

Dynamics of News Events and Social Media Reaction

Mikalai Tsytsarau
University of Trento

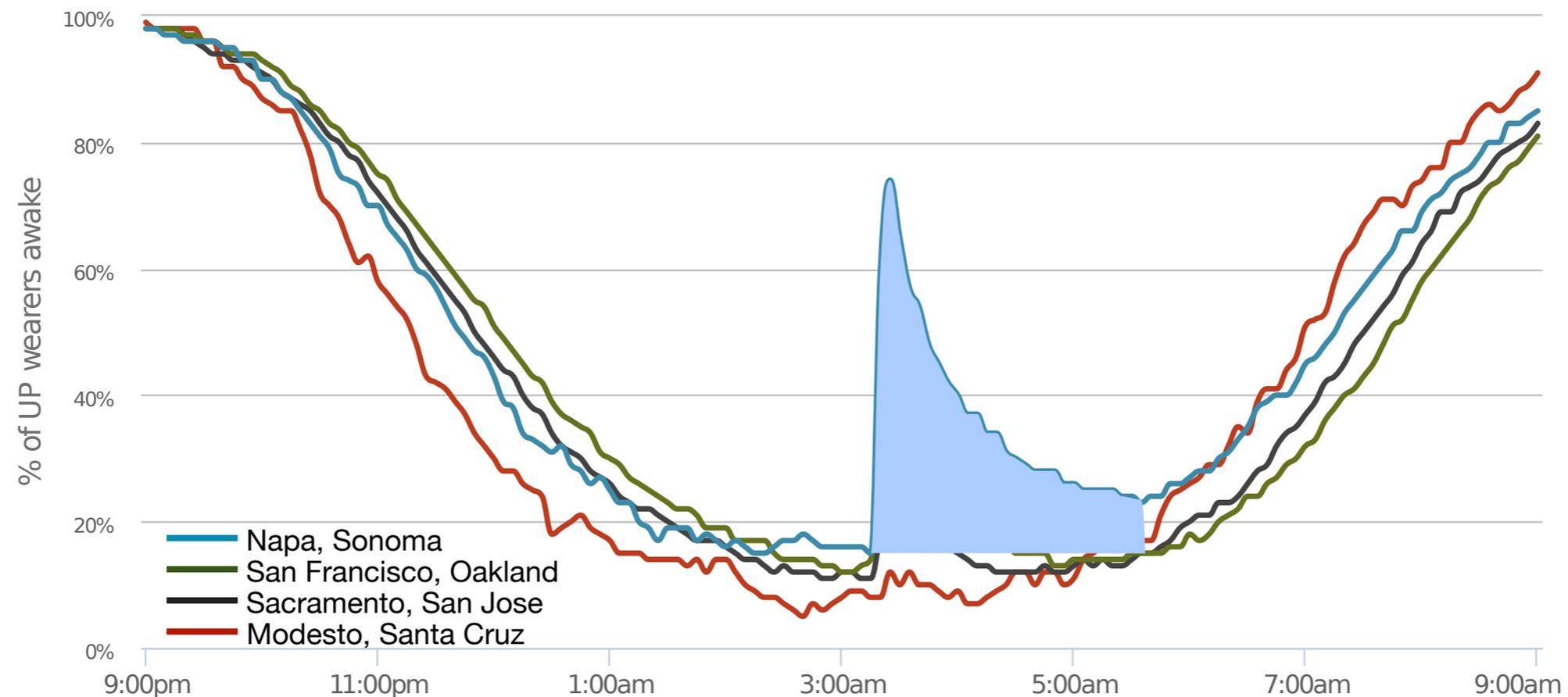
Themis Palpanas
Paris Descartes University

