In Video Veritas – Verification of Social Media Video Content for the News Industry

Video Retrieval for Multimedia Verification of Breaking News on Social Networks

Lyndon Nixon, MODUL Technology GmbH

www.invid-project.eu

Our role in the InVID project

- Our goals are to:
  - identify new news events occurring in social media
  - rank best candidate social media video for the news event
  - analyse selected video in preparation for verification
- Why?
  - Social media like Twitter is quicker than the traditional news outlets in breaking news stories
  - The scale of social media is huge so we need to focus our social network queries on retrieving content about the current news
  - Journalists need quick and easy access to what are the current news stories discussed on social media and what video is being posted around those stories

Structure of this presentation

- 1. How we detect stories from the social media stream

- 2. How we use story detection for information retrieval from social networks

- 3. How guided video retrieval supports the news verification process
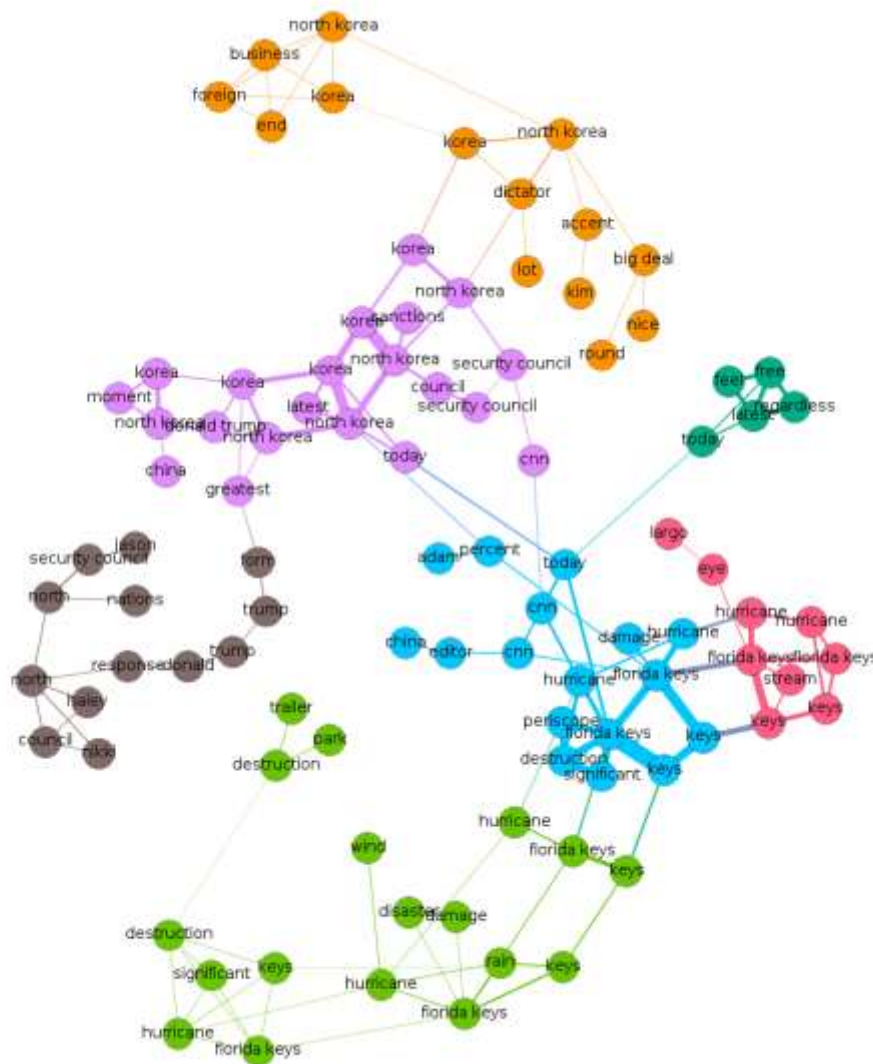
The Louvain community algorithm partitions the keyword graph into sub-graphs with optimized modularity.

There are limitations in smaller datasets which lead to imperfect partitioning.

