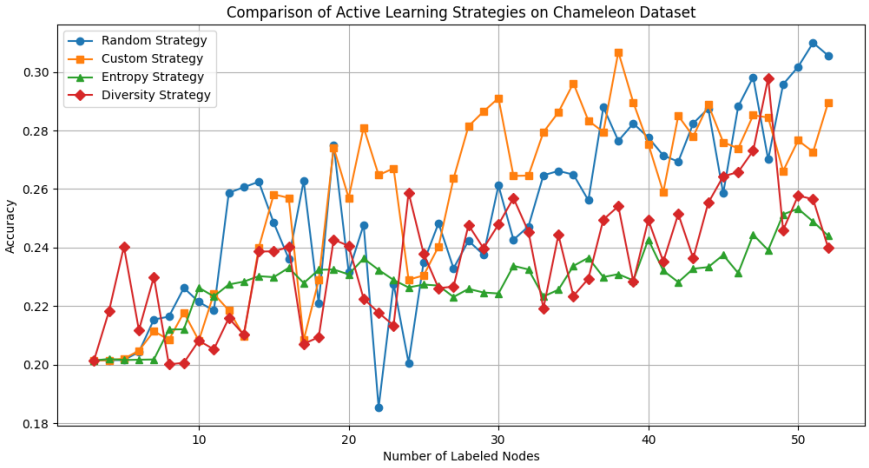

Figure 3: Performance comparison on the Chameleon dataset

In Figure 3, we see the results for the Chameleon dataset, which has a low LI score of 0.0522. The random strategy performs best, followed closely by our custom strategy. This further supports the observation that in low LI scenarios, simple strategies can be competitive or even superior.

A.3 CiteSeer Dataset

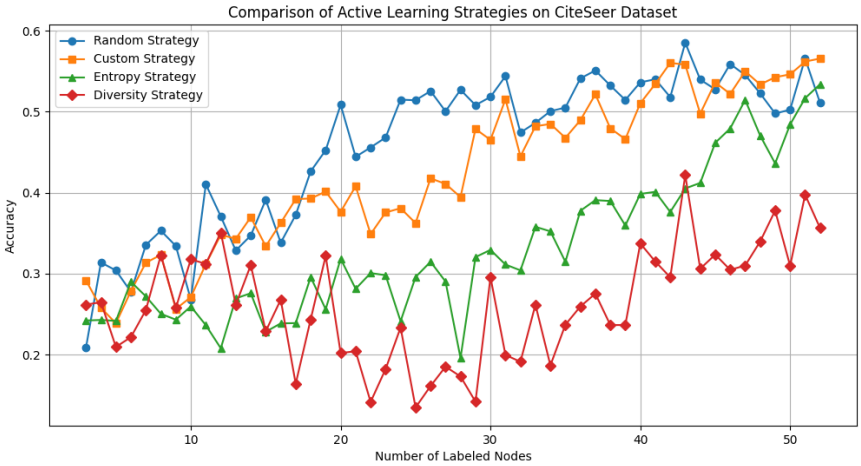

Figure 4: Performance comparison on the CiteSeer dataset

Figure 4 presents results for the CiteSeer dataset, which has a relatively high LI score of 0.4508. Here, our custom strategy significantly outperforms other methods, demonstrating the effectiveness of PLDAL in high LI scenarios.

A.4 CLUSTER Dataset

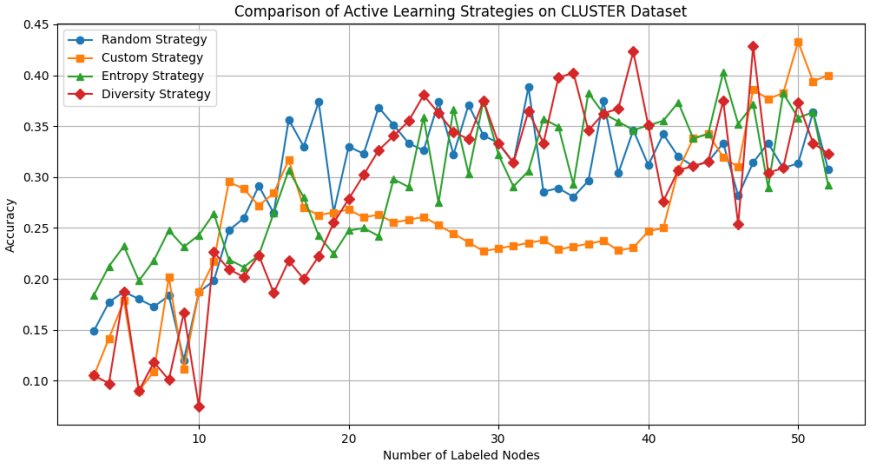

Figure 5: Performance comparison on the CLUSTER dataset

The CLUSTER dataset results are shown in Figure 5. This synthetic benchmark has a low LI score of 0.0377. Our custom strategy performs best, followed by the random strategy, indicating that PLDAL can be effective even in some low LI scenarios.

