Big Data Set : Gap threshold = 97 | foldChange = -Inf 1000 -750 **Dimensions** 500 2(Original data) - 5 250 0 80 85 90 95 threshold

Frequency of similarities Original Random 100000 1000 -Frequency 10 -140 100 120 160 20 40 . 08 .09 80 20 4 9 0 0 Similarity