cos similarity histogram - layer 12 (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 -0.50-0.250.00 0.25 0.50 -1.00-0.750.75 1.00