

Revisit Behavior in Social Media: The Phoenix-R Model and Discoveries

**Flavio Figueiredo, Yasuko Matsubara, Bruno Ribeiro,
Jussara M. Almeida, Christos Faloutsos**

Institute for Web Research (InWeb) @ DCC-UFMG

Databases Group @ CMU

How should we account and model information popularity online?



A screenshot of a YouTube video player showing the official trailer for X-Men: Days of Future Past. The video is paused at 0:33 of a 2:41 duration. The video content shows several characters in a dark, industrial setting. Below the video player, the title is "X-Men: Days of Future Past | Official Trailer 3 [HD] | 20th Centur...". The channel name is "X-Men Movies" with 166 videos. The video has 14,429,124 views, 45,306 likes, and 2,134 dislikes. A red "Subscribe" button is visible with 142,424 subscribers.

X-Men: Days of Future Past | Official Trailer 3 [HD] | 20th Centur...

X-MEN MOVIES - 166 videos

14,429,124 views

45,306 likes 2,134 dislikes

Subscribe 142,424

How should we account and model information popularity online?



The image shows a YouTube video player interface. The video is titled "X-Men: Days of Future Past | Official Trailer [HD] | 20th Century...". The video player shows a scene from the movie with several characters in a dark, industrial setting. The video progress bar is at 0:33 / 2:41. Below the video player, the channel name "X-Men Movies" is visible with a checkmark and "166 videos". A red "Subscribe" button is present with "142,424" subscribers. The view count is "14,429,124 views" with a thumbs up icon for "45,306" likes and a thumbs down icon for "2,134" dislikes. A green callout arrow points from the text "Over 14 million views" to the view count.

Over 14 million views

X-Men: Days of Future Past | Official Trailer [HD] | 20th Century...

X-MEN
MOVIES

X-Men Movies ✓ - 166 videos

Subscribe 142,424

14,429,124 views

45,306 2,134

Audience: Unique users



X-Men Movies

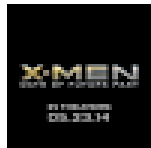


via Google+ 4 weeks ago

The final X-Men trailer is here! How many times have you watched it?

Reply · 845  

Audience vs Visits

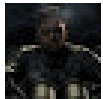


X-Men Movies

via Google+ 4 weeks ago

The final X-Men trailer is here! How many times have you watched it?

Reply - 845



Tristan Nicoll 4 weeks ago
about 13 times lol

Reply - 18



Charlie Davis 4 weeks ago
27 times

Reply - 10



methazza 4 weeks ago
56 times.... im pathetic
Translate

Reply - 16



OutOfRegs 4 weeks ago
over 9000 times

Reply - 19



Measuring both visits and audience (unique users) have their benefits

- How many users watched my ad?
 - Exposure
 - Revenue
- How many times was my ad watched?
 - Caching
 - Sharding and content provisioning
- However...
 - Understanding and modeling both effects is still an open issue

Our Study

- **Understanding and modeling** revisit behavior in social media
- **Understanding**
 - Characterization of millions of user activities
 - User played/watched/visited a social media object at a certain time
- **Modeling**
 - The **Phoenix-R** model for popularity time series

Datasets

- User Activity
 - User, Object (song/tweet/video), Time stamp
- All of the datasets range from months to years

Dataset	User Activities	Description
MMTweet (Million Musical Tweets)	Little over 1 million	Tweets declaring songs which users listen to
Twitter	576 million	Hashtags
LastFM	19 million	Plays on artists and songs
YouTube	-	3 million daily time series

Discoveries

Discoveries

- Relationships between audience (unique users) and revisits

Dataset	Median $\frac{\#Revisits}{\#Audience}$	Median $\frac{\#Revisits}{\#Total\ Visits}$	% of cases $\#Revisits > \#Audience$
MMTtweet	0.68	0.40	33%
Twitter	1.70	0.62	66%
LastFM	25.39	0.96	100%

