Australian National University

Visual Memes in Social Media
Tracking real-world news on YouTube

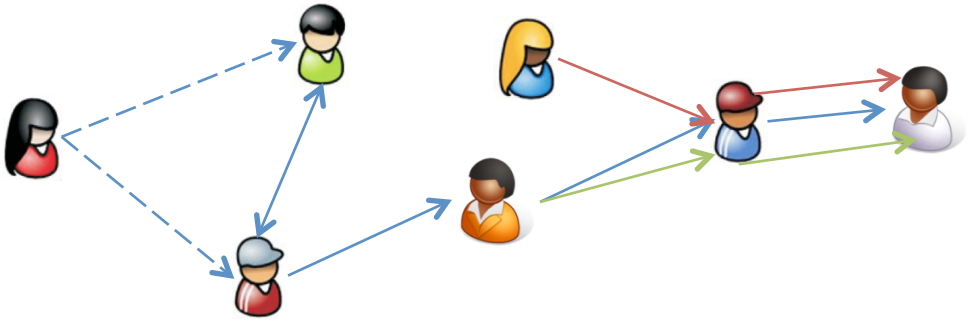
Lexing Xie, Apostol Natsev, John R Kender,
Matthew Hill, John R Smith

Australian National University, IBM Research,
Columbia University

@ACM MM Nov 29, 2011

The problem

- Information diffusion on YouTube
  - Can we track the video-equivalent of “RT”, “@” … ?



- Understanding real-world events
  - Can studying YouTube tell us what content are interesting, and who are important?

Outline

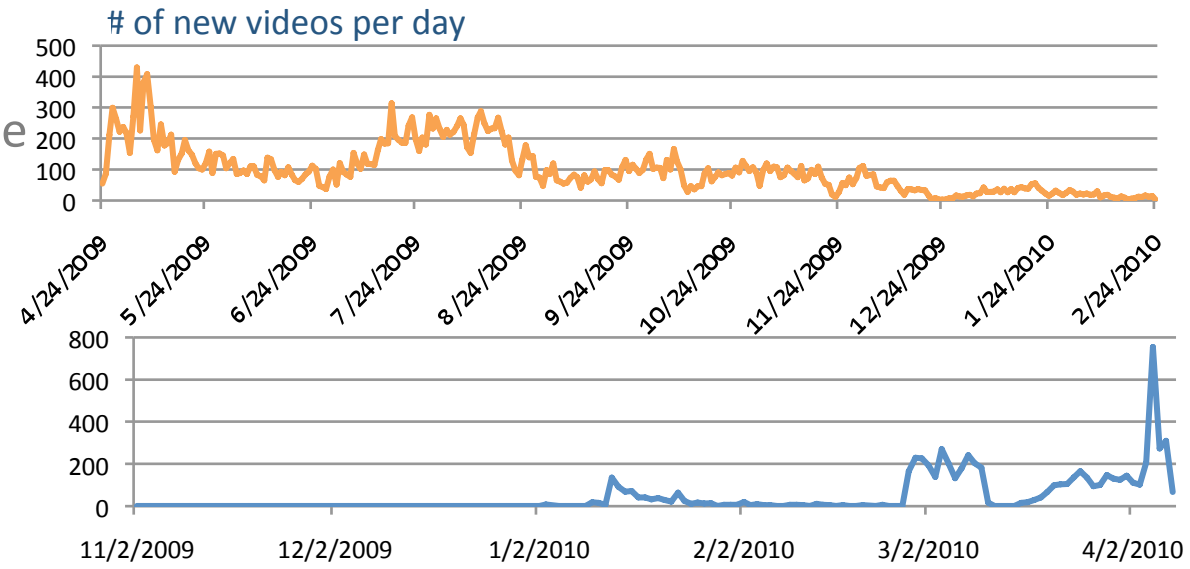
- Events on YouTube
- Quoting and remixing with videos

- Visual Meme detection
- Observations, influence and importance

Event buzz on YouTube

- YouTube: too much content?
  - 48 hrs uploaded every minute, >=10% of internet traffic
- How are real-world events reflected on YouTube?

Pop quiz:
which trace is about swine flu? earthquake ?

