cos similarity histogram - layer 7 pert-large-uncased uniform-tokens (Qx,Ky) trained-params dense-on 24-lay 8 7 6 5 4 3 2

0.00

0.25

0.50

0.75

1.00

1

-1.00

-0.75

-0.50

-0.25