THE 18TH ACM SIGKDD CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING
KDD 2012 BEIJING
Beijing, China August 12-16, 2012

Rise and Fall Patterns of Information Diffusion: Model and Implications

Yasuko Matsubara
Kyoto University
y.matsubara@db.soc.i.kyoto-u.ac.jp

Yasushi Sakurai
NTT Communication Science Labs
yasushi.sakurai@acm.org

B. Aditya Prakash
Carnegie Mellon University
badityap@cs.cmu.edu

Lei Li
University of California, Berkeley
leili@cs.berkeley.edu

Christos Faloutsos
Carnegie Mellon University
christos@cs.cmu.edu

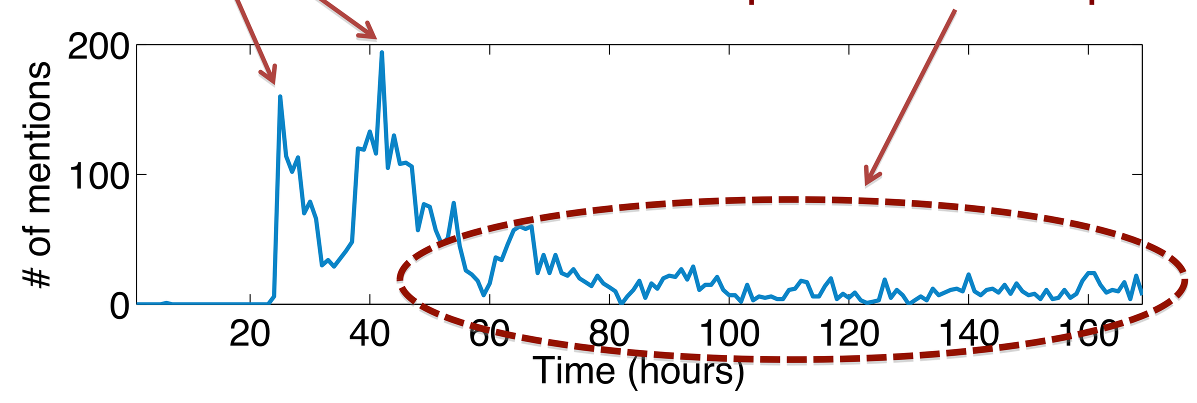
Motivation

Social media facilitates faster diffusion of news, rumors

Q. How do news & rumors spread in social media?

Rise-and-fall patterns
in information diffusion through online media

e.g., Meme (# of mentions in blogs) [KDD'09]



"lipstick on a pig" : Short phrases sourced from U.S. politics in 2008

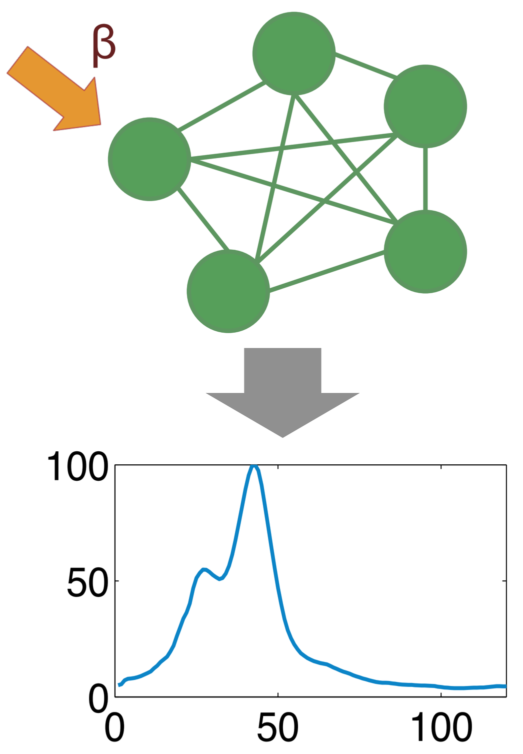
Q. Do the rise-fall patterns follow a simple law?