Malu Castellanos
Hewlett Packard

dbTrento

26 August 2014
KDD

Breaking News

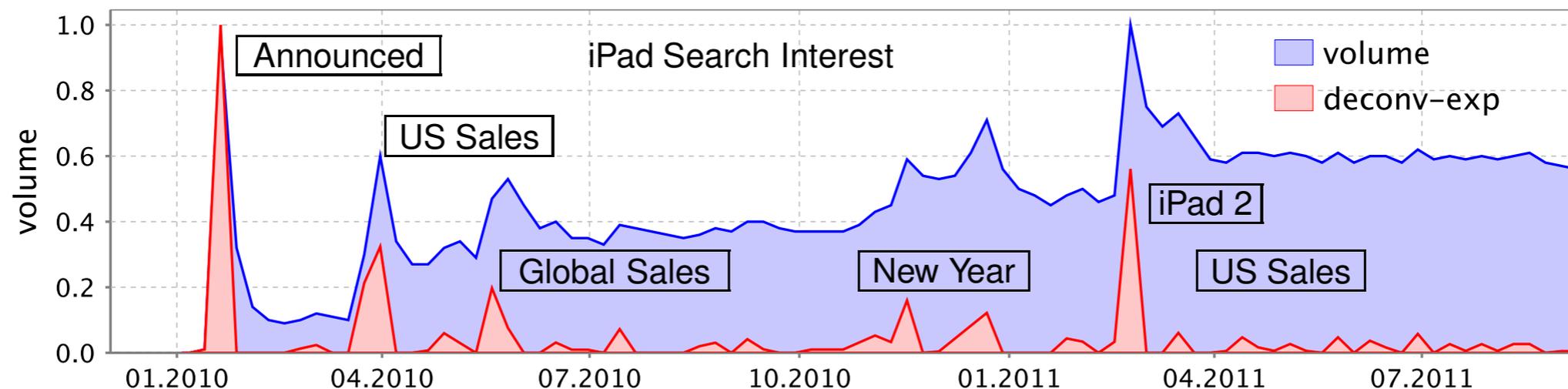


On August 25, earthquake in Napa (California) woke people up

- ▶ The number of awake people was proportional to earthquake magnitude
- ▶ It was also exponentially decreasing with time, like in a Poisson process
- ▶ **So.. can social media dynamics be modeled as a physical process?**

Our goals:

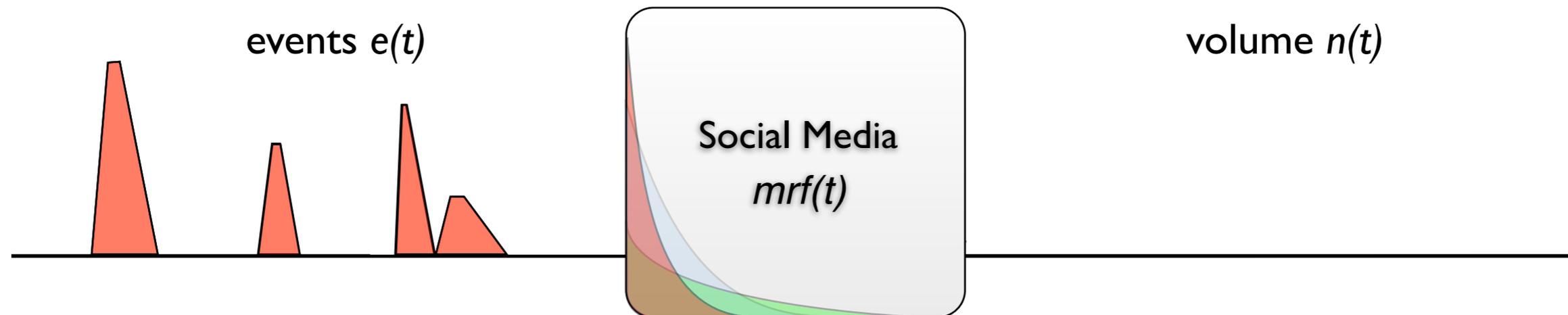
- ▶ Constructing robust model of publication dynamics
- ▶ Extracting event importance by only observing the volume
- ▶ Correlating impacting events to sentiment shifts



Reconstructing event properties is difficult task:

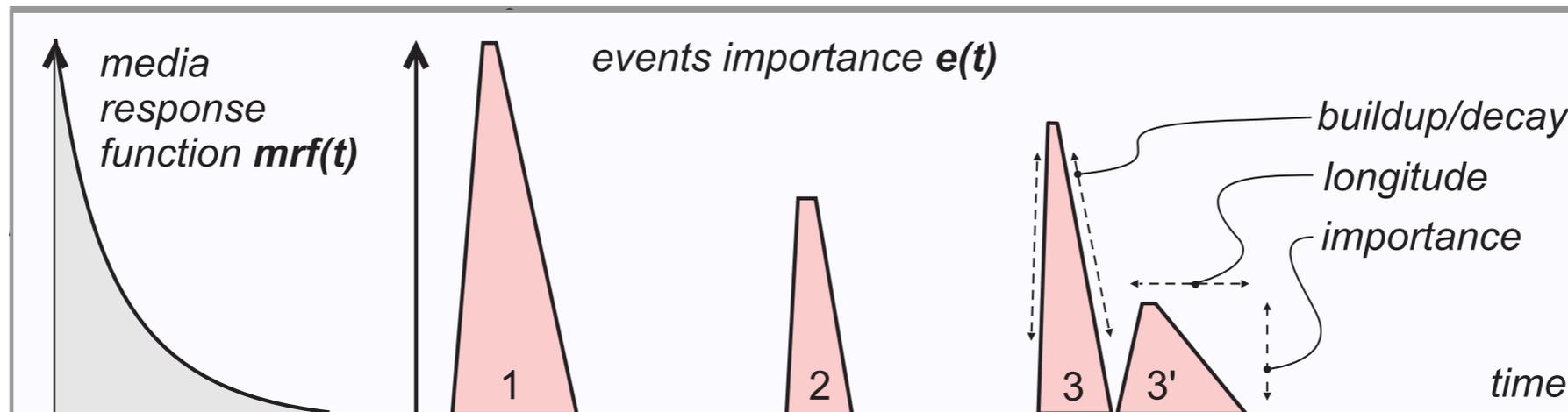
- ▶ Events importance is masked by the trend and overlapping events
- ▶ Beginnings of events and their true longitudes are not observable
- ▶ Events may have different dynamics in different sources / topics

Method Highlights



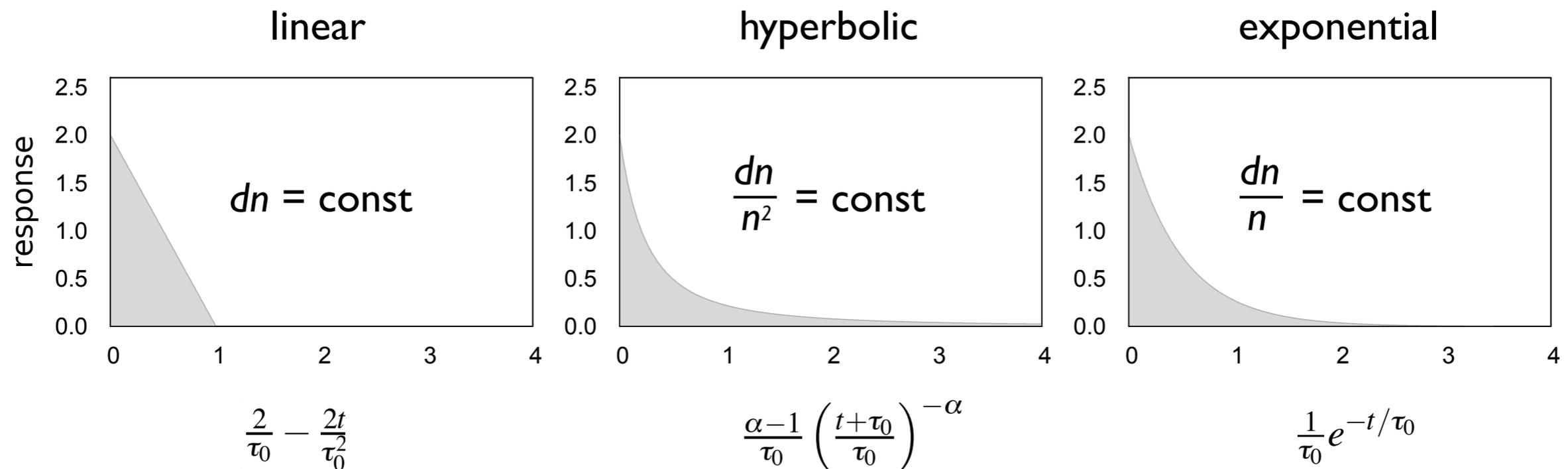
- ▶ We model social- and news media as a *dynamic system*, taking events as input and outputting news publications
- ▶ Output volume of publications depends on the shape of media response function $mrf(t)$, different for various events

