Social Media Cocktail

Scott Hendrickson
Data Scientist
Gnip Inc.
@DrSkippy27

March 1, 2013

Getting the right mix for data-driven social marketing

why social media for breaking stories?

1. audience, perspective and coverage

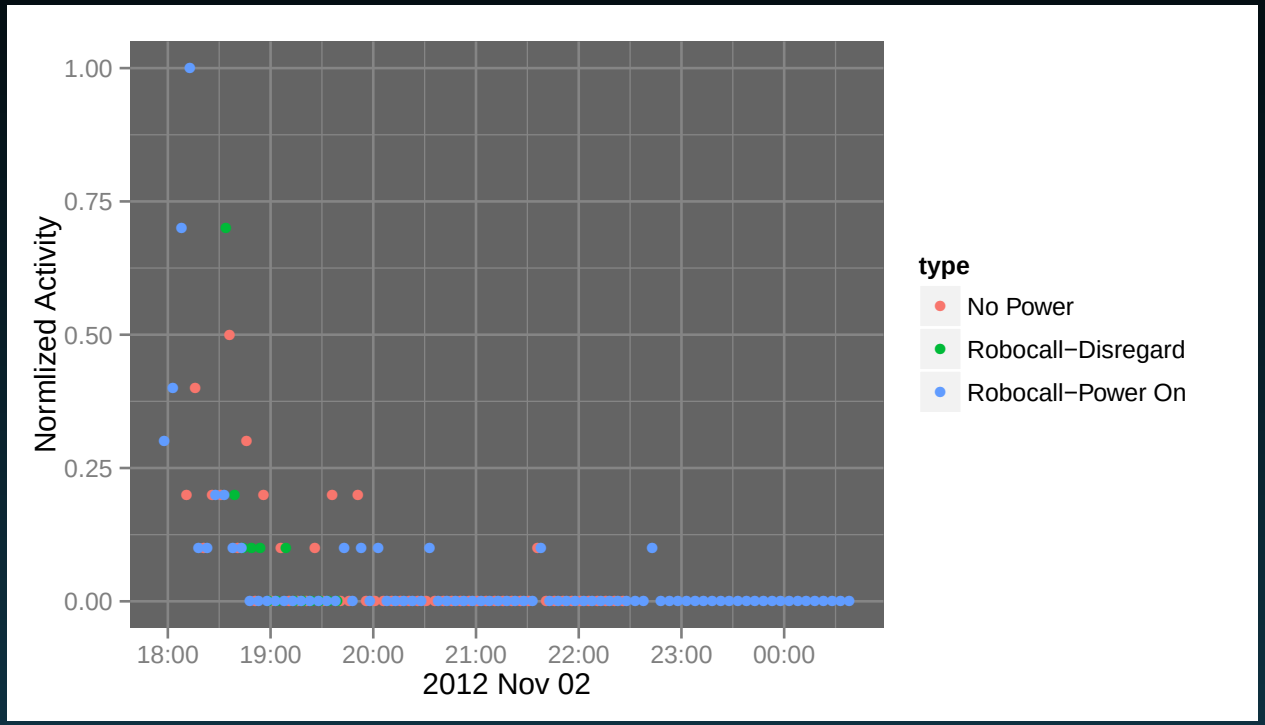
2. speed

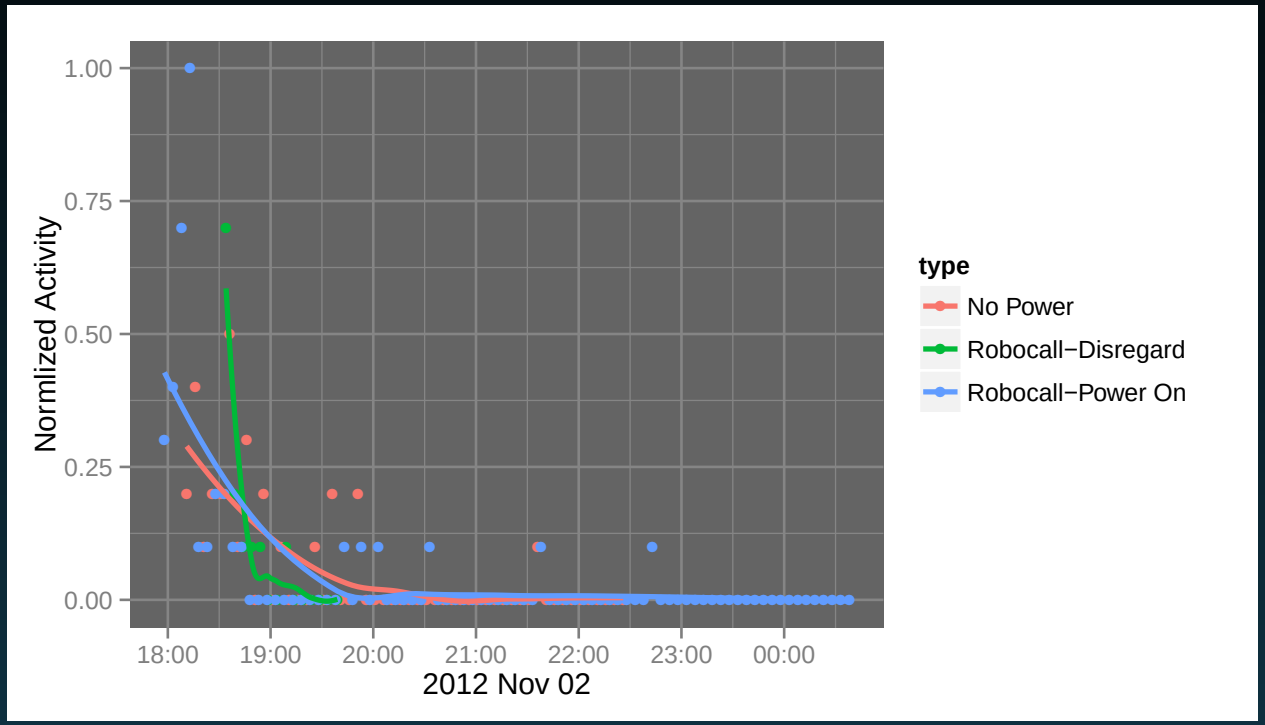
3. richness, diversity

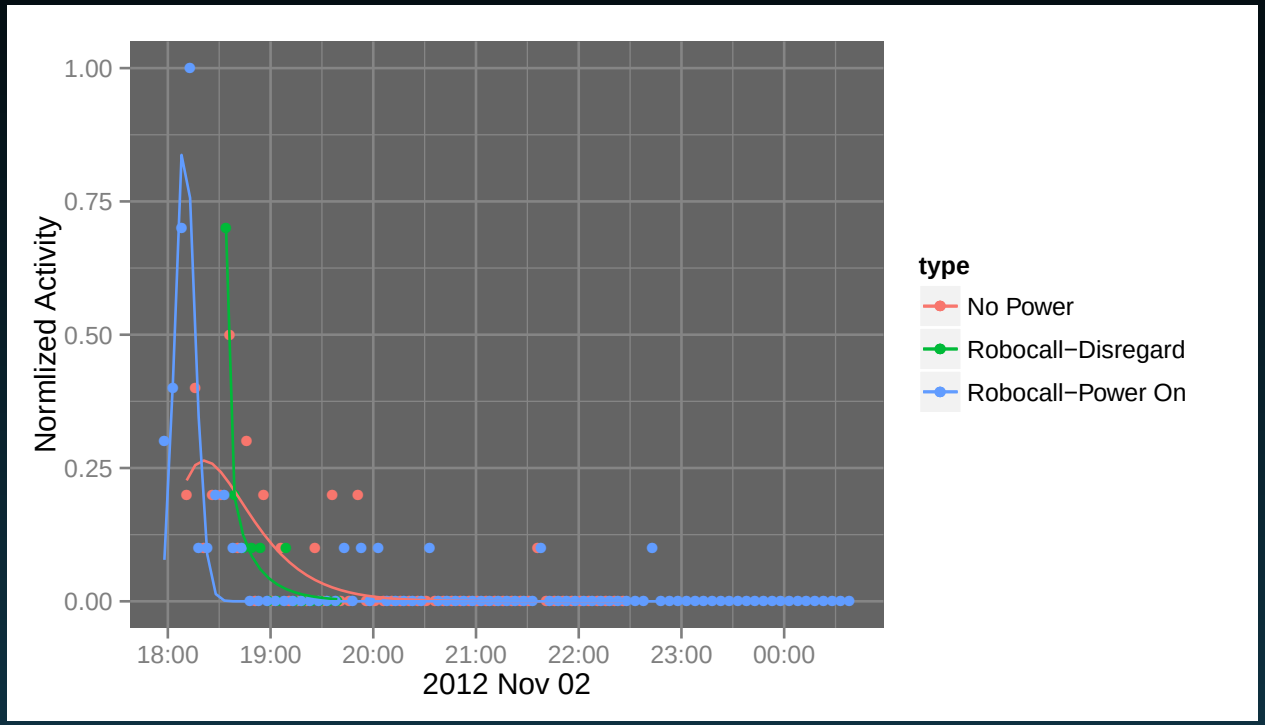
many publishers: audience, perspective and coverage