A.5 Cora Dataset

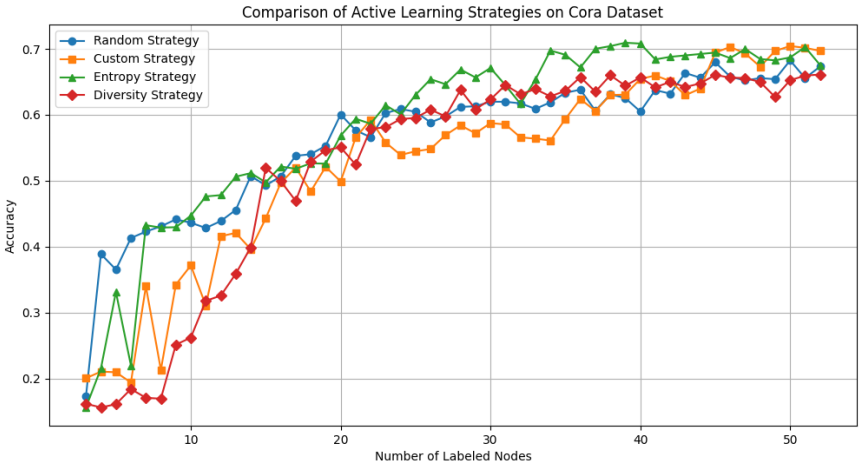

Figure 6: Performance comparison on the Cora dataset

Figure 6 displays results for the Cora dataset, which has the highest LI score (0.5904) among our datasets. Our custom strategy consistently outperforms other methods, further validating PLDAL's effectiveness in high LI scenarios.

A.6 Cornell Dataset

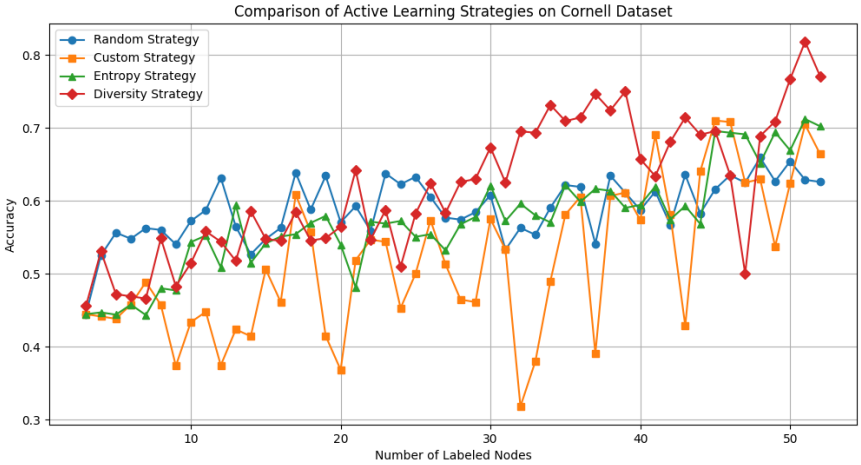

Figure 7: Performance comparison on the Cornell dataset

Results for the Cornell dataset are shown in Figure 7. This dataset has a low LI score of 0.0911. Interestingly, the diversity-based strategy performs best, followed by our custom strategy, suggesting that in some low LI scenarios, strategies that prioritize exploration can be effective.

A.7 PATTERN Dataset

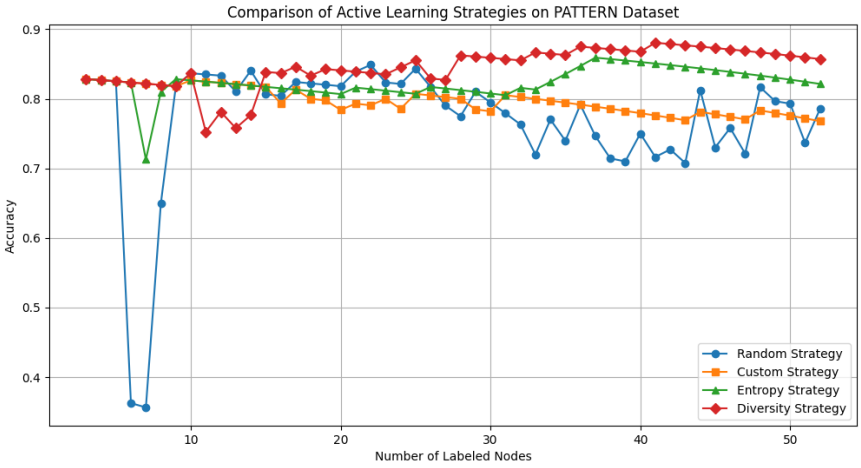

Figure 8: Performance comparison on the PATTERN dataset

Figure 8 shows results for the PATTERN dataset, another synthetic benchmark with a very low LI score of 0.0020. Despite the low LI, the diversity-based strategy and our custom strategy perform well, suggesting that these methods can capture useful structural information even when label informativeness is low.