Discoveries on Smaller time Scales

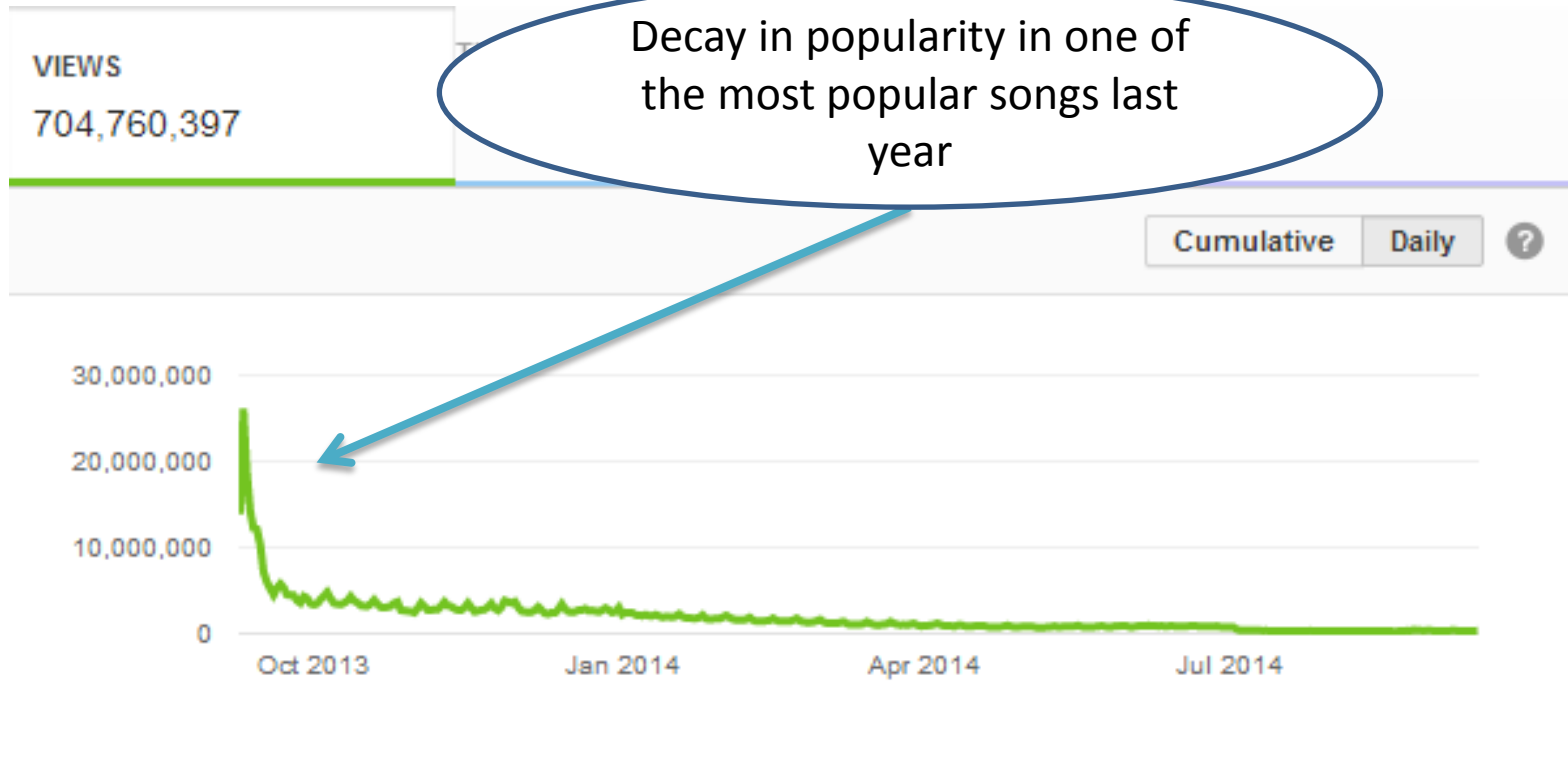
- Isolate the effect of users coming back to the datasets after long periods
- Daily Time Windows

Dataset	$\frac{\text{Median } \#Revisits}{\#Audience}$
MMTweet	0.83
Twitter	2.50
LastFM	28.0