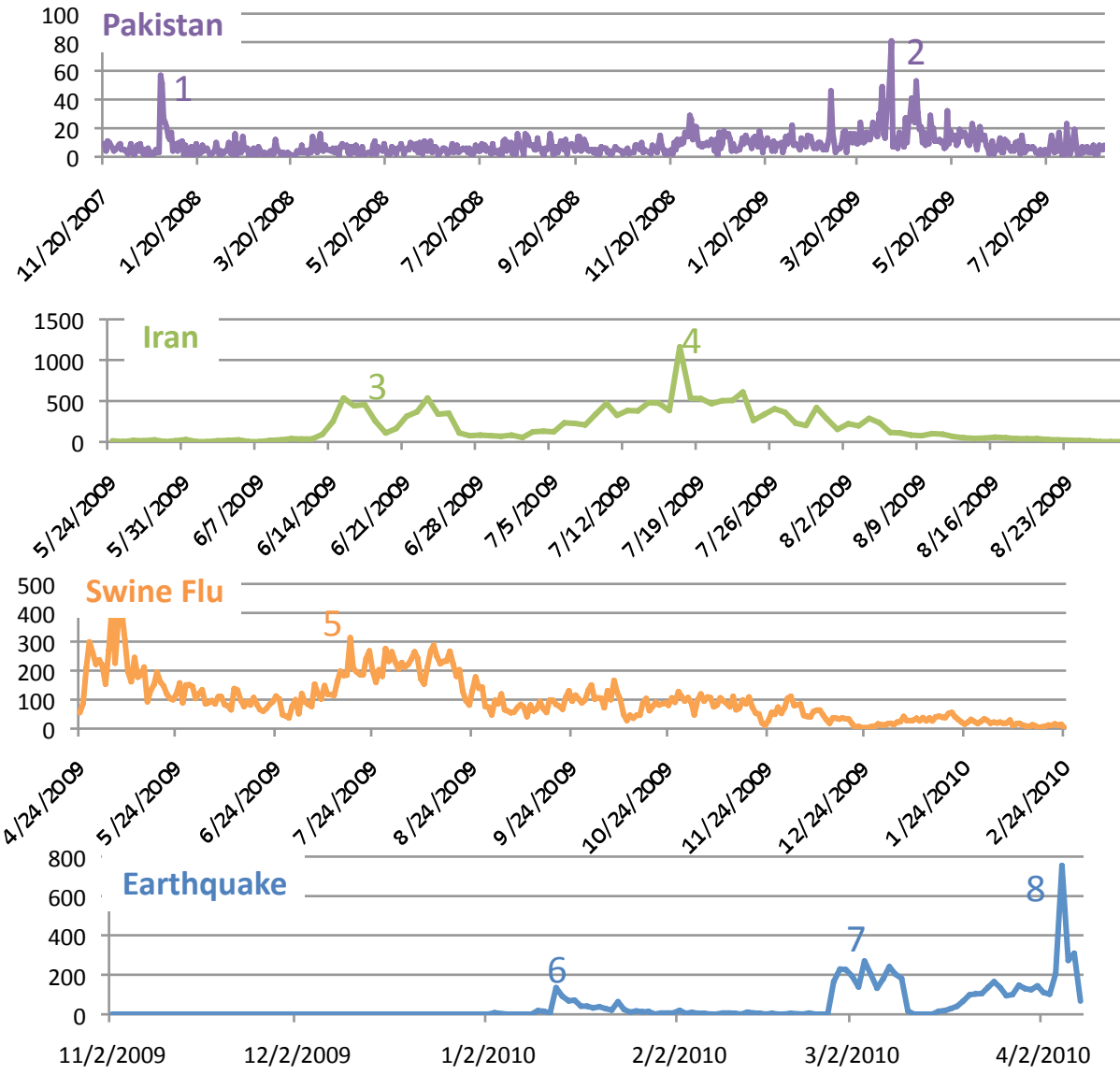


Australian National University

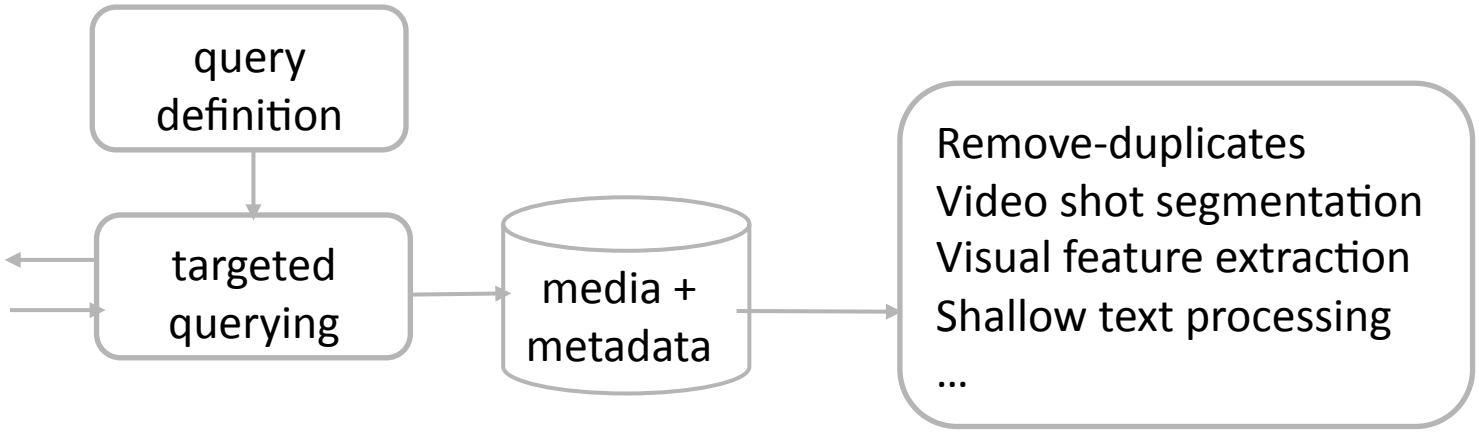
Event buzz on YouTube

- Event Buzz: noisy activity trace left by real-world events

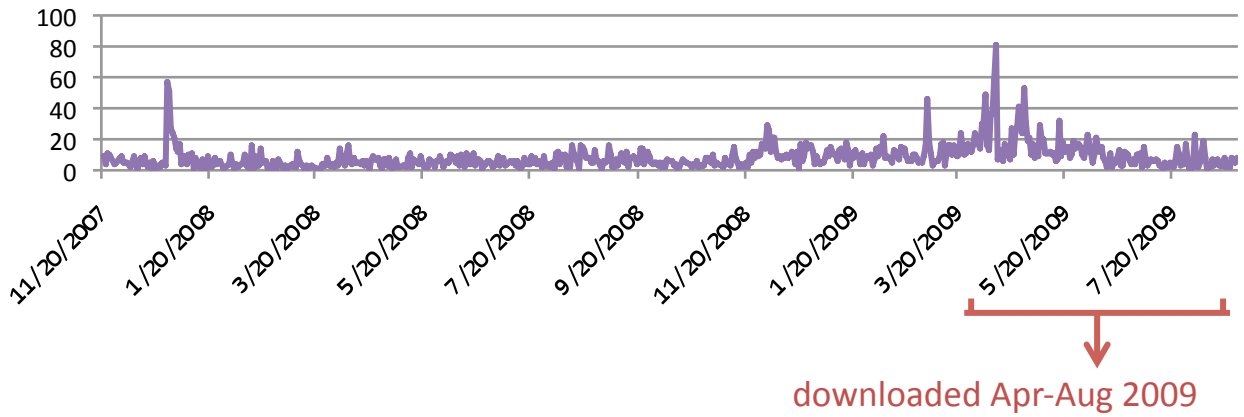
1. Bhutto assassination
2. Cricket bombing + protest
3. Post-election protests
4. Friday Prayer in Tehran
5. H1N1 Human trial begin
6. Haiti earthquake
7. Chili earthquake
8. California earthquake



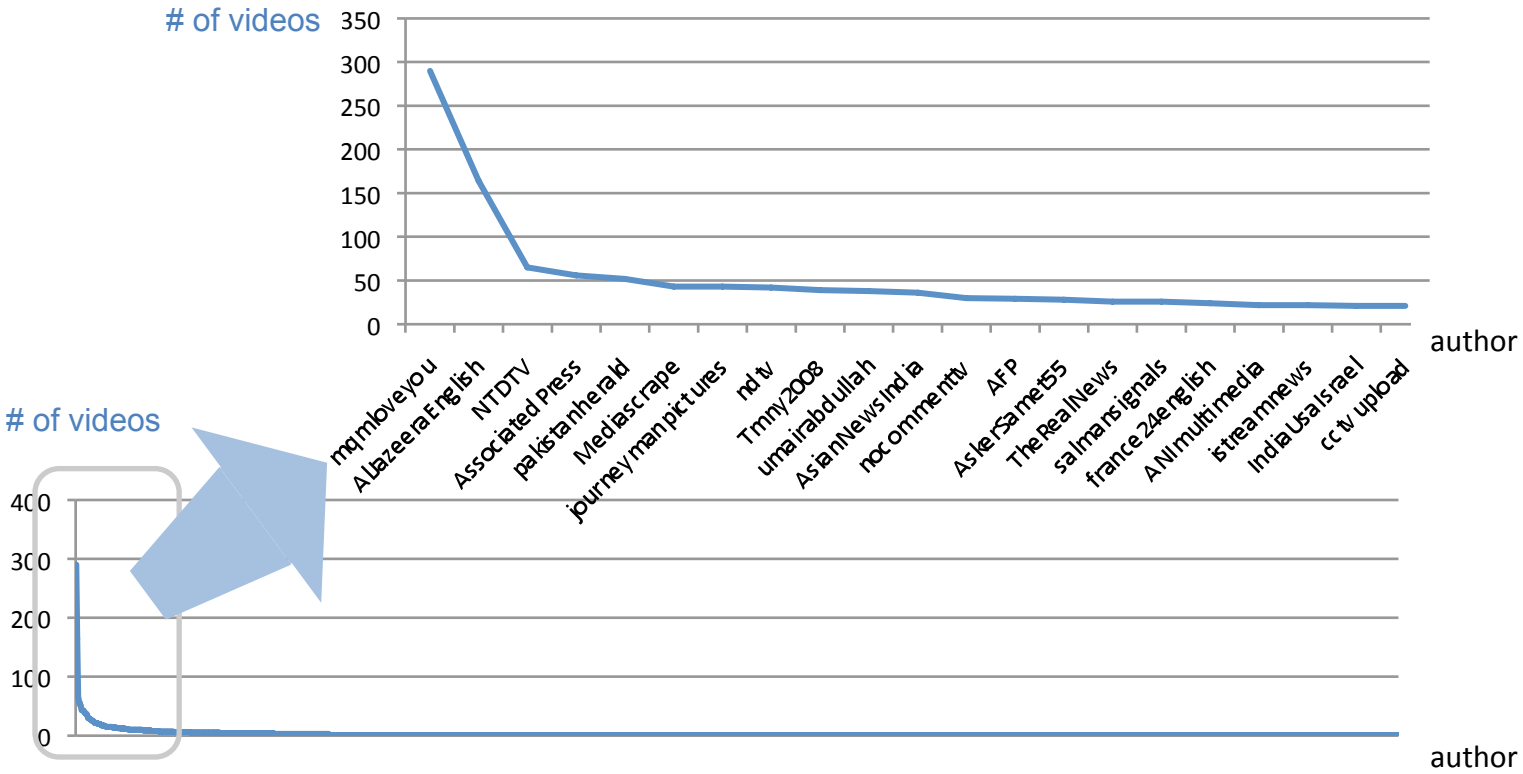
Monitoring event buzz on YouTube



pakistan military
pakistan india tension
pakistan taliban
pakistan bhutto
pakistan ppp
pakistan terrorist



Who is Reporting Events on YouTube?

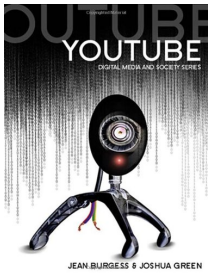


Content authored or remixed by:

- Professional news agencies: *AssociatedPress, AlJazeeraEnglish, BBCWorldwide, PakistanHerald, …*
- News enthusiasts: *mqmloveyou, askersamet55, tmny2008, …*
- General users (1000+)

What is YouTube?

