

JONATHON HARE

MINING EVENTS FROM
MULTIMEDIA STREAMS

ACKNOWLEDGEMENTS

Sina Samangooei, Nick Gibbins, Mahesan Niranjan,
David Dupplaw, Paul Lewis, Kirk Martinez, Jamie Davies,
Neha Jain, John Preston, Laurence Wilmore, Ash Smith,
Heather Packer and other members of the WAIS & VLC
research groups at the University of Southampton



“Life is full of special moments. So is your photo library.”

APPLE WWDC KEYNOTE 2013

BUT WHAT IF WE GO BEYOND
PERSONAL MEDIA LIBRARIES?

CAN WE AUTOMATICALLY
FIND MEANINGFUL EVENTS
AND TRENDS IN STREAMS
OF SOCIAL MULTIMEDIA?

CONTENTS

- Research challenges
- Case studies using shared social media:
 - Monitoring Twitter's visual pulse
 - Detecting social events
- Open questions & future directions



flickr

RESEARCH CHALLENGES 1

- How do we deal with **massive** amounts of data?
 - There is a **temporal** aspect to the data we're dealing with...
 - Develop **streaming algorithms** that can be:
 - Incrementally **updated**
 - Allowed to **forget**

RESEARCH CHALLENGES 2

- How can we effectively make use of **contextual data** and **metadata** from different modalities?
 - Develop techniques for **exploiting all modalities** and **fusing features** from each modality effectively
 - Develop techniques that are robust to **missing** or **inaccurate** features

MONITORING TWITTER'S VISUAL PULSE

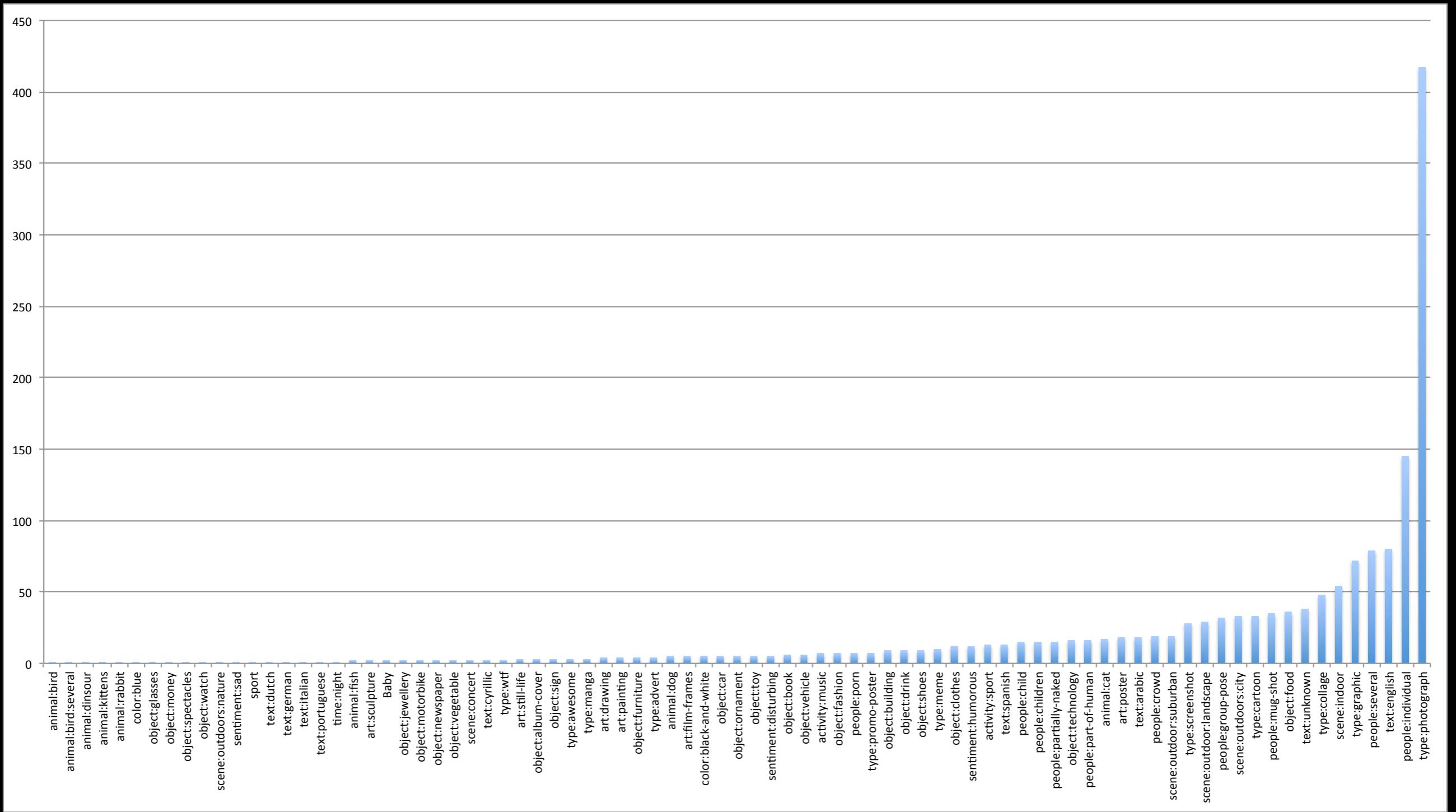


PHOTO SHARING (BROADCASTING?) ON TWITTER

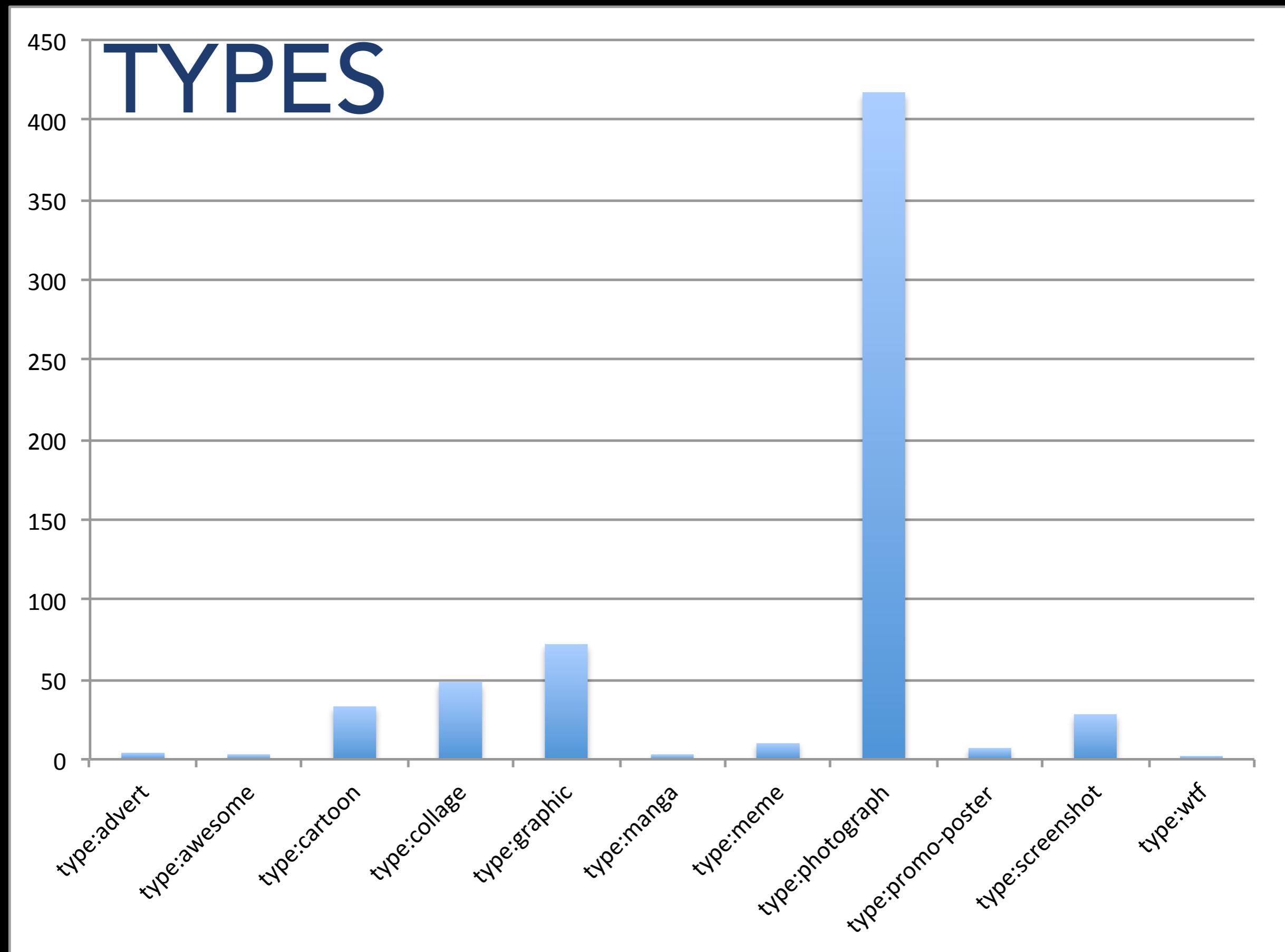
- Photos are heavily shared on Twitter
 - In 2012 we were capturing 100,000 images per day from the sample stream
 - Equates to 70GB of data
 - Sample stream is supposedly ~1% of all tweets
 - Implying on the firehose upwards of 10,000,000 images per day
 - Thats 7TB of data per day

WHAT TYPES OF PHOTOS
ARE SHARED ON TWITTER?

