cos similarity histogram - layer 23 (albert-xlarge-v2\_wikitext\_(x,y)\_trained-params\_dense-on\_24-layers) 5 4 3 2 1 -0.50-0.250.00 0.25 0.50 -1.00-0.750.75 1.00