Social Media Cocktail

Scott Hendrickson
Data Scientist, Gnip Inc.
@DrSkippy27

March 5, 2013

social data has unlimited value and near limitless application

why mix social media data?

<one>
audience, perspective,
coverage

Audience – Volume

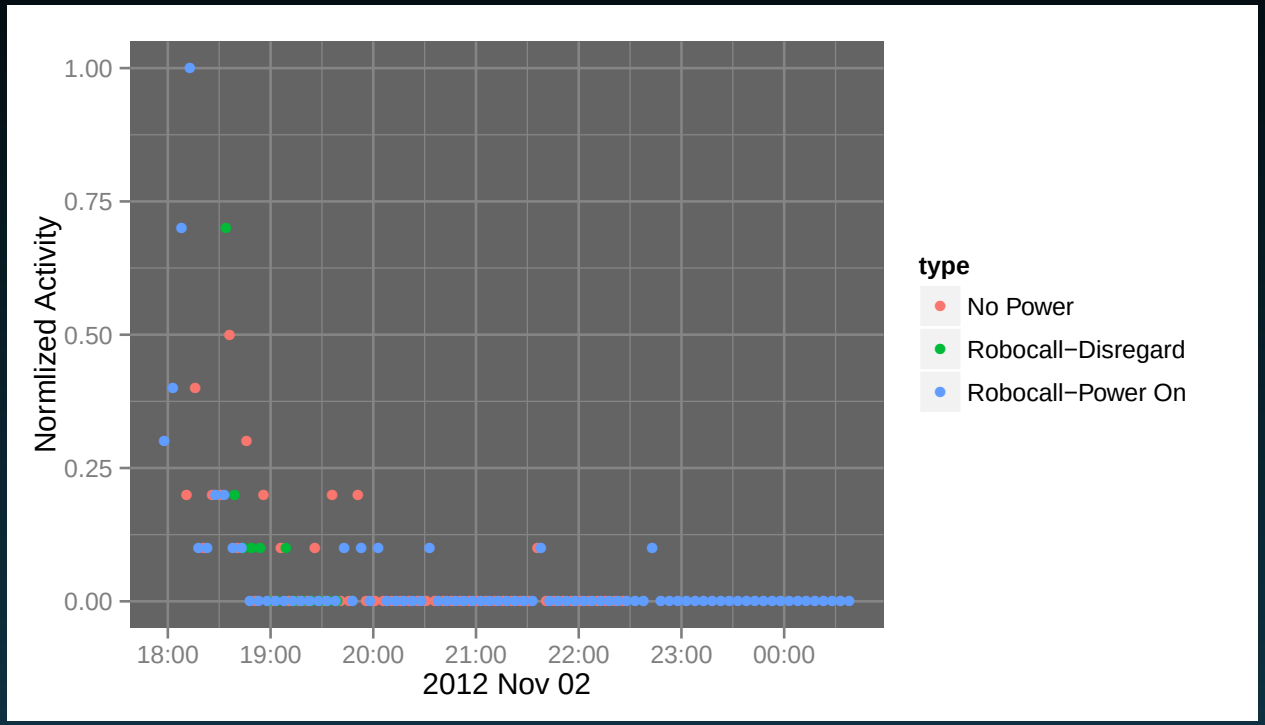
Publisher	Daily Activity
Twitter	400M
Tumblr	75M
Wordpress Posts	615k
Wordpress Comments	1.1M
Disqus	1.3M
Engagement (likes, votes)	2.4M

