$cosine \ T-SNE \ visualisations \\ (bert-large-uncased_wikitext_\langle x,y\rangle_trained-params_dense-on_24-layers)$ layer 0 layer 1 layer 2 layer 3 100 100 100 100 50 50 50 50 0 -50-50-50-50 -100 -100-1000 50 100 0 50 100 0 50 100 0 50 100 -50 -100 -50-50-50 -100 -100-100 layer 4 layer 5 layer 6 layer 7 100 100 100 100 50 50 50 50 0 0 0 -50-50 -50 **-**50 -100-100-100-50 0 50 100 -100-500 50 100 -100 -50 0 50 100 0 50 100 -100 -100-50 layer 8 layer 9 layer 10 layer 11 100 100 100 100 50 50 50 50 0 0 0 -50-50-50-100 -100 -100-10050 100 50 100 0 50 100 0 50 100 -100-50 -100-50 -100-50-100-50 layer 12 layer 13 layer 15 layer 14 100 100 100 100 50 50 50 50 -50**-**50 -50 -50-100-100-100-100 -50 50 -100 -50 0 50 -50 0 50 -50 0 50 -100 0 100 100 -100100 -100100 layer 16 layer 17 layer 18 layer 19 100 100 100 100 50 50 50 50 0 -50 -50 --50 · -50 --100 | -100 -100 --100 | -100 -100 Ò -50 50 100 -100 -50 0 50 100 -50 0 50 100 -50 0 50 100 -100 layer 20 layer 22 layer 23 layer 21 100 100 100 100 50 50 50 50 0 0 0 0 -50-50 -50 -50

-100

-100

-50

100

50

-100 -

-100

-50

50

100

100

50

-100 -

-100

-50

0

50

100

-100 -

-100

-50