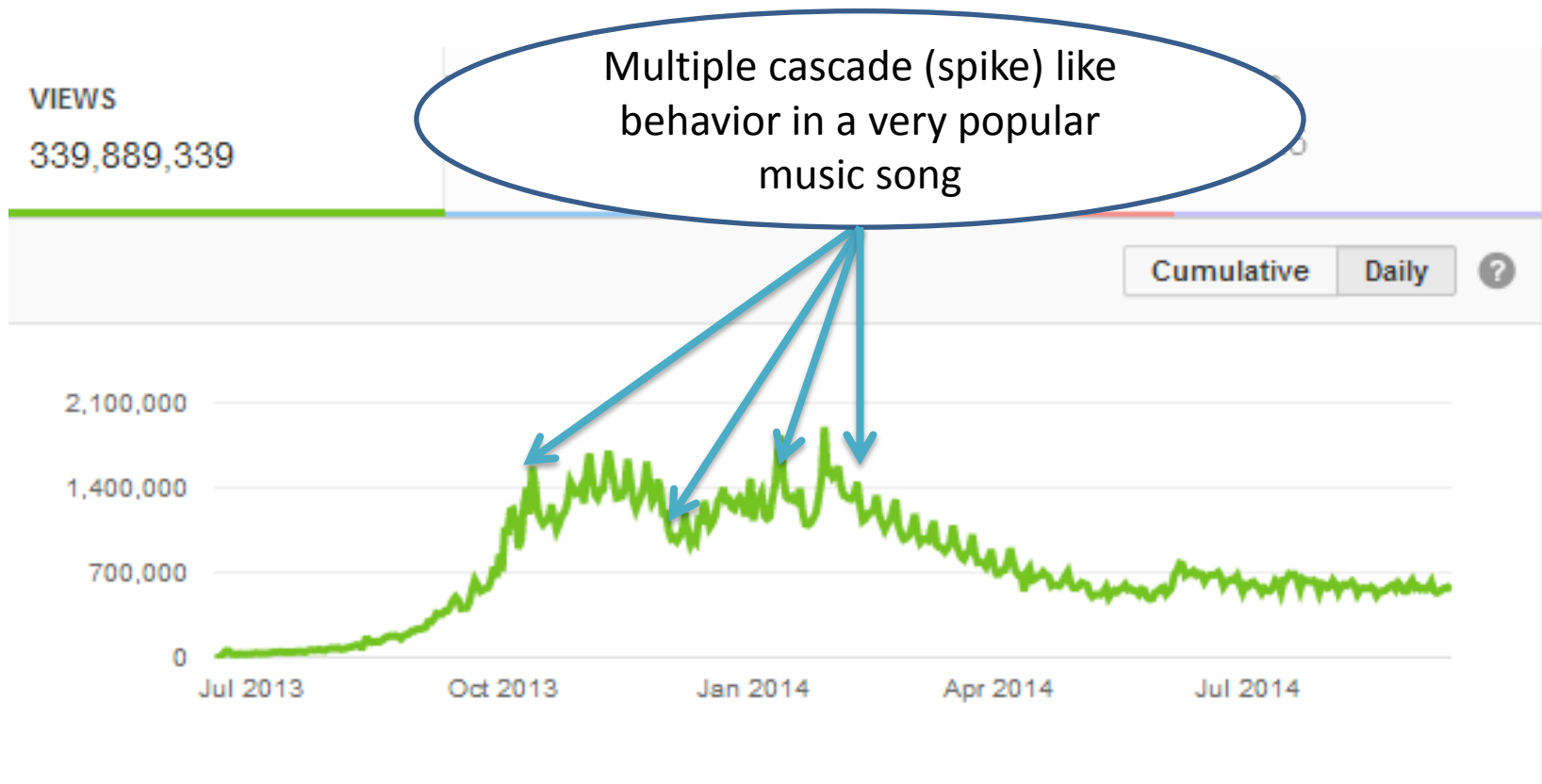
What we know so far

- Users revisit the same object
 - On some datasets (LastFM and Twitter) most of visits are returning users
- Revisits are common on small time scales
 - Above results hold
 - Complements [Anderson2014]
- Users abandon content but it may take a long time
 - Preying behavior from [Ribeiro2014]

Users eventually stop visiting



Some objects behave like a sum of multiple cascades



How de we model these time series?

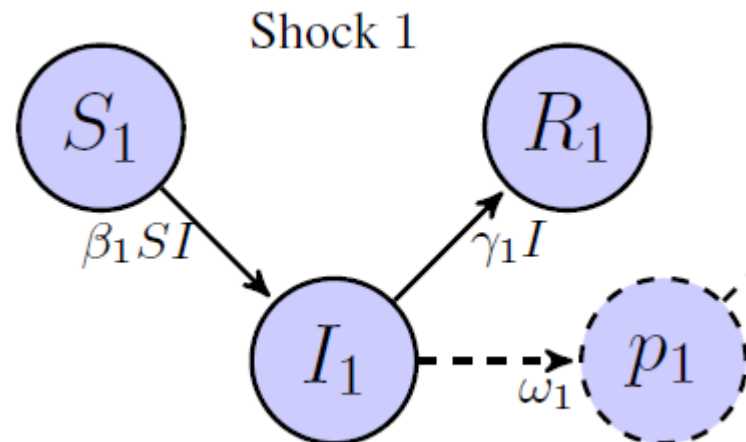
The Phoenix-R Model!

Table 1: Comparison of PHOENIX-R with other approaches

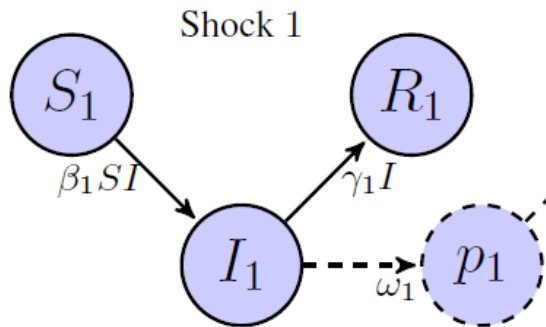
	Revisits	Non-Linear	Forecasting	Multi Cascade
SI [12]		✓		
SpikeM [18]		✓	✓	
TemporalDynamics [21]			✓	
PHOENIX-R	✓	✓	✓	✓

