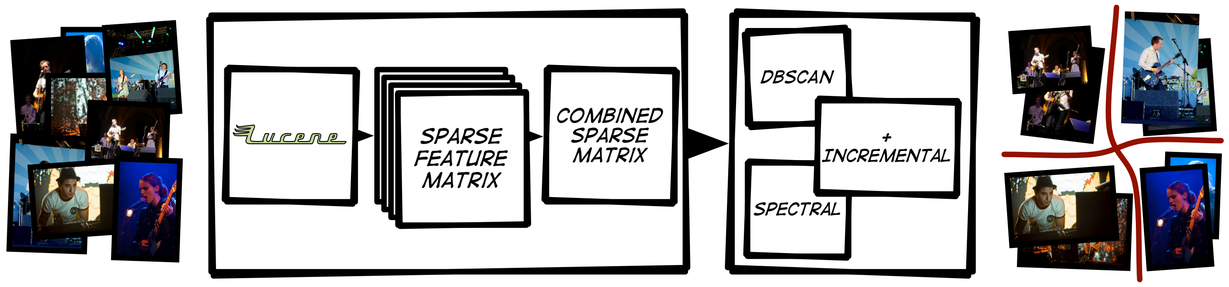
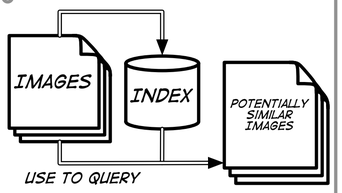
OVERALL APPROACH



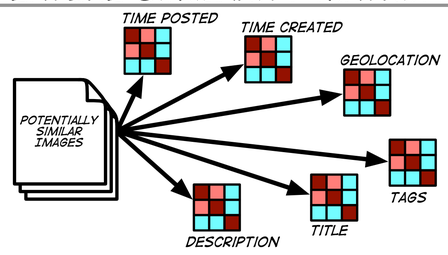
SPARSE SIMILARITY MATRIX

LUCENE INDEX



... CONSTRUCT A LUCENE INDEX OF ALL THE IMAGES WE HOPE TO CLUSTER. CONSTRUCT A QUERY FOR EACH IMAGE AND USE ONLY RETURNED RESULTS FOR THE NEXT STAGE...

FEATURE SIMILARITY MATRIX



... EXTRACT FEATURES FOR EACH IMAGE. USING CUSTOM DISTANCE METRICS COMPUTE THE SIMILARITY OF TWO IMAGES. CONSTRUCT A SIMILARITY MATRIX FOR EACH FEATURE FOR THE NEXT STAGE...

MATRIX COMBINATION

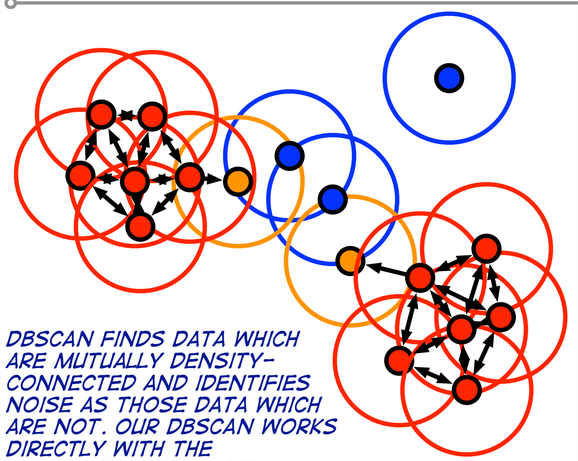
$$\sum W_{\blacksquare} = \blacksquare$$

$$\sum W_{\blacksquare} = 1$$

... WEIGHT EACH FEATURE SIMILARITY MATRIX BY SOME LEARNT WEIGHT AND SUM TOGETHER TO FORM A SINGLE FEATURE FUSED SIMILARITY MATRIX READY FOR CLUSTERING...

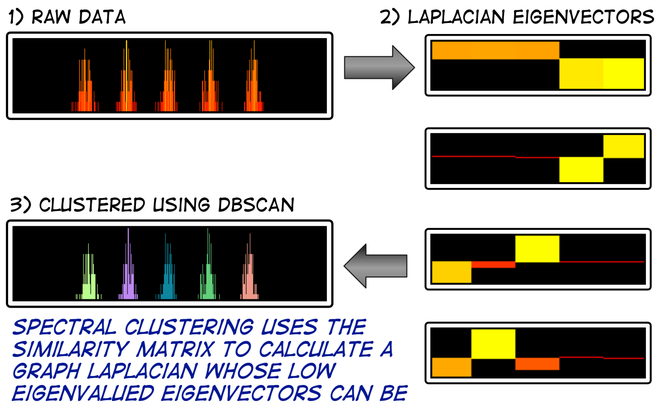
CLUSTERING

DBSCAN



DBSCAN FINDS DATA WHICH ARE MUTUALLY DENSITY-CONNECTED AND IDENTIFIES NOISE AS THOSE DATA WHICH ARE NOT. OUR DBSCAN WORKS DIRECTLY WITH THE AGGREGATED SIMILARITY MATRIX AS WELL AS WITH STANDARD GEOMETRIC POINTS

