The Social Cocktail

Scott Hendrickson
Data Scientist, Gnip

@gnip
@DrSkippy27

March 6, 2013

Social data has unlimited value and near limitless application

Why mix social 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

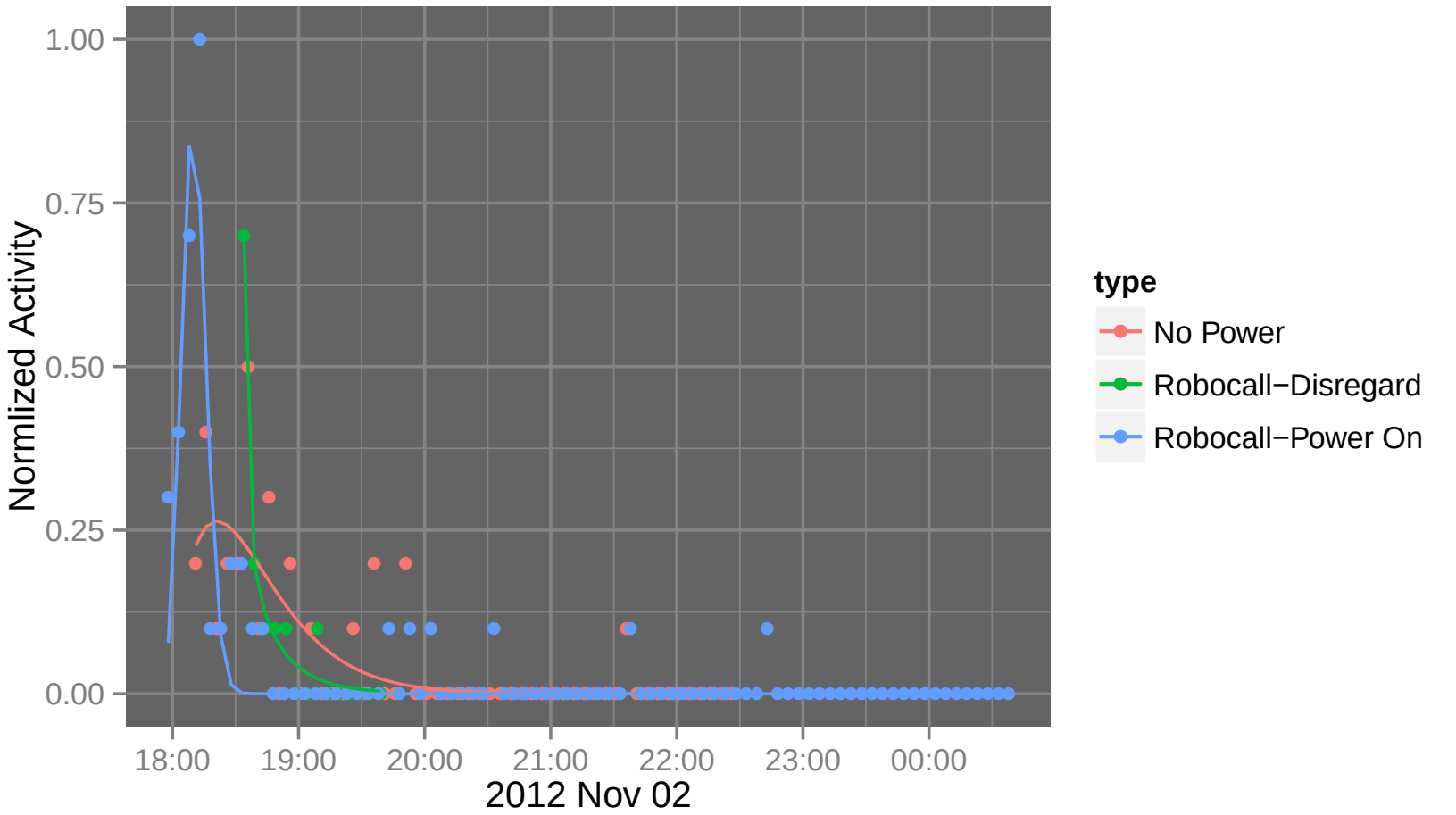
Gnip

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

3B+ activities/day

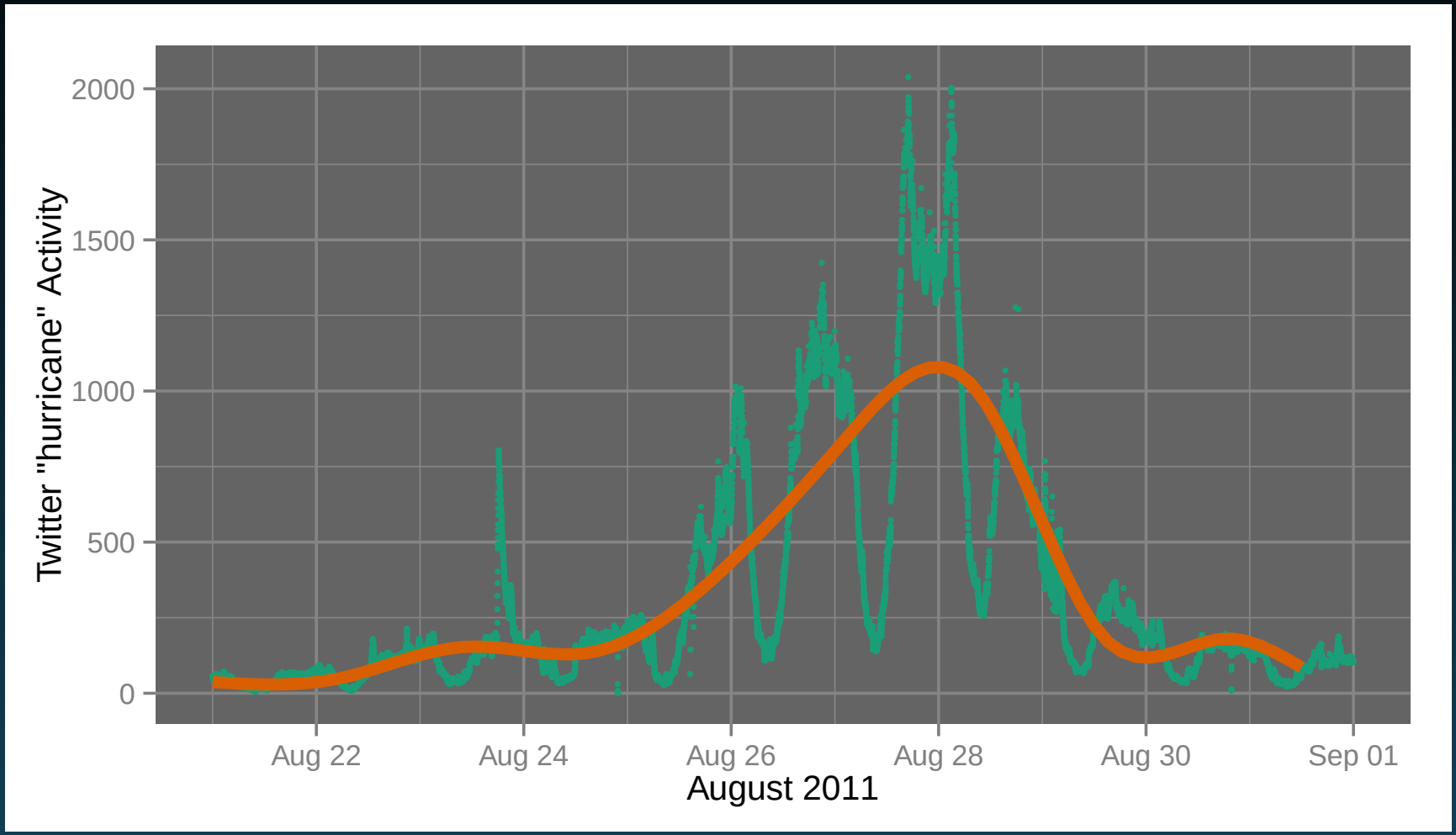
Signal or noise?