Phoenix-R Explained

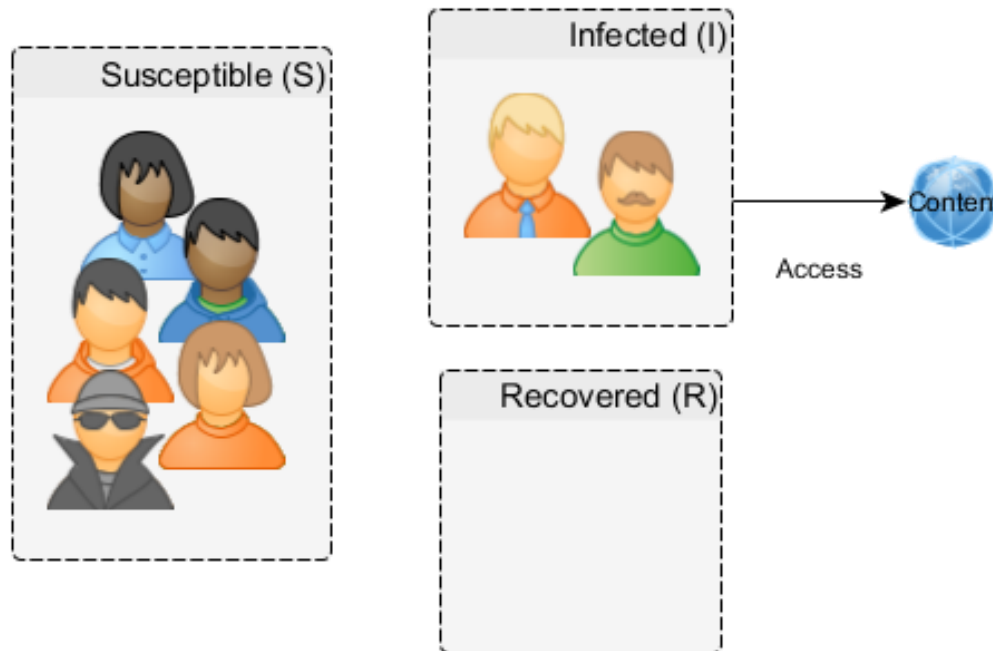
- Single shock (cascade) model
- Epidemiology model



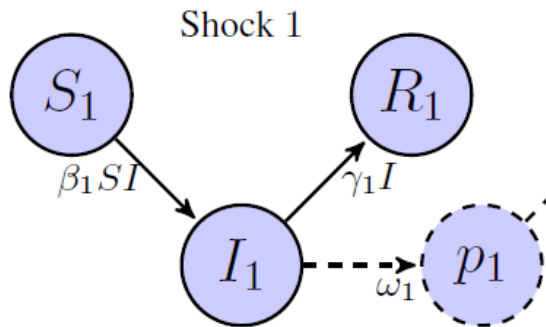
Single Shock



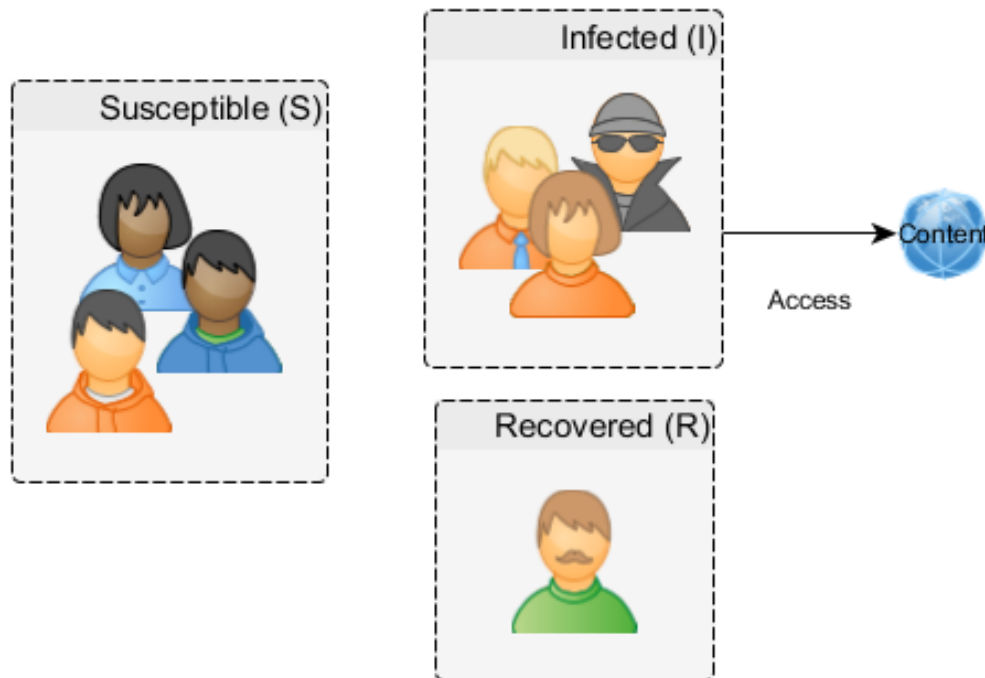
- Starting with some Susceptible and Infected Individuals
- The Infected access the content