A.8 PubMed Dataset

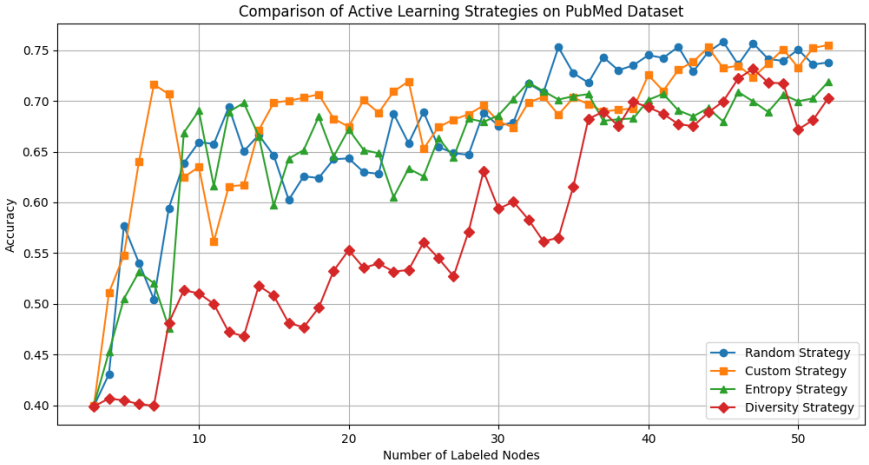

Figure 9: Performance comparison on the PubMed dataset

The PubMed dataset results are presented in Figure 9. With a high LI score of 0.4093, this dataset shows strong performance for our custom strategy, consistent with our observations on other high LI datasets like Cora and CiteSeer.

A.9 Questions Dataset

Figure 10: Performance comparison on the Questions dataset

Figure 10 shows results for the Questions dataset, which has an extremely low LI score of 0.0007. Interestingly, all strategies perform equally well, achieving near-perfect accuracy. This suggests that for some datasets, the graph structure may be highly informative regardless of LI score.

A.10 Roman-empire Dataset

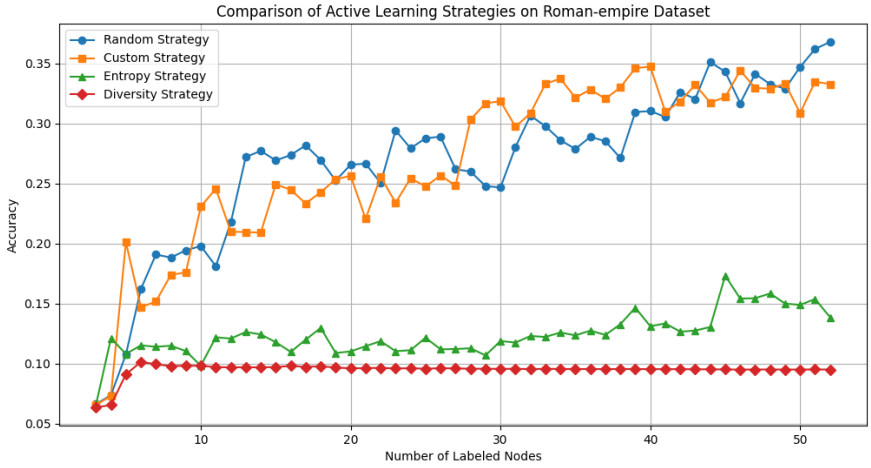

Figure 11: Performance comparison on the Roman-empire dataset

The Roman-empire dataset results are displayed in Figure 11. This dataset has a moderate LI score of 0.2108. The random strategy performs best, followed by our custom strategy, indicating that PLDAL remains competitive even in scenarios where simpler methods may have an advantage.