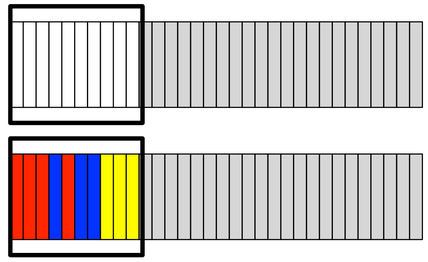
SPECTRAL CLUSTERING



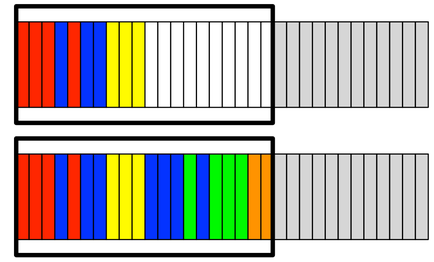
SPECTRAL CLUSTERING USES THE SIMILARITY MATRIX TO CALCULATE A GRAPH LAPLACIAN WHOSE LOW EIGENVALUED EIGENVECTORS CAN BE USED AS A SPACE WITHIN WHICH ANOTHER CLUSTERING ALGORITHM CAN BETTER CLUSTER DATA. WE USE DBSCAN FOR THIS.

INCREMENTAL CLUSTERING

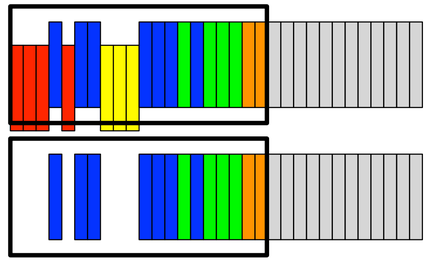
1) CLUSTER INITIAL WINDOW



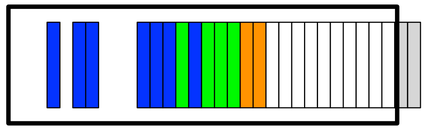
2) GROW WINDOW, CLUSTER AGAIN



3) IDENTIFY STABLE CLUSTERS



4) CONTINUE IGNORING STABLE CLUSTERS



...

RESULTS

WEIGHT SELECTION

A SIMPLEX SEARCH WAS USED TO FIND THE BEST WEIGHTING OF FEATURES. HOWEVER THE DIFFERENCE BETWEEN THE TOP 1000 POINTS ON THE SIMPLEX WERE SMALL SO AN AVERAGE WEIGHTING COMBINING THE TOP 1000 WEIGHTS WAS ALSO USED

SETTING	TIME TAKEN	TIME POSTED	LOCATION	TEXT DESC	TEXT TITLE	TEXT TAGS
BEST	2	0	1	1	0	3
AVERAGE	2.1	1.8	1.4	0.7	0.3	1.7

OFFICIAL RESULTS

IN ALL EXPERIMENTAL SETTINGS SUBMITTED THE INCREMENTAL ALGORITHM WAS USED. THOUGH OUR INCREMENTAL TECHNIQUE ALLOWED FOR SPECTRAL CLUSTERING ACROSS A RELATIVELY LARGE DATASET, WE FOUND DBSCAN PERFORMED BEST OVERALL

	F1	NMI	F1(DIV)	RB F1	DIV F1
DBSCAN (BEST)	0.945	0.985	0.935	0.059	**0.887**
SPECTRAL (BEST)	0.911	0.977	0.882	0.058	0.853
DBSCAN (AVG)	**0.946**	**0.985**	**0.936**	0.060	0.886
SPECTRAL (AVG)	0.902	0.974	0.866	0.057	0.846

TITLE

SOCIAL EVENT DETECTION VIA SPARSE MULTI-MODAL FEATURE SELECTION AND INCREMENTAL DENSITY BASED CLUSTERING

CLIENT
MEDIAEVAL 2013
SOCIAL EVENT DETECTION FOR SOCIAL MULTIMEDIA

POWERED BY
Lucene
OpenIMAJ Intelligent Multimedia Analysis in Java

TEAM
SOTON-WAIS
(UNIVERSITY OF SOUTHAMPTON)

MEMBERS
SINA SAMANGOOEI, JONATHON HARE, DAVID DUPPLAW, MAHESAN NIRANJAN, NICHOLAS GIBBINS, PAUL LEWIS, NEHA JAIN, JOHN PRESTON, JAMIE DAVIES

CONTACT
SINA SAMANGOOEI
SS@ECS.SOTON.AC.UK

SPONSORS
University of Southampton
arcomem
TrendMiner
SEVENTH FRAMEWORK PROGRAMME