Response Dynamics Model



- ▶ The output of the system is modeled as the convolution between events importance series $e(t)$ and media response function $mrf(t)$
- ▶ Events importance $e(t)$ is modeled as piecewise linear function: events have *buildup / decay* rates, *longitude* and *importance*
- ▶ We reconstruct $e(t)$ and its parameters by deconvolution of $n(t)$

Media Response Functions

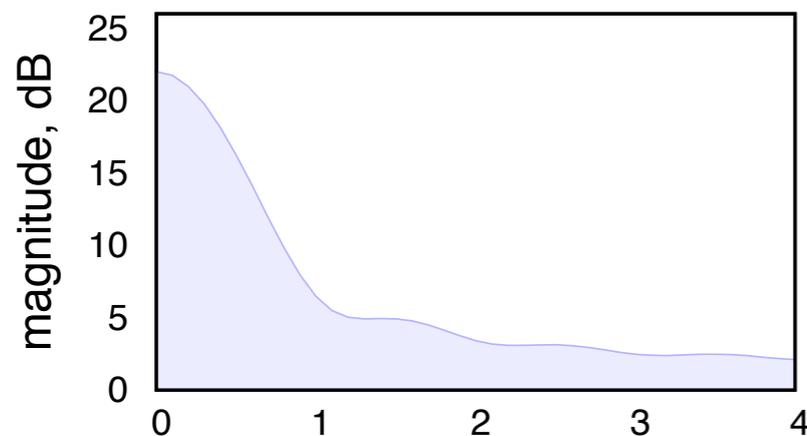


We evaluate the three media response functions:

- ▶ linear response decays in a finite time
- ▶ exponential response's decay is proportional to volume
- ▶ hyperbolic response's decay is proportional to squared volume

