

Social Media Analytics

Data Mining Applied to Insurance Twitter Posts

Commitment Beyond Numbers



Gary C. Wang, FCAS, MAAA

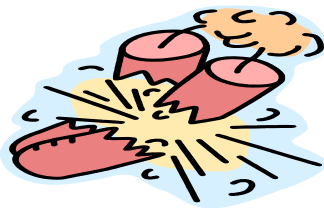
April 4th, 2013

Social Media Analytics

- The growth in social media
- Data
- General descriptive statistics
- Processing the data
- Analysis – identifying the themes
- Analysis challenges
- Application of social media analytics



The Growth in Social Media



Commitment Beyond Numbers

Social Media Defined

- **Social media:** a group of Internet-based applications that build on the ideological and technological foundations of [Web 2.0](#), and that allow the creation and exchange of [user-generated content](#)
- Building blocks
 - Identity
 - Conversations
 - Sharing
 - Presence
 - Relationships
 - Reputation
 - Groups

Kaplan, Andreas M.; Michael Haenlein (2010). "Users of the world, unite! The challenges and opportunities of Social Media". *Business Horizons*

Social Media Platforms



I need to eat



I ate



This is where I ate



Why am I eating?



This is a review of where I ate



Watch me eat



I have five connections that recommend me because I eat so well



Let's all eat together

Social Media – Explosive Growth

- 91% of all adults use social media regularly
- YouTube users watch more than 3 billion hours of video per month
- 93% of US adult internet users are on Facebook
- Every minute of the day:
 - 100,000 tweets are sent
 - 684,478 pieces of content are shared on Facebook
 - 48 hours of video are uploaded to YouTube
 - 47,000 apps are downloaded

“216 Social Media and Internet Statistics.” thesocialskinny.com. September, 2012.

Social Media is Impacting Business

- Consumers spend an average of \$391 million per day online
- 70% of social networkers shop online
- 79% of online shoppers spend 50% of their online shopping time researching products
- 44% of businesses acquired customers on Twitter
- Companies that blog have 55% more website visitors

"216 Social Media and Internet Statistics." thesocialskinny.com. September, 2012.

Insurance Facebook Fans

Page	Fan Count (February, 2013)	Percent Growth	NI	NPTA
Flo, The Progressive Girl	4,889,250	0.9	4.6	5.2
Farmers Insurance	2,337,697	-0.3	0.5	0.9
Mayhem (Allstate)	1,521,093	0.9	22.4	10.4
State Farm Nation	1,464,196	0.3	0.3	0.9
AARP	1,013,237	0.2		25.7
Metlife	554,562	0	18.1	6.4
USAA	504,213	2.2	11.2	8.6
Gerber Life	457,629	13.6	0.7	10.9
Liberty Mutual Insurance	384,903	21.8	5.2	22.1
New York Life	370,480	6.4	16.5	17.9

Insurance Twitter Followers

Company	Followers (February, 2013)	Percent Change	Tweets	Tweets (3M Average)
New York Life	57,381	13.9	187	202
USAA	40,553	4.2	51	159
Allstate Insurance	31,018	2.0	242	196
State Farm Insurance	30,820	2.7	509	403
UHC	30,016	0.4	15	39
State Farm Nation	28,541	0.5	46	39
VPI	21,273	2.3	158	168
Progressive - Its Flo	19,273	1.4	30	31
AFLAC Duck	18,327	3.0	46	121
Nationwide Insurance	16,856	2.2	96	106

Source: Customer Respect Group. "Social Eyes: The Insurer's View of Social Media."

Are You Taking Advantage of Social Media?

- Insurance companies are investing significant resources in a social media presence
- Current and potential customers are voluntarily sharing intimate details of their life with the world
- Current and potential customers are interacting with companies on a very personal level
- This information can be applied in different ways (service, marketing, competitive monitoring)



Data

Commitment Beyond Numbers

Data

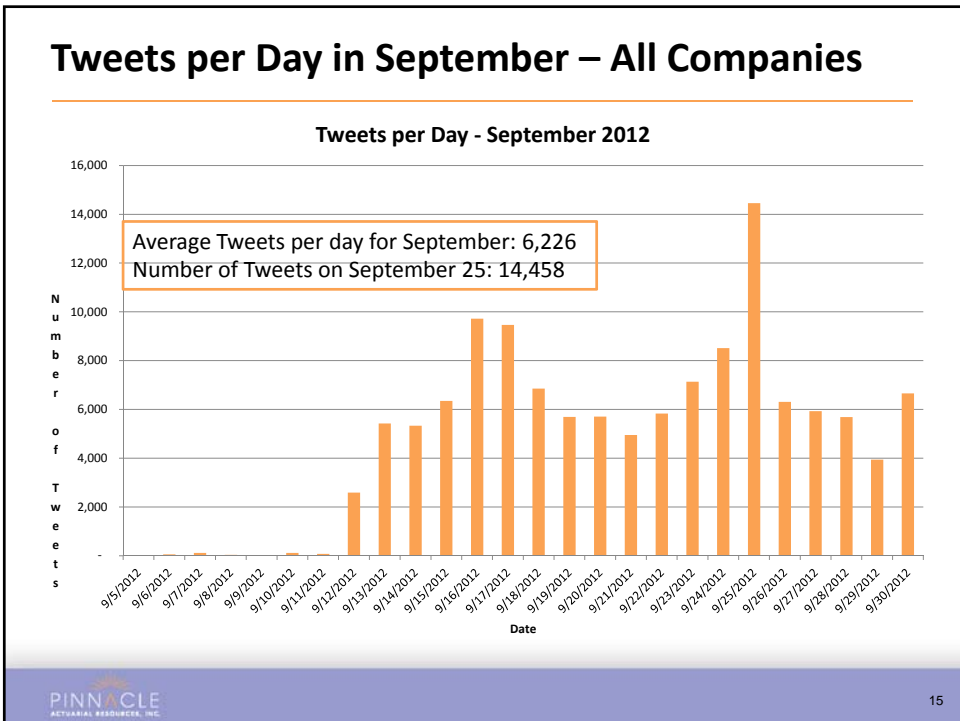