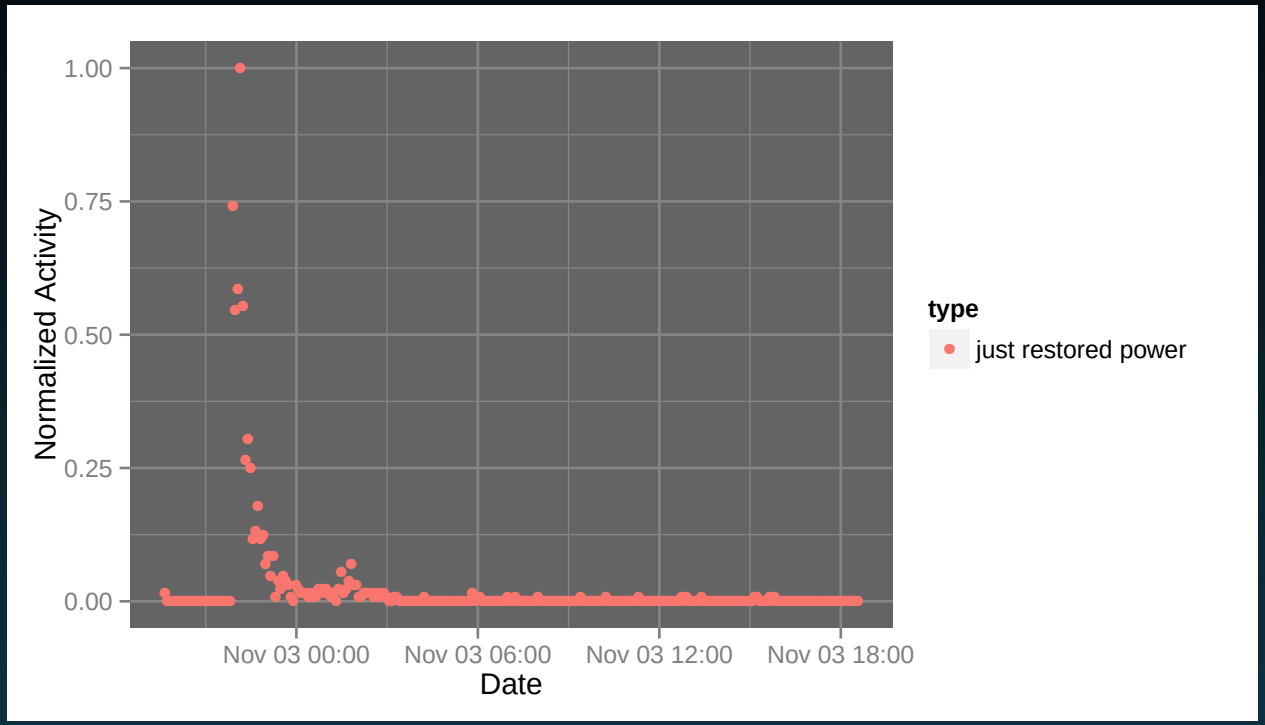
Publisher	Daily Activity
Twitter	400M
Tumblr	75M
Wordpress Posts	615k
Wordpress Comments	1.1M
Disqus	1.3M
Engagement (likes, votes)	2.4M

