cos similarity histogram - layer 23 (albert-xlarge-v2_uniform-tokens_(Qx,Ky)_trained-params_dense-off_24-layer 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0 -0.500.00 0.25 0.50 -1.00-0.75-0.250.75 1.00