

A Dataset Statistics

In Table 2 we provide the key statistics of the four datasets used during our experiments. This includes the number of samples and number of classes each dataset consists of. In addition we report on the average number of tokens per sample in each dataset. In this case we refer to *#tokens* as the number of tokens a PLM produces by feeding the samples through its tokenizer.

| Dataset | #samples | #tokens | #classes |
|-----------|----------|---------|----------|
| Twitter | 9,983 | 22.44 | 2 |
| MR | 10,426 | 25.23 | 2 |
| Snippets | 12,161 | 23.64 | 8 |
| TagMyNews | 32,591 | 10.64 | 7 |

Table 2: Dataset statistics.

B Influence of n -hop Neighborhood

| n | Twitter | MR | Snippets | TMN |
|-----|--------------|--------------|--------------|--------------|
| 1 | 0.814 | 0.711 | 0.821 | 0.667 |
| 2 | 0.811 | 0.686 | 0.828 | 0.658 |
| 3 | 0.814 | 0.673 | 0.822 | 0.649 |
| 4 | 0.822 | 0.707 | 0.827 | 0.636 |

Table 3: Macro F1 scores for different n -hops of the fully connected graphs.

C Influence of Hidden Dimension

These experiments are run with the 1-hop connected graph which has demonstrated robust performance in the n -hop experiments.

| GNN dim | Twitter | MR | Snippets | TMN |
|---------|--------------|--------------|--------------|--------------|
| 32 | 0.762 | 0.667 | 0.799 | 0.651 |
| 64 | 0.733 | 0.684 | 0.808 | 0.660 |
| 128 | 0.814 | 0.711 | 0.818 | 0.667 |
| 256 | 0.739 | 0.686 | 0.818 | 0.661 |
| 512 | 0.741 | 0.668 | 0.813 | 0.661 |

Table 4: Macro F1 scores for different hidden dimensions of the GNN.

D Performance of GNN Layer Types

We evaluate three different GNN layer types on the graphs. This includes Graph Convolutional Neural Networks (GCN) (Kipf and Welling, 2017), GraphSAGE (Hamilton et al., 2017) and Graph Attention

Layers (GAT) (Veličković et al., 2018). The experiments were conducted with a hidden dimension of 128 and 1-hop fully connected graphs.

| layer type | Twitter | MR | Snippets | TMN |
|------------|--------------|--------------|--------------|--------------|
| GCN | 0.783 | 0.686 | 0.780 | 0.646 |
| SAGE | 0.794 | 0.685 | 0.782 | 0.647 |
| GAT | 0.814 | 0.711 | 0.818 | 0.667 |

Table 5: Macro F1 scores for different GNN layer types.