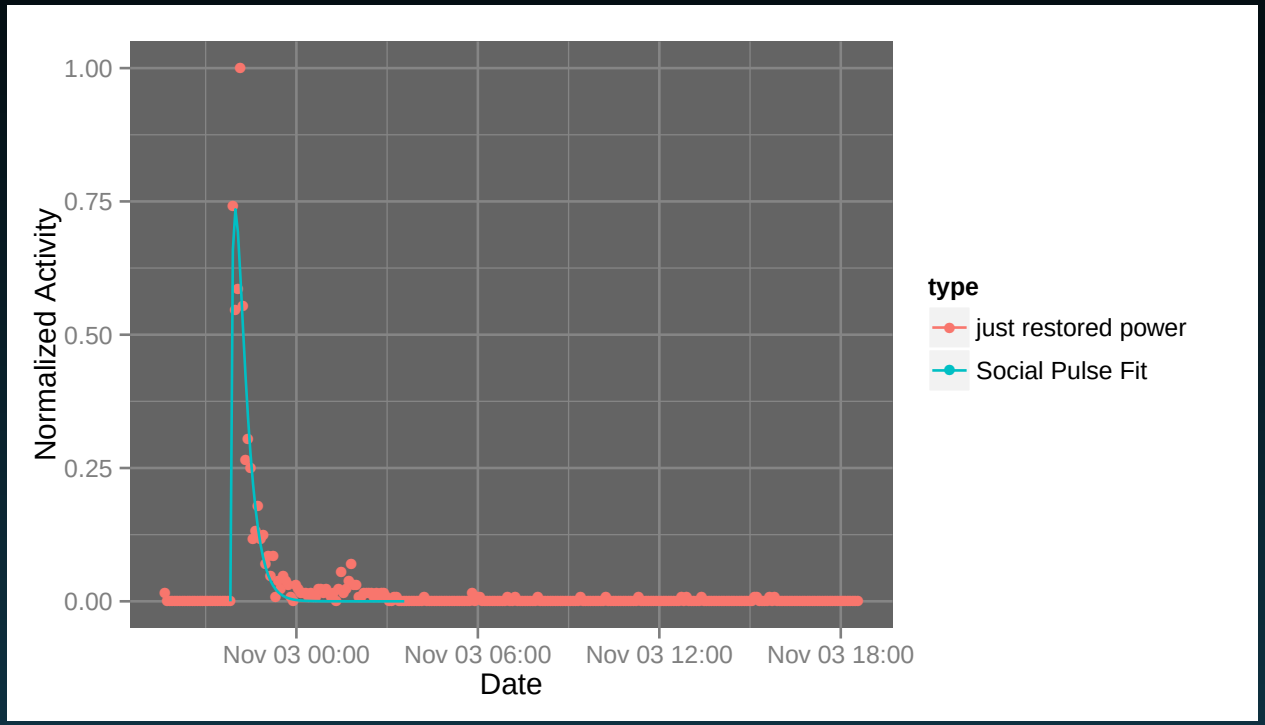
noise or signal?


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
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