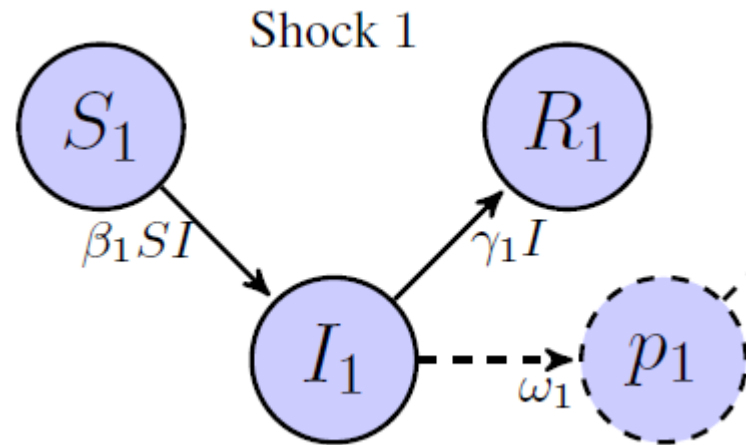
Single Shock



- At the next time tick some Infected recover
- Some Susceptible are infected by the previous infected
- We now expect more visits (more infected)



Single Shock Equations



$$S(t) = S(t - 1) - \beta S(t - 1)I(t - 1)$$

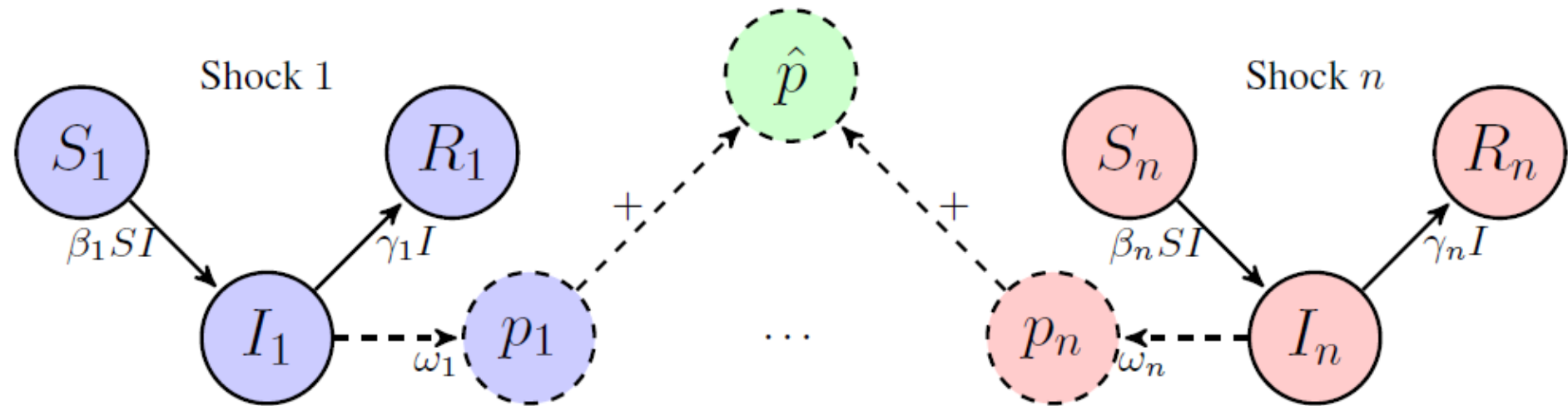
$$I(t) = I(t - 1) + \beta S(t - 1)I(t - 1) - \gamma I(t - 1)$$

$$R(t) = R(t - 1) + \gamma I(t - 1)$$

$$p(t) = \omega I(t).$$

Multiple Shocks

- Simplifying assumption that each shock is a new population (set of users)



$$\hat{p}(t) = \sum_{i, s_i \in \mathcal{S}} p_i(t - s_i) \mathbb{1}[t > s_i]$$

How many shocks to add?

- A perfect model (zero error) can be created by
 - Letting each access be a single user which immediately recovers
 - However, lot's of parameters (cost)
- Using Minimum Description Length (MDL)

$$Cost(\mathbf{t}; \mathcal{P}) = \log^* n + Cost(\mathcal{P}) + Cost(\mathbf{t} | \mathcal{P})$$

How do we fit a time series?