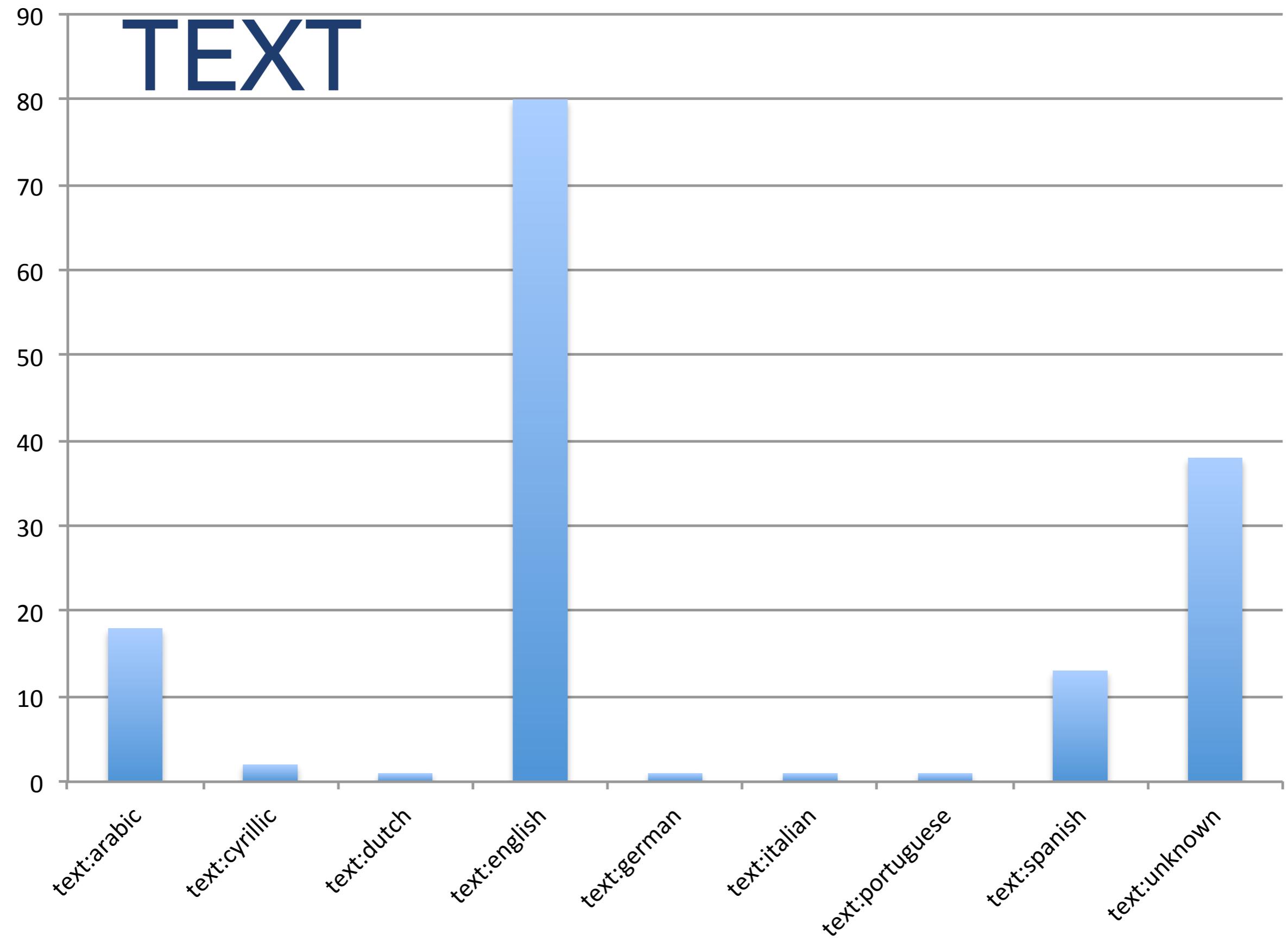
- Small study performed in September 2012
 - Aim to get a *feel* for what is being Tweeted
 - ~1 Day's worth of data chosen (81368 images)
 - 14 *trusted* Human annotators from WAIS
 - Uniform random sample of 667 images annotated



