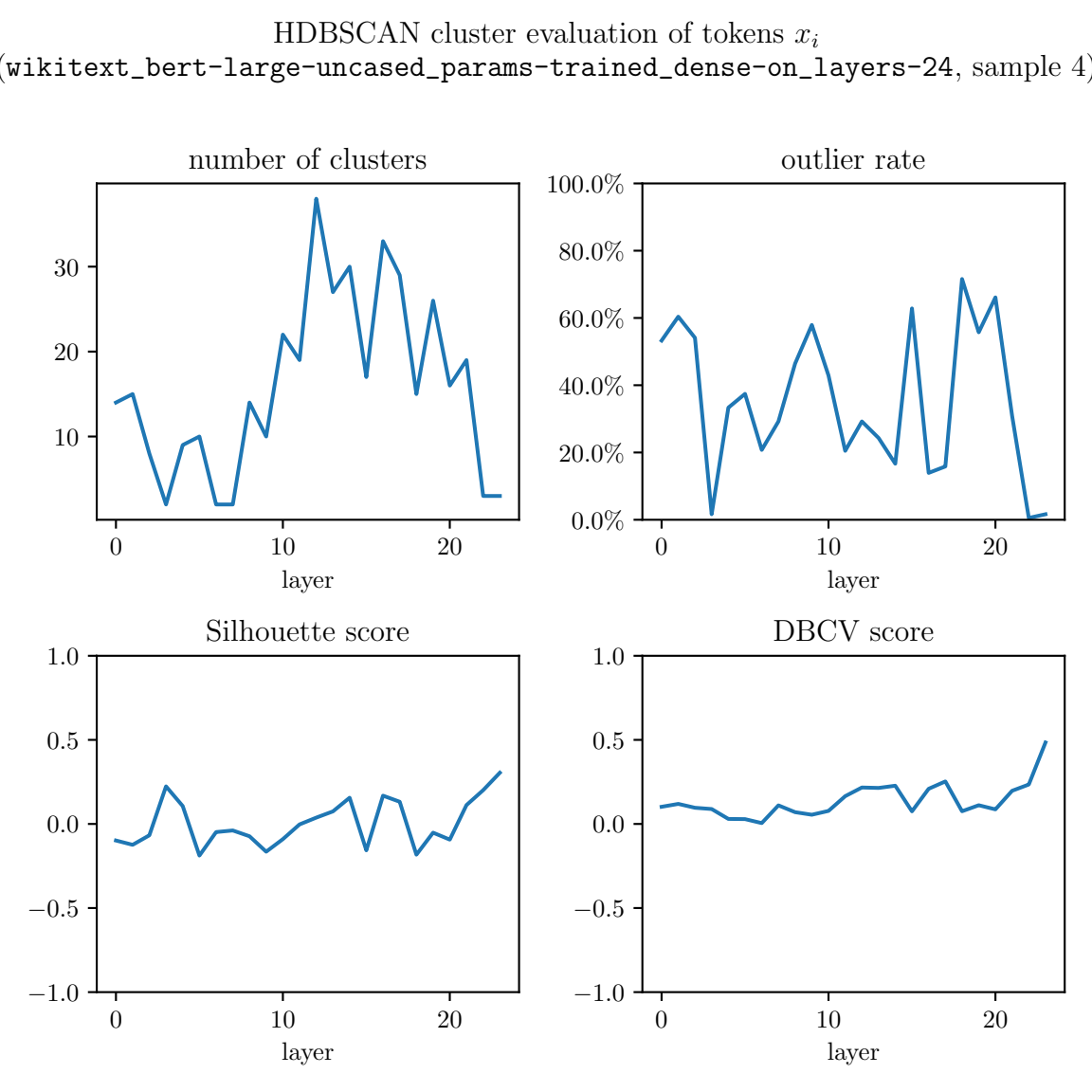
HDBSCAN cluster evaluation of tokens $x_i$
(wikitext_bert-large-uncased_params-trained_dense-on_layers-24, sample 4)
number of clusters
outlier rate
Silhouette score
DBCV score
layer
layer
layer
layer
100.0%
80.0%
60.0%
40.0%
20.0%
0.0%
1.0
0.5
0.0
−0.5
−1.0
0
10
20
30