HDBSCAN cluster sizes of tokens x_i
(imdb_bert-large-uncased_params-trained_dense-off_layers-24, sample 3)