- Step 1:
 - Identify Peaks using Wavelets
 - Intuitively, each peak is a candidate shock (cascade)
 - Linear
- Step 2:
 - Add each peak sorted by height to the model
 - If the MDL decreases, accept peak
- Step 3:
 - Stop when the MDL stops decreasing

Linear runtime (time series length) and parameter free algorithm

Algorithm 1 Fitting the PHOENIX-R model. Only the time series is required as input.

```
1: function FITPHOENIXR(t)
2:    $\epsilon = 0.05$ 
3:    $s \leftarrow \{\}$ 
4:    $p, s' \leftarrow FindPeaks(t)$ 
5:    $s[1] = 0$ 
6:    $s \leftarrow append(s')$ 
7:    $\mathcal{P} \leftarrow \{\}$ 
8:    $min\_cost \leftarrow \infty$ 
9:   for  $i \leftarrow 1$  to  $|s|$  do
10:     $\mathcal{F} \leftarrow LM(t, s(:i))$ 
11:     $m \leftarrow PhoenixR(\mathcal{F})$ 
12:     $mdl\_cost \leftarrow Cost(m, t, \mathcal{F})$ 
13:    if  $mdl\_cost < min\_cost$  then
14:       $min\_cost \leftarrow mdl\_cost$ 
15:       $\mathcal{P} \leftarrow \mathcal{F}$ 
16:    end if
17:    if  $mdl\_cost > min\_cost * (1 + \epsilon)$  then
18:      break
19:    end if
20:  end for
21:  return  $\mathcal{P}$ 
22: end function
```

Find Peaks

Adding shocks

Exit

How good is Phoenix-R?

- Comparing Phoenix-R with two state of the art alternatives
 - RMSE (smaller is better)

	PHOENIX-R vs. TemporalDynamics (daily series)		PHOENIX-R vs. SpikeM (hourly series)	
	RMSE PHOENIX-R	RMSE TemporalDynamics	RMSE PHOENIX-R	RMSE SpikeM
MMTweet	2.93 (± 0.23)	4.18 (± 0.49)	-	-
LastFM	7.09 (± 0.23)	8.31 (± 0.32)	-	-
Twitter	72.05 (± 6.08)	194.79 (± 20.49)	10.83 (± 1.61)	9.77 (± 2.24)
YouTube	280.58 (± 29.29)	3429.19 (± 577.76)	-	-

- Phoenix-R is always better or just as good

How good is Phoenix-R?

- Comparing Phoenix-R with two state of the art alternatives
 - RMSE (smaller is better)

	PHOENIX-R vs. TemporalDynamics (daily series)		PHOENIX-R vs. SpikeM (hourly series)	
	RMSE PHOENIX-R	RMSE TemporalDynamics	RMSE PHOENIX-R	RMSE SpikeM
MMTweet	2.93 (± 0.23)	4.18 (± 0.49)	-	-
LastFM	7.09 (± 0.23)	8.31 (± 0.32)	-	-
Twitter	72.05 (± 6.08)	194.79 (± 20.49)	10.83 (± 1.61)	9.77 (± 2.24)
YouTube	280.58 (± 29.29)	3429.19 (± 577.76)	-	-

- Phoenix-R is always better or just as good

Phoenix-R is also good at forecasting

- RMSE (smaller is better)
- 1, 7 or 30 days ahead forecasting
- Ties on very linear time series

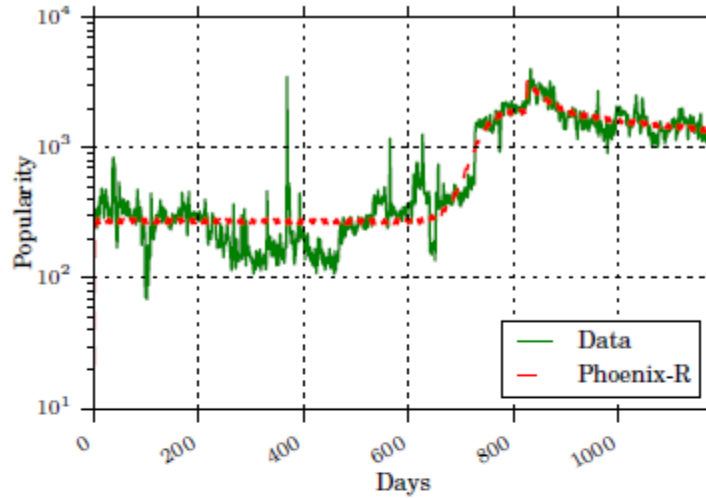
		5%			25%			50%		
		1	7	30	1	7	30	1	7	30
MMTweet	PhoenixR	11.61	12.78	15.15	8.67	6.74	8.82	4.08	6.87	13.58
	TempDynamics	17.07	17.41	16.52	9.63	10.78	14.46	25.19	23.08	30.39
Twitter	PhoenixR	53.68	60.78	215.76	132.21	135.15	210.30	75.58	229.59	254.93
	TempDynamics	104.45	129.36	255.69	643.39	643.83	786.50	420.74	587.86	598.75
LastFM	PhoenixR	2.37	3.97	5.71	8.60	12.06	14.66	11.34	15.03	15.43
	TempDynamics	6.47	7.03	8.00	11.15	14.62	17.86	14.91	18.15	18.80
YouTube	PhoenixR	91.62	106.38	138.88	83.76	113.14	147.04	127.53	97.97	115.97
	TempDynamics	3560.65	3631.09	3661.81	5091.82	5107.82	5143.70	4136.14	4139.73	4169.26

