cosine T-SNE visualisation – layer 9 (albert-xlarge-v2\_wikitext\_ $\langle Qx,Ky \rangle$ \_trained-params\_dense-off\_24-layers) 75 50 25 0 -25-50-75-10050 -100-75-50-2525 75 100