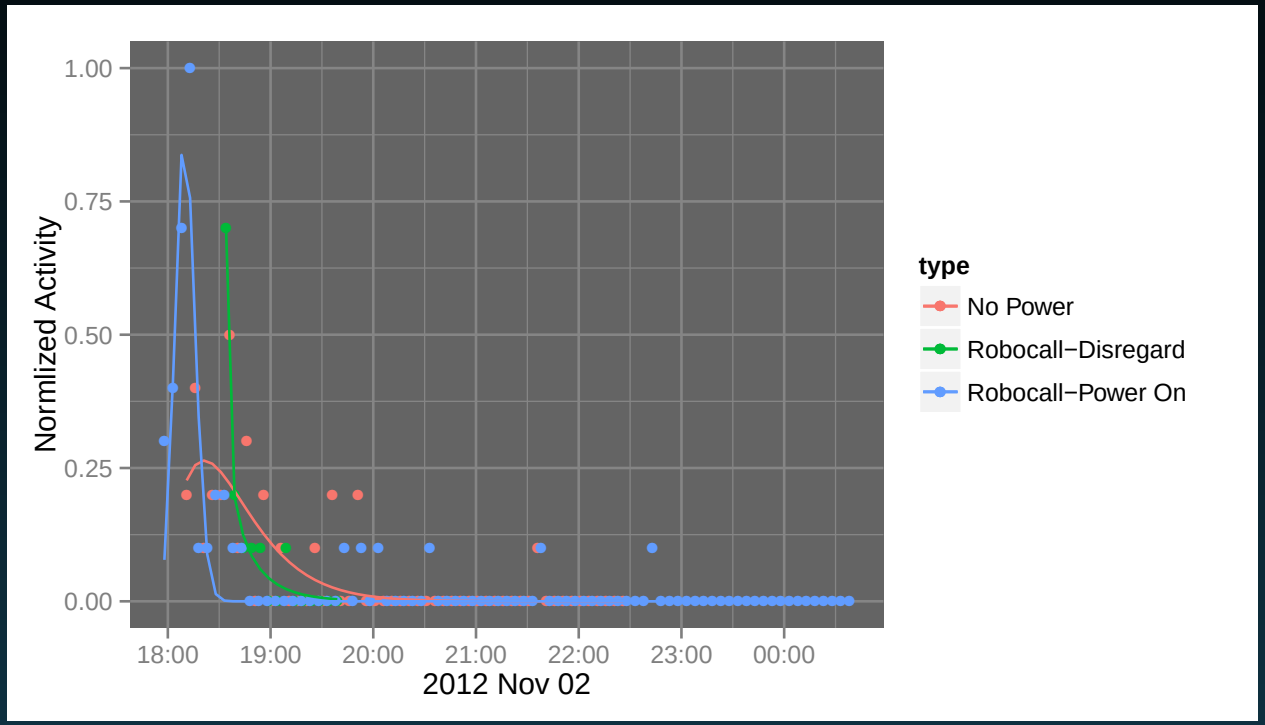
Gnip

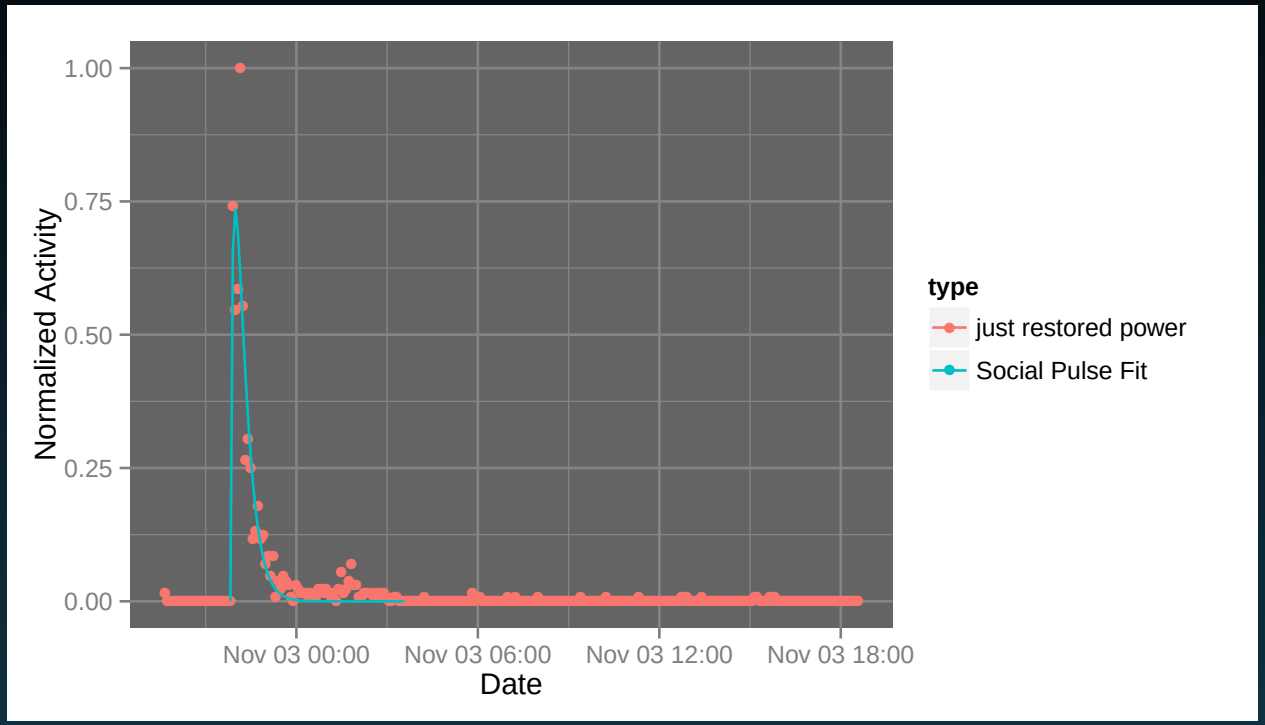
- 4,600 Tweets/second
- 1/2M unique Tumblr users/hour
- PowerTrack filtering on data and metadata, PowerTrack Replay, Historical…

