package com.twitter.simclusters\_v2.scalding

package multi\_type\_graph.assemble\_multi\_type\_graph

import com.twitter.simclusters\_v2.thriftscala.RightNodeType

object Config {

val User = System.getenv("USER")

val RootPath: String = s"/user/$User/manhattan\_sequence\_files/multi\_type\_simclusters/"

val RootThriftPath: String = s"/user/$User/processed/multi\_type\_simclusters/"

val AdhocRootPrefix = s"/gcs/user/$User/adhoc/multi\_type\_simclusters/"

val HalfLifeInDaysForFavScore = 100

val NumTopNounsForUnknownRightNodeType = 20

val GlobalDefaultMinFrequencyOfRightNodeType = 100

val TopKRightNounsForMHDump = 1000

// the topK most frequent nouns for each engagement type

val TopKConfig: Map[RightNodeType, Int] = Map(

RightNodeType.FollowUser -> 10000000, // 10M, current simclusters\_v2 has this value set to 20M, providing this the most weight

RightNodeType.FavUser -> 5000000,

RightNodeType.BlockUser -> 1000000,

RightNodeType.AbuseReportUser -> 1000000,

RightNodeType.SpamReportUser -> 1000000,

RightNodeType.FollowTopic -> 5000,

RightNodeType.SignUpCountry -> 200,

RightNodeType.ConsumedLanguage -> 50,

RightNodeType.FavTweet -> 500000,

RightNodeType.ReplyTweet -> 500000,

RightNodeType.RetweetTweet -> 500000,

RightNodeType.NotifOpenOrClickTweet -> 500000,

RightNodeType.SearchQuery -> 500000

)

val SampledEmployeeIds: Set[Long] =

Set()

}