[Burgess and Green 2009]
[Snickars and Vonderau 2010]



- Is it
  - A media outlet ?
  - A place to look at pointless monologues, babies and dogs?
- YouTube is
  - Is a “platform for participatory culture”
  - it is “ruled by the clip and the quote” – remixing is popular!

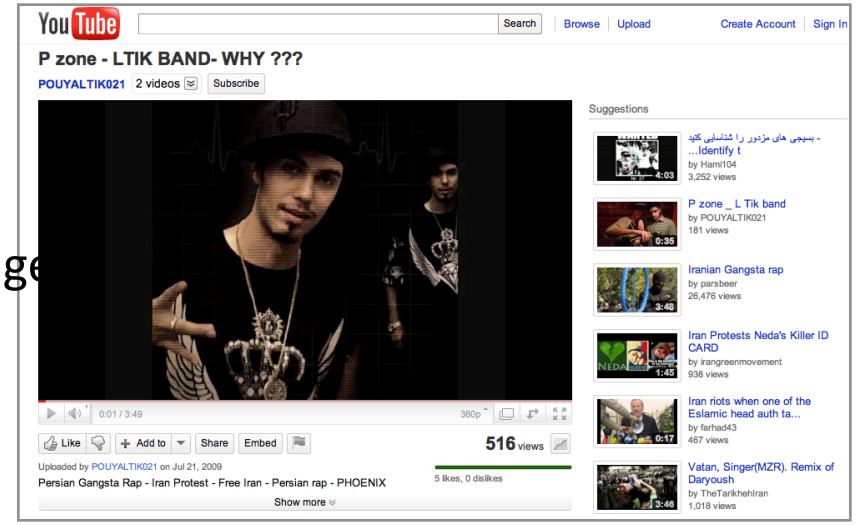
YouTube
Broadcast Yourself

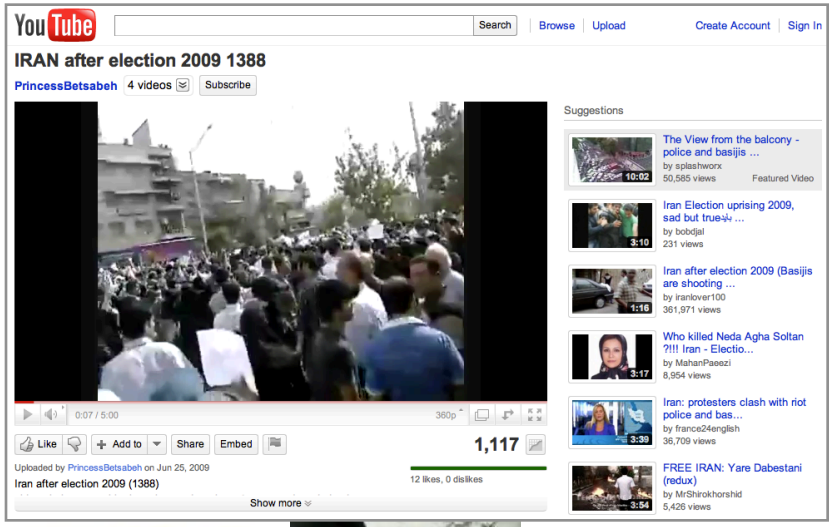
Remixing on Youtube

Video A

Video B

Youtube Video page

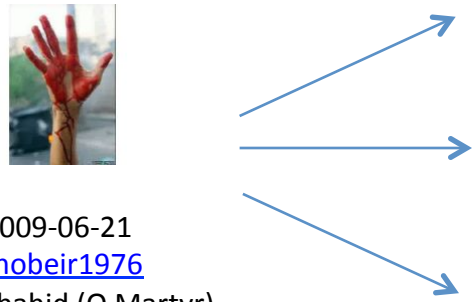




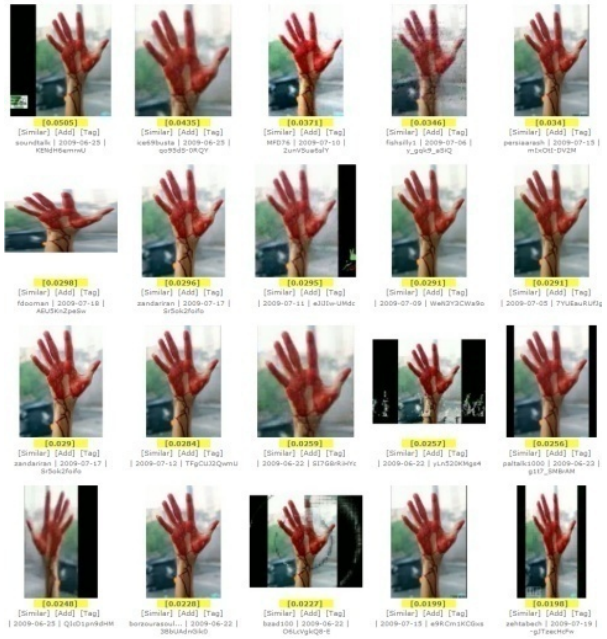
Meme shot examples

Visual memes

- Meme := a cultural unit (an idea or value or pattern of behavior) that is passed from one person to another by social means
- Visual meme := frequently re-posted visual units -- image or short video segments



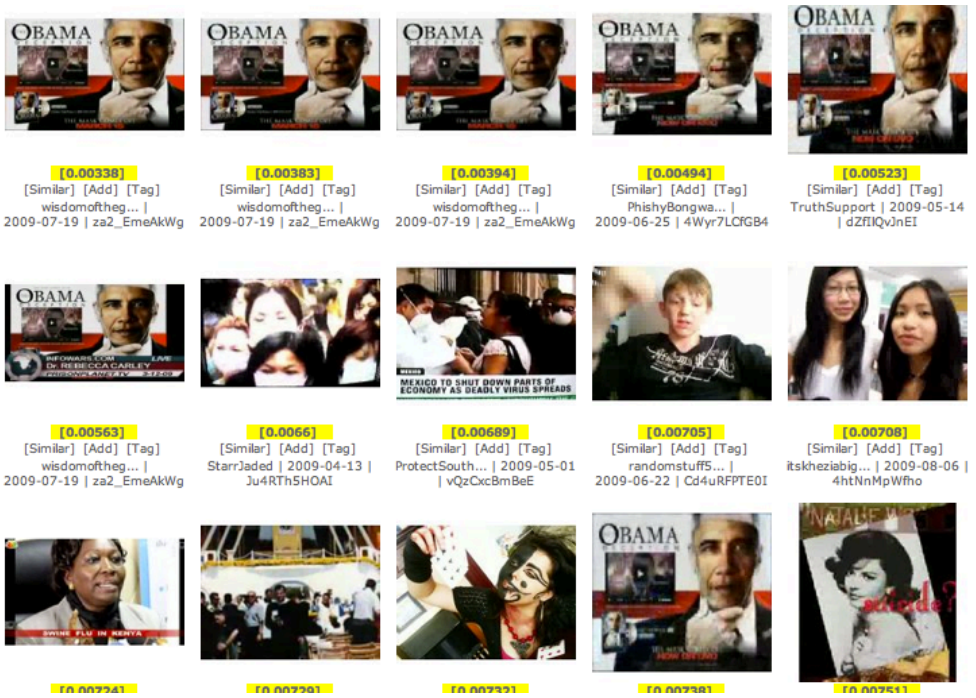
upload: 2009-06-21
author: shobeir1976
title: Ey Shahid (O Martyr)



