Active Learning for Node Classification

DGL

Cora dataset: 2708 nodes, 5429 links, 1433 features, 7 classes, avg. degree 4

No. of queries: 100

Query batch size: 1

Size of initial labelled dataset: 35 (5 instances/class \* 7 classes)

Validation size: 14 (2 instances/class \* 7 classes)

Test size: 700 (100 instances/class \* 7 classes)

GNN layers: 2

Hidden dimension: 48

Training batch size: 5

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Query Strategy** | **Initial Classification Accuracy** | **Classification Accuracy after no. of queries** | | | |
| **25** | **50** | **75** | **100** |
| Degree | 29.43% | 46.42% | 58.57% | 68.28% | 75.14% |
| Betweenness | 30.42% | 54.71% | 67.85% | 70% | 80.57% |
| Closeness | 29.71% | 42% | 50.28% | 54.42% | 60.97% |