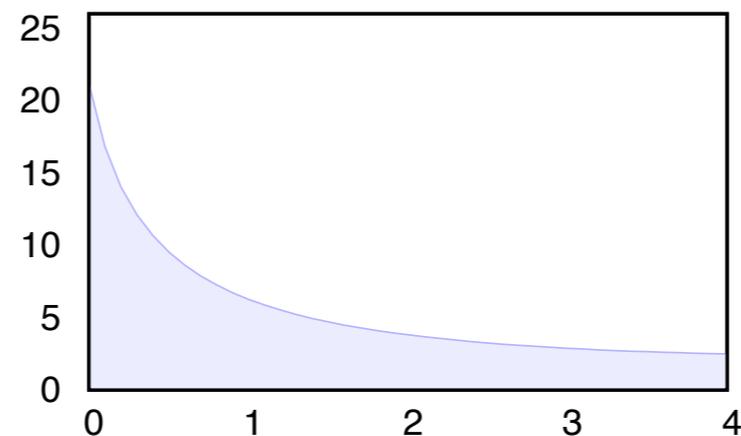
Media Response Functions

linear



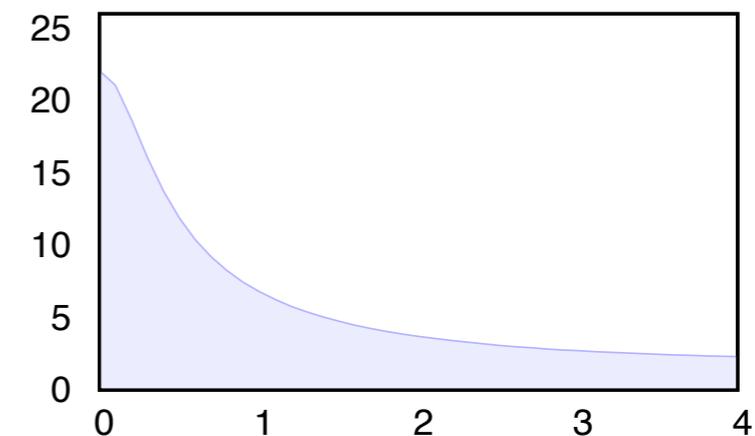
$$\frac{2}{\omega^2 \tau_0^2} (1 - e^{-j\omega \tau_0}) + \frac{2}{j\omega \tau_0}$$

hyperbolic



$$\Gamma(1 - \alpha, j\omega \tau_0) (\alpha - 1) (j\omega \tau_0)^{\alpha - 1} e^{j\omega \tau_0}$$

exponential



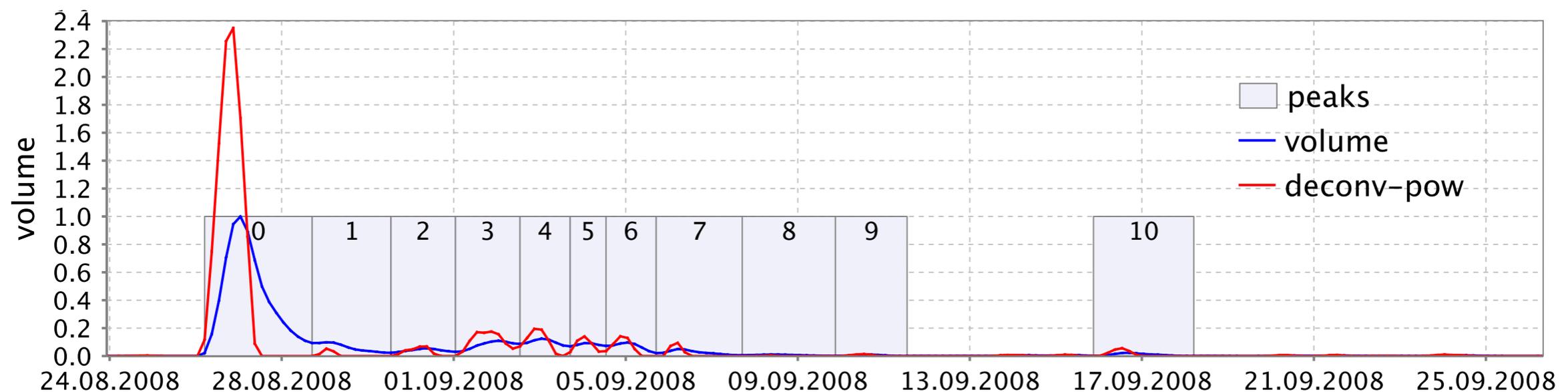
$$\frac{1}{1 + j\omega \tau_0}$$

Convolution Theorem allows to deconvolve in the frequency domain:

$$\mathcal{F}\{n(t)\} = \mathcal{F}\{e(t) * mrf(t)\} = \mathcal{F}\{e(t)\} \cdot \mathcal{F}\{mrf(t)\}$$

$$e(t) = \mathcal{F}^{-1}\{e(\omega)\} = \mathcal{F}^{-1}\{n(\omega) / mrf(\omega)\}$$

Deconvolution Method



Our method reconstructs events importance, longitude and delay:

- ▶ It automatically extracts peaks by tracking the derivative and volume
- ▶ For every peak, the descending slope is analyzed using regression
- ▶ Model's parameters are obtained by averaging peaks parameters
- ▶ Following deconvolution, events are represented as linear shapes

Experimental Setup

Meme dataset (Leskovec et al. 2009)

contains social media phrases during US elections (Aug 2008 - Apr 2009)

- ▶ We used volume time series of top 100 memes from the dataset
- ▶ We processed volume using the granularity of 1/6 day with smoothing

