The Lifecyle of a Youtube Video: Phases, Content and Popularity

Honglin Yu, Lexing Xie, Scott Sanner

Australian National University, NICTA

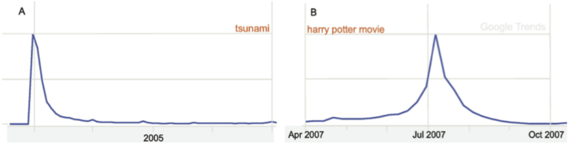
May 22, 2015

Overview

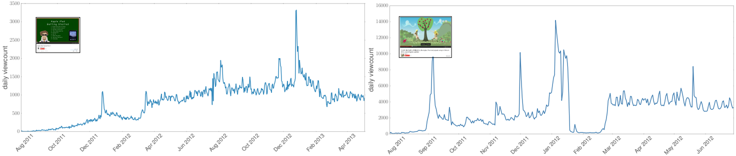
"The scarce, and therefore valuable, resource is now attention"

— B. A. Huberman

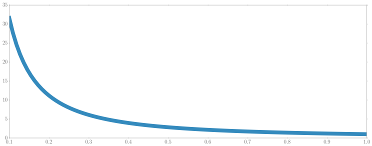
- Previous: Crane and Sornette's model (PNAS 2008)

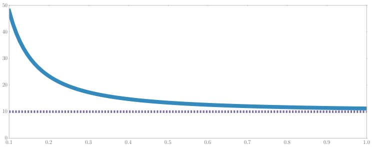


- But, in reality ...



Generalized Power-law Phases





[CS08]: $x[t] \sim t^b$

- result of epidemic branching processes

[Ours]: $x[t] = at^b + c$

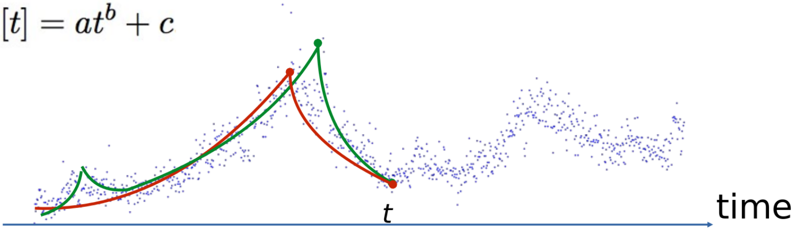
- sufficiently expressive for monotonic curves
- model multiple phases
- account for different background processes

Both are efficient to fit

The Phase-finding Algorithm

$$\min.\ \sum_{i=1}^{n} \overbrace{E_i\{x[\,\underbrace{t_i^s : t_i^e}_{\text{boundary}}\,], \underbrace{a_i, b_i, c_i}_{\text{paremeter}}\}}^{\text{fitting error}} + \overbrace{\eta(n-1)}^{\text{Regularizer}}$$

$$x[t] = at^b + c$$



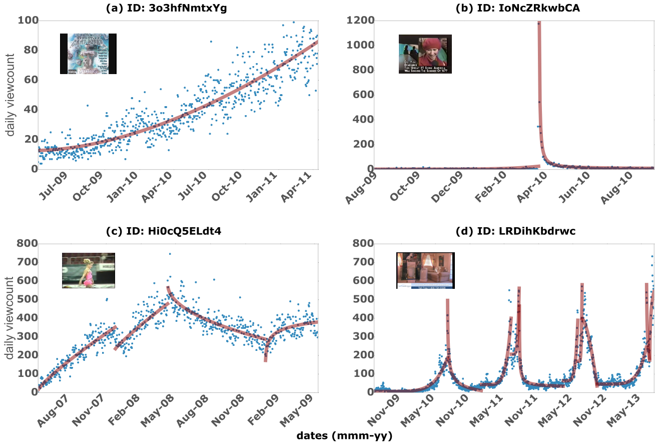
- Try all the possible segmentation
- Dynamic programming with fitting in loop