Phoenix-R is also good at forecasting

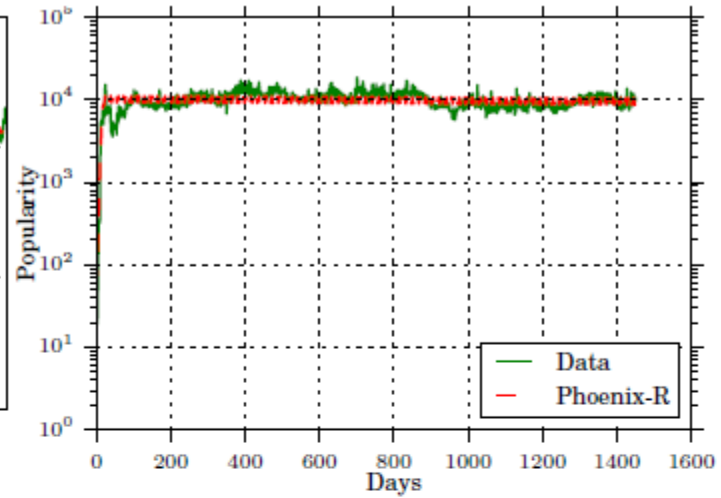
- RMSE (smaller is better)
- 1, 7 or 30 days ahead forecasting
- Ties on very linear time series

		5%			25%			50%		
		1	7	30	1	7	30	1	7	30
MMTweet	PhoenixR	11.61	12.78	15.15	8.67	6.74	8.82	4.08	6.87	13.58
	TempDynamics	17.07	17.41	16.52	9.63	10.78	14.46	25.19	23.08	30.39
Twitter	PhoenixR	53.68	60.78	215.76	132.21	135.15	210.30	75.58	229.59	254.93
	TempDynamics	104.45	129.36	255.69	643.39	643.83	786.50	420.74	587.86	598.75
LastFM	PhoenixR	2.37	3.97	5.71	8.60	12.06	14.66	11.34	15.03	15.43
	TempDynamics	6.47	7.03	8.00	11.15	14.62	17.86	14.91	18.15	18.80
YouTube	PhoenixR	91.62	106.38	138.88	83.76	113.14	147.04	127.53	97.97	115.97
	TempDynamics	3560.65	3631.09	3661.81	5091.82	5107.82	5143.70	4136.14	4139.73	4169.26

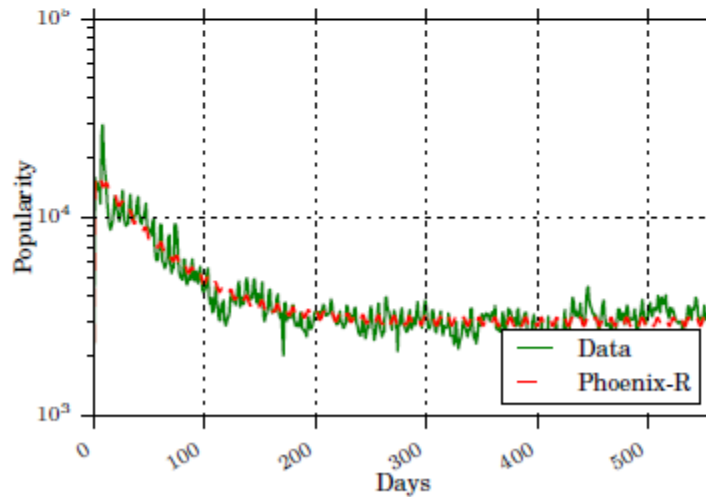
Examples of Phoenix-R at work



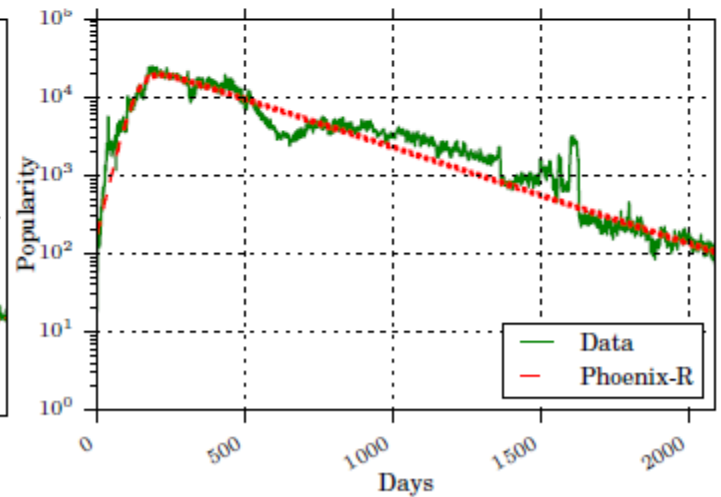
(a) Rock Song (growth in popularity)



