token\_similarity histograms (wikitext\_albert-large-v2\_params-random\_dense-on\_layers-48) after layer 1 after layer 3 after layer 4 after layer 2 50 50 50 50 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 1.0 -0.50.5 1.0 -0.51.0 -1.0-0.5-1.0-1.00.5 after layer 7 after layer 8 after layer 5 after layer 6 50 50 50 50 40 40 40 40 30 30 30 30 20 20 20 20 10 10 10 10 0.0 0.0 0.0 -0.50.0 0.5 1.0 -1.0-0.50.5 1.0 -1.0-0.50.5 1.0 -1.0-0.50.5 -1.0after layer 10 after layer 9 after layer 11 after layer 12 50 50 50 50 40 40 40 40 30 30 30 30 20 20 20 20 10 10 10 10 0 0.0 0.0 -0.50.0 0.0 -1.0-0.50.5 1.0 -1.0-0.50.5 1.0 -1.00.5 1.0 -1.0-0.50.5 1.0 after layer 13 after layer 14 after layer 15 after layer 16 50 50 50 50 40 40 40 40 30 30 30 30 20 20 20 20 10 10 10 10 0.0 -0.50.0 -0.50.0 -0.50.5 -0.50.0 0.5 0.5 -1.0-1.01.0 -1.01.0 -1.01.0 0.5 1.0 after layer 17 after layer 18 after layer 19 after layer 20 50 50 50 50 40 40 40 40 30 30 30 30 20 20 20 20 10 10 10 10 0.0 0.5 0.0 -1.00.0 0.0 -1.0-0.51.0 -1.0-0.50.5 1.0 -0.50.5 1.0 -1.0-0.5 0.5 1.0 after layer 21 after layer 22 after layer 23 after layer 24 50 50 50 50 40 40 40 40 30 30 30 30 20 20 20 20 10 10 10 10 0 -0 0 0

-0.5

-1.0

0.0

0.5

-1.0

1.0

-0.5

0.0

0.5

-1.0

1.0

-0.5

0.0

0.5

-1.0

1.0

-0.5

0.0

0.5

1.0