Work on improving the disambiguation of stories:
- *Merge* based on overlap of keywords in story clusters with weighting
- *Split* through keyword co-occurrence matrices
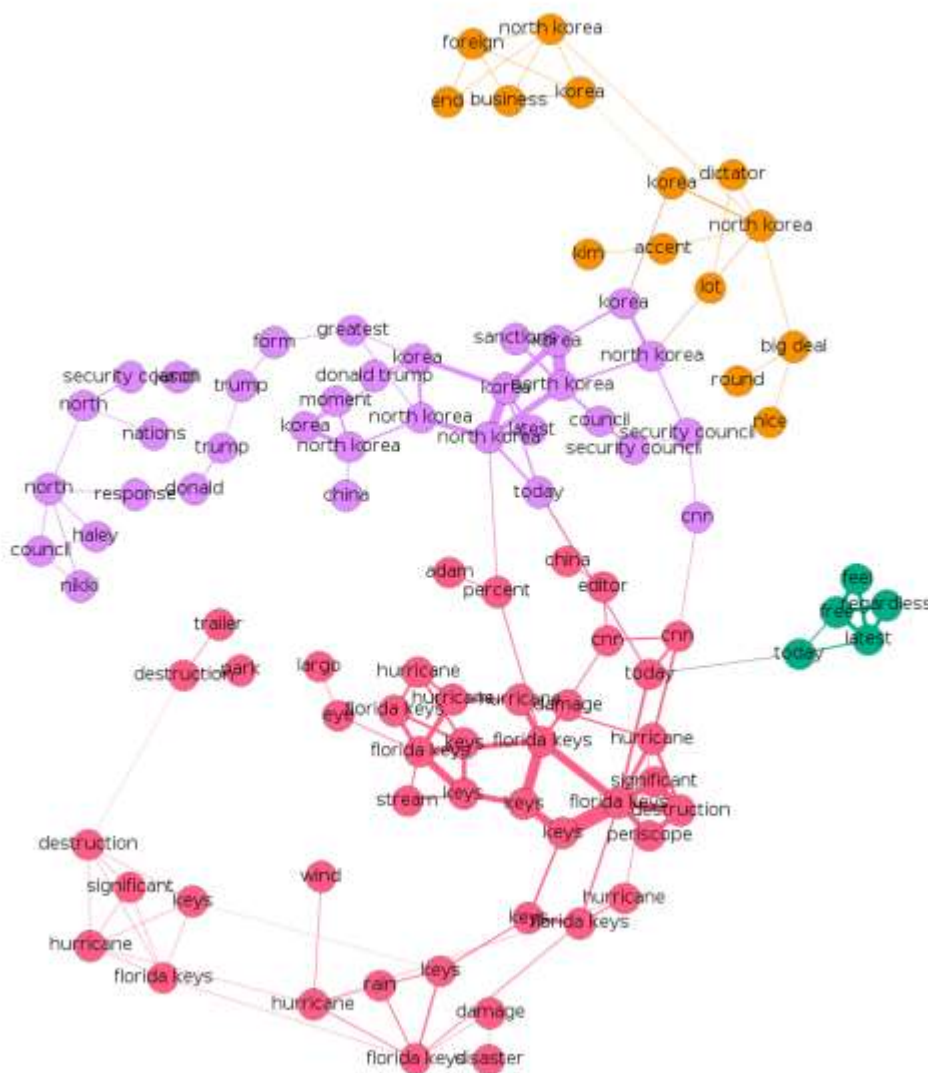
Same stories are partitioned into different clusters whenever the connections between the nodes are weak or there are no connections at all when those components are located in different time range (e.g. New report about a same event across several days). Those components need to be merged.

