HDBSCAN attention\_logit cluster sizes (wikitext\_bert-large-uncased\_params-trained\_dense-on\_layers-24) after layer 1 after layer 3 after layer 4 after layer 2 60 50 80 0.04 40 cluster size 0.02 40 30 0.00 20 20 -0.0220 10 -0.040 0 0 k-th largest cluster k-th largest cluster k-th largest cluster k-th largest cluster after layer 6 after layer 7 after layer 5 after layer 8 30 0.04 30 25 cluster size 0.02 20 20 0.00 15 10 -0.0210 5 -0.040 0 k-th largest cluster 2 1 2 3 1 4 1 3 4 k-th largest cluster k-th largest cluster k-th largest cluster after layer 12 after layer 9 after layer 10 after layer 11 40 0.04 0.04 15 cluster size 0.02 0.02 30 10 0.00 0.00 20 -0.02-0.02 5 10 -0.04-0.040 0 k-th largest cluster k-th largest cluster 1 2 3 k-th largest cluster k-th largest cluster after layer 13 after layer 14 after layer 15 after layer 16 40 60 50 40 cluster size 40 40 30 30 20 20 20 10 10 0 0 0 0 -3 4 3 1 k-th largest cluster k-th largest cluster k-th largest cluster k-th largest cluster after layer 17 after layer 18 after layer 19 after layer 20 30 60 125 25 100 20 cluster size 40 75 15 50 10 20 25 5 -0 -0 -2 1 1 2 k-th largest cluster k-th largest cluster k-th largest cluster k-th largest cluster after layer 23 after layer 21 after layer 22 after layer 24 15 125 125 15 cluster size of 01 100 100 75 10 75 50 50 5 25 25 0 0 2 5 1 1 2 1 3 4 1 2 3 4 5 6 3 k-th largest cluster k-th largest cluster k-th largest cluster k-th largest cluster