Con Edison Chelsea

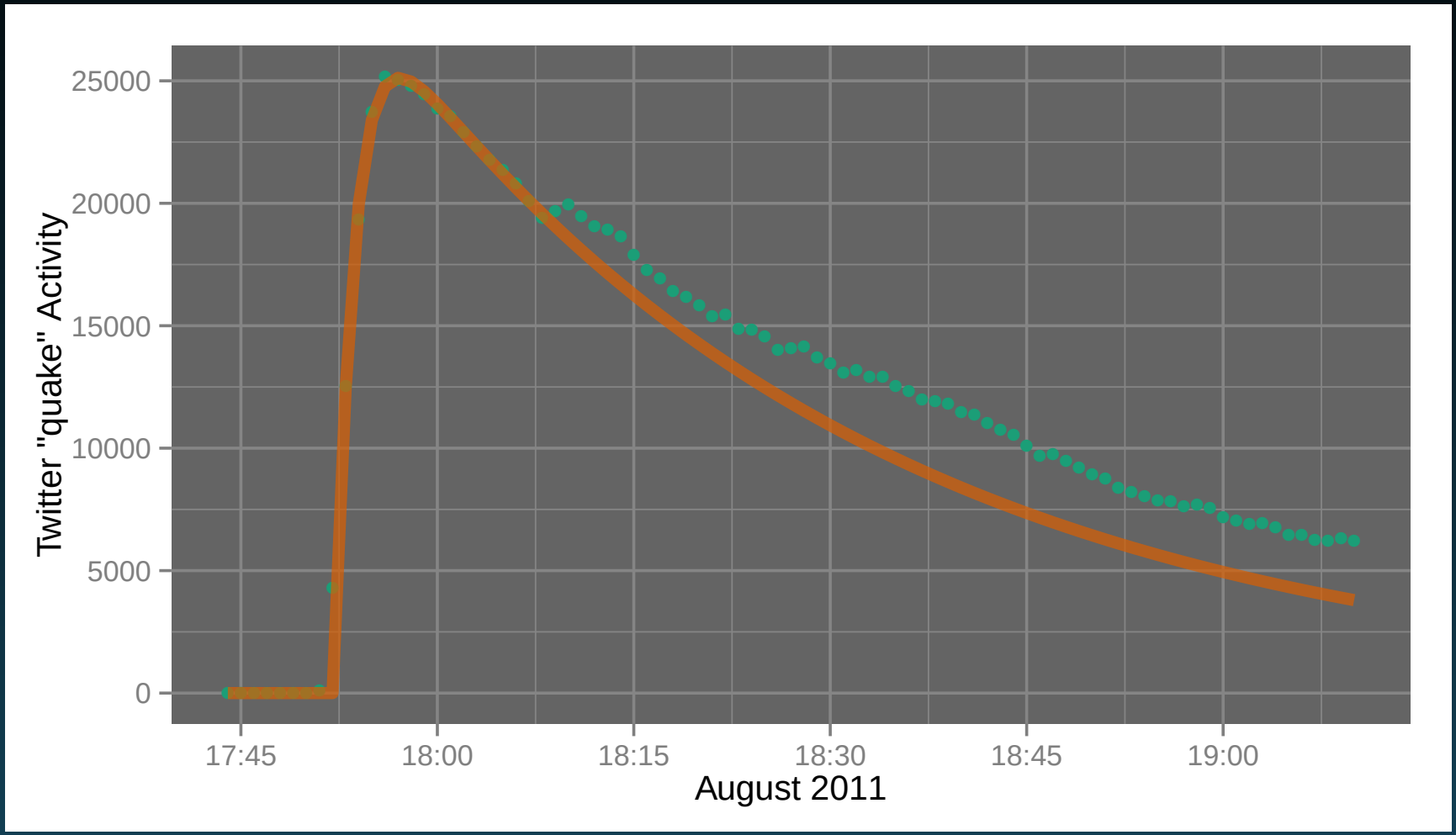


<Two>
Timing, evolution

Expected: Hurricane



Unexpected: Earthquake



Classifying Events

Type	Response	Examples
Expected	Approx. Symmetric	Hurricane Sandy Olympics
Unexpected (many obs.)	Social Media Pulse	Beyoncé 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,