3B+ activities/day

signal or noise?

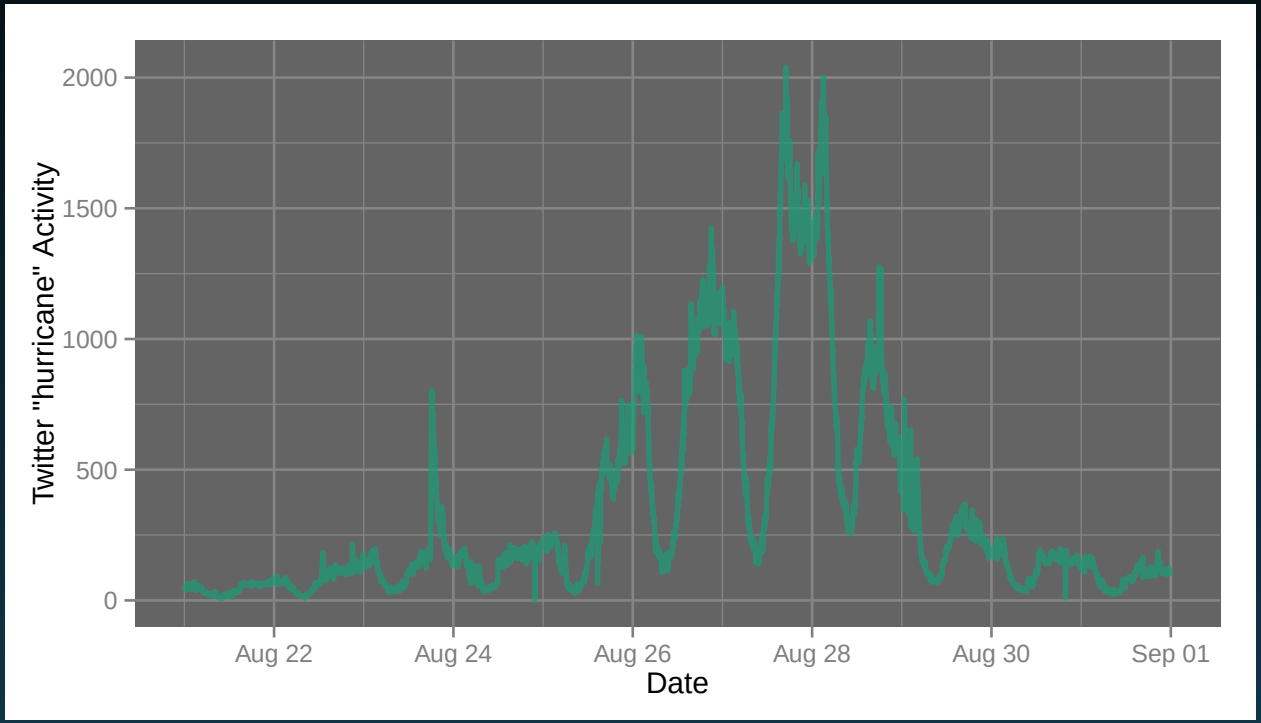

1.00
0.75
0.50
0.25
0.00
Normlized Activity
18:00
19:00
20:00
21:00
22:00
23:00
00:00
2012 Nov 02
type
No Power
Robocall−Disregard
Robocall−Power On