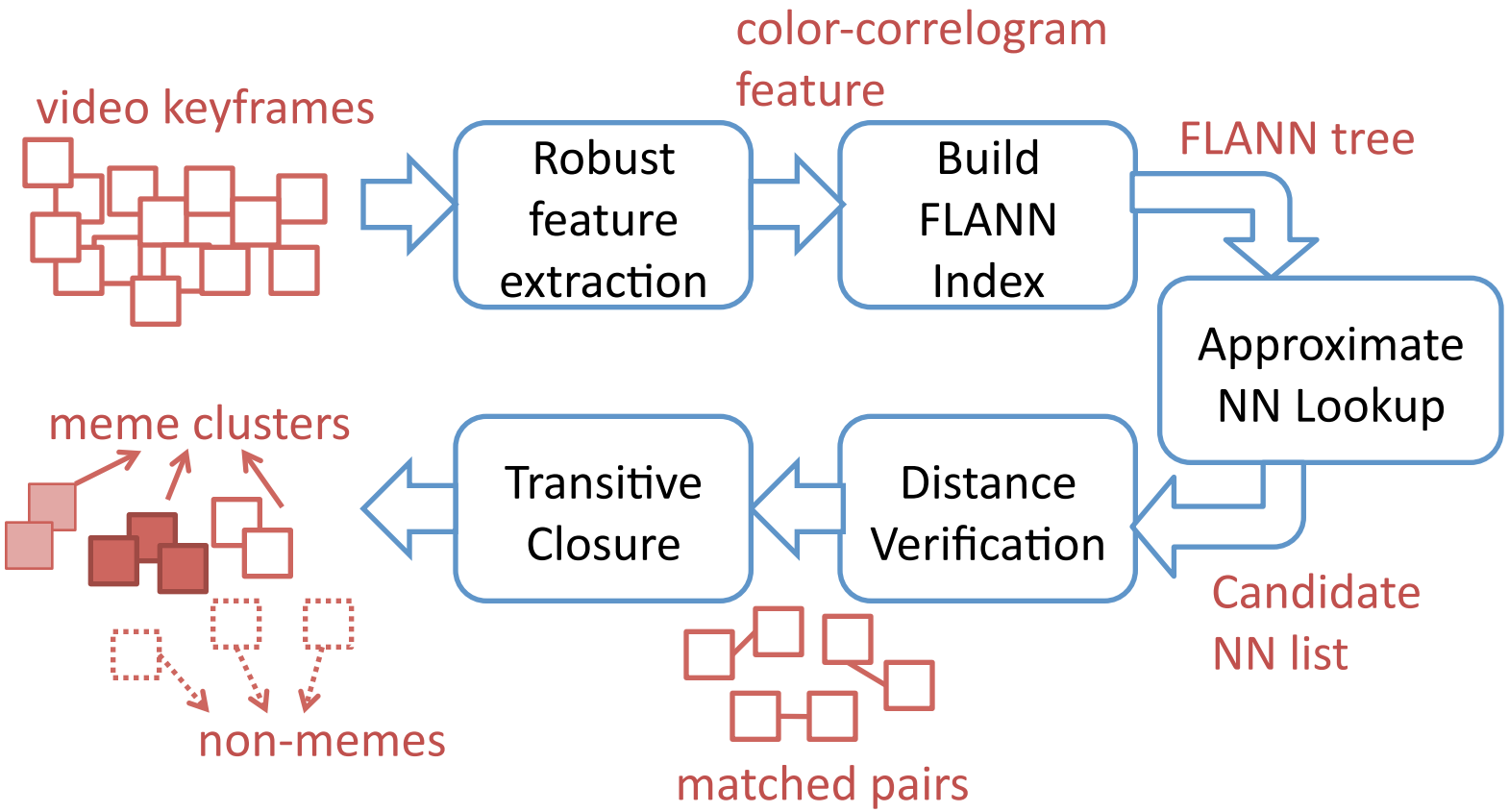
51 other videos, 2009-06 ~ 08

How to detect visual memes?

- Nearest neighbor search?
- Challenges:
  - Appearance variations: size, quality, color, gamma, overlay, borders …
  - Do this for 1 Million images, potentially O (N^2)



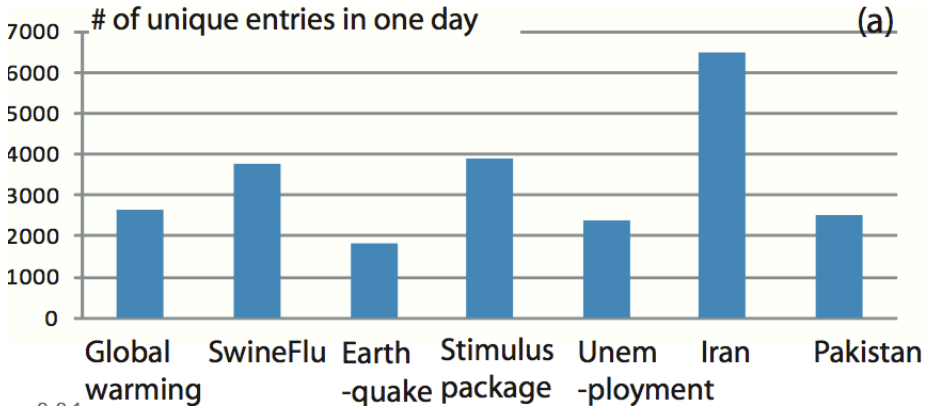
Visual meme detection

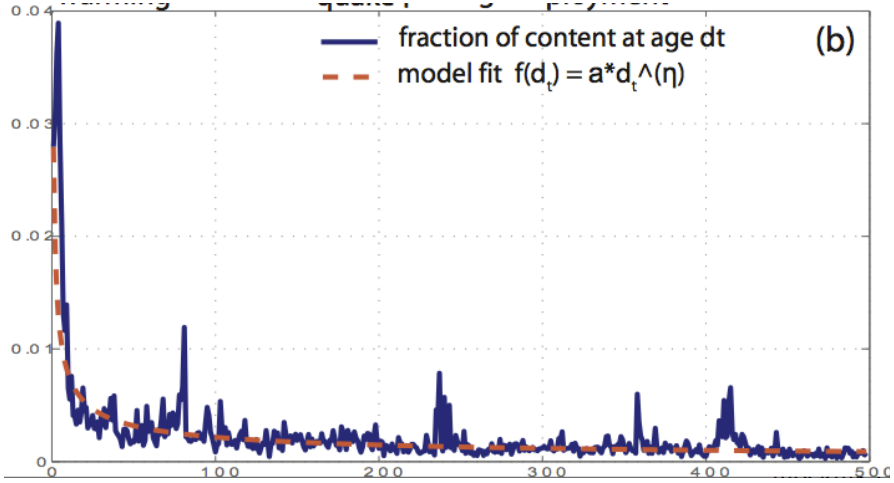


Pilot datasets

- Continuously collected from YouTube over 3-9 months time.

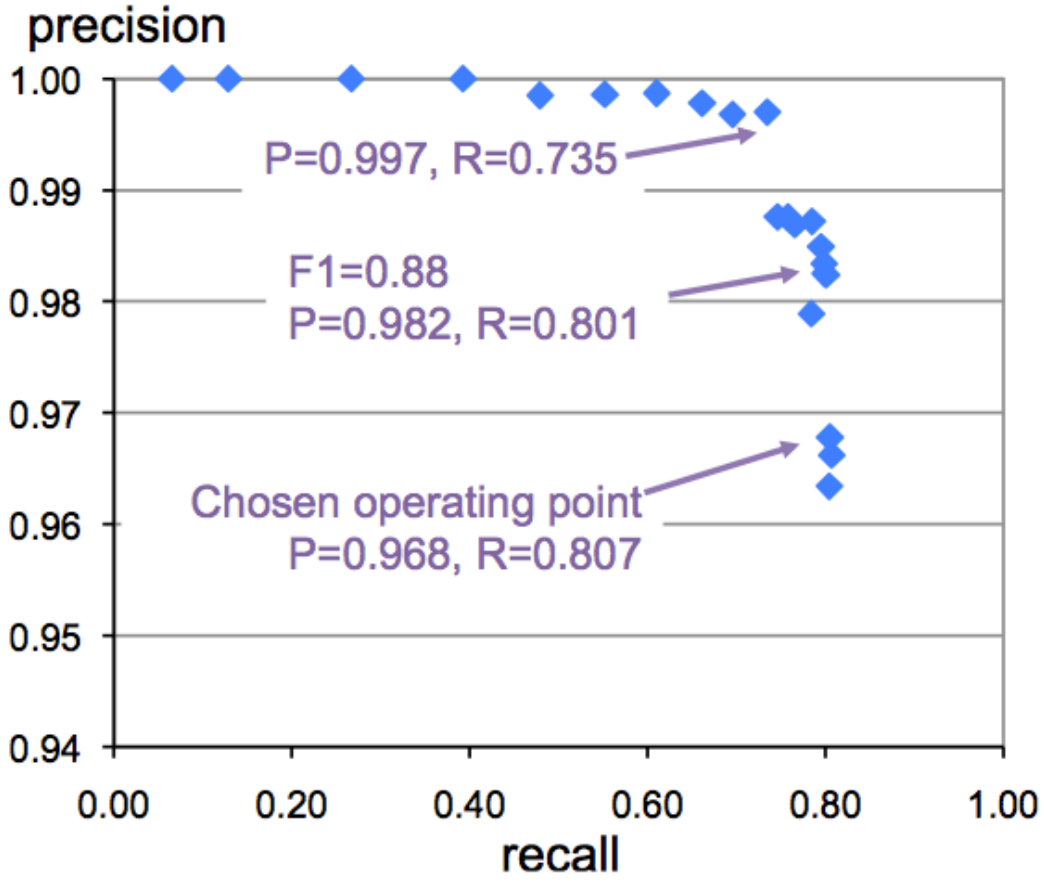
ID	Topic	Videos	Authors	Shots	Upload time
A	Swine flu	31,488	10,804	1,202,479	04/09∼03/10
B	Iran	23,049	4,681	1,255,062	08/07∼08/09
B1	Iran 06/09	5,429	2,393	210,259	09/07∼07/09
C	Housing	2,446	654	71,872	08/07∼08/09



