cosine T-SNE visualisation – layer 17 (albert-xlarge-v2_wikitext_ $\langle Qx,Ky \rangle$ _trained-params_dense-off_24-layers) 75 50 25 0 -25-50-75-100-100-75-50-2525 50 75 100