(b) Flashdance (80's movie) clip (revisits)

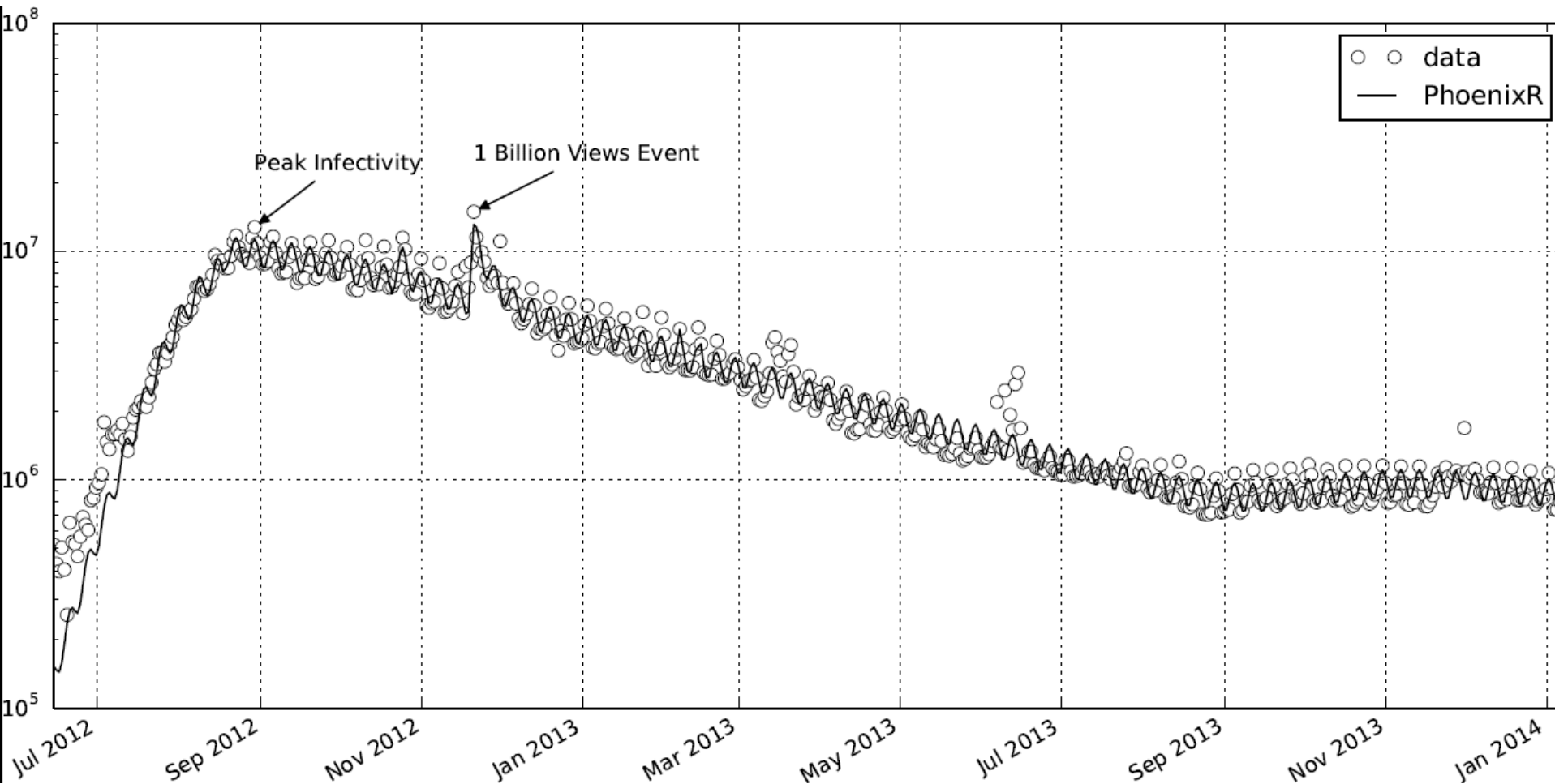


(c) Korean Music Video (single cascade)



(d) User Dancing Video (single cascade)

Examples of Phoenix-R at work



Conclusions

- Phoenix-R model for revisits and multiple cascades
- Based on discoveries from real data
- Scalable linear fitting algorithm
 - On time series length
- Useful for predictions