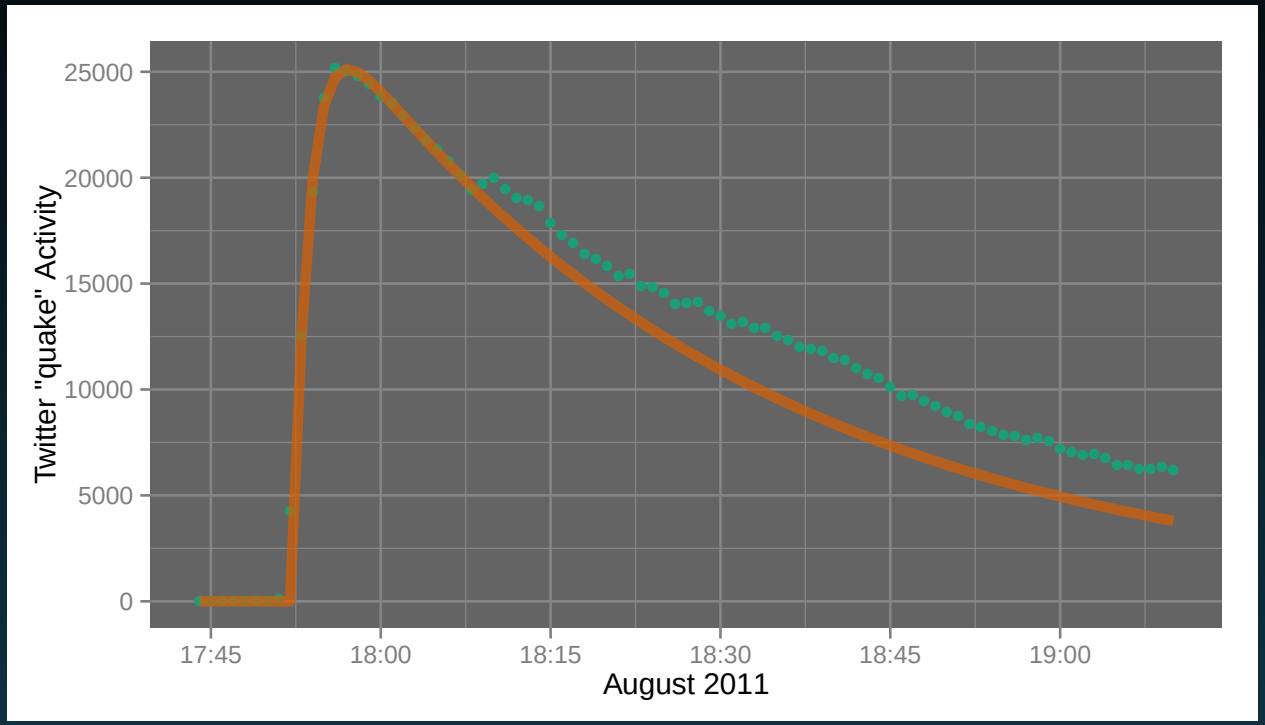
$$f(t) = \lambda \exp(-\lambda t), \text{ for } t \geq 0.$$

Sum random variables $S = X_1 + \ldots + X_n$ gives probability distribution (PDF),