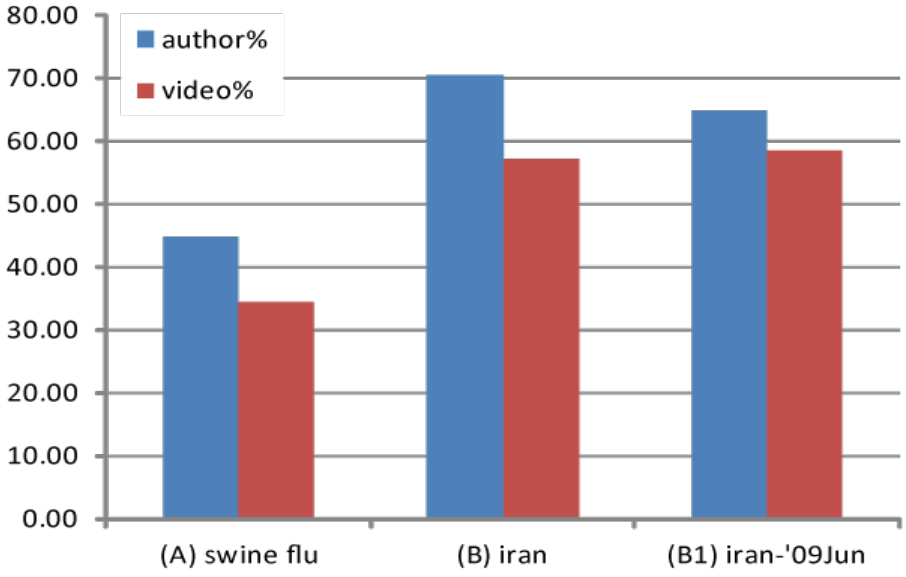


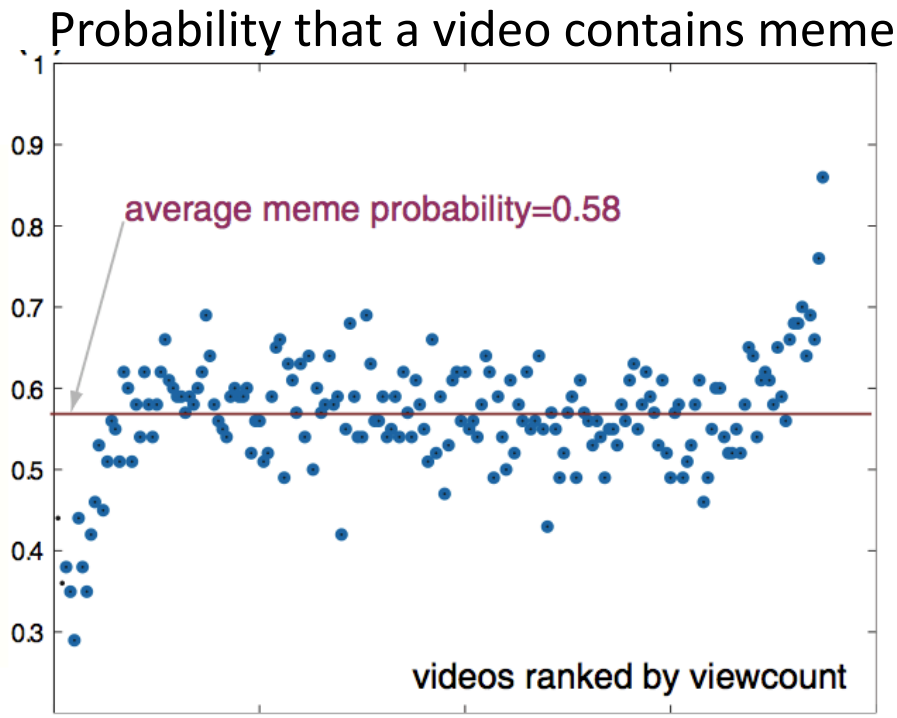
Visual meme detection

- 1.2 M^2 comparisons in ~5-7 hours
- Precision >96%
  - "housing" data set
  - ~15K positive pairs
  - ~25K negative pairs



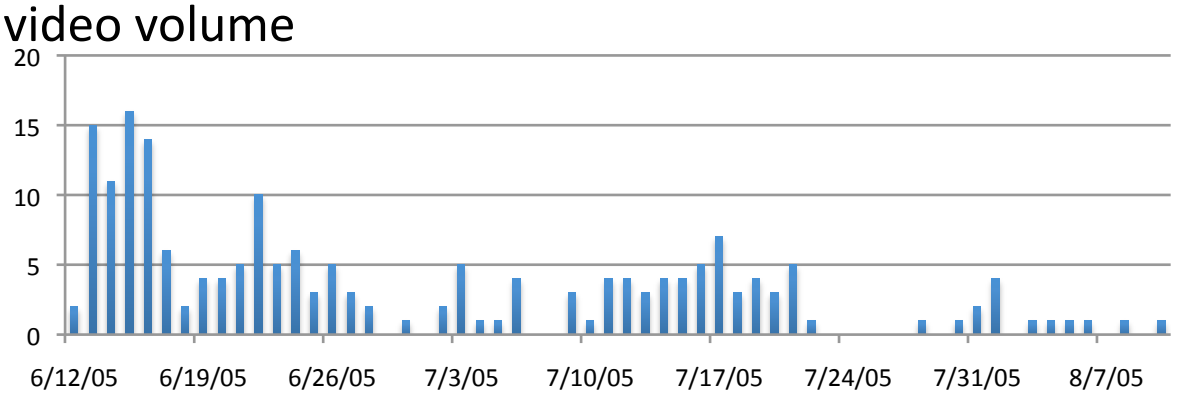
How prevalent are visual memes?





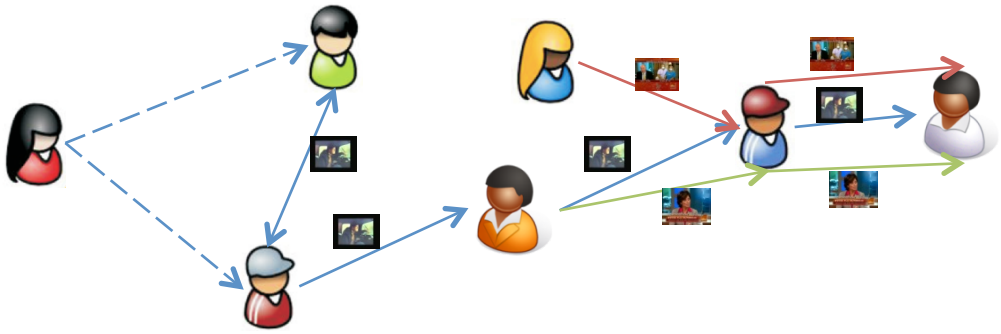
- >50% video contain memes, ~70% authors participate in producing and disseminating memes.
- Video popularity (viewcount) can be inversely correlated with being meme-videos !

