HDBSCAN attention_logit cluster sizes (wikitext_bert-large-uncased_params-trained_dense-on_layers-24) after layer 4 after layer 1 after layer 2 after layer 3 0.04 0.02 0.00 -0.02-0.04k-th largest cluster k-th largest cluster k-th largest cluster k-th largest cluster after layer 5 after layer 6 after layer 7 after layer 8 0.04 cluster size 0 0 0.02 0.00 -0.02-0.04k-th largest cluster k-th largest cluster k-th largest cluster k-th largest cluster after layer 9 after layer 10 after layer 11 after layer 12 0.04 12.5 oluster size 7.5 5.0 0.02 0.00 -0.022.5 -0.040.0 k-th largest cluster k-th largest cluster k-th largest cluster k-th largest cluster after layer 13 after layer 14 after layer 15 after layer 16 cluster size 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 cluster size 0 7 0 -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 24 100 -cluster size 10 k-th largest cluster k-th largest cluster k-th largest cluster k-th largest cluster