Twitter dataset

contains US tweets extracted for the period (Jul 2009 - Dec 2009)

- ▶ We extracted 30 time series for most impacting events during that period
- ▶ We processed the data using the granularity of 1/4 day and smoothing

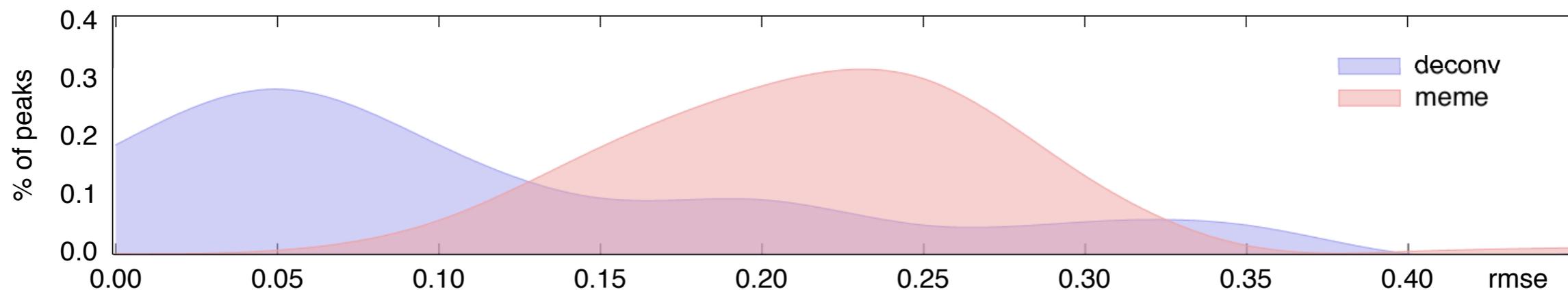
We measured the accuracy of our models for these media as follows:

1. For every volume time series, we extracted peaks and *mrf* parameters
2. We deconvolved volume and approximated event importance as triangular
3. We convolved events series back, measuring RMSE versus the original

Meme Dataset

We compared accuracy of our model and meme model (Leskovec et al.)

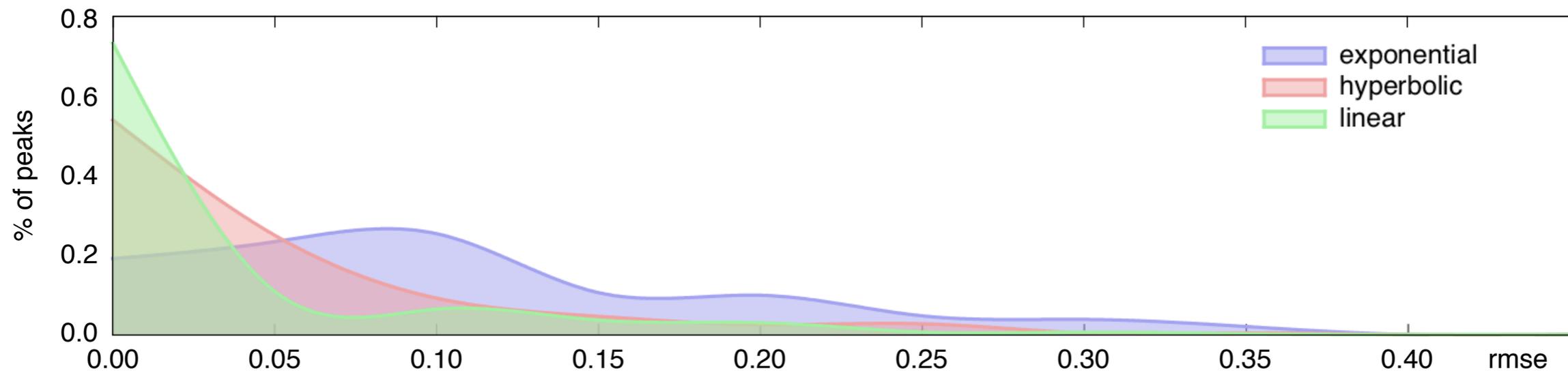
- ▶ Deconvolution parameters were *fixed* ($\tau=0.8$, $\alpha=2.0$) for *all time series*
- ▶ Meme model's parameters were *fitted* using linear regression *per peak*
- ▶ Meme model demonstrated the error of 0.22 ± 0.09
- ▶ Deconvolution model demonstrated the error of 0.11 ± 0.09