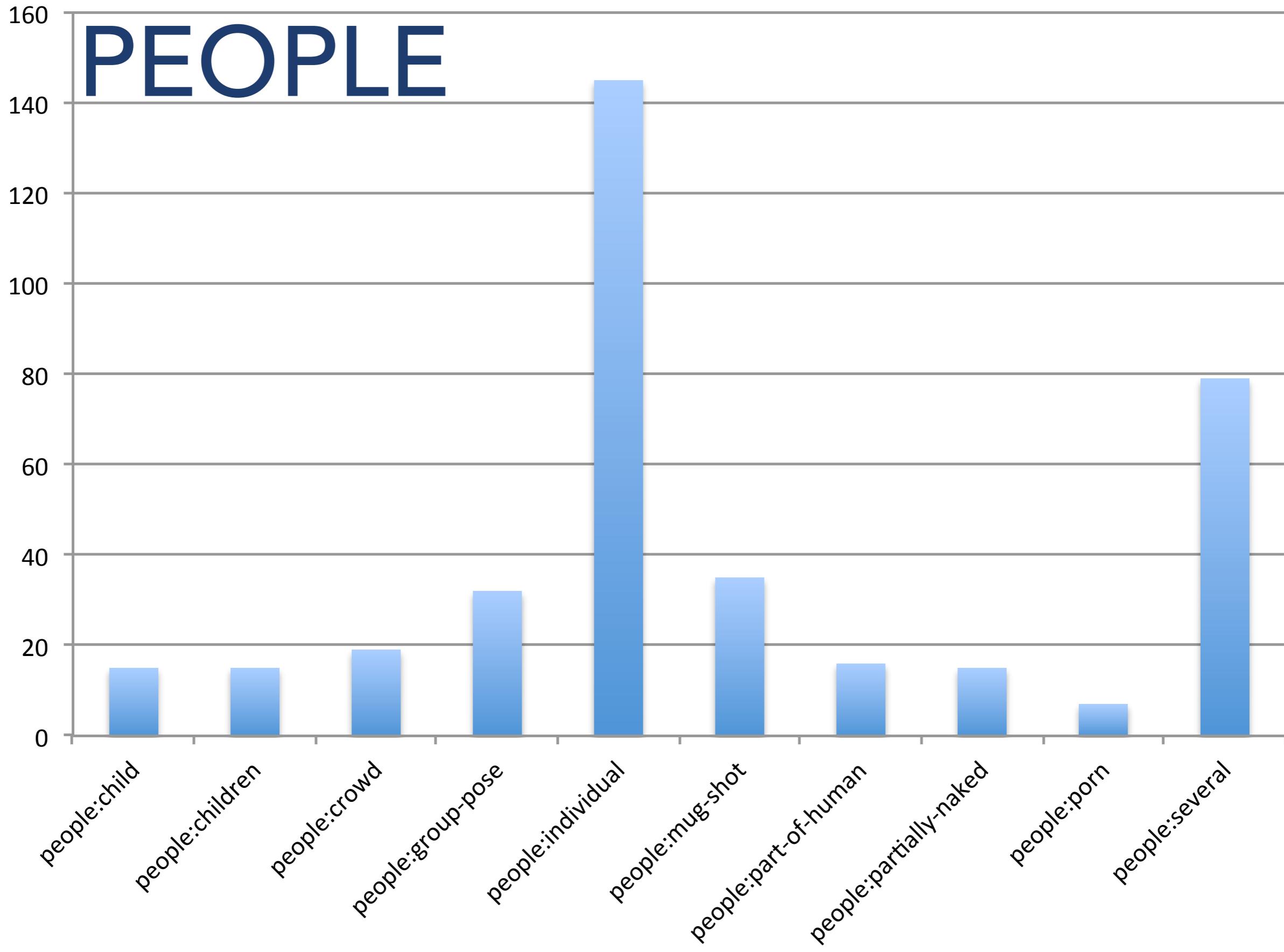
TYPES



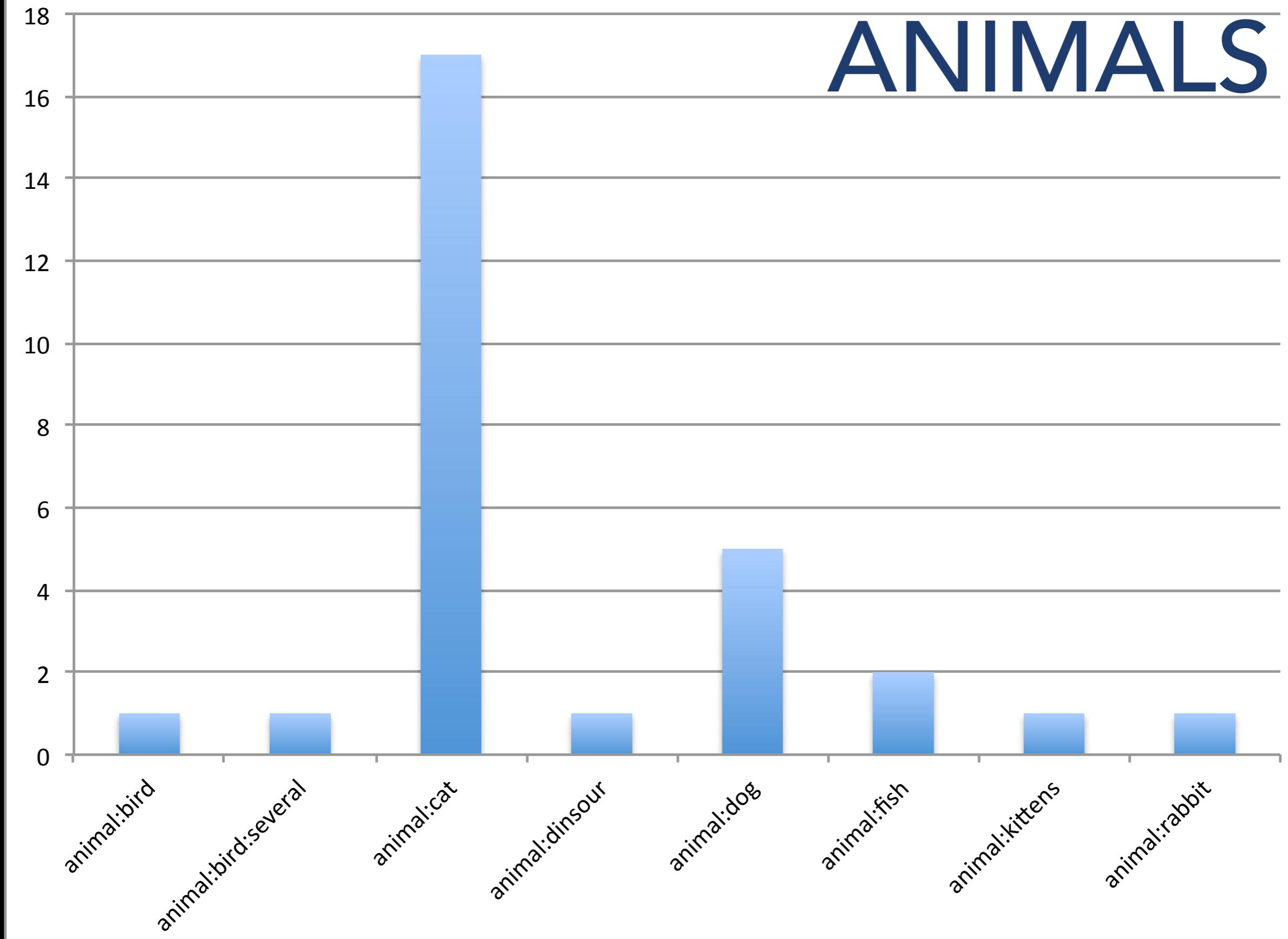
TEXT



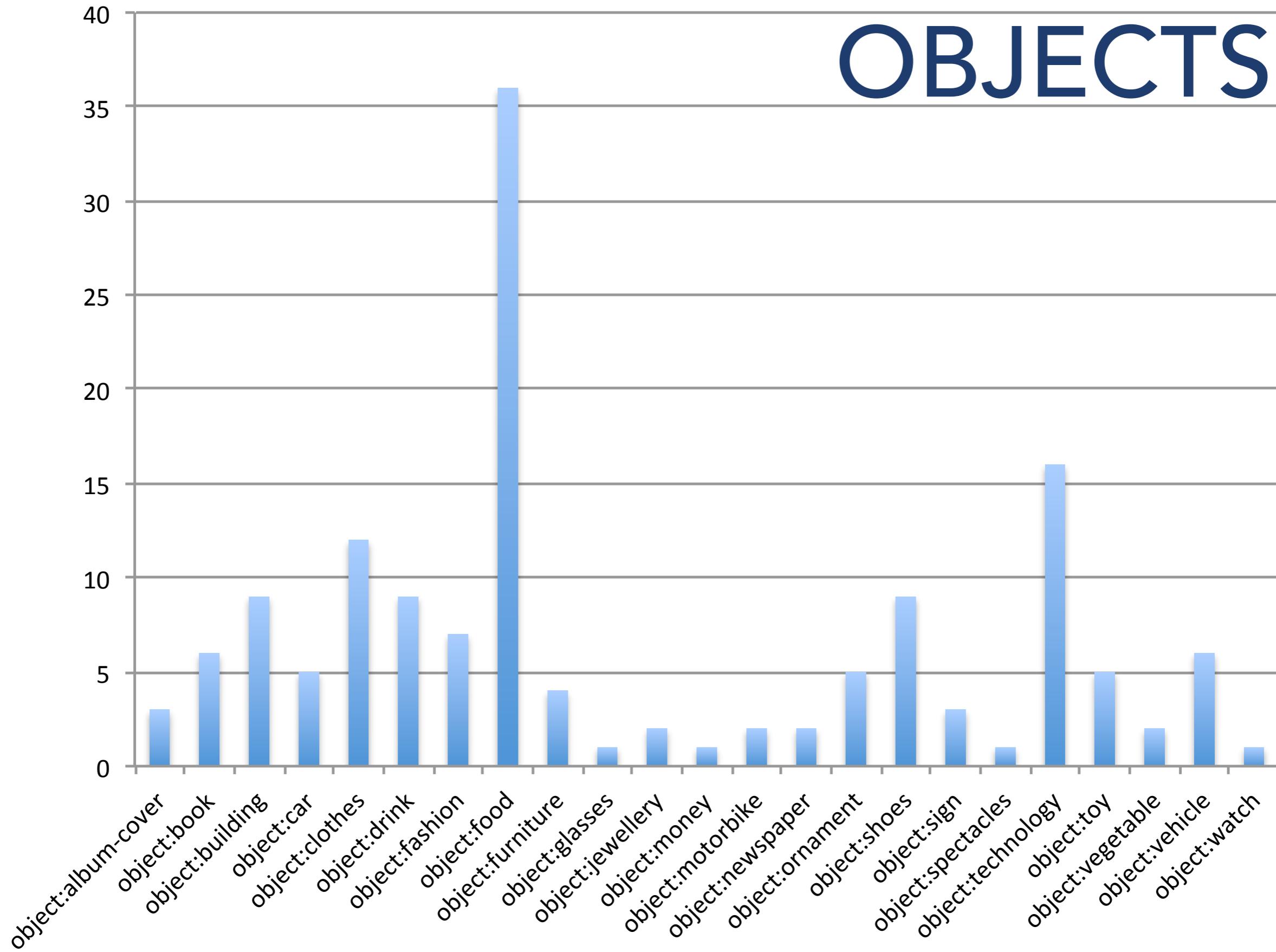
PEOPLE



ANIMALS



OBJECTS

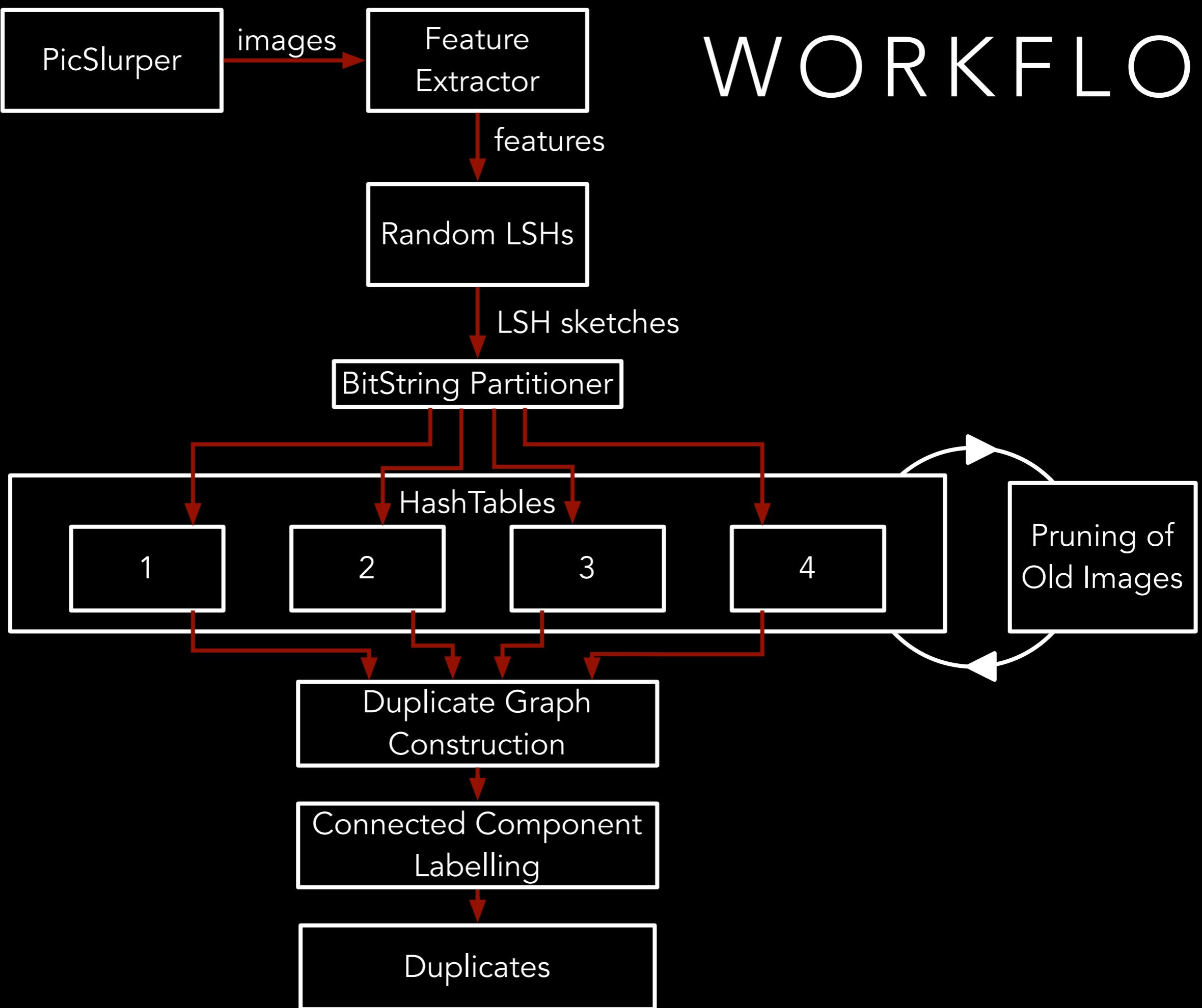


CAN WE FIND WHAT IMAGES
ARE TRENDING ON TWITTER?

- Look for near duplicates of images we've seen before
 - can't just use the URL...
 - the same image appears at many different URLs
 - potentially with slight changes
 - compression
 - added text
 - scaling
 - rotation/warps
 - Don't need to consider all images we've seen in the past...
 - maybe just those in a sliding time window

- Need a fast and robust near duplicate detection technique that can be adapted to working with a stream.
 - Want to use a local feature (i.e. DoG-SIFT) for robustness
 - BoVW+Inverted index, VLAD+PQ not ideal
 - ...use locality sensitive hashing...
 - Idea inspired by method of Dong et al at ICMR'12

WORKFLOW



4 HASH TABLES

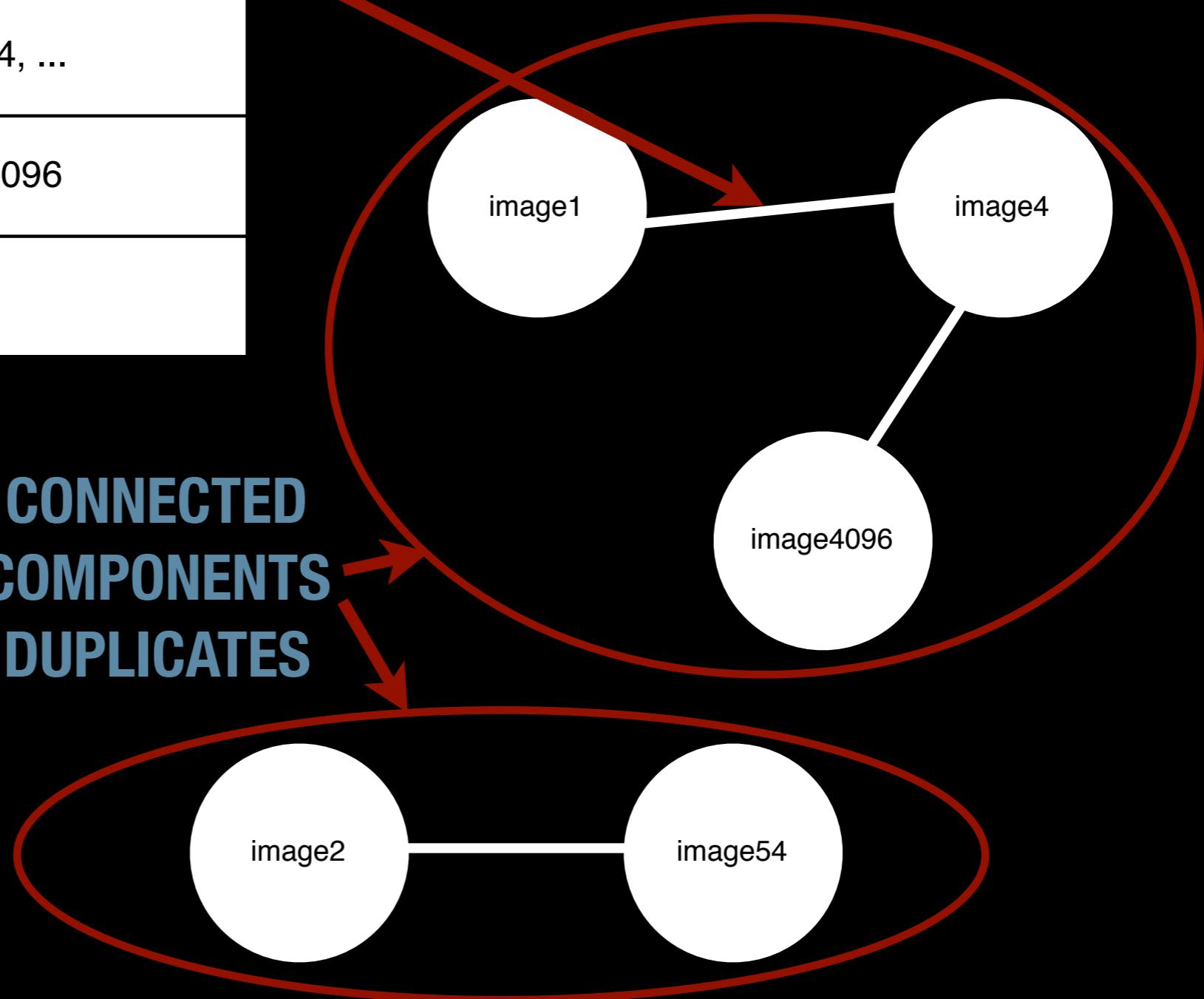
HashTable	
1	image1, image4
212	image2, image54, ...
783	image4, image4096
...	...