1.00
0.75
0.50
0.25
0.00
Normlized Activity
18:00
19:00
20:00
21:00
22:00
23:00
00:00
2012 Nov 02
type
No Power
Robocall−Disregard
Robocall−Power On

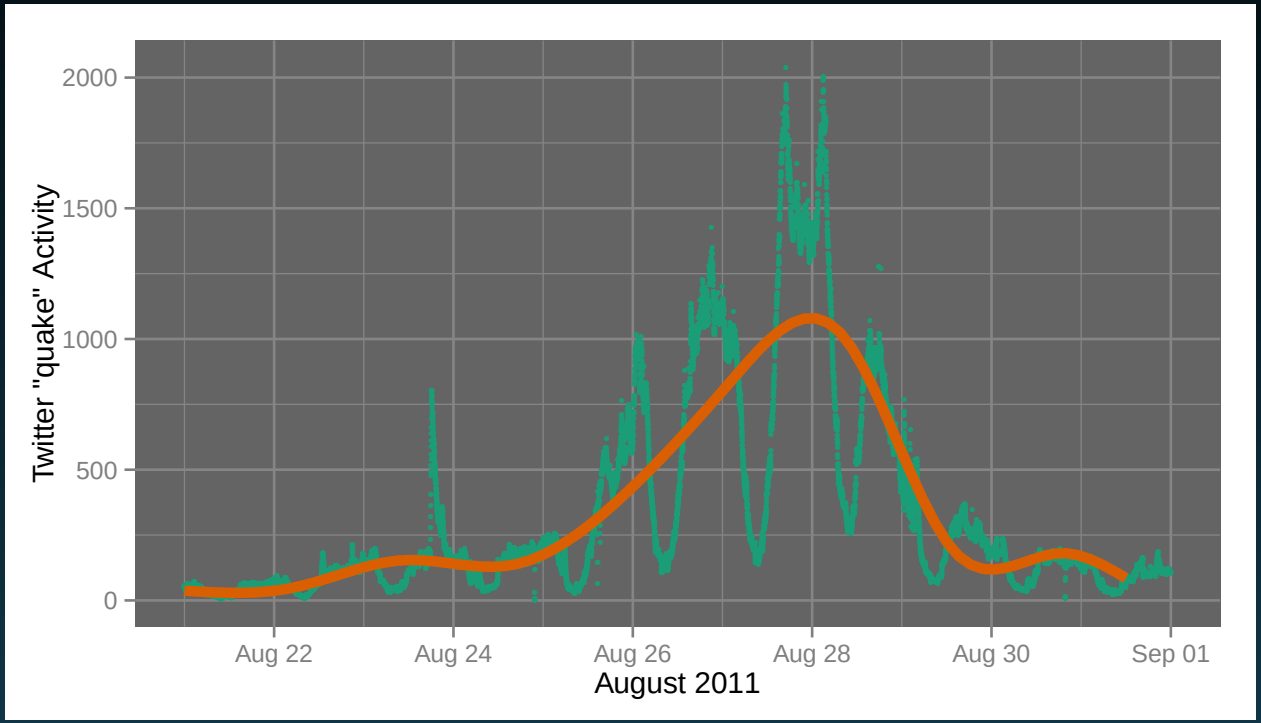

1.00
0.75
0.50
0.25
0.00
Normalized Activity
Nov 03 00:00
Nov 03 06:00
Nov 03 12:00
Nov 03 18:00
Date
type
just restored power
Social Pulse Fit