Problem definition

Goal: predict/model social activity

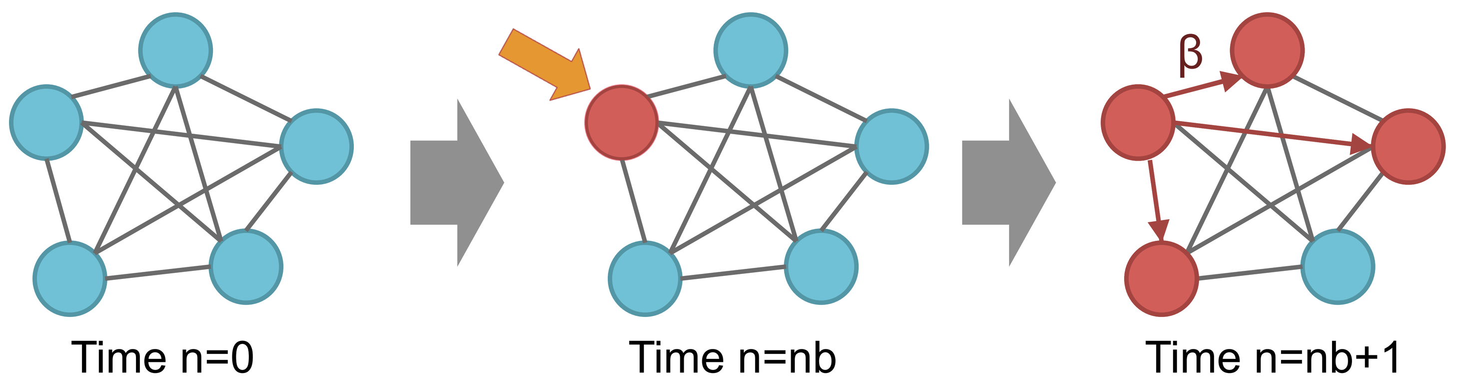
Given:
- Network of bloggers/users
- External shock/event
- Quality of the event β

Find:
- How blogging activity will evolve over time



Proposed method: SpikeM

Main idea:
1. Un-informed bloggers (uninformed about rumor)
2. External shock at time n_b (e.g, breaking news)
3. Infection (word-of-mouth)



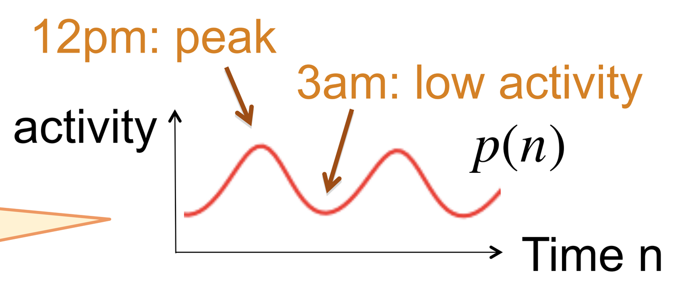
U – Un-informed node (bloggers)
B – informed, and Blogged about rumor

External shock
- S_b – Strength of external shock at birth
- n_b – Starting time of breaking news

Infectiveness of a blog-post at age n : $f(n) = \beta * n^{-1.5}$
- β – Strength of infection (quality of news)
- $f(n)$ – Decay function (how infective a blog posting is)

SpikeM – with periodicity

Bloggers change their activity (e.g., daily, weekly, yearly)



Code: http://www.kecl.ntt.co.jp/csl/sirg/people/yasuko/software.html

Equations of SpikeM

$$\Delta B(n+1) = p(n+1)\cdot\left[U(n)\cdot\sum_{t=n_b}^{n}(\Delta B(t)+S(t))\cdot f(n+1-t)+\varepsilon\right]$$