The "Tweeted Video" Dataset

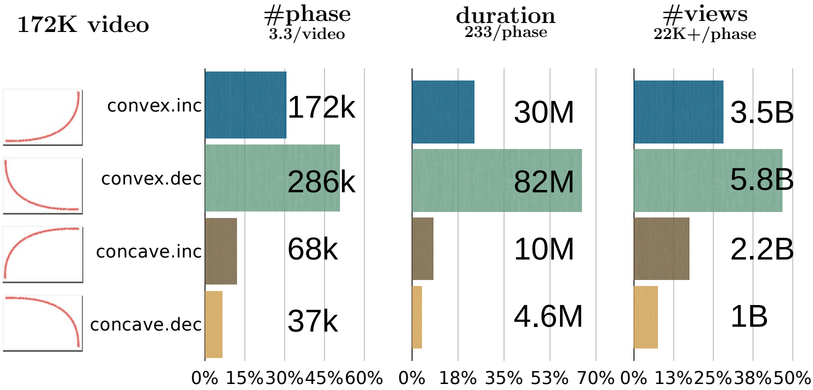
Category	**#videos**	**Category**	**#videos**
Music	64096	Howto	4357
Entertainment	26602	Travel	3379
Comedy	14616	Games	3299
People	12759	Nonprofit	2672
News	10422	Autos	2398
Film	8356	Animals	2375
Sports	7872	Shows	407
Tech	4626	Movies	15
Education	4577	Trailers	13
Total number: 172841			

- Unique longitudinal popularity history for a large+diverse set of videos
- From 20-30% sample of tweets 2009.06-07

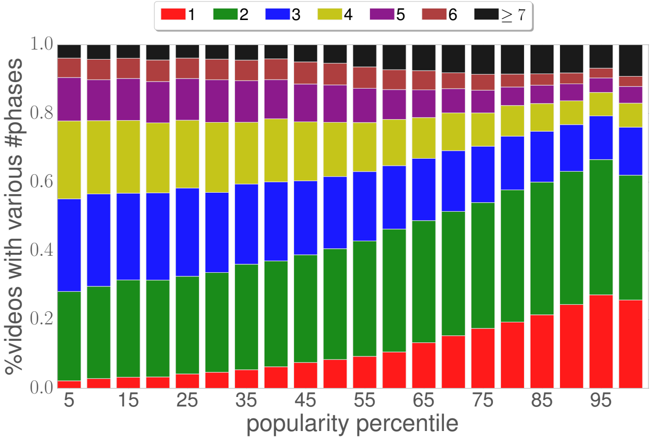
Examples of Segmentation Result



Four Types of Phases

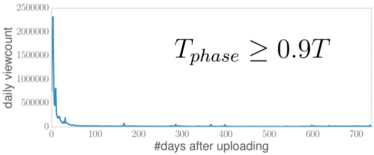


#Phase v.s. Video Popularity

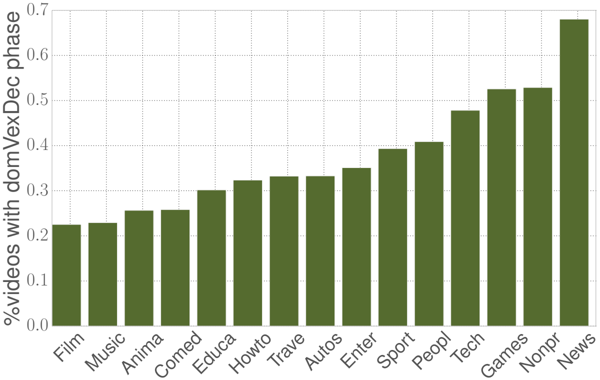


- Popular videos have more phases.