<two>
timing, evolution

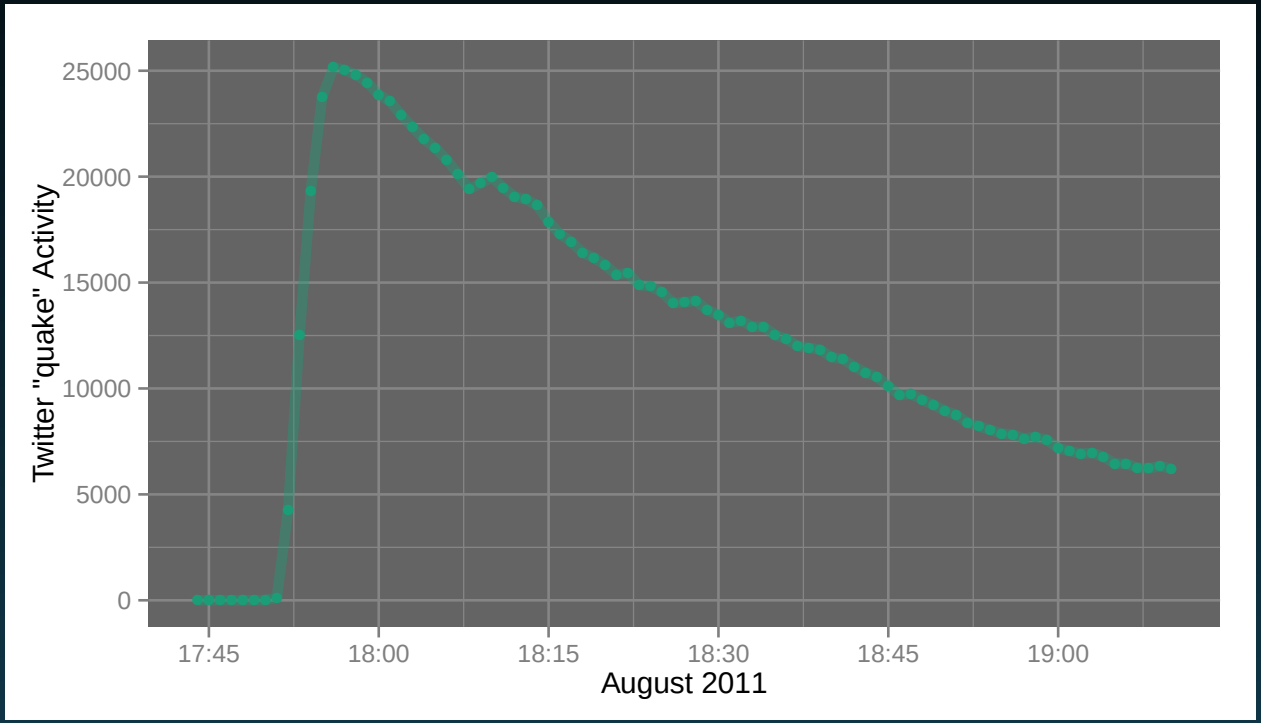
Expected: Hurricane



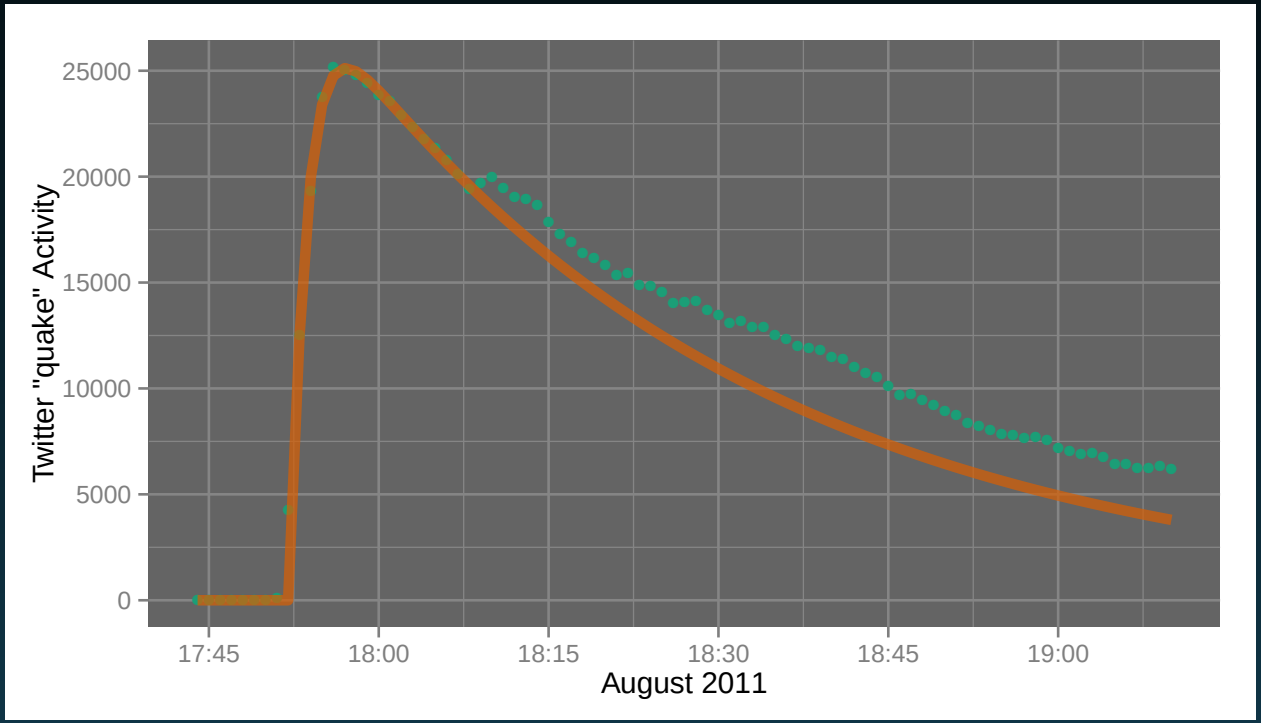
Expected: Hurricane



Unexpected: Earthquake



Unexpected: Earthquake



Classifying Events

Type	Response	Examples
Expected	Approx. Symmetric	Hurricane Sandy Olympics
Unexpected (many obs.)	Social Media Pulse	Beyonce' VMAs Mexico earthquake Steve Jobs