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

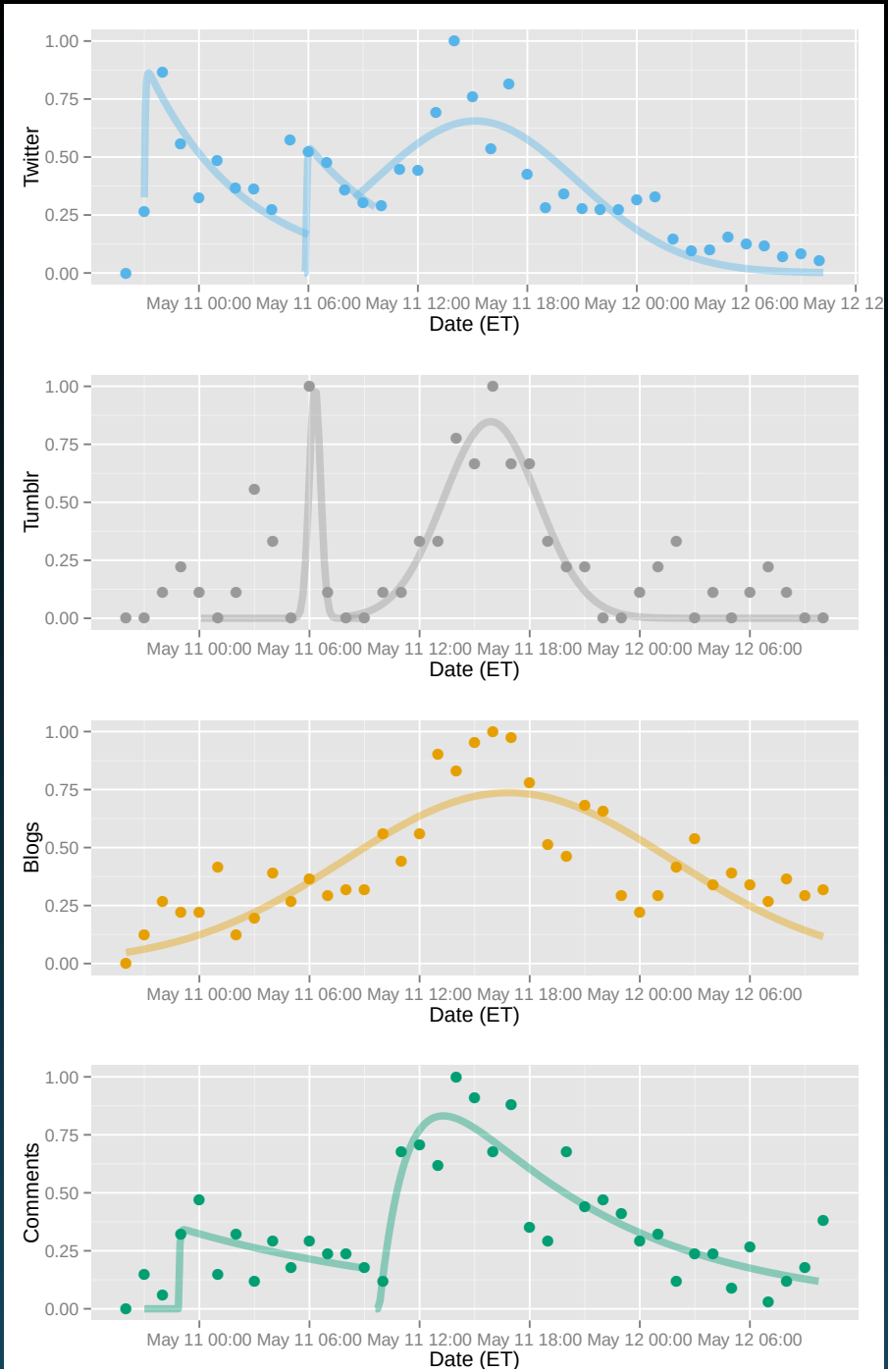
Why model half-life?

- 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	Medium
WordPress Posts	Deliberate
WordPress Comments	Medium
Disqus	Medium
Engagement (likes, votes)	Medium