Blogged … count of informed bloggers at time **n**

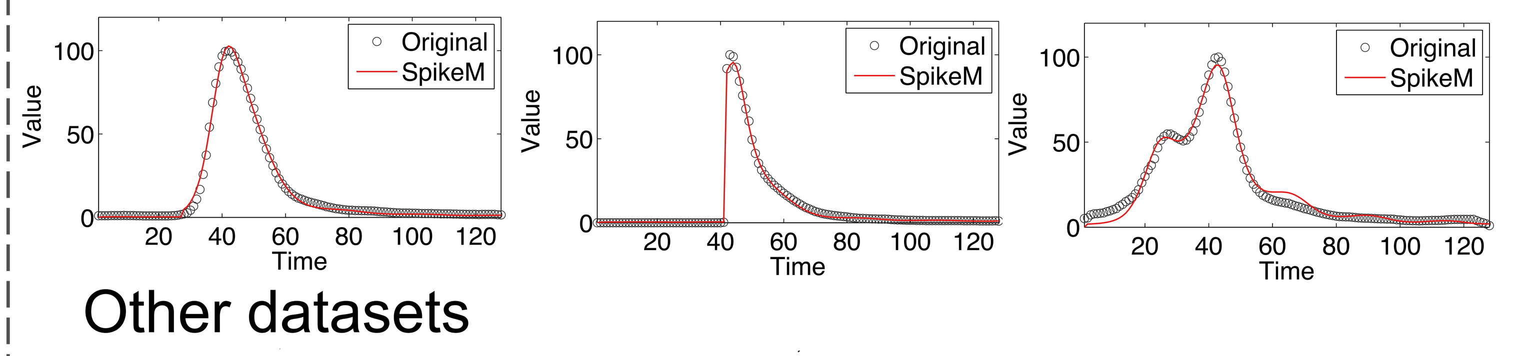
$$U(n+1) = U(n) - \Delta B(n+1)$$

Un-informed … count of un-informed bloggers

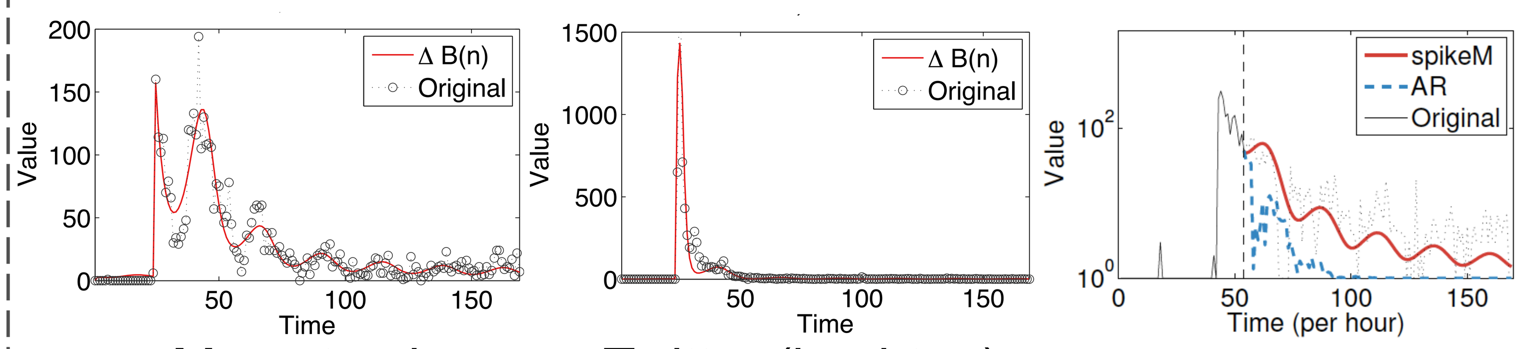
- N – Total population of available bloggers
- β – Strength of infection/news
- n_b, S_b – External shock S_b at birth (time n_b)
- $p(n)$ – Periodicity

Experiments – fitting results

K-SC [WSDM'11] … spikes of online media (memes)