Timeline of a visual meme cluster



Meme graphs

- Memes as links



Video graph $G = \{\mathcal{D}, \mathcal{E}_G\}$

$$\omega^*_{ij} \propto \nu_{ij}$$

$$\omega'_{ij} \propto \nu_{ij} \triangle t^{\eta}_{ji}$$

Author graph $H = \{\mathcal{A}, \mathcal{E}_H\}$

$$\tau_{rs} = \Sigma_{\{i,a(d_i)=a_r\}} \Sigma_{\{j,a(d_j)=a_s\}} \omega_{ij}$$

$$r, s \in \mathcal{A}, \;\; i, j \in \mathcal{D}$$

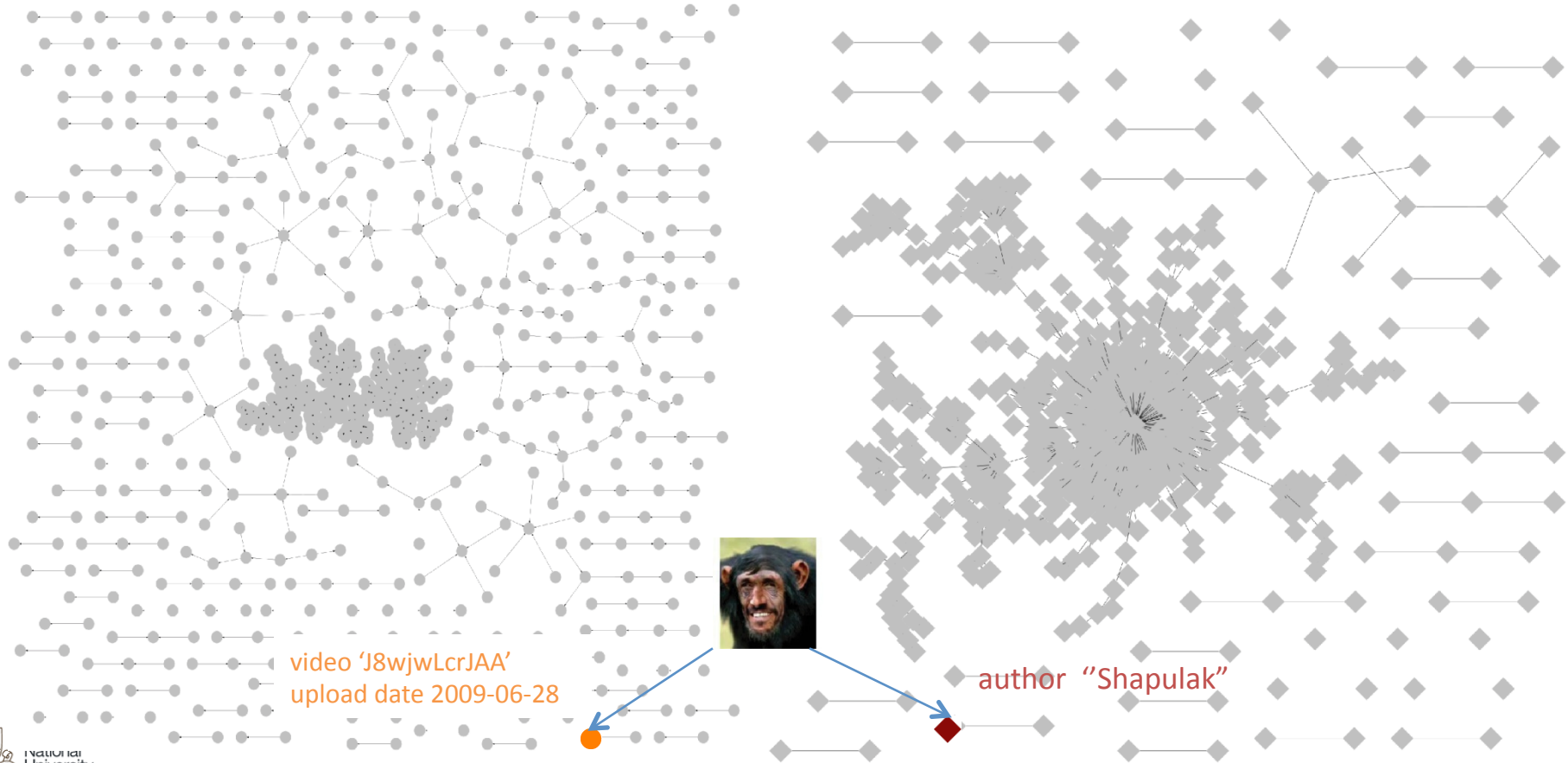
Meme graphs

Video graph $G = \{\mathcal{D}, \mathcal{E}_G\}$

$$\omega^*_{ij} \propto \nu_{ij}$$
$$\omega'_{ij} \propto \nu_{ij} \triangle t^{\eta}_{ji}$$

Author graph $H = \{\mathcal{A}, \mathcal{E}_H\}$

$$\tau_{rs} = \Sigma_{\{i,a(d_i)=a_r\}} \Sigma_{\{j,a(d_j)=a_s\}} \omega_{ij}$$
$$r, s \in \mathcal{A}, \; i, j \in \mathcal{D}$$



