Part 1: Node2Vec and Logistic Regression

Evaluation Results:

Accuracy: 0.8088929889298892

Precision: 0.8117210268179236

Recall: 0.7808057009069094

Macro F1-score: 0.7935703554385498

Confusion Matrix:

[[52 3 4 2 1 3 1]

 [0 73 6 1 2 0 0]

 [1 3 129 8 1 1 11]

 [2 0 12 65 0 0 9]

 [2 1 8 2 39 0 1]

 [2 0 2 0 1 20 4]

 [4 0 6 4 1 0 55]]

Classification Report:

	precision	recall	f1-score	support
0	0.83	0.79	0.81	66
1	0.91	0.89	0.90	82
2	0.77	0.84	0.80	154
3	0.79	0.74	0.76	88
4	0.87	0.74	0.80	53
5	0.83	0.69	0.75	29
6	0.68	0.79	0.73	70
accuracy			0.80	542
macro avg	0.81	0.78	0.79	542
weighted avg	0.80	0.80	0.80	542

I have increased the number of epochs and accuracy also getting improved:

Accuracy: 0.8133579335793358

Precision: 0.798845873421475

Recall: 0.8057571155677504

Macro F1-score: 0.784223544479778

Confusion Matrix:

[[35   1   6   1   0   1   1]
 [2 83   5   0   2   0   0]
 [4   5 141   7   2   0   5]
 [1   1   9 59   0   0   4]
 [0   2   9   3 44   0   2]
 [5   0   3   0   0 25   4]
 [5   1   9   4   1   7 43]]

Classification Report:

	precision	recall	f1-score	support
0	0.77	0.78	0.72	45
1	0.89	0.90	0.90	92
2	0.77	0.86	0.82	164
3	0.80	0.80	0.80	74
4	0.90	0.73	0.81	60
5	0.76	0.68	0.71	37
6	0.73	0.61	0.67	70
accuracy			0.81	542
macro avg	0.80	0.77	0.77	542
weighted avg	0.80	0.79	0.79	542

Analysis:

- By increasing the number of epochs during training, the LR model's accuracy showed a slight improvement from approximately 80.9% to 81.3%.
- This observation suggests that allowing the model to train for more epochs enabled it to better capture the underlying patterns in the data, resulting in a marginal enhancement in classification performance.
- The LR model, being a simple linear classifier, exhibits relatively lower computational complexity compared to more complex models like Graph Convolutional Networks (GCNs).
- Despite its simplicity, the LR model achieves reasonable performance, demonstrating its effectiveness as a baseline model for node classification tasks on graph-structured data.
- However, the LR model's performance may plateau or reach a limit due to its linear nature, especially when dealing with complex relationships and dependencies among nodes.