Other datasets

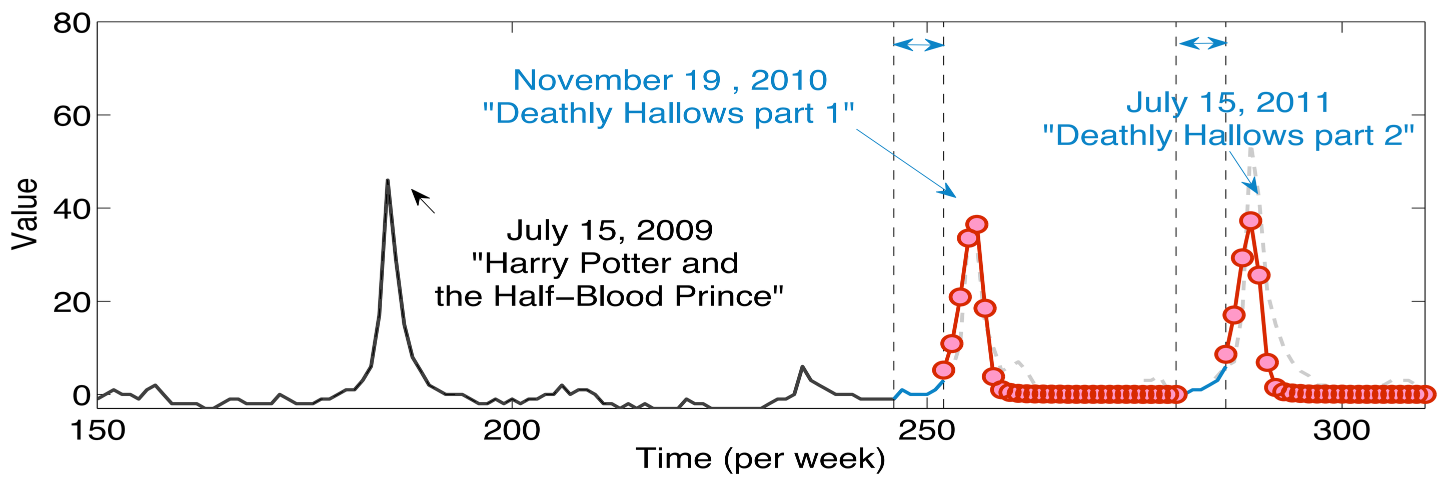


Memetracker "lipstick on a pig"   Twitter (hashtag) "#stevejobs"   **Tail-part forecasts** SpikeM vs. AR

SpikeM – at work

(a) "What-if" forecasting:

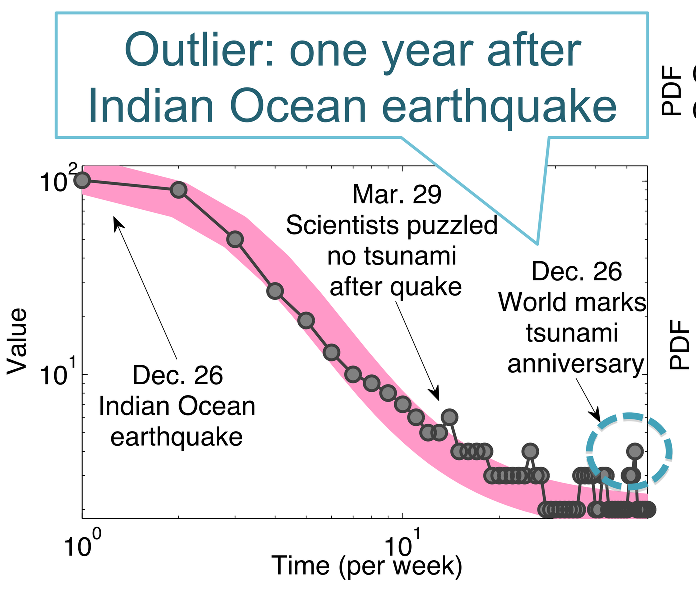
Given (1) first spike (2) release date (3) two weeks before release
Forecast: upcoming spikes (red)



"Harry potter" (Google trends)

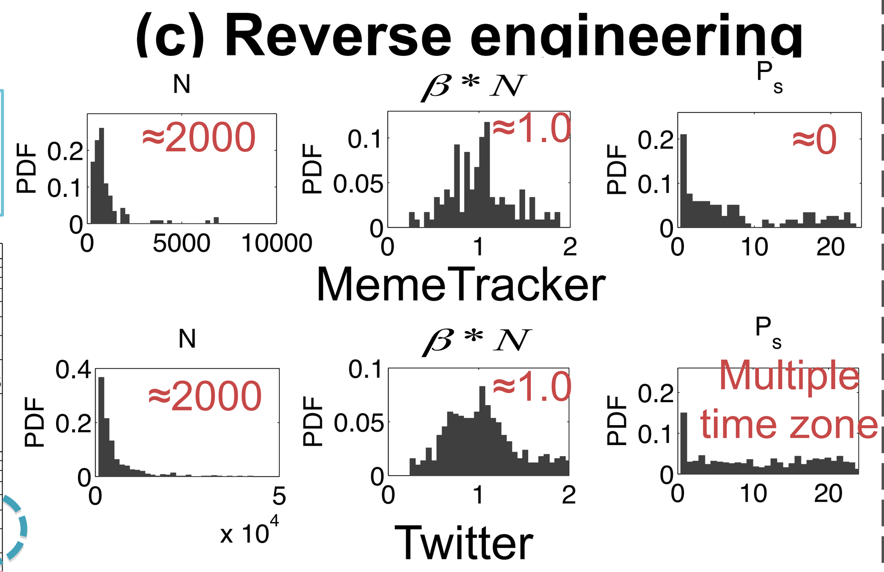
(b) Outlier detection

Outlier: one year after Indian Ocean earthquake



"Tsunami" (Google trends)

(c) Reverse engineering



PDF of SpikeM parameters over 1,000 memes/hash tags

Conclusions

SpikeM has following advantages:
- Unification power: includes earlier patterns/models
- Practicality: matches behavior of real datasets
- Parsimony: requires only 7 parameters
- Usefulness: answers what-if scenarios, spot outliers, forecasting, reverse engineering