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

realtime firehose: speed

Events

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

Expected: Hurricane

Unexpected: Earthquake

Half-life

time to observe
$\frac{1}{2}$ of the activities
for an event

Social media pulse

Given an event, the probability of a activity from one person,

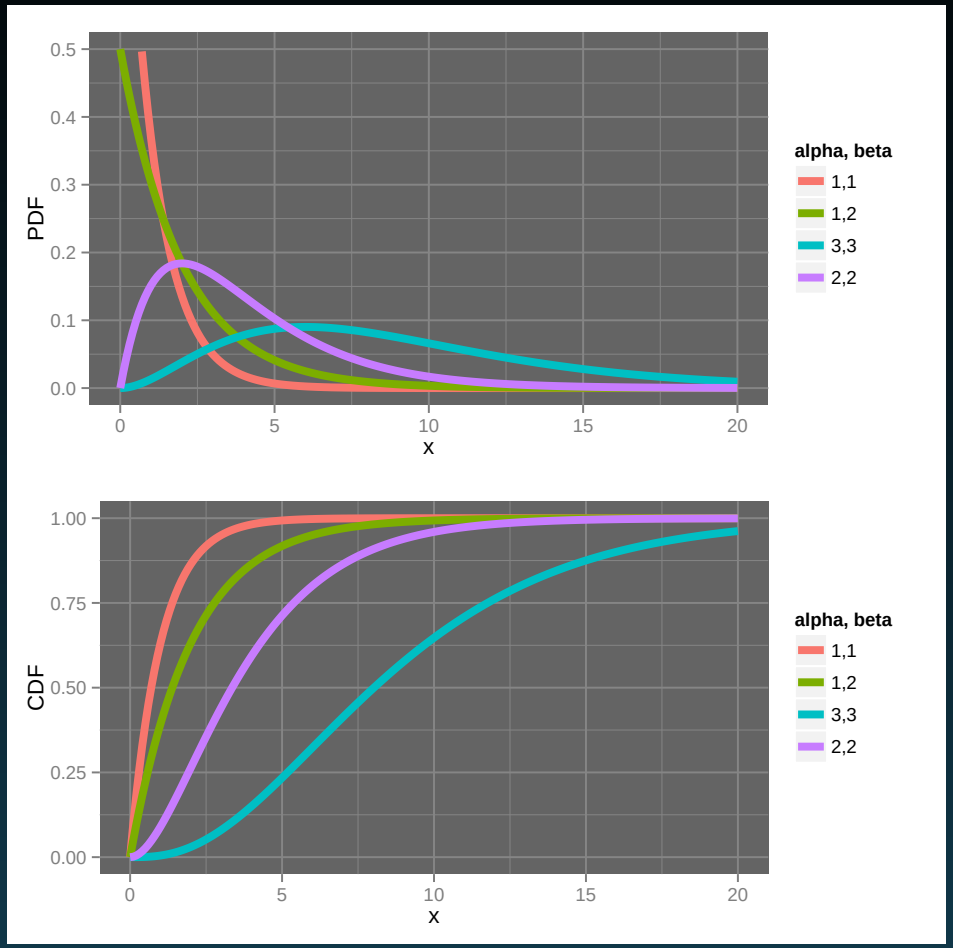
$$f(t) = \lambda \exp(-\lambda t), \text{ for } t \geq 0.$$

Many people posting, so sum of random variables
$S = X_1 + X_2 + \ldots + X_{n \text{ posters}}$.
Probability distribution function,

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

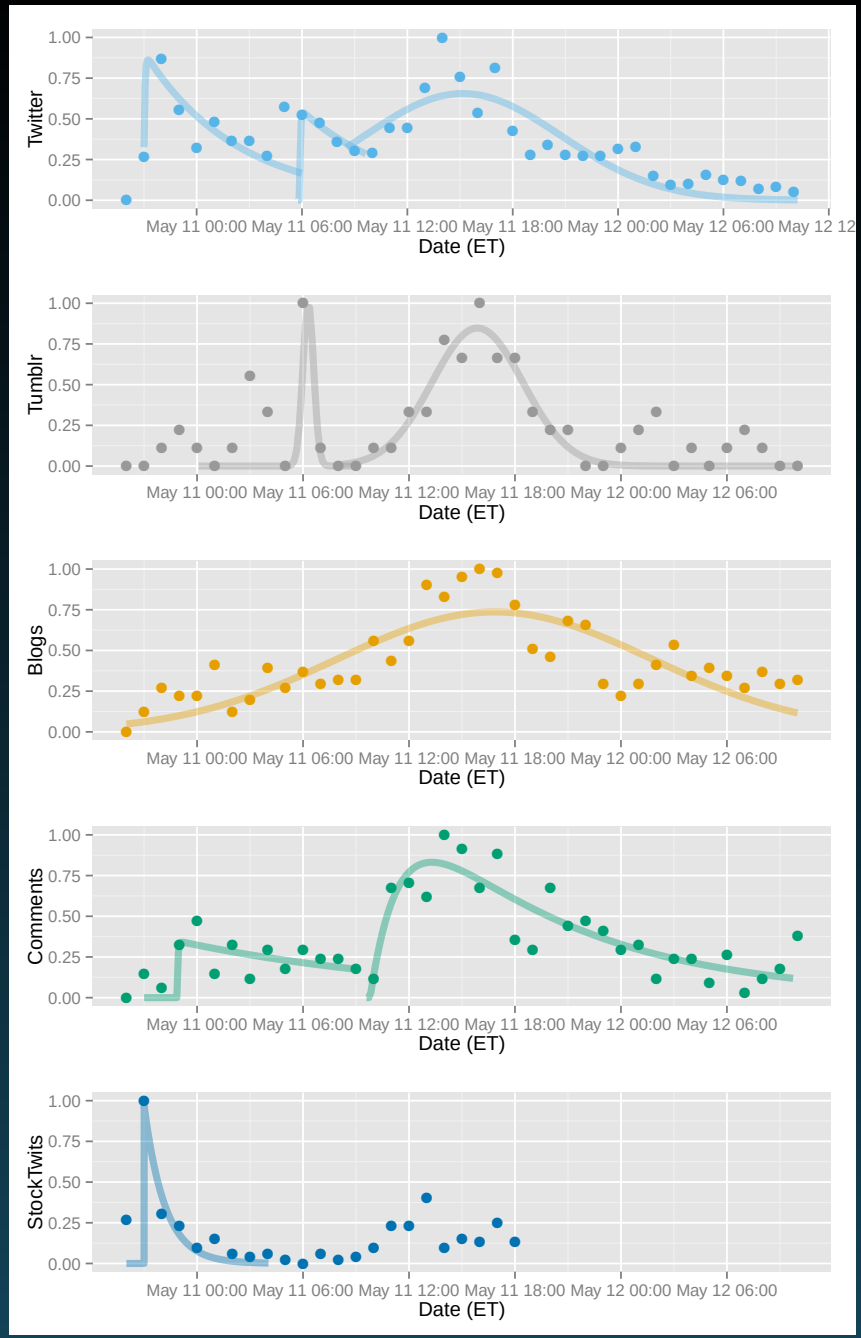
Cumulative distribution is the “generalized regularized incomplete gamma function”,

