cos similarity histogram - layer 10 ert-large-uncased_uniform-tokens_(Qx,Ky)_trained-params_dense-on_24-lay 8 7 6 5 4 3 2 1 -0.500.00 0.25 0.50 -1.00-0.75-0.250.75 1.00