token\_similarity heatmaps (imdb albert-large-v2 params-random dense-on layers-48) after layer 3 after layer 1 after layer 2 after layer 4 0 0.5 token\_similarity 0.5 token\_similarity 0.5 token\_similarity 0.5 100 100 100 100 token 500 token i 0.0 200 0.0 0.0 200 0.0 300 300 300 300 400 400 400 400 200 200 200 200 400 0 400 0 -1.0-1.0-1.0-1.0token j token j token j token j after layer 5 after layer 6 after layer 7 after layer 8 0 0 0 0 token\_similarity 0.5 token\_similarity token\_similarity 0.5 100 100 100 100 token 500 token 200 token 0.0 200 0.0 0.0 200 0.0 300 300 300 300 400 400 400 400 200 200 200 200 0 40 40 40 -1.0-1.0-1.0-1.0token j token i token j token j 1.0 1.0 1.0 after layer 9 after layer 10 after layer 11 after layer 12 0 0 0 0 0.C co.c co.c similarity token\_similarity 0.5 0.5 token\_similarity 0.5 0.5 100 100 100 100 token 500 token i 200 token i 200 0.0 0.0 0.0 200 300 300 300 300 -0.5 -0.5400 400 -400 400 0 200 0 200 200 400 200 400 400 0 0 -1.0-1.0-1.0-1.0token j token j token j token j 1.0 1.0 1.0 after layer 13 after layer 14 after layer 15 after layer 16 0 0 0.5 token similarity 0.5 0.5 token\_similarity 0.5 token\_similarity token\_similarity 100 100 100 100 token i 200 token i 200 token i token 200 0.0 200 0.0 0.0 0.0 300 300 300 300 400 400 -400 400 200 200 40( 200 40( 200 40 0 40 -1.0-1.0-1.0-1.0token j token j token j token j 1.0 1.0 1.0 1.0 after layer 17 after layer 18 after layer 19 after layer 20 0 o c. 2 token\_similarity o G G token\_similarity 0.5 token\_similarity 0.5 0.5 0.5 token\_similarity 100 100 100 100 token 200 token i token i 0.0 0.0 0.0 200 0.0 300 300 300 300 400 -400 400 400 200 200 200 200 40 40 40 0 40 -1.0-1.0-1.0-1.0token j token j token j token j after layer 23 after layer 21 after layer 22 after layer 24 0 0 o-0 20 10ken\_similarity 0.5 0.5 100 100 100 100 to 500 token i to ken : token i 0.0 300 300 300 300 400 400 · 400 400

200

token j

-1.0

200

token j

400

-1.0

200

token j

-1.0

200

token j