Meme Dataset

We observe good fitness of linear and hyperbolic response models

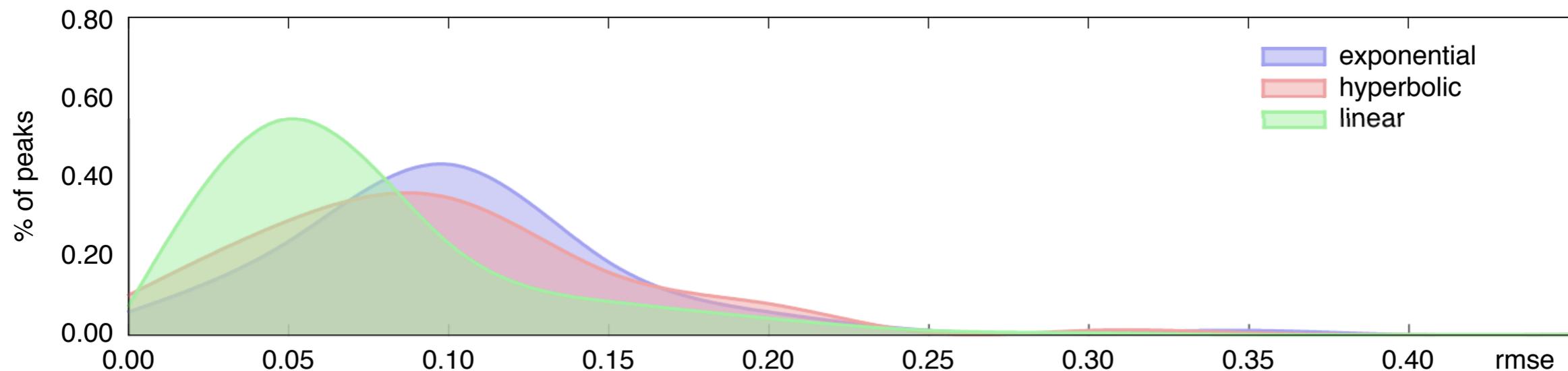
- ▶ The average error of linear model was 0.03 ± 0.06 ($\tau = 0.5$ day)
- ▶ Hyperbolic model had almost equal error of 0.05 ± 0.06 ($\tau = 0.4$, $\alpha = 2.8$)
- ▶ Exponential model demonstrated worse error of 0.11 ± 0.09 ($\tau = 0.4$)