32-BIT KEY

**VALUES ARE
LISTS OF
IMAGES**

**ANY COLLISION RESULTS IN A
GRAPH EDGE**

**CONNECTED
COMPONENTS
DUPLICATES**



SOCIAL EVENT DETECTION



MediaEval Benchmark



<http://www.multimediaeval.org/>

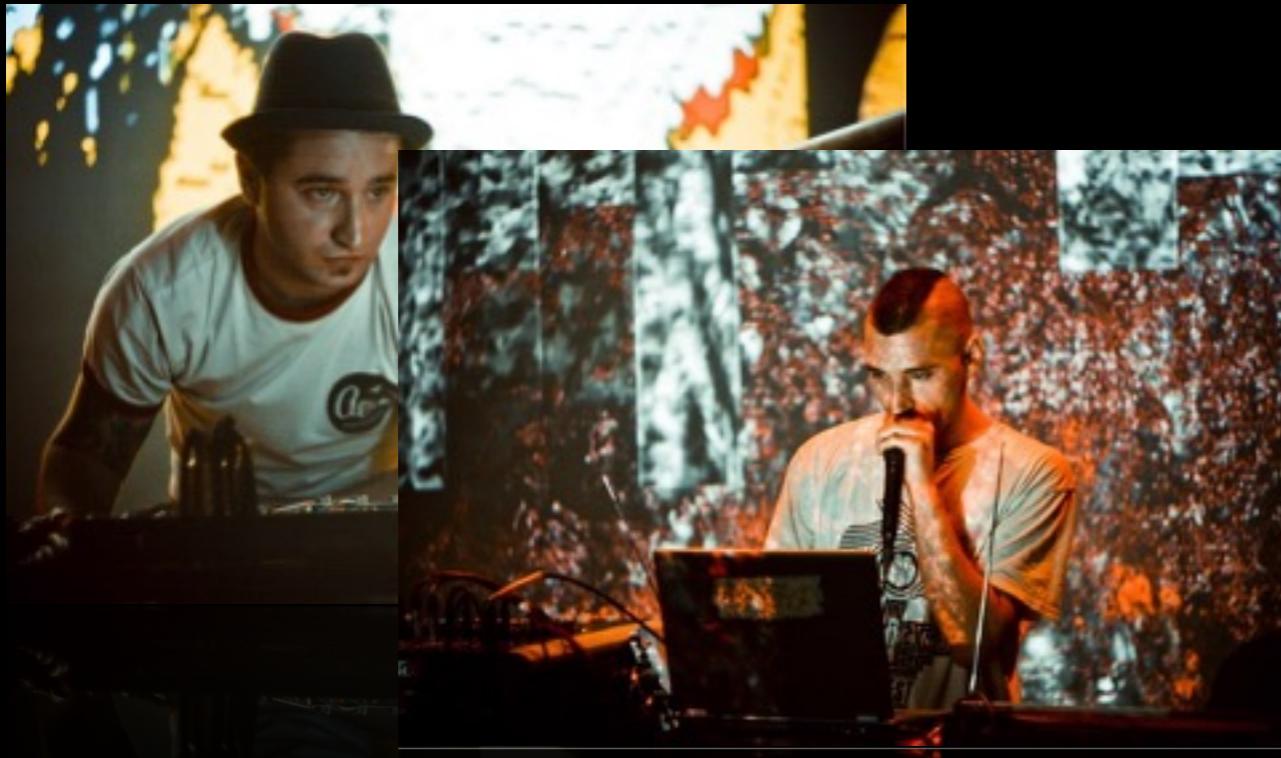
MEDIAEVAL 2013: SED TASK

- Open evaluation challenge at MediaEval 2013:
 - Event clustering of multimodal social streams
- Specifically:
 - Detect social events...
 - Given 500k flickr images with
 - image, tags, (some) geo, (some) time take, time posted etc.

WHAT ARE SOCIAL EVENTS?

We define ***social events*** as events that are ***planned*** by people, ***attended*** by people and the media illustrating the events are ***captured*** by people

SED2013 EXAMPLES



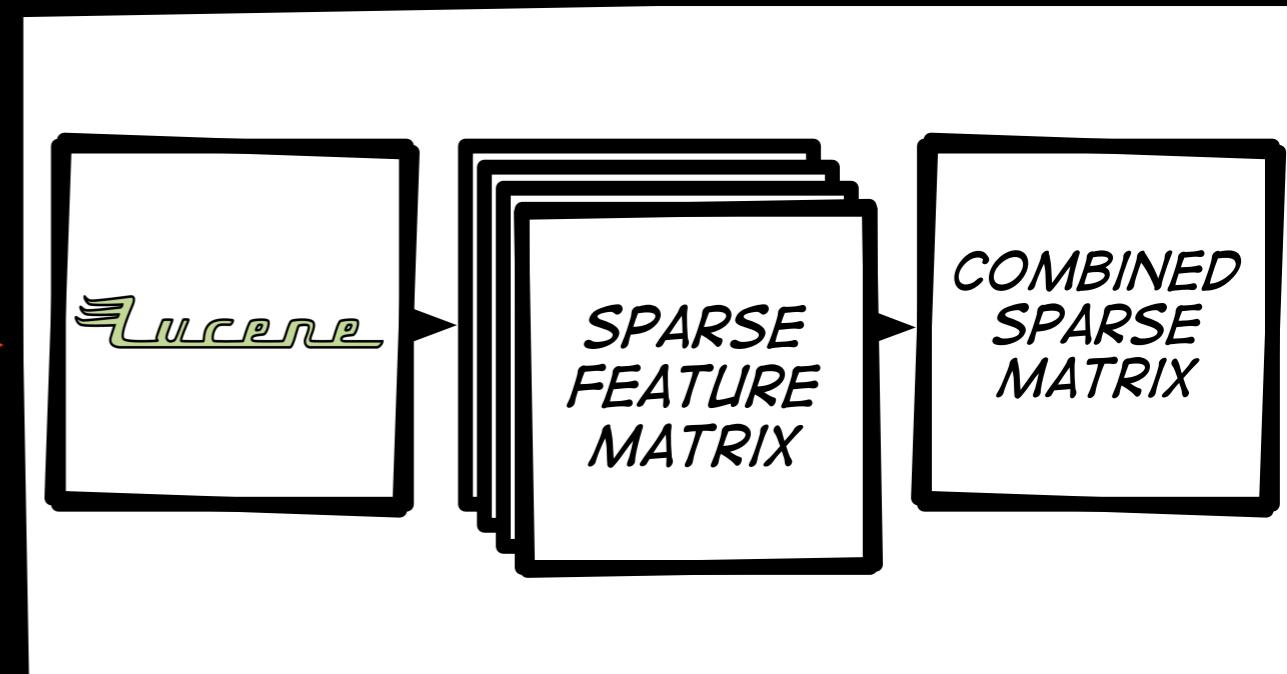
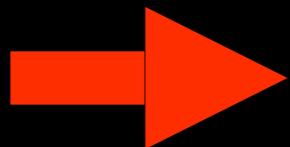
```
<PHOTO ID="4302746429" PHOTO_URL="HTTP://  
FARM5.STATICFLICKR.COM/  
4068/4302746429_F8CD7F2582.JPG" USERNAME="AT THE  
FOOT OF THE HILL" DATETAKEN="2010-01-23  
21:36:05.0" DATEUPLOADED="2010-01-25 10:43:34.0">  
<TITLE>POO, DEAD VOICES ON AIR @ A4</TITLE>  
<DESCRIPTION>WHOLE SET ...</DESCRIPTION>  
<TAGS>  
<TAG>2/58</TAG>  
<TAG>58MM</TAG>  
<TAG>CONCERT</TAG>  
<TAG>F2</TAG>  
<TAG>POO</TAG>  
</TAGS>  
</PHOTO>
```



```
<PHOTO ID="2280060852" PHOTO_URL="HTTP://  
FARM3.STATICFLICKR.COM/  
2110/2280060852_16539F5F0D.JPG"  
USERNAME="PILOT_10" DATETAKEN="2008-02-19  
23:34:09.0" DATEUPLOADED="2008-02-20 18:39:50.0">  
<TITLE>JENS LEKMAN @ TEATRO RASI, RAVENNA</TITLE>  
<DESCRIPTION>19 FEBBRAIO 2008</DESCRIPTION>  
<TAGS>  
<TAG>ALL RIGHTS RESERVED</TAG>  
<TAG>CONCERTI</TAG>  
<TAG>COOLPIX_4300</TAG>  
<TAG>EMILIAROMAGNA</TAG>  
</TAGS>  
<LOCATION LATITUDE="44.4153"  
LONGITUDE="12.2052"></LOCATION>  
</PHOTO>
```

