cos similarity heatmap - layer 20 -large-uncased_uniform-tokens_(x,y)_trained-params_dense-off_24-layers) 0.75 100 0.50 0.25 0.00 cosine similarity 0.25 200 token i 0.00 300 -0.50400 -0.75500 -1.00100 200 300 400 500 token j