Unexpected (spread)	Network Models	Osama Bin Laden Whitney Houston Syrian dissidents

Social Media Pulse Half-life

time to observe
half of the activities

Social media pulse

Probability of an activity from one person,

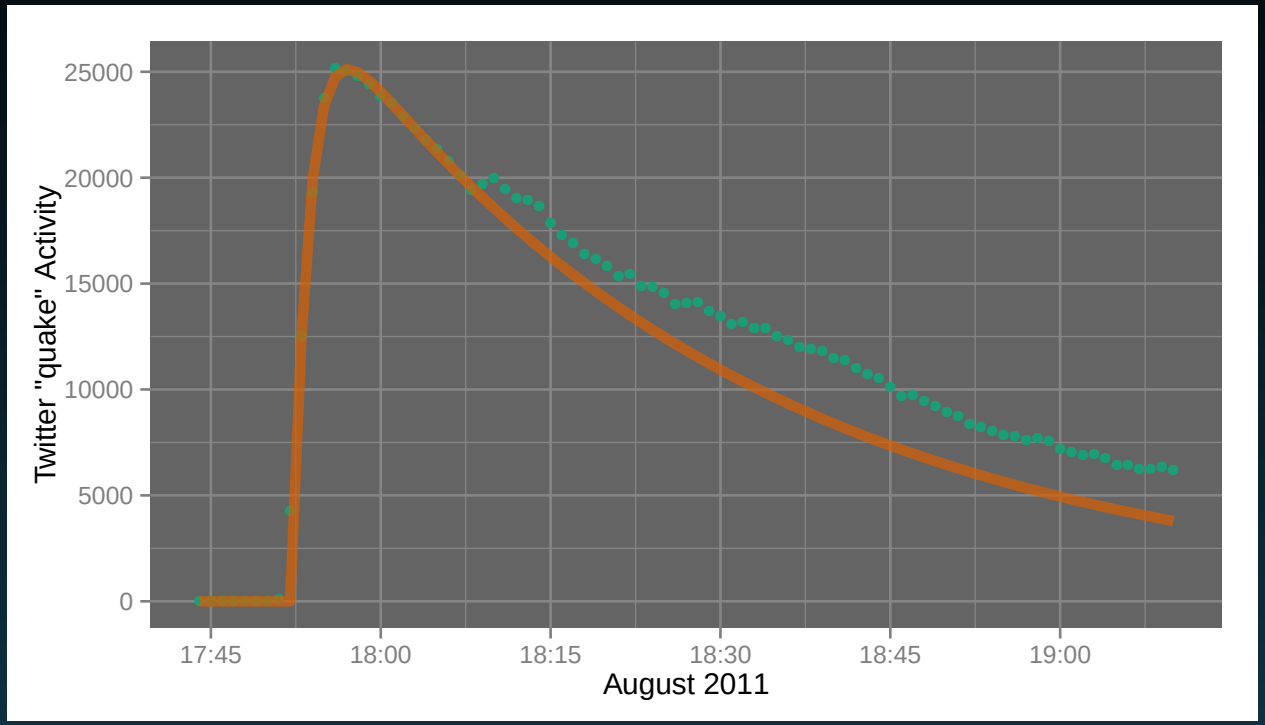
$$f(t) = \lambda \exp(-\lambda t), \text{ for } t \geq 0.$$