Cluster labels are sufficient for a first merge. We consider the keywords as a bag of words and merge iff 2 cluster labels overlap by at least half.
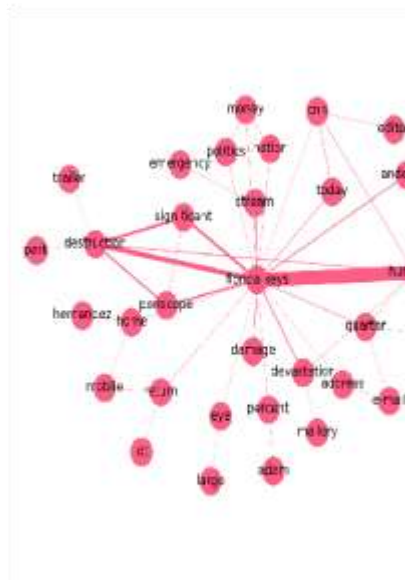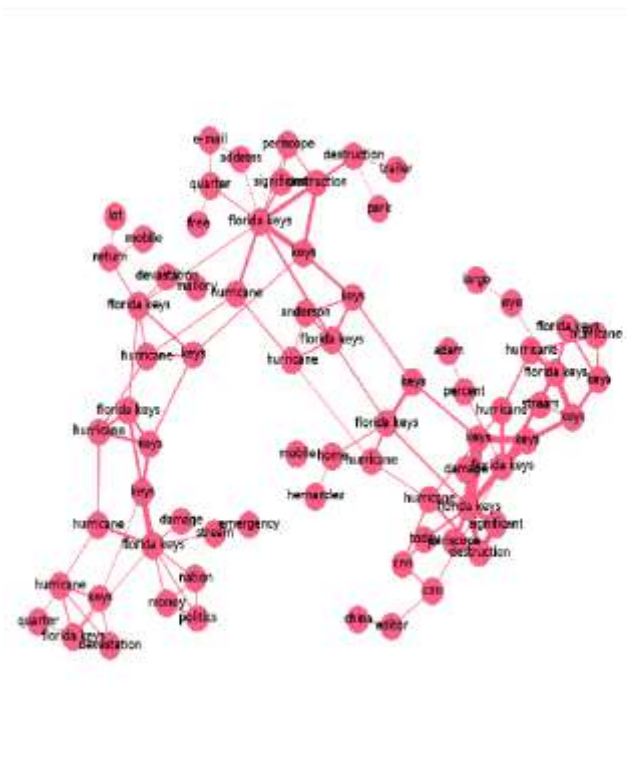
Within partitions, there may be keywords which associate together to represent different stories, requiring an approach to split partitions.

First, we remove redundant keywords, i.e. repetitions or keywords which are sub-sets of other keywords (e.g. "keys" together with "florida keys")

1. News story detection

Story: attack + isis + cruise missile + finsbury park + raqqa + coalition

Keyword co-occurrence matrix for the current time slice

	attack	isis	cruise missile	finsbury park	raqqa	coalition
attack		2	2	4		
isis	2		2		4	4
cruise missile	2	2				
finsbury park	4					
raqqa		4				5
coalition		4			5	

First experiments with keywords of the clusters where stories are merged indicated that there is a weaker co-occurrence between the keywords from different stories than between the keywords from the same story.

In the matrix, "finsbury park" is only associated with "attack" while "raqqa" and "coalition" are only associated with "isis" (apart from one another) which is also associated with "attack" and "cruise missile".

