HDBSCAN attention_logit cluster sizes (wikitext_albert-large-v2_params-trained_dense-on_layers-24) after layer 1 after layer 3 after layer 4 after layer 2 1 2 4 5 6 7 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 cluster size 05 04 k-th largest cluster k-th largest cluster k-th largest cluster k-th largest cluster after layer 9 after layer 11 after layer 10 after layer 12 cluster size 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 20 20 0 -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 -0 7 k-th largest cluster k-th largest cluster k-th largest cluster k-th largest cluster after layer 22 after layer 21 after layer 23 after layer 24 100 -150 · cluster size 20 001 k-th largest cluster k-th largest cluster k-th largest cluster k-th largest cluster