cos similarity histogram - layer 16 (bert-large-uncased\_wikitext\_(x,y)\_trained-params\_dense-on\_24-layers) 10 8 6 4 2 -0.50-0.250.00 0.25 0.50 -1.00-0.750.75 1.00