Many people, so sum random variables $S = X_1 + \ldots + X_n$.
Probability distribution function,

$$f_S(t) = \frac{\beta^{-\alpha} t^{\alpha-1} \exp(\frac{-t}{\beta})}{\Gamma(\alpha)}$$

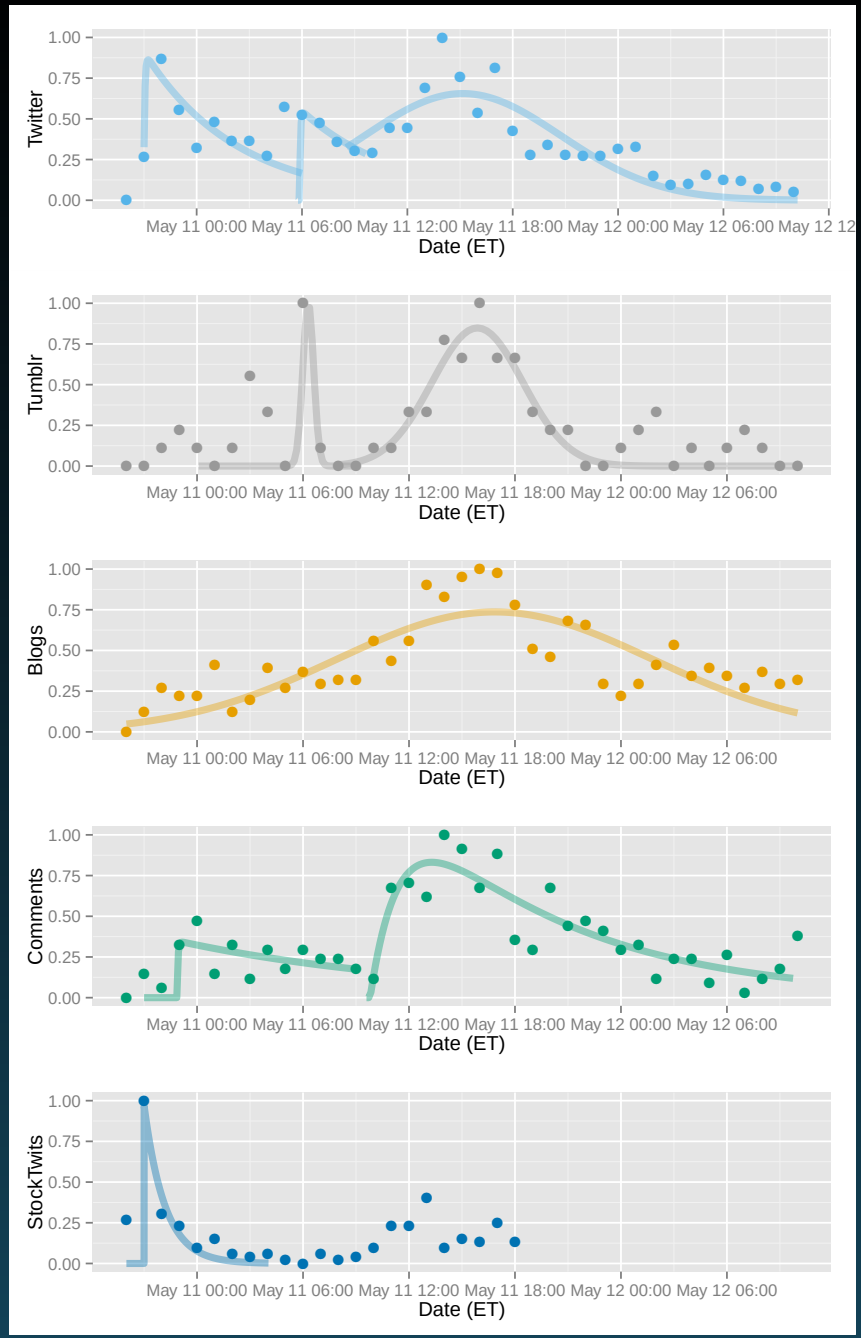
Why model half-life?

