SOTON-WAIS @ SED2013

Sparse Features and Incremental Density Based Clustering

Sina Samangooe
BUILD SPARSE MATRIX

Lucene → SPARSE FEATURE MATRIX → COMBINED SPARSE MATRIX
BUILD SPARSE MATRIX

Lucene → SPARSE FEATURE MATRIX → COMBINED SPARSE MATRIX

CLUSTER

DBSCAN

+ INCREMENTAL

SPECTRAL
MAIN ACHIEVEMENTS

• CREATE **BIG** SPARSE SIMILARITY MATRICES
  – **(QUICKLY)**

• **OPEN SOURCE**
  – CLUSTERING, EXPERIMENTS... **THE LOT**
  – ...**OPENIMAJ**... **WOO!**

• **INCREMENTAL CLUSTERING**
  – SPECTRAL CLUSTERING WITH 300K ITEMS, 0.01 DENSITY.
  – HARD TO DO NAIVELY, TRY IN MATLAB OR PYTHON:

    eigsh(srand(300000,300000,0.001),1000)

assuming you have **128GB RAM** it’ll be done soon(ish)