cos similarity histogram - layer 23 (albert-xlarge-v2 wikitext (x,y) trained-params dense-off 24-layers) 10 -8 6 4 2 0.25 0.50 -1.00-0.75-0.50-0.250.00 0.75 1.00