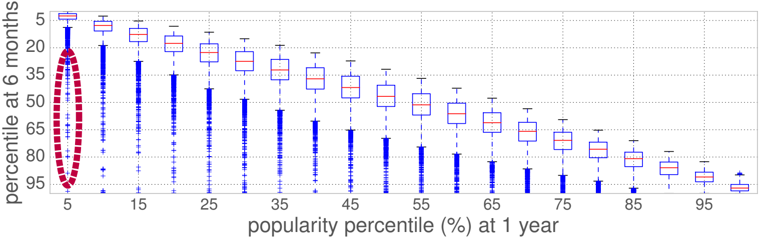
Dominant Convex Decreasing Phases

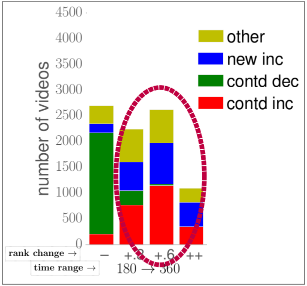


- Novelty is the (only) most important factor
- Do not revive



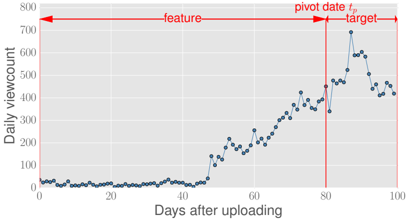
How do popularity change?





- Many videos go through a jump in popularity.
- They have been in a continuously increasing phase, or have at least one new phase.

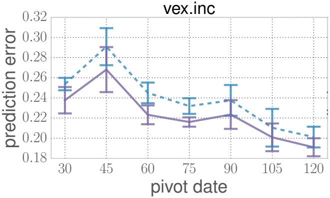
Phase-aware Viewcount Prediction

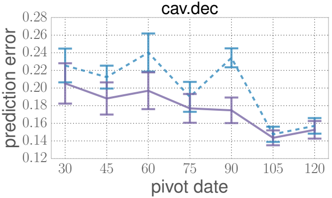


- Target: $\chi^* = \sum_{\tau=1}^{\Delta t} x[t_p + \tau]$
- Prediction: $\hat{\chi} = \sum_{\tau=1}^{t_p} \alpha_\tau x[\tau]$
- Measure: normalized MSE,

$$\epsilon = \frac{1}{\Delta t |\mathcal{V}|} \sum_{v \in \mathcal{V}} (\chi^* - \hat{\chi})^2$$

		#phase $\lesssim$ 4 (79.5% videos)			
		vex.inc	**vex.dec**	**cav.inc**	**cav.dec**
15	baseline	0.2450 ± 0.0103	0.0370 ± 0.0038	0.2745 ± 0.0447	0.2402 ± 0.0216
	phase	$\mathbf{0.2232 \pm 0.0093}^{*\dagger}$	$\mathbf{0.0337 \pm 0.0037}^{\dagger}$	0.2614 ± 0.0432	$\mathbf{0.1969 \pm 0.0208}^{*\dagger}$
30	baseline	0.5013 ± 0.0386	0.0852 ± 0.0027	0.5953 ± 0.0562	0.5085 ± 0.0552
	phase	$\mathbf{0.4642 \pm 0.0373}^{*\dagger}$	$\mathbf{0.0771 \pm 0.0011}^{*\dagger}$	0.5734 ± 0.0598	$\mathbf{0.4241 \pm 0.0428}^{*\dagger}$





Summary

- Main contribution
  - New representation: popularity phases.
  - New method: phase extraction algorithm.
  - A large-scale measurement study.
  - Better viewcount prediction.
- Links
  - Segmentation Algorithm: `https://github.com/yuhonglin/segfit`
  - Dataset: `https://github.com/yuhonglin/ytphasedata`
  - Data crawler: `https://github.com/yuhonglin/YTCrawl`
- Our on-going work: generative model of popularity

Thank you!

Riley Crane and Didier Sornette.
Robust dynamic classes revealed by measuring the response function of a social system.
Proceedings of the National Academy of Sciences, 105(41):15649–15653, 2008.