cos similarity histogram - layer 1 (albert-xlarge-v2 imdb (Qx,Ky) trained-params dense-on 24-layers) 8 7 6 5 4 3 2 1 -0.50-0.250.00 0.25 0.50

0.75

1.00

-1.00

-0.75