token\_similarity histograms (randomtext\_albert-large-v2\_params-trained\_dense-on\_layers-48) after layer 4 after layer 1 after layer 3 after layer 2 40 40 40 40 30 30 30 30 20 20 20 20 10 10 10 10 0.0 0.5 0.0 0.5 0.0 0.0 -0.51.0 -0.51.0 -0.50.5 1.0 -0.5-1.0-1.0-1.0-1.01.0 after layer 6 after layer 5 after layer 7 after layer 8 40 40 40 40 30 30 30 30 20 20 20 20 10 10 10 10 0.0 0.0 0.0 0.0 -0.50.5 1.0 -1.0-0.50.5 1.0 -1.0-0.51.0 -1.0-0.51.0 -1.0after layer 9 after layer 11 after layer 12 after layer 10 40 40 40 40 30 30 30 30 20 20 20 20 10 10 10 10 0 0 -1.0-0.50.0 0.5 1.0 -1.0-0.50.0 0.5 1.0 -1.0-0.50.0 0.5 1.0 -1.0-0.50.0 1.0 after layer 13 after layer 14 after layer 15 after layer 16 40 40 40 40 30 30 30 30 20 20 20 20 10 10 10 10 0.0 0.0 -0.50.0 -1.0-0.50.5 -0.50.0 0.5 -1.0-0.50.5 -1.01.0 -1.01.0 1.0 0.5 1.0 after layer 17 after layer 18 after layer 19 after layer 20 40 40 40 40 30 30 30 30 20 20 20 20 10 10 10 10 -1.00.0 0.0 -0.50.0 -0.50.0 -0.50.5 1.0 -1.0-0.50.5 1.0 -1.00.5 1.0 -1.00.5 1.0 after layer 21 after layer 22 after layer 23 after layer 24 40 40 40 40 30 30 30 30 20 20 20 20 10 10 10 10

0

-1.0

-0.5

0.0

0.5

1.0

0

-1.0

-0.5

0.0

0.5

1.0

1.0

0 -

-1.0

-0.5

0.0

0.5

0

1.0

-1.0

-0.5

0.0

0.5