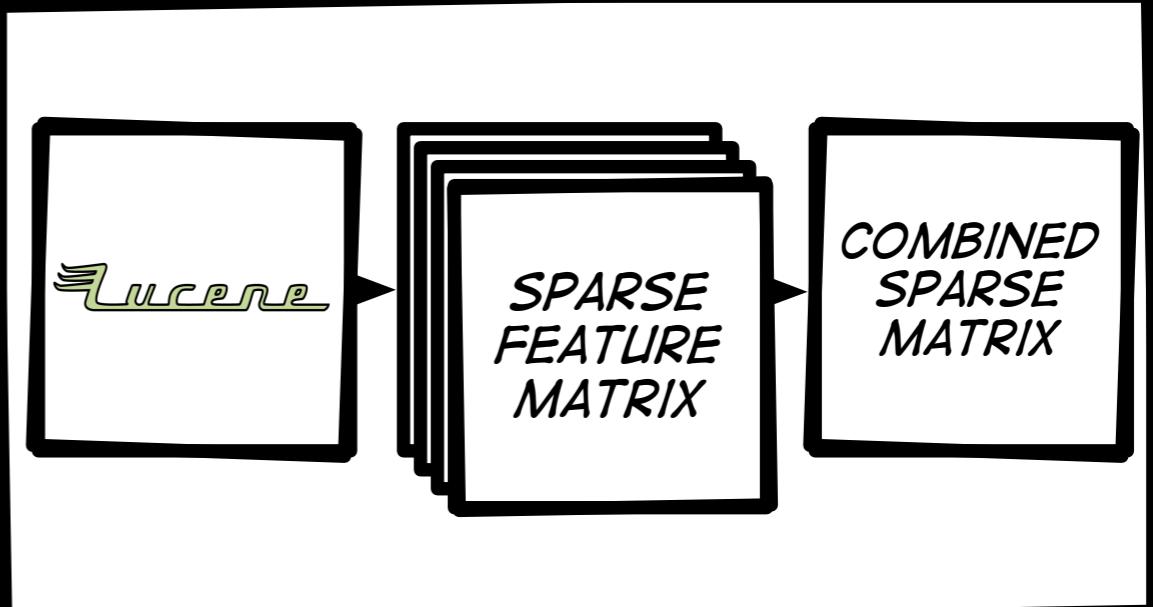


BUILD SPARSE MATRIX

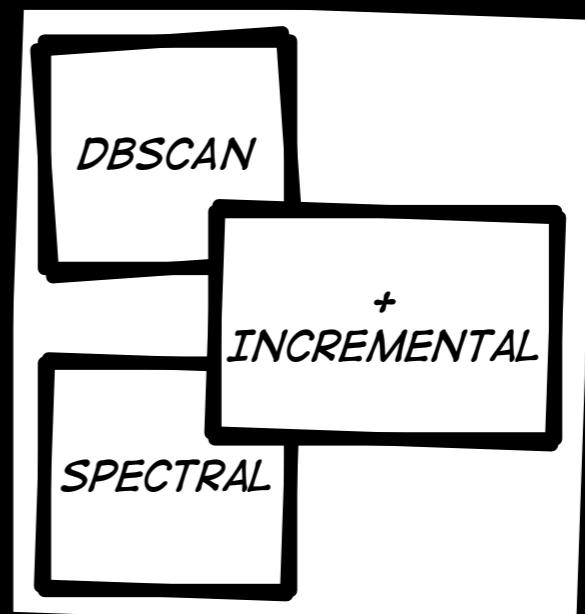




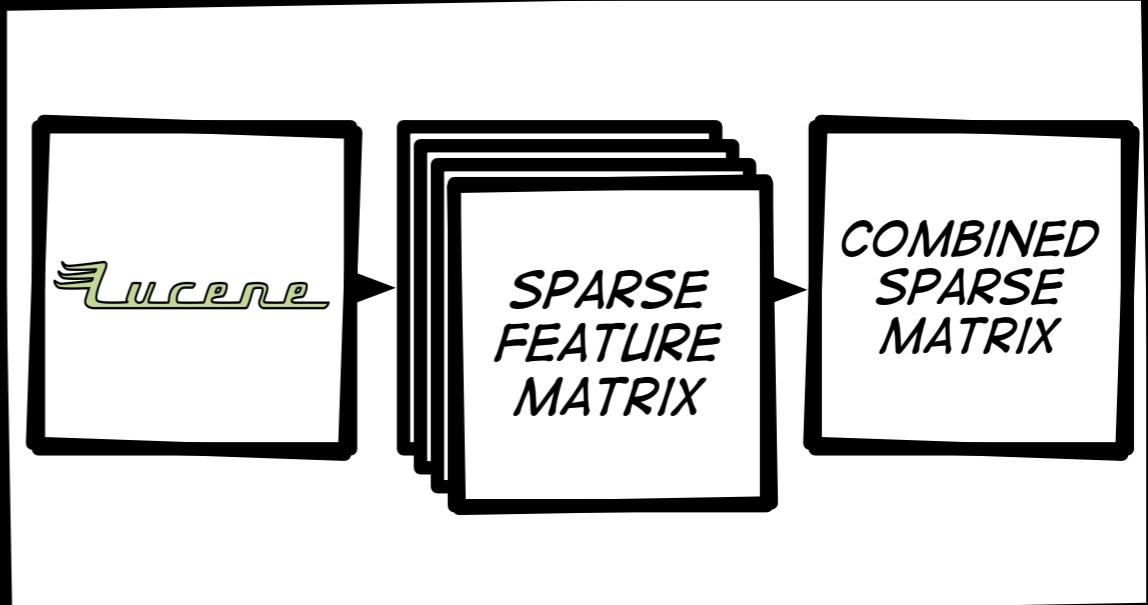
BUILD SPARSE MATRIX



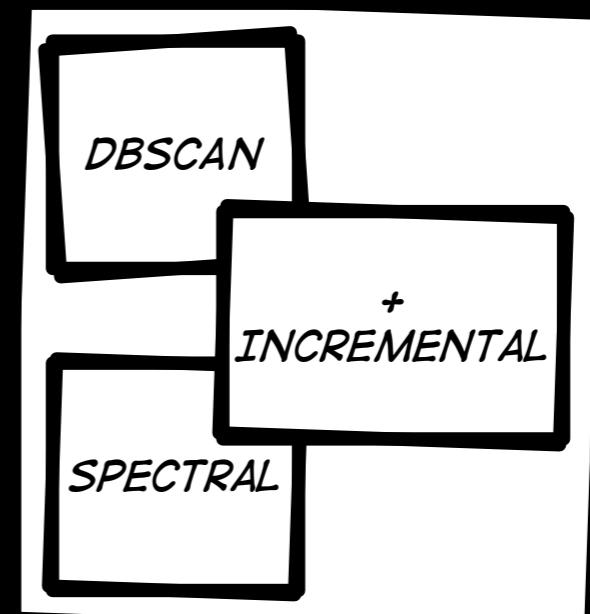
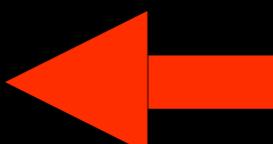
CLUSTER



BUILD SPARSE MATRIX



CLUSTER



FEATURES

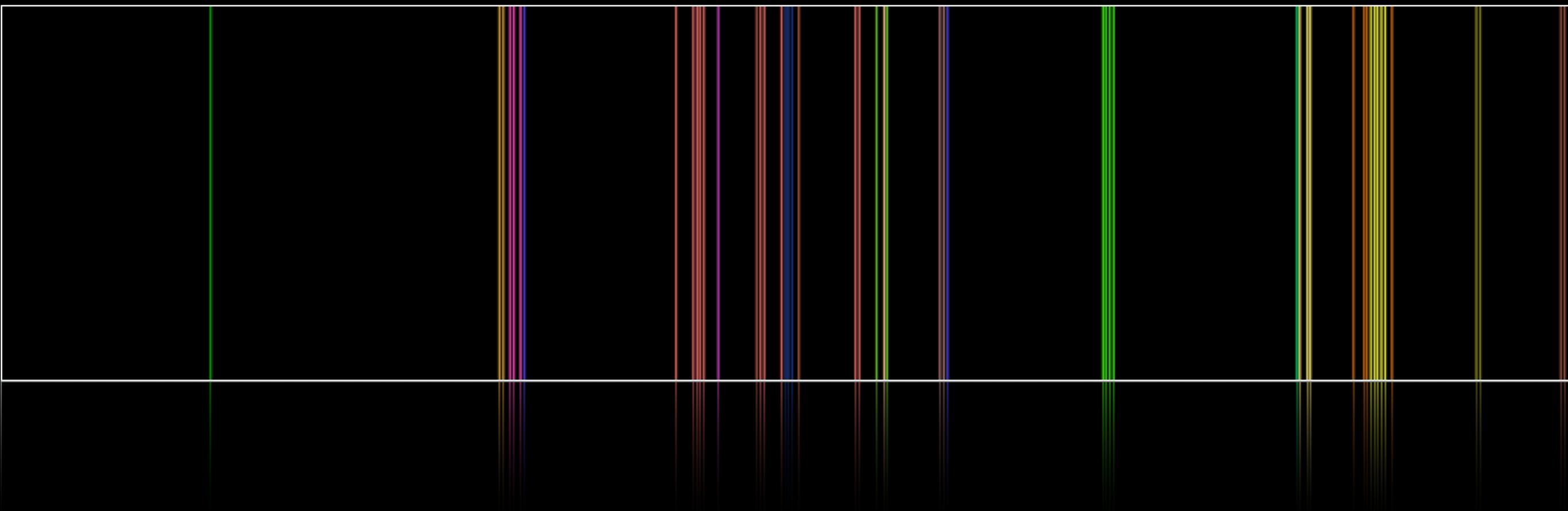
- Events potentially separable using:
 - **Images**: should look similar?
 - **Time**: should be temporally close?
 - **Location**: should be geographically close?
 - **Text**: should be described similarly?
- Flickr multimodal social stream contains:
 - Time taken (potentially inaccurate or missing entirely)
 - Time posted (accurate, though may be event agnostic)
 - Geo (often inaccurate, sparse)
 - Tags, title, description (multi-tag, spelling etc.)
 - Image features (these didn't work so well!)

FEATURES

- Events potentially separable using:
 - **Images**: should look similar?
 - **Time**: should be temporally close?
 - **Location**: should be geographically close?
 - **Text**: should be described similarly?
- Flickr multimodal social stream contains:
 - Time taken (potentially inaccurate or missing entirely)
 - Time posted (accurate, though may be event agnostic)
 - Geo (often inaccurate, sparse)
 - Tags, title, description (multi-tag, spelling etc.)
 - Image features (these didn't work so well!)