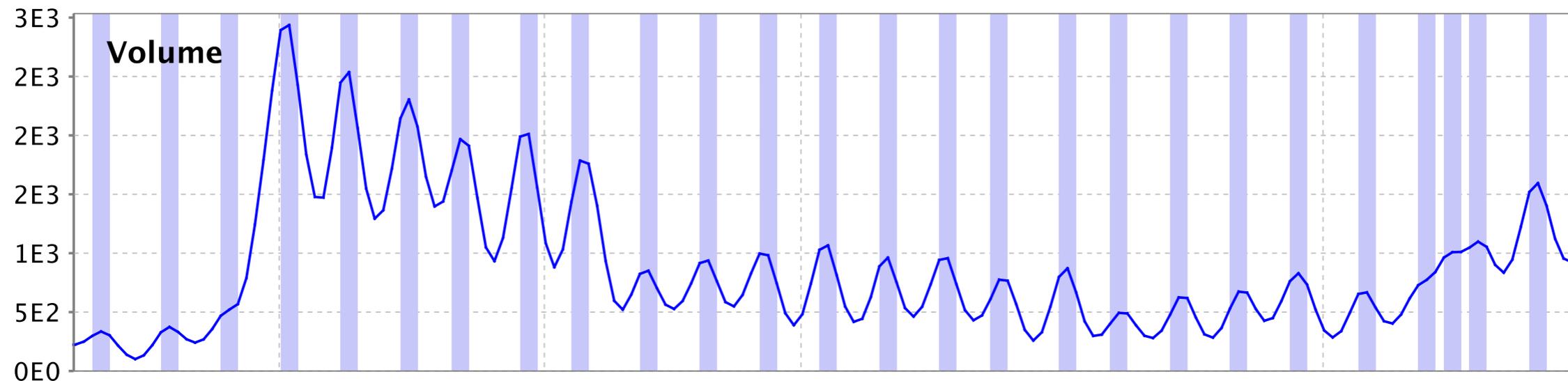
Twitter Dataset

In Twitter dataset, the linear model demonstrated the best performance

- ▶ The average error of linear model was 0.06 ± 0.06 ($\tau = 2.5 \pm 0.9$ day)
- ▶ Hyperbolic model had higher error of 0.09 ± 0.07 ($\tau = 1.3$, $\alpha = 2.4$)
- ▶ Exponential model demonstrated the worst error of 0.10 ± 0.07 ($\tau = 1.5$)



Correlation Evaluation - Method



We use Cosine and Jackard similarity to measure correlation:

$$\rho(s, n, \delta) = \frac{|S_{t+\delta} \cap N_t|}{||S_{t+\delta}|| ||N_t||} \quad \text{Cosine}$$

$$\rho(s, n, \delta) = \frac{|S_{t+\delta} \cap N_t|}{|S_{t+\delta} \cup N_t|} \quad \text{Jackard}$$

Time lag δ is adjusted until the correlation becomes the largest:

$$\Delta = \arg \max_{\delta} [\rho(s, n, \delta)]$$

Correlation Evaluation - Results

Topic Name	Type	Num	Contradiction		Pos Volume		Neg Volume	
			ρ_C	lag	ρ_C	lag	ρ_C	lag
Hangover	lin	20	0.68	0.0	0.65	0.0	0.60	2.0
Ice Age	exp	22	0.58	0.6	0.66	-1.0	0.65	1.6
Harry Potter	exp	13	0.42	1.7	0.57	0.0	0.45	1.7
LCROSS	hyp	12	0.65	-0.9	0.73	-1.5	0.70	-1.4
CERN LHC	hyp	21	0.68	1.0	0.61	1.0	0.54	-1.4
NASA	hyp	6	0.67	-1.5	0.72	1.5	0.74	-1.5
Michael Jackson	hyp	14	0.54	-1.3	0.49	1.5	0.51	-1.1
Eclipse	hyp	6	0.67	-1.4	0.42	2.1	0.50	1.9
Swine Flu	exp	19	0.58	0.6	0.59	0.0	0.56	-0.5
Barack Obama	hyp	17	0.60	-1.6	0.55	0.0	0.47	0.6
Iran Election	hyp	17	0.51	-0.5	0.47	0.0	0.53	1.6
Iraq	hyp	22	0.61	-1.6	0.58	-1.6	0.74	0.6
Follow Friday	lin	20	0.83	-1.3	0.62	0.7	0.75	-1.4
"Aquila Summit"	exp	30	0.79	0.5	0.83	0.9	0.96	-0.5
"Beer Summit"	hyp	6	0.71	0.8	0.62	0.9	0.71	-0.8
Transformers	exp	16	0.56	-1.8	0.58	0.6	0.51	-1.9
Facebook	hyp	15	0.60	-1.2	0.50	0.8	0.56	-1.3
TwitterPeek	hyp	6	0.72	-1.3	0.72	0.0	0.77	0.0
Leica	hyp	22	0.70	0.0	0.60	-1.7	0.80	-0.5
Gmail	hyp	20	0.71	0.6	0.55	-1.9	0.67	0.6
Super Bowl	lin	20	0.71	0.0	0.53	-1.6	0.71	0.0
Google Wave	hyp	11	0.50	0.5	0.67	-1.0	0.51	0.5



Thanks

Questions are welcome!