package com.twitter.simclusters\_v2.common.clustering

import com.twitter.simclusters\_v2.common.SimClustersEmbedding

/\*\*

\* SimilarityFunctions provide commonly used similarity functions that this clustering library needs.

\*/

object SimilarityFunctions {

def simClustersCosineSimilarity: (SimClustersEmbedding, SimClustersEmbedding) => Double =

(e1, e2) => e1.cosineSimilarity(e2)

def simClustersMatchingLargestDimension: (

SimClustersEmbedding,

SimClustersEmbedding

) => Double = (e1, e2) => {

val doesMatchLargestDimension: Boolean = e1

.topClusterIds(1)

.exists { id1 =>

e2.topClusterIds(1).contains(id1)

}

if (doesMatchLargestDimension) 1.0

else 0.0

}

def simClustersFuzzyJaccardSimilarity: (

SimClustersEmbedding,

SimClustersEmbedding

) => Double = (e1, e2) => {

e1.fuzzyJaccardSimilarity(e2)

}

}