- predict total story volume
- compare half-lives
- anomalous story evolution


25000
20000
15000
10000
5000
0
Twitter "quake" Activity
17:45
18:00
18:15
18:30
18:45
19:00
August 2011

Story Timing

Publisher	Speed
Twitter	Fast
Tumblr	Fast and Slow
Wordpress Posts	Fast and Medium
Wordpress Comments	Fast
Disqus	Fast
Engagement (likes, votes)	Fast

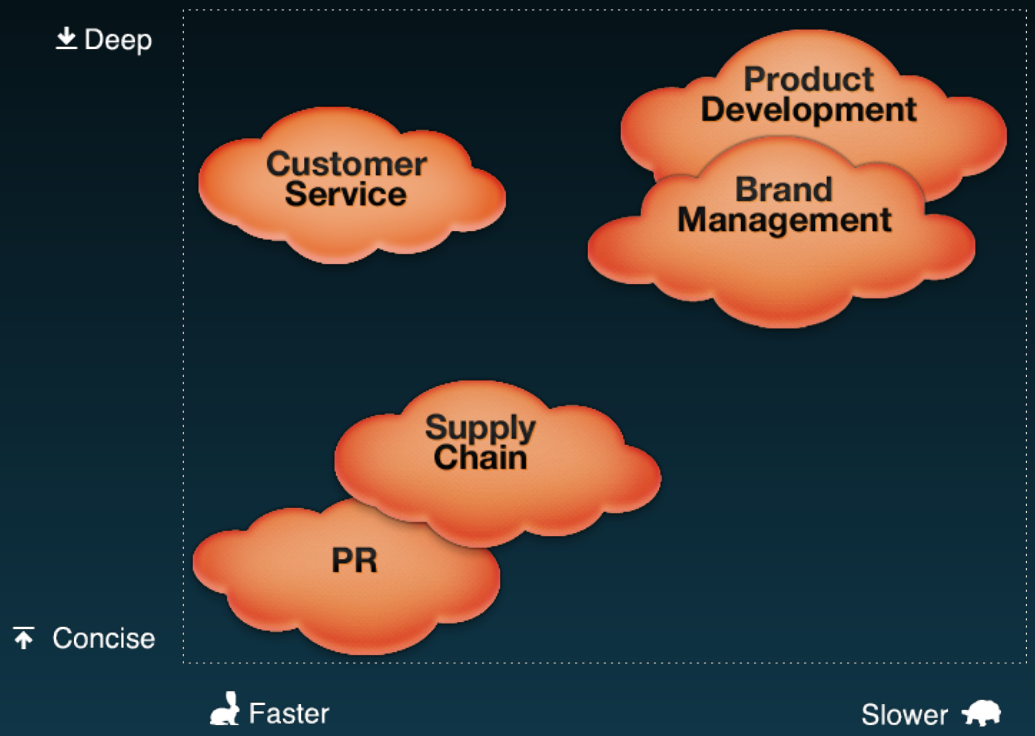

Twitter
1.00
0.75
0.50
0.25
0.00
May 11 00:00 May 11 06:00 May 11 12:00 May 11 18:00 May 12 00:00 May 12 06:00 May 12 12
Date (ET)
Tumblr
1.00
0.75
0.50
0.25
0.00
May 11 00:00 May 11 06:00 May 11 12:00 May 11 18:00 May 12 00:00 May 12 06:00
Date (ET)
Blogs
1.00
0.75
0.50
0.25
0.00
May 11 00:00 May 11 06:00 May 11 12:00 May 11 18:00 May 12 00:00 May 12 06:00
Date (ET)
Comments
1.00
0.75
0.50
0.25
0.00
May 11 00:00 May 11 06:00 May 11 12:00 May 11 18:00 May 12 00:00 May 12 06:00
Date (ET)
StockTwits
1.00
0.75
0.50
0.25
0.00
May 11 00:00 May 11 06:00 May 11 12:00 May 11 18:00 May 12 00:00 May 12 06:00
Date (ET)

<three>
content richness

Speed and Richness

Publisher	Speed	Richness
Twitter	Fast	Concise
Tumblr	Fast, Slow	Rich, multimedia
Wordpress Posts	Fast, Medium	Rich, text
Wordpress Comments	Fast	Reactive, small-to-medium
Disqus	Fast	Reactive, small-to-medium
Engagement	Fast	Terse

Social Cocktail



Thank you!

Presentation, data, code at:
github.com/DrSkippy27/SBS2013