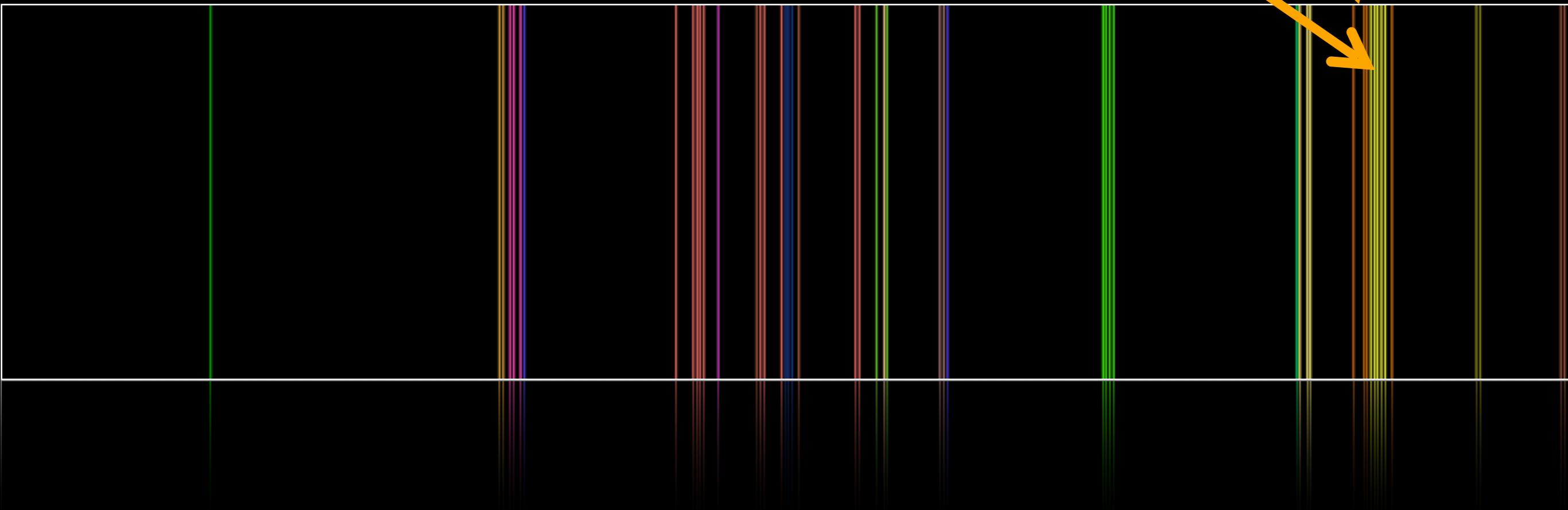
EVENT SEPARATION WITH FEATURES

- January 2007 until February 2007
- Random color assigned to clusters



EVENT SEPARATION WITH FEATURES

EVENTS LIKE THIS COULD EASILY BE CONFUSED



EVENT SEPARATION WITH FEATURES

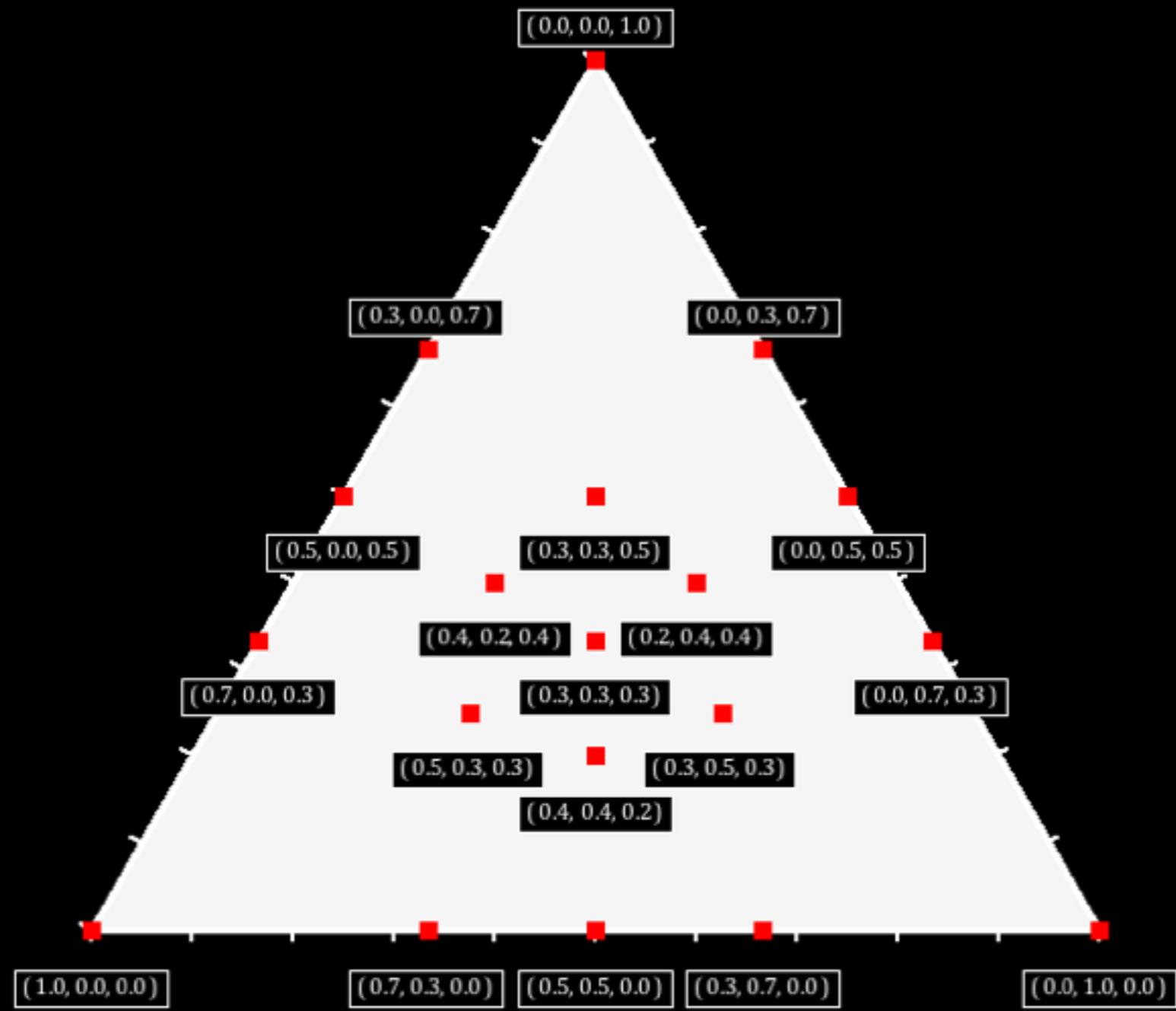
MORE FEATURES MAY HELP SEPARATE EVENTS



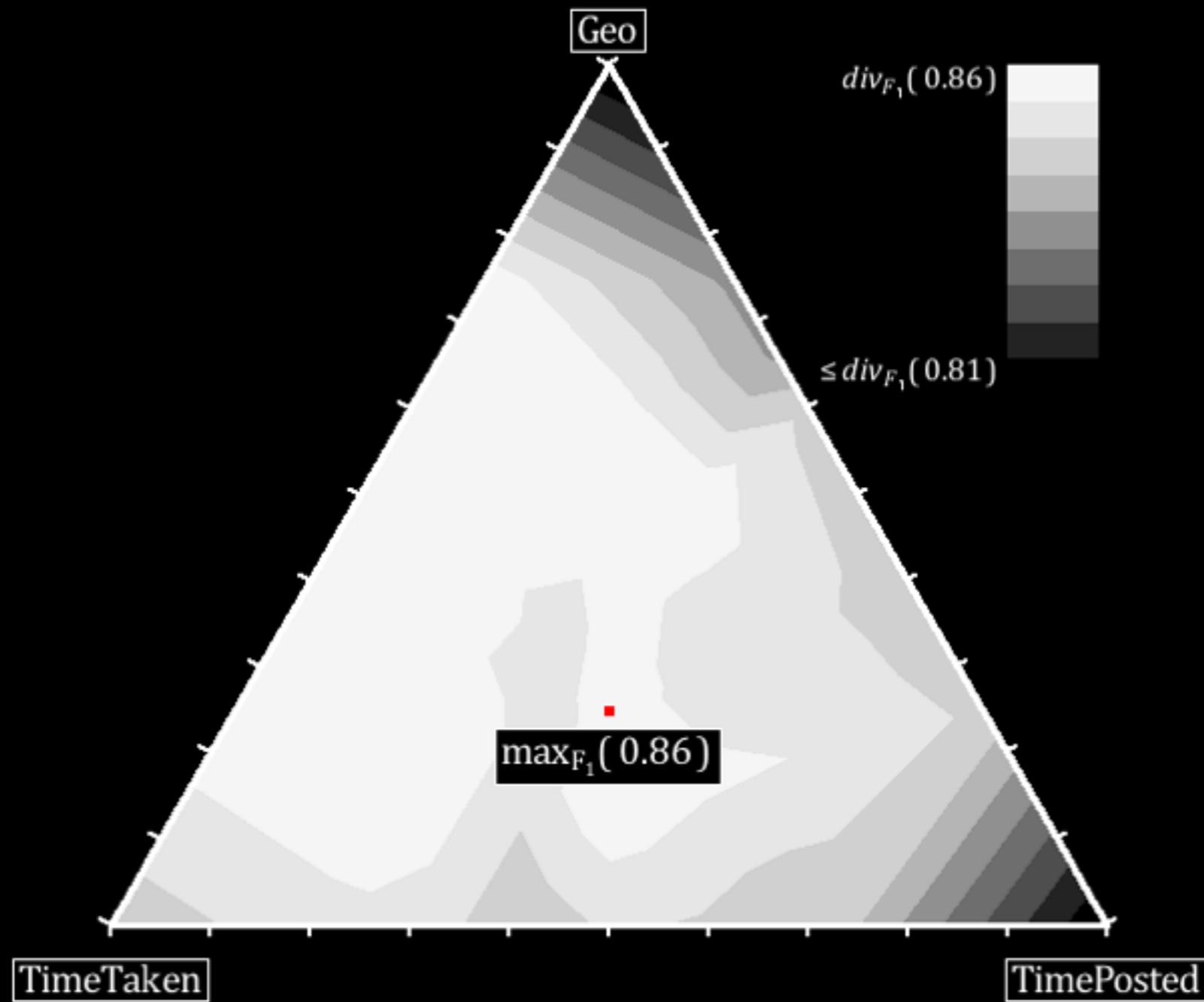
FEATURE WEIGHTS

- The **features matter** for different reasons
 - Some are more important than others
 - *This is a feature fusion problem!*
 - We went with a simple strategy: **weighted summation**
 - Use simplex search to find what feature weights give the best cluster quality