Two stories are in this cluster: ATTACK + FINSBURY PARK, ATTACK + ISIS + CRUISE MISSILE + RAQQA + COALITION
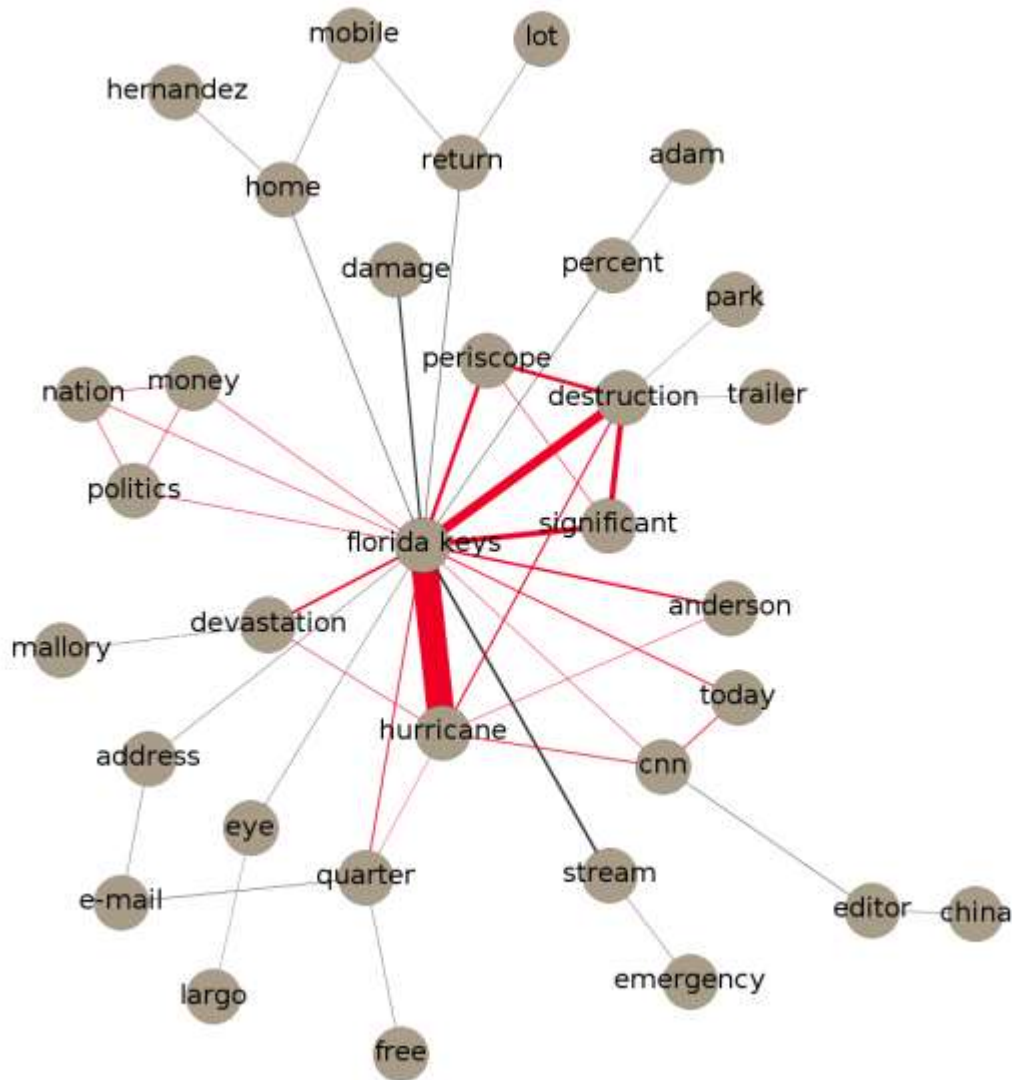
We formalized the idea of identifying keywords which co-occur significantly as **triads**. These are pair-wise associations between keywords. Keywords not part of a triad are **leaf nodes**. Triads are grouped when each triad has at least one common node with another triad in the group.