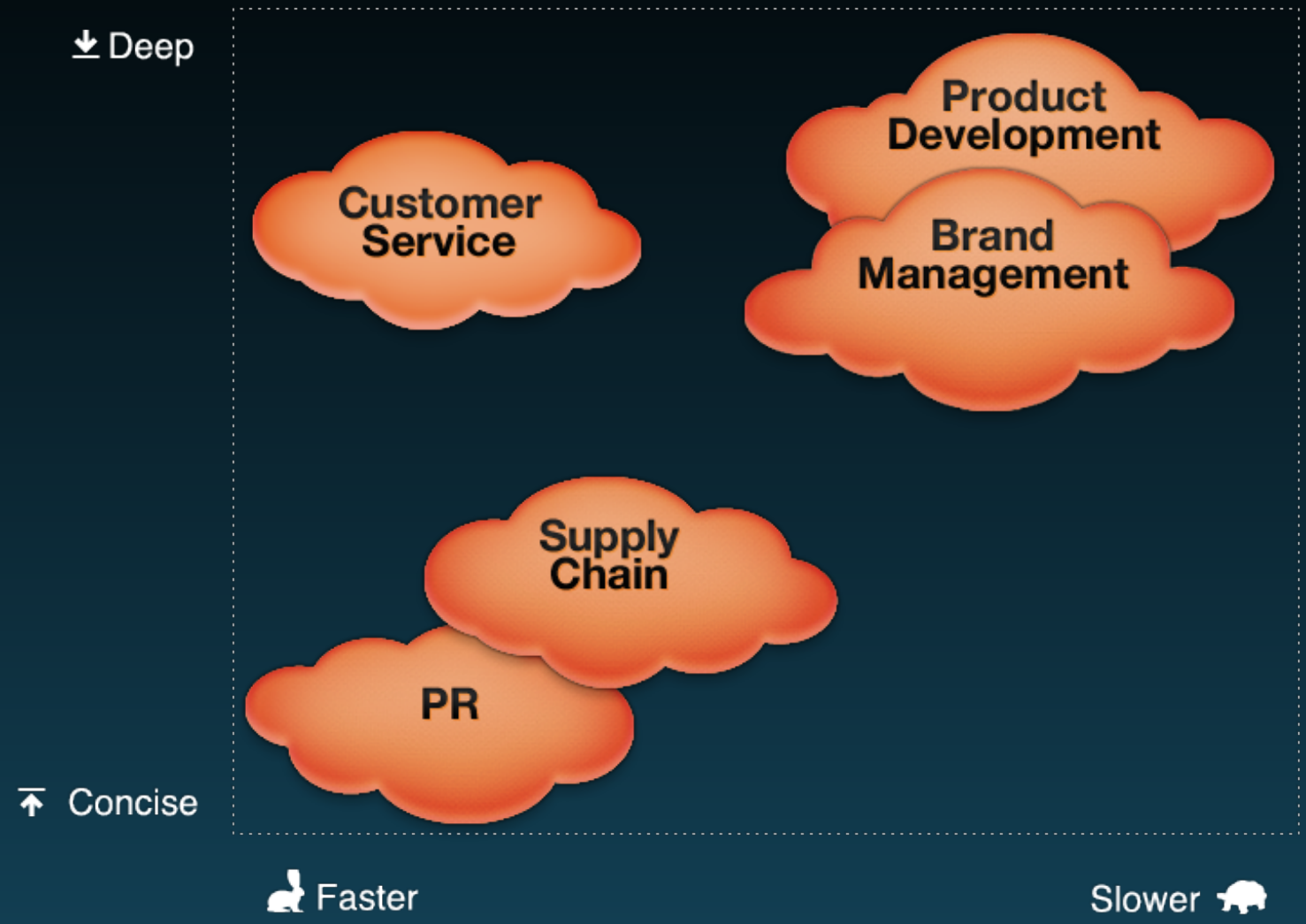

Twitter
1.00
0.75
0.50
0.25
0.00
May 11 00:00 May 11 06:00 May 11 12:00 May 11 18:00 May 12 00:00 May 12 06:00 May 12 12
Date (ET)
Tumblr
1.00
0.75
0.50
0.25
0.00
May 11 00:00 May 11 06:00 May 11 12:00 May 11 18:00 May 12 00:00 May 12 06:00
Date (ET)
Blogs
1.00
0.75
0.50
0.25
0.00
May 11 00:00 May 11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

<Three>
Content richness

Speed and Richness

Publisher	Speed	Richness
Twitter	Fast	Concise text, links
Tumblr	Medium	Multimedia, text, reblogging
WordPress	Deliberate	Rich text, multimedia
Comments	Medium	Small, medium text, threading
Engagement and Votes	Medium	Specific values

Social Cocktail



Thank you!

Follow us @gnip

Presentation at:
github.com/DrSkippy27/SBS2013