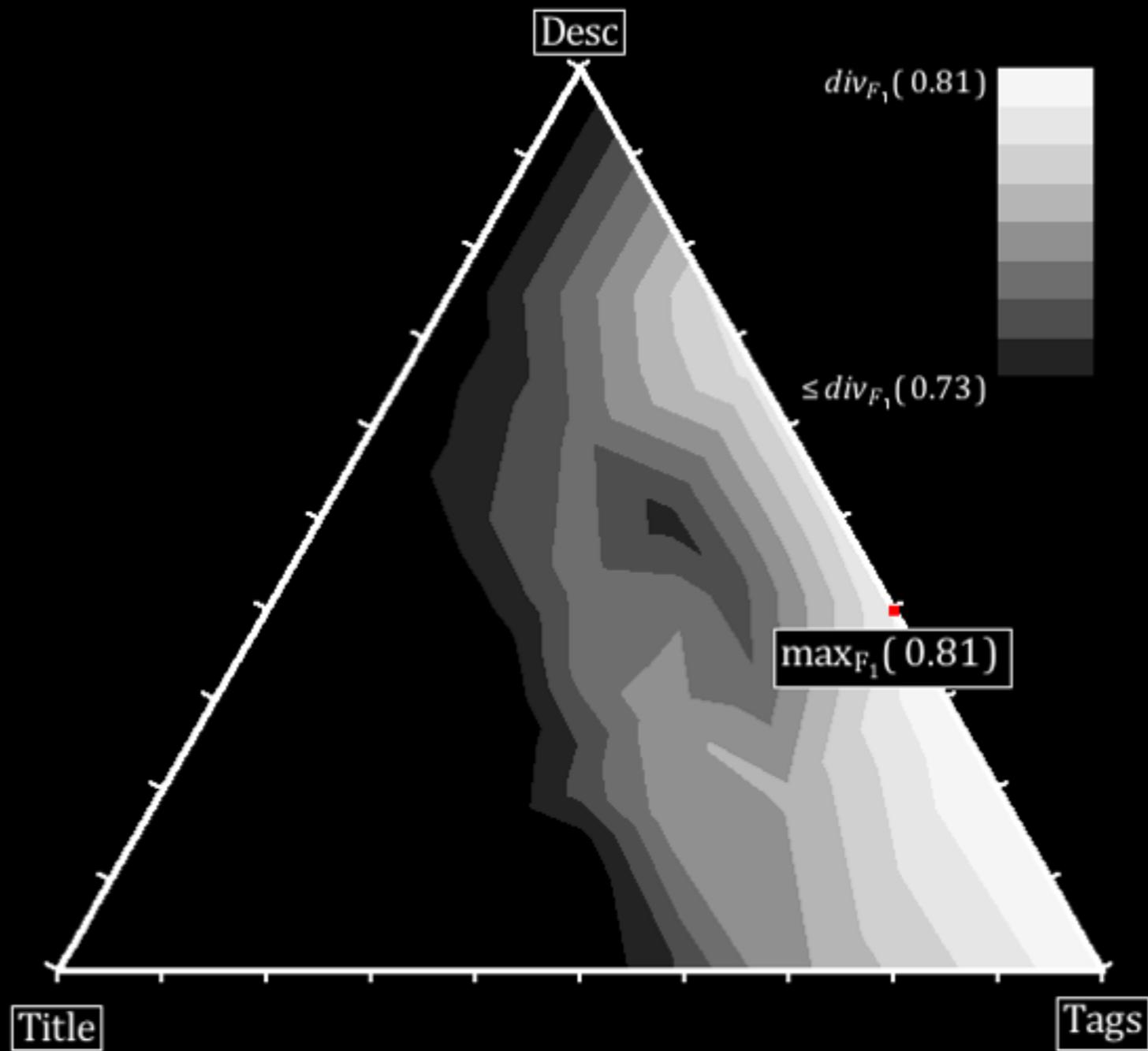
FEATURE WEIGHT SIMPLEX



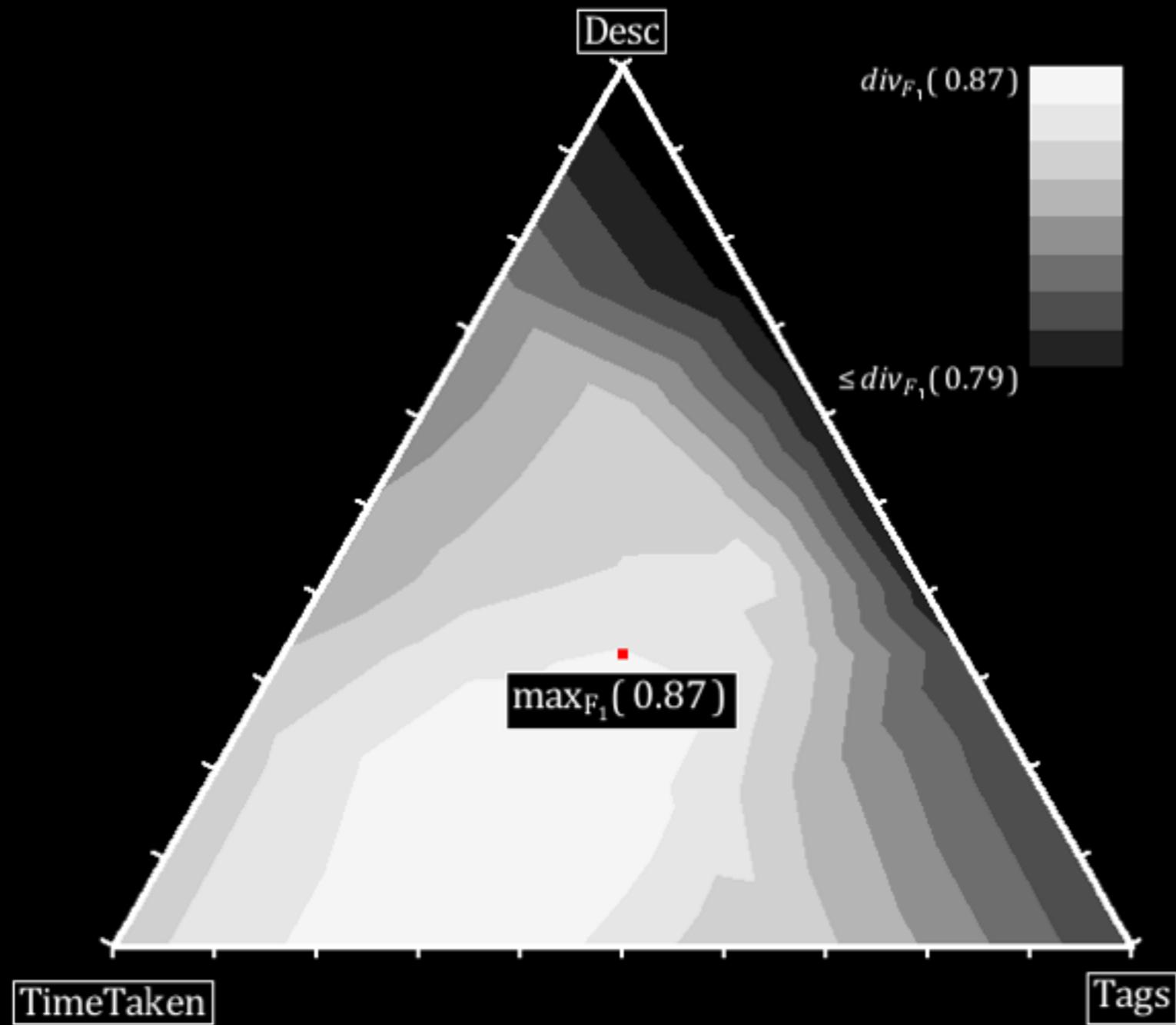
FEATURE WEIGHT SIMPLEX



FEATURE WEIGHT SIMPLEX



FEATURE WEIGHT SIMPLEX



FEATURE WEIGHTS

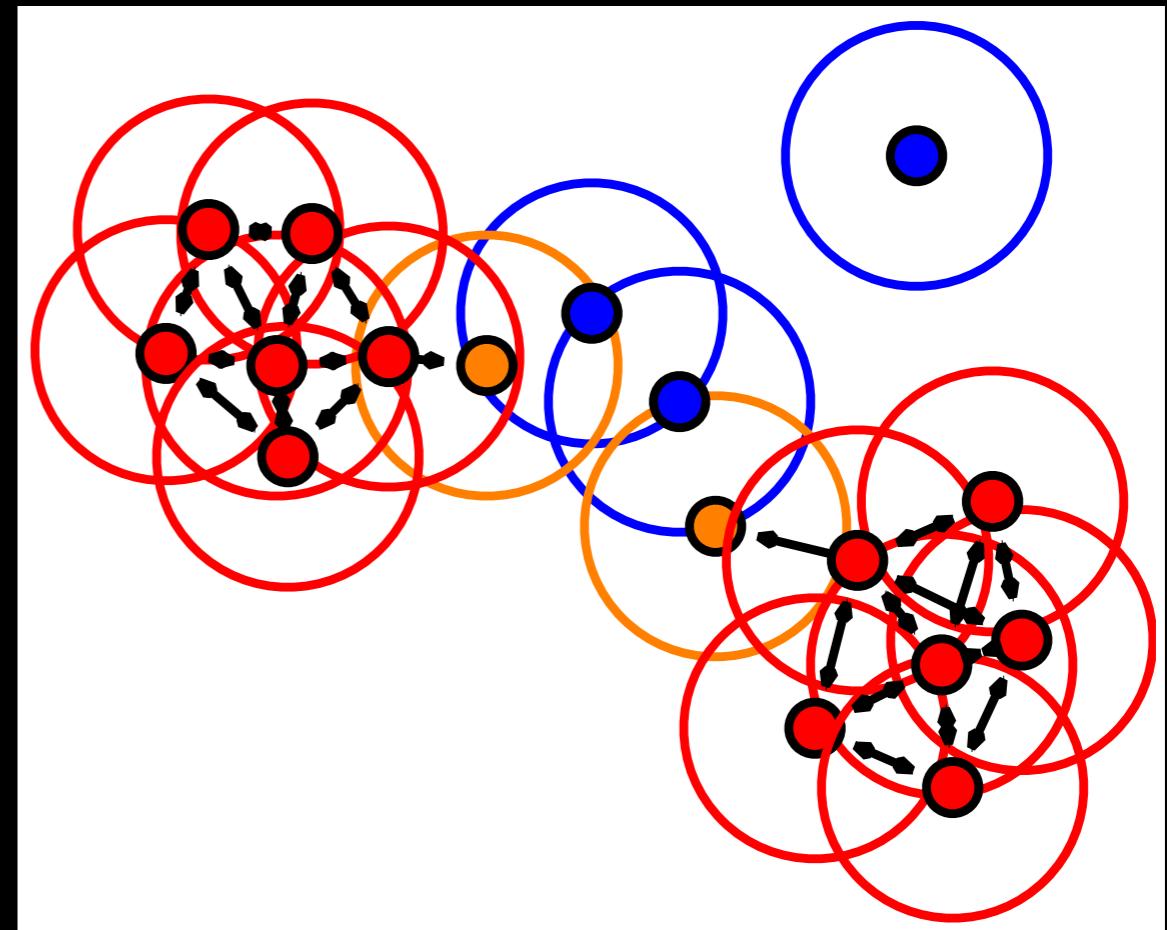
- **Time taken** is apparently most important
- **Time posted + geo** seem to hold the same information
- **Tags** beat titles and descriptions

CLUSTERING – FINDING EVENTS

- Clustering specific challenges
 - Number of clusters hard to estimate
 - But, it's a parameter in many clustering algorithms
 - Many noise points
 - 2% clusters with 1 member
 - Clustering a stream
 - The size of the streams
 - Updating clusters