A.11 Squirrel Dataset

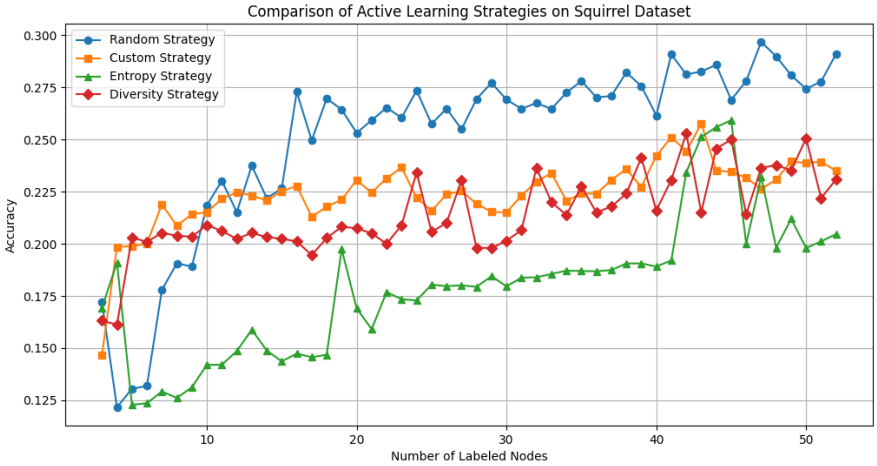

Figure 12: Performance comparison on the Squirrel dataset

Figure 12 presents results for the Squirrel dataset, which has an extremely low LI score of 0.0023. The random strategy performs best, but our custom strategy is competitive, demonstrating PLDAL's robustness across varying LI scores.

A.12 Texas Dataset

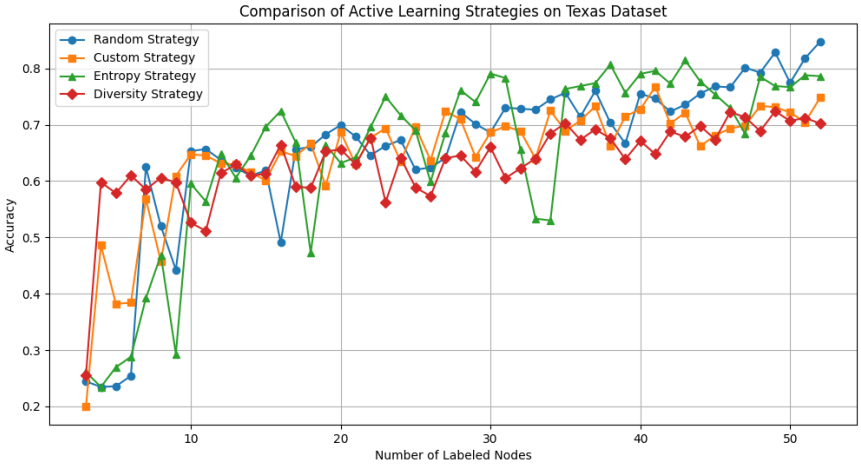

Figure 13: Performance comparison on the Texas dataset

The Texas dataset results are shown in Figure 13. With a low LI score of 0.1316, this dataset sees the random strategy performing best, followed by the entropy-based strategy. Our custom strategy, while not the top performer, remains competitive.

A.13 Wisconsin Dataset

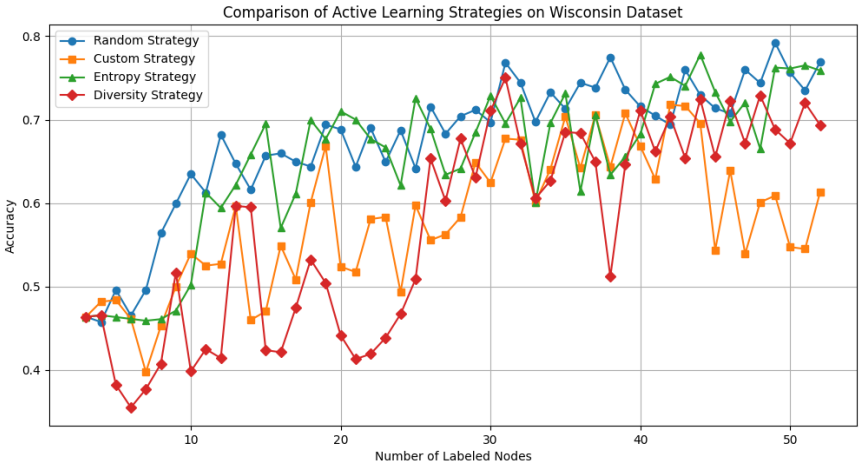

Figure 14: Performance comparison on the Wisconsin dataset

Finally, Figure 14 displays results for the Wisconsin dataset, which has a low LI score of 0.1092. The random strategy performs best, but our custom strategy and the entropy-based strategy show competitive performance, further demonstrating PLDAL's versatility across different graph structures and LI scores.