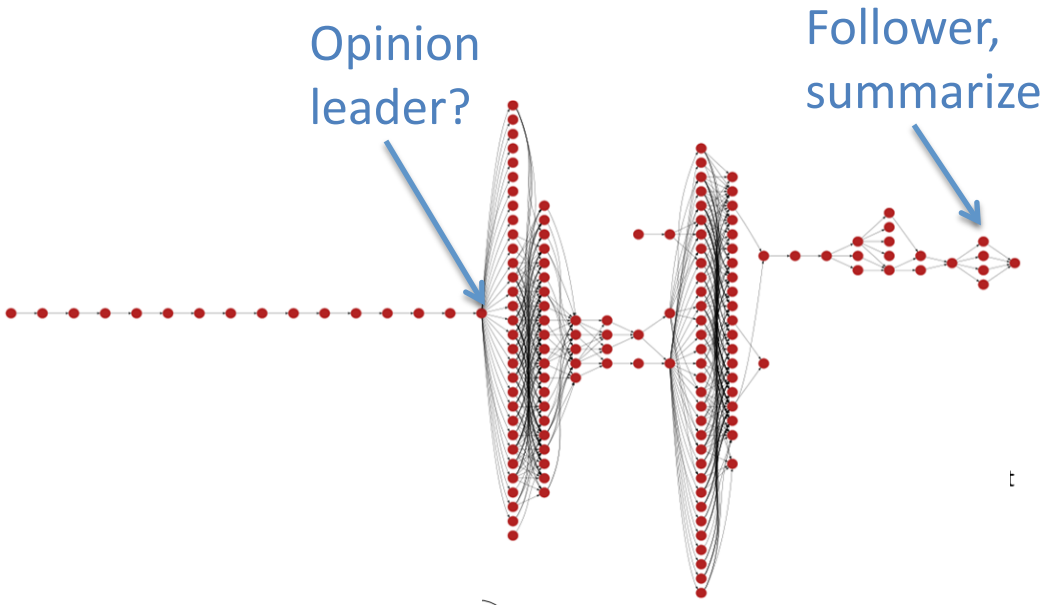
National University

Diffusion influence index

- Defined for video *i* and author *r*

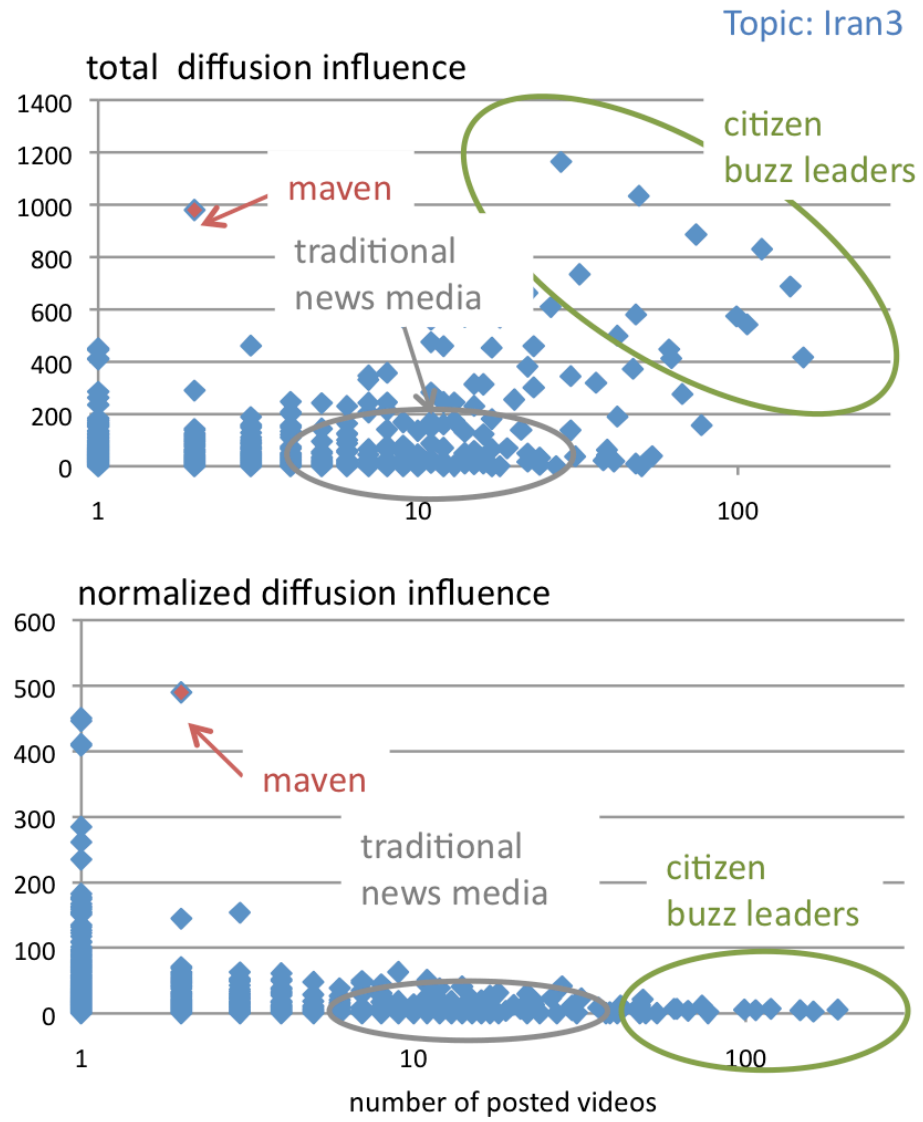
$$\chi_i = \Sigma_{w_v} \frac{\zeta^{in}_{i,w_v}}{1+\zeta^{out}_{i,w_v}} \qquad \chi r = \Sigma_{\{i,a(d_i)=a_r\}}\ \chi i$$

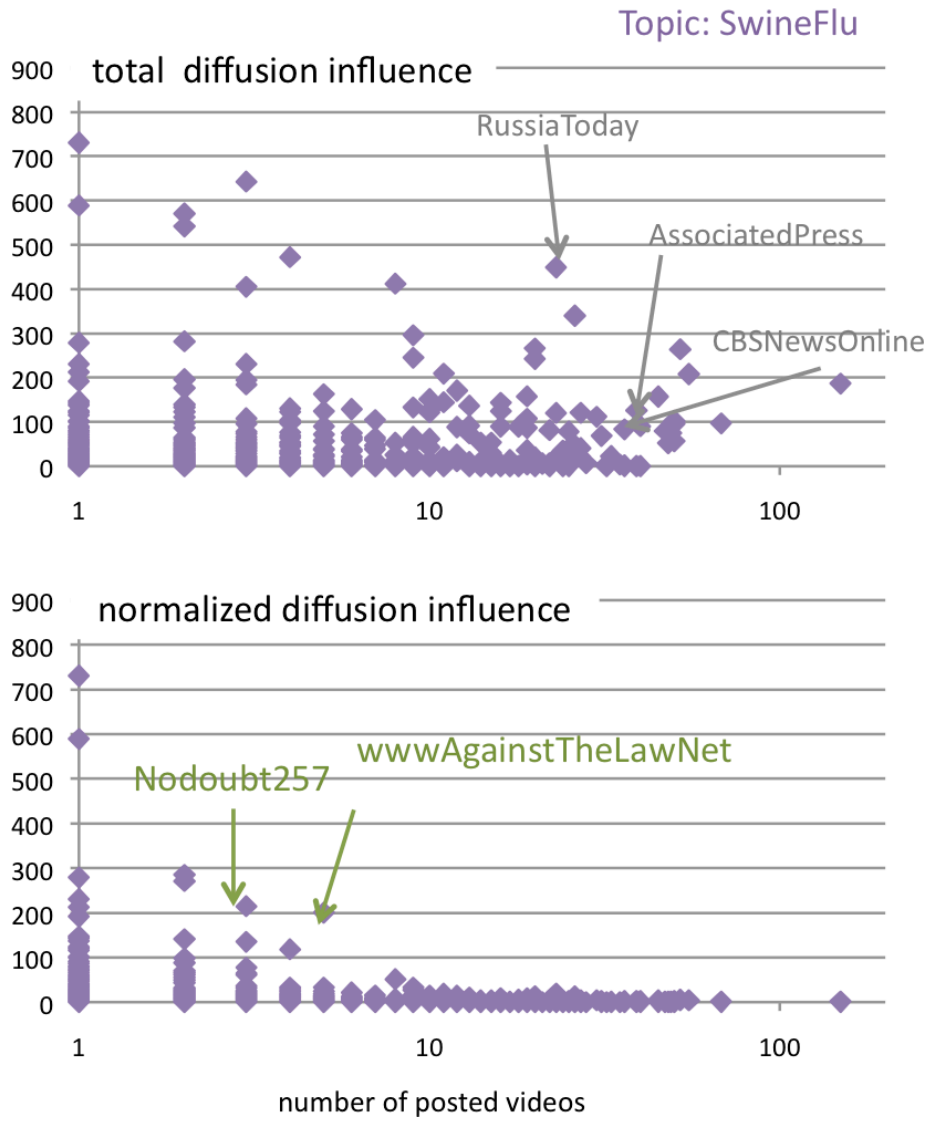
Social network



Diffusion cascade

Diffusion influence of authors

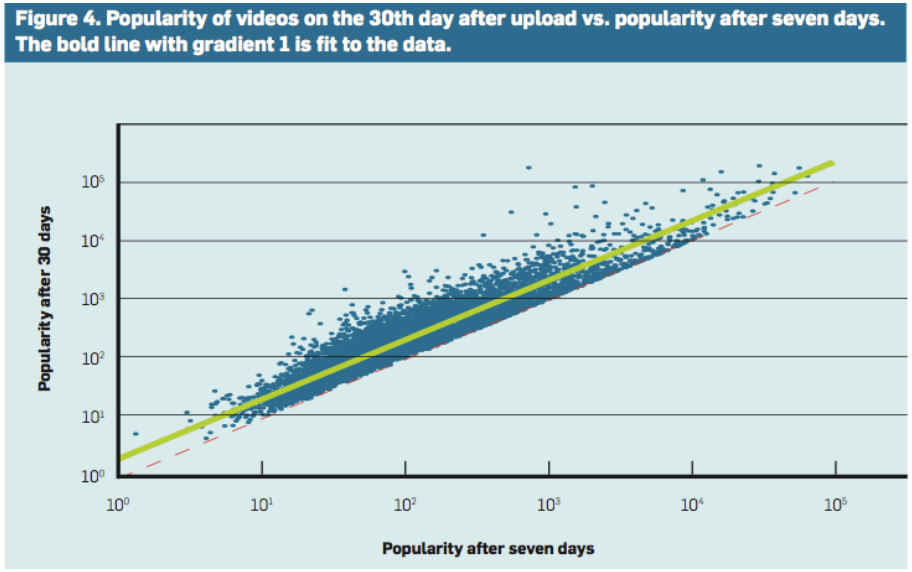




Can we predict virality?

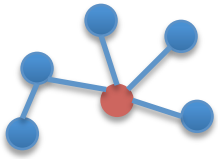
- On YouTube: early view-count correlate well with ultimate view count.
  [Szabo and Huberman, CACM2010]

- What should be the dimensions of meme virality?
  - Volume: how many people remix with this visual meme?
  - Longevity: when is the last remix (among the observed)?