$$F_S(t) = Q(\alpha, 0, \frac{t}{\beta})$$


PDF
0.5
0.4
0.3
0.2
0.1
0.0
0
5
10
15
20
x
alpha, beta
1,1
1,2
3,3
2,2
CDF
1.00
0.75
0.50
0.25
0.00
0
5
10
15
20
x
alpha, beta
1,1
1,2
3,3
2,2

Publishers

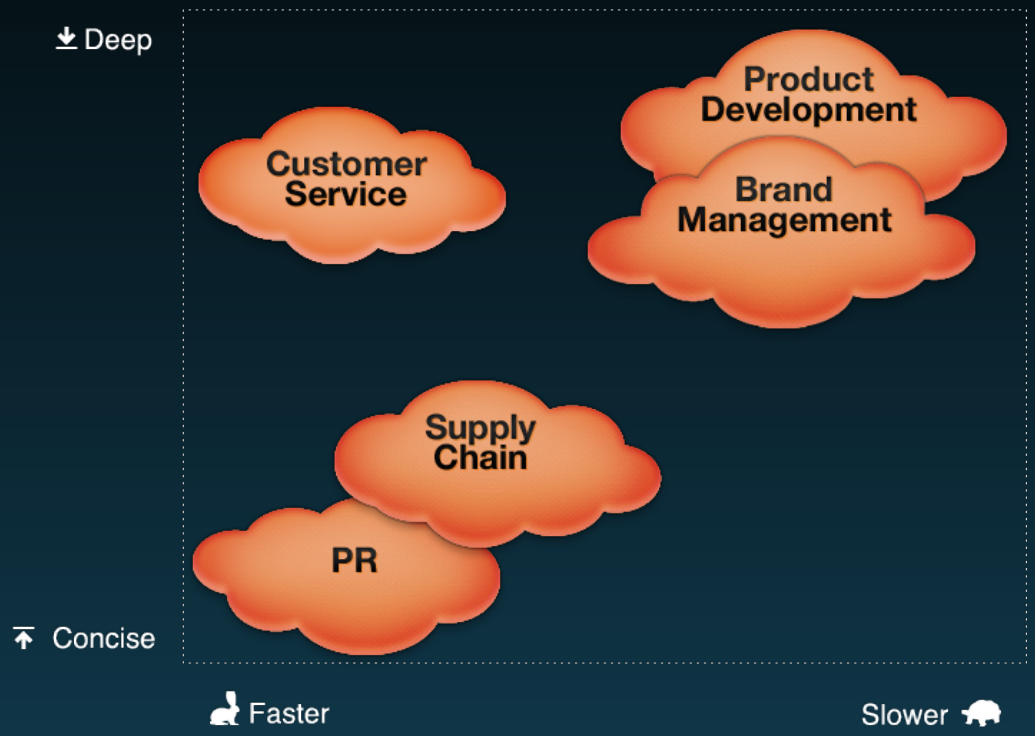
Publisher	Speed
Twitter	Fast
Tumblr	Fast and Slow
Wordpress Posts	Fast and Medium
Wordpress Comments	Fast
Disqus	Fast
Engagement (likes, votes)	Fast


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)

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/SMS2013