1. News story detection
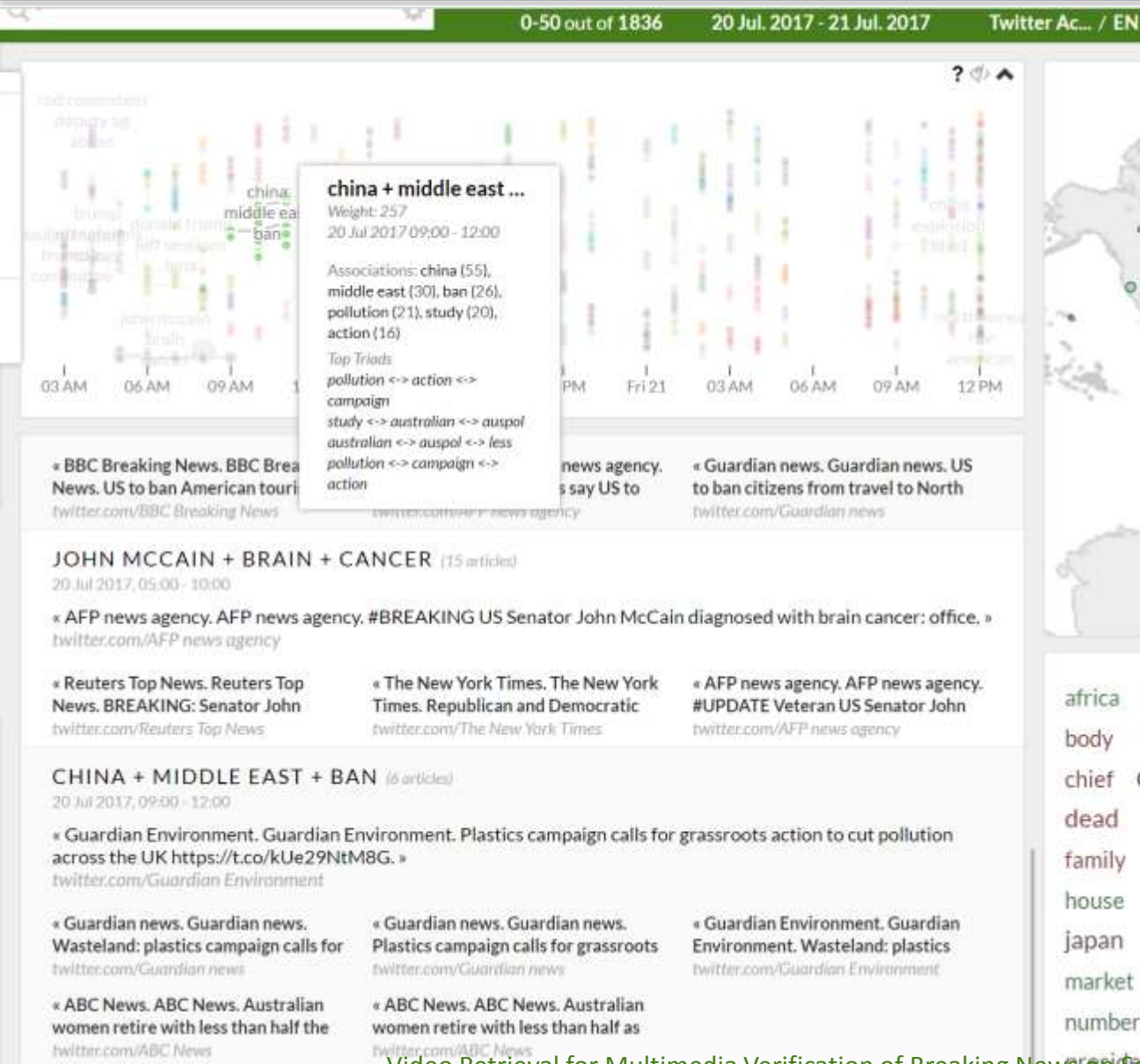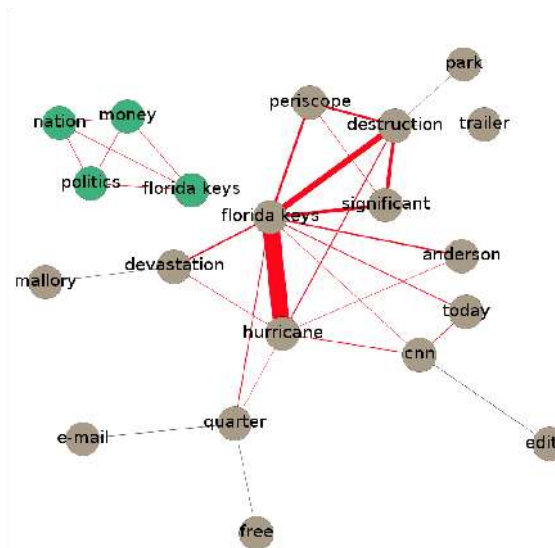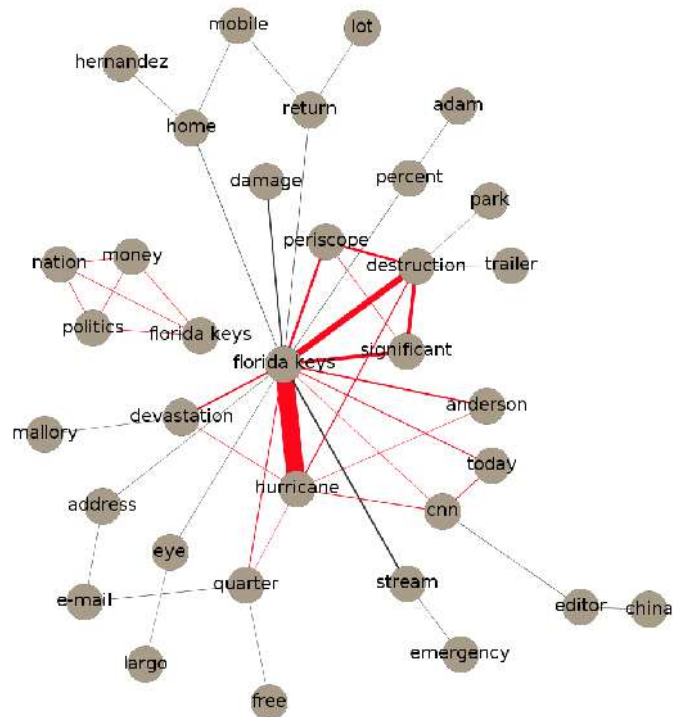
An example of a story with the top triads shown.