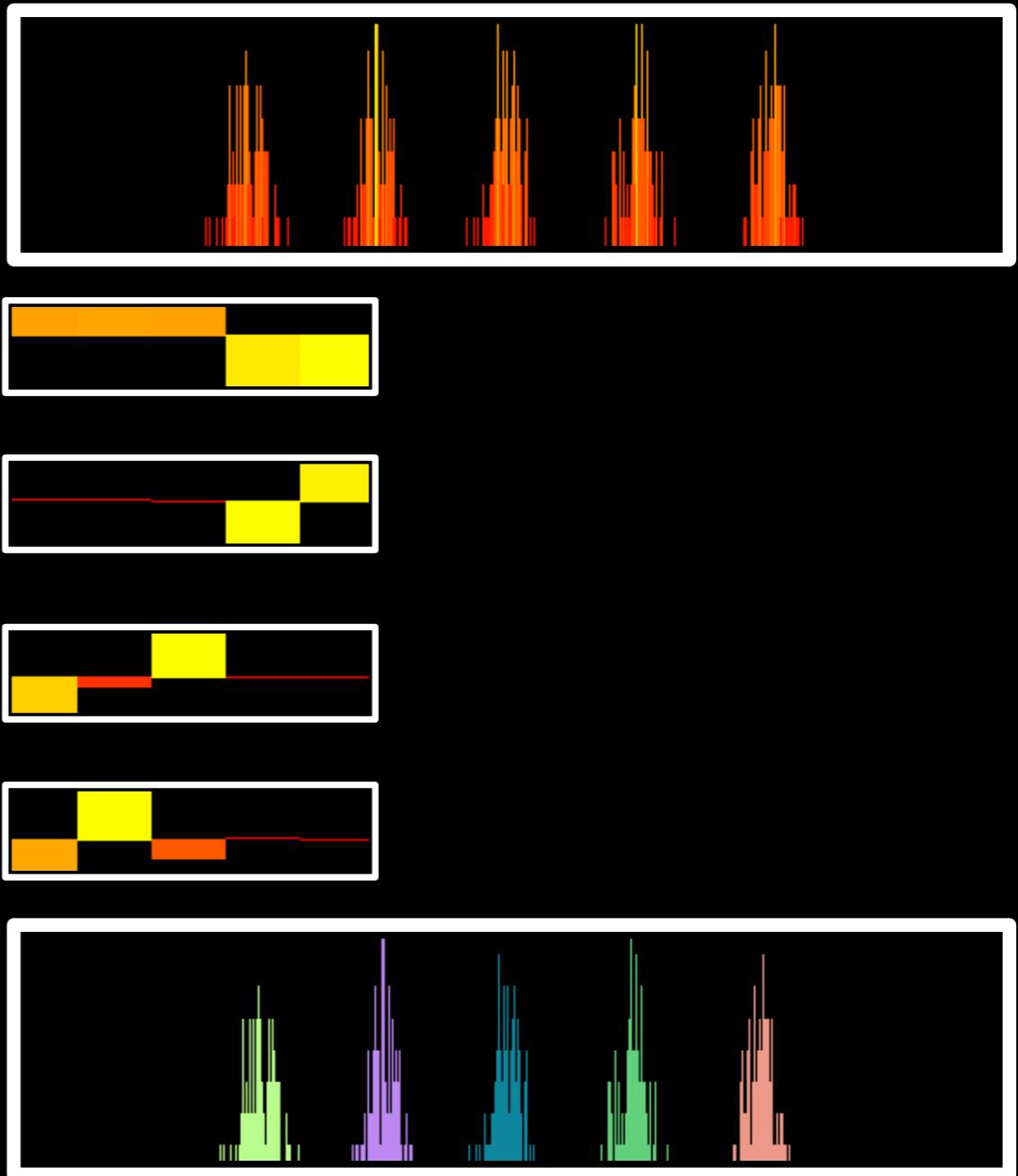
CLUSTERING – DBSCAN

- DBSCAN is an old, well studied clustering algorithm
- Detects clusters and identify noise
- No knowledge of cluster count needed
- Requirements:
 - Neighbourhood function (e.g. thresholded sparse similarity matrix)
 - Neighbourhood density counts
- Find mutually *density-connected* items



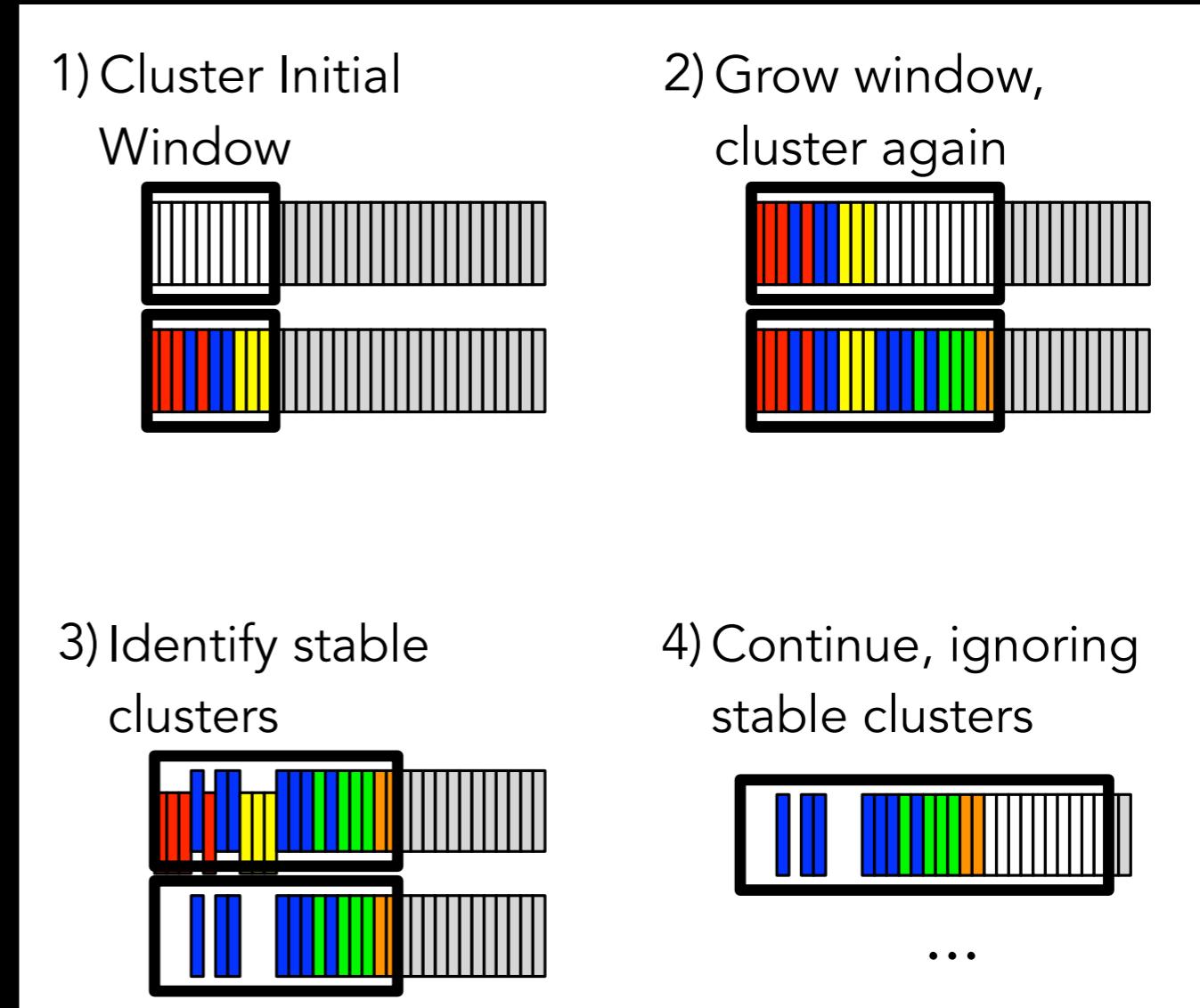
CLUSTERING - SPECTRAL

- Theoretically appealing non-parametric clustering algorithm
 - Rooted in graph theory
 - Potentially auto detects cluster count
- Basic premise is:
 - *Use the smallest (**near zero**) eigenvalues eigenvectors of the **graph laplacian** of the **similarity matrix** of some data as a space within which to apply **another clustering algorithm***



CLUSTERING - INCREMENTAL

- Practical restrictions of spectral clustering mean we can't apply it to the whole dataset
- Make an assumption about the data style
- Images likely to be clustered together will appear sequentially in terms of upload time
- Leverage this to cluster sub windows of data



MEDIAEVAL SED TASK RESULTS

- Calibration on 300k training items to optimise:
 - Feature weightings
 - Clustering parameters
- Clustering performed on a 200k item test set
 - In the 2013 SED task, this technique:
 - produced the best F1 scores;
 - was one of the only streaming approaches;
 - was one of the few open-source approaches

POWERED BY



RESULTS

Group/Technique	F₁	NMI	<i>divF₁</i>
CERTH-ITI(1) [13]	0.8865	0.8739	0.5701
CERTH-ITI(2) [14]	0.7031	0.9131	0.6367
UPC [15]	0.8798	0.9720	0.8268
UNITN [16]	0.9320	0.9849	0.8793
TUWIEN [17]	-	0.94	0.78
ADMRG [18]	0.812	0.954	0.758
ISMLL [19]	0.8755	0.9641	-
NTUA [20]	0.2364	0.6644	-
VIT [21]	0.1426	0.1802	0.0724
QM [22]	-	0.94	0.78
DBSCAN (best-weight)	0.9454	0.9851	0.8865
Spectral (best-weight)	0.9114	0.9765	0.8534
DBSCAN (average-weight)	0.9461	0.9852	0.8864
Spectral (average-weight)	0.9024	0.9737	0.8455

OPEN QUESTIONS AND FUTURE DIRECTIONS

IS EVENT DETECTION SOLVED?

- Scores in the 98% range certainly indicate so!
- But that's a bit of a fallacy - the data was biased
 - It only contained images of social events
 - We need to do more exploration of real data streams which also contain noise in the form of images that don't belong to a social event

WHY DON'T IMAGE FEATURES WORK WELL IN THE SED TASK?

- The features we tried using were too strict.
 - Designed for near duplicate detection
 - but, social events are highly visually diverse
 - highly visually similar images already being clustered through the other features
 - Need to look at better engineered or more specific features
 - Biometrics, facial matching, text, colour...

WHERE NEXT WITH THE MULTIMEDIA TWITTER ANALYSIS?

- Involve multimodal features
 - Tweets are rather rich in data that can be combined with the image analysis
- Applications:
 - Meme-tracking
 - Detection of world events
 - Exploiting multimodal feature co-occurrence

ANY QUESTIONS?