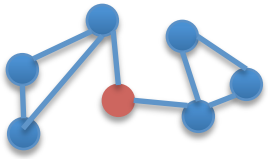


Predicting importance with meme features

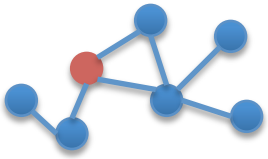
- Content volume up to 24 hours
- Author graph centrality features



Degree centrality

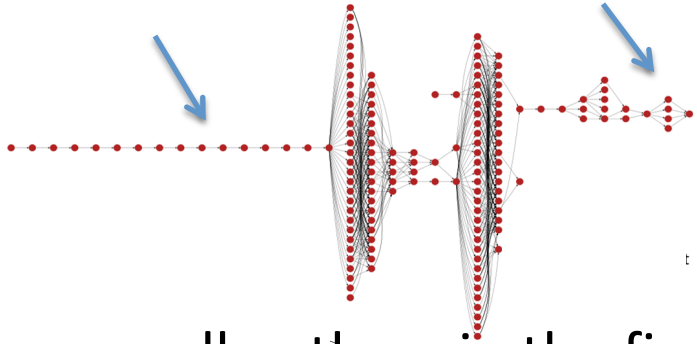


Betweeness centrality



Closeness centrality

- Author influence index



- The Mean+Std of the above features over all authors in the first 24 hours

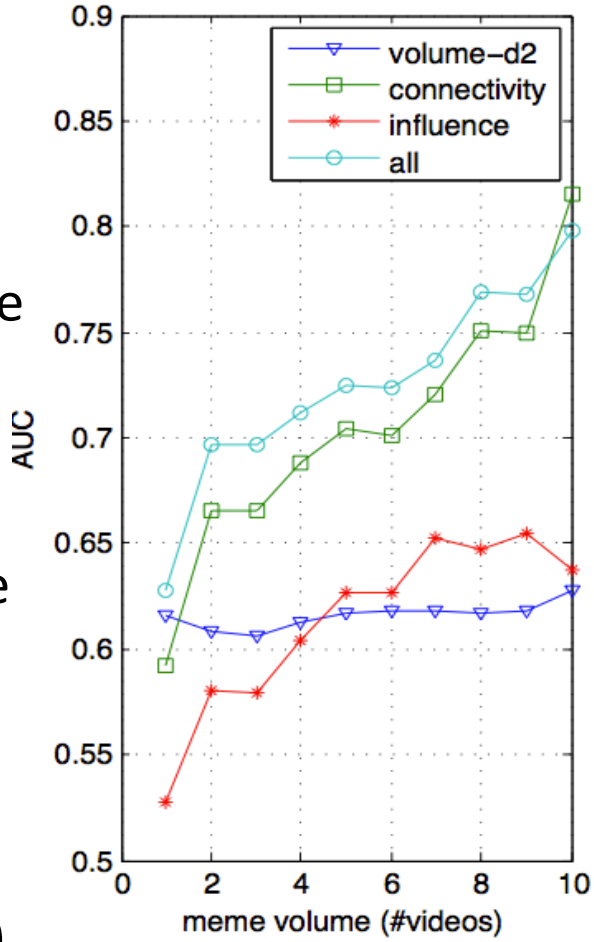
Our Results: Predicting Meme Importance

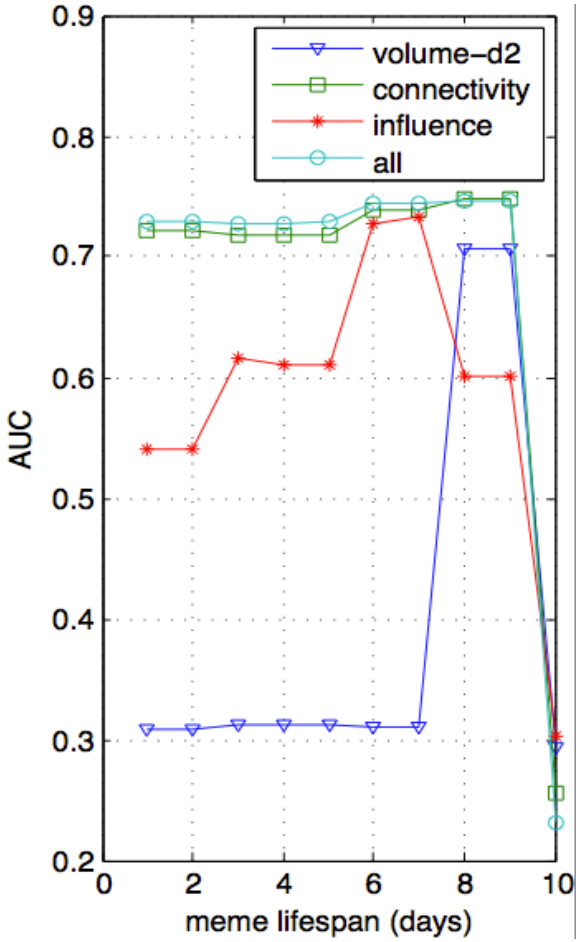
Evaluation :

- Make binary prediction tasks by thresholding meme volume and longevity at various levels

- Train on first month of the Iran3 topic
- Predict on the next two months

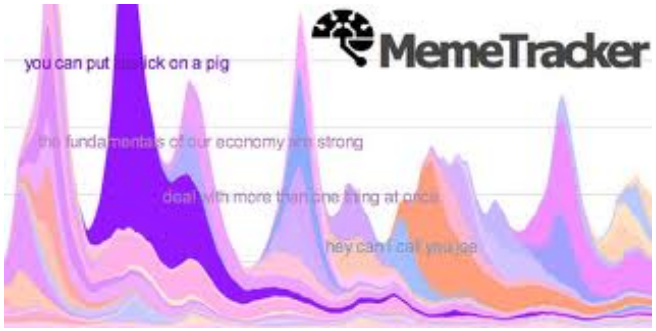
- Using SVM (RBF and D=3)

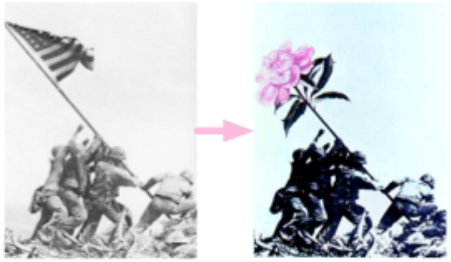


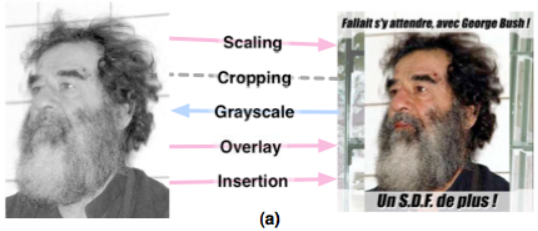


Related work

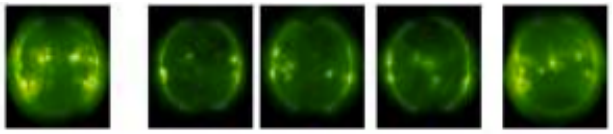
- MemeTracker and the dynamics of news cycle [Leskovec et al KDD'09]
- Web-scale image clustering [Liu and Rosenberg, 2007]
- Internet Image Archeology [Kennedy and Chang, MM'08]







860 images



Our contribution is to reliably extract and analyze "quoting" on Youtube -- visual memes for understanding video sharing behavior and information dissemination on social video sites.

Summary

- YouTube is a living lab of large-scale social behavior
- We propose visual memes as a tool to track large-scale remixing and quoting in videos
- We observed event dynamics and user influence on real-world news events

  Website with data release http://cecs.anu.edu.au/~xlx/proj/visualmemes.html

- This study is a start: more exciting questions ahead
  - Tracking other video genres and domains
  - Does shot sequences matter?
  - Is the associated text useful?
  - How can we use this?
  - …

Thank You !

- For listening ☺

Questions welcome

Info and contact:
http://cecs.anu.edu.au/~xlx/proj/visualmemes.html

lexing.xie@anu.edu.au , @lexing