Note how here two triad groups have formed:

POLLUTION + ACTION + CAMPAIGN

STUDY + AUSTRALIAN + AUSPOL + LESS

These match well the two stories within this cluster, as reflected in the documents shown.

After the triad groups are detected, they are split, keeping the leaf nodes directly connected to the triads when they belong to a single group. Those leaf nodes which are connected to triads in multiple groups are removed, because it is hard to determine which group they might correctly belong to.

After splitting, a content-based merge is performed to correct splits within a single news story (if the keywords used for the story differ sufficiently within the documents, the triad approach will likely split those partitions).

For each cluster the top n keywords (n is configurable) with the highest frequency are chosen to represent the cluster. The remaining low frequency keywords, which may contain noise, are not taken into consideration. Overlapping rate is computed between each cluster pair. When one story overlaps significantly with another story (threshold of overlapping rate is configurable), these are merged.

Examples of merges:
"north korea,nuclear,president donald trump" is merged with "north korea,donald trump,big deal"
"security council,north,nikki,haley" is merged with "north korea,security council,sanctions,today,latest,moment,greatest,china,donald trump"

Evaluating the story detection

Story evaluation (a framework for evaluating the accuracy of the topic detection algorithm)

- ~~**Ground truth evaluation.** Comparison of stories occurring in dashboard with stories on aggregation sites, e.g. Current News Portal of Wikipedia (similar to SNOW2014 Challenge)~~

- **Story correctness.** Boolean test for whether a cluster unambiguously identifies a newsworthy story

- **Story distinctiveness.** Boolean test for whether a cluster uniquely refers to a distinct news story

- **Story completeness.** Measure of the extent to which documents for a single news story are found within a single cluster.

- **Story homogeneity.** Measure of the extent to which each cluster contains only documents relevant to the same news story.

Evaluating the story detection

Story evaluation set-up: Twitter Accounts in EN during 19-23 June, daily stories from all tweets + 5 most popular topics

Evaluating the story detection

	Correctness	Distinctiveness	Completeness	Homogeneity
Aggregated 5-day	0.90	0.60	0.93	0.93

Correctness could have been 100% if not for one Twitter account which we have now removed.

Also the figures were similar across the individual days and the individual topics.

We also tested the Twitter News source (which is UGC for "breaking news". While correctness was unsurprisingly lower it was still good (0.76) and some stories were detected which did not appear in the professional sources, reflecting "public interest" stories like 'Two elephants work together to save a baby elephant from drowning'.

2. News video retrieval from social media
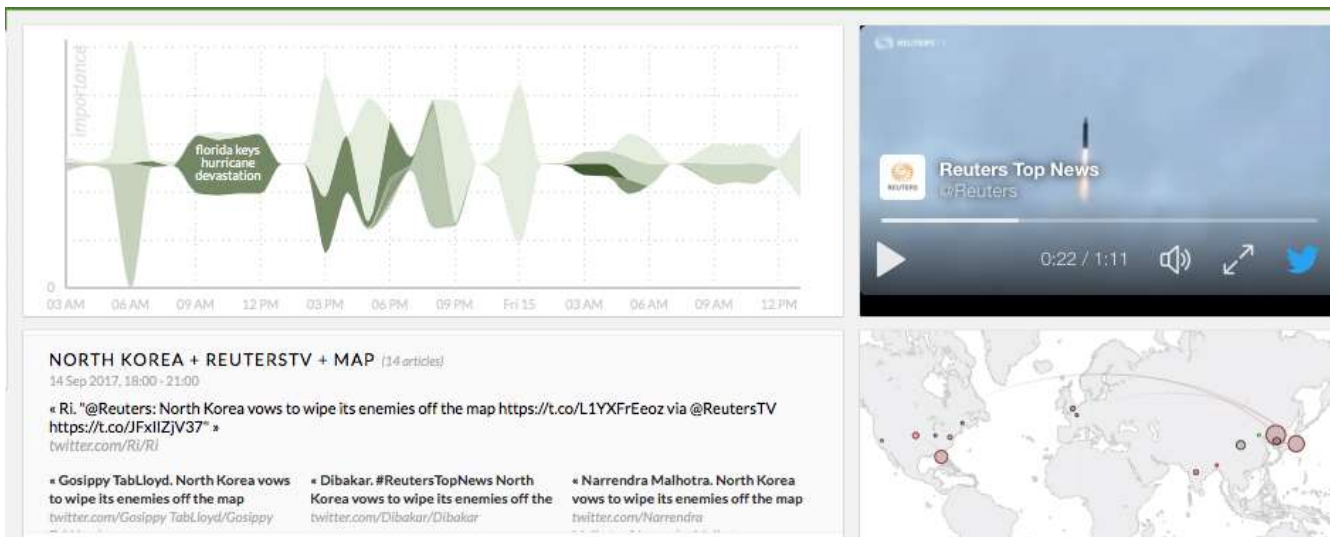
How to surface the appropriate video content for each story?

- Using the story detection, we generate a set of new queries every 6 hours for the social network APIs (Twitter Video, YouTube, Vimeo, DailyMotion)
- We create for the top max n results per query a document per response with the video link and extracted metadata
- We also cluster the collected video documents around stories

Evaluating the news video retrieval

Metric	value
avg precision	0.36
avg accuracy	0.84
avg recall	0.42
f-measure	0.425
story breadth	9
story depth	0.17

Table 2: Results 13Jun-current

Metric	value
avg precision	0.89
avg accuracy	0.97
avg recall	0.91
f-measure	0.92
story breadth	2
story depth	0.85

Table 4: Results 10May-current

Metric	value
avg precision	0.54
avg accuracy	0.82
avg recall	0.64
f-measure	0.59
story breadth	18
story depth	0.3

Table 3: Results 13Jun-proposed

Metric	value
avg precision	0.52
avg accuracy	0.69
avg recall	0.59
f-measure	0.57
story breadth	15
story depth	0.35

Table 5: Results 10May-proposed

Findings suggest querying by topic and using stories labels (instead of keyword associations) will significantly improve the Video document collection (breadth of stories)

3. How this supports news video verification

Public demo at https://invid.weblyzard.com

Project consortium and funding agency

InVID is funded by the EU Horizon 2020 Programme under the project number ICT-2015-687786

Thank you for your attention!

Find our dashboard for news story detection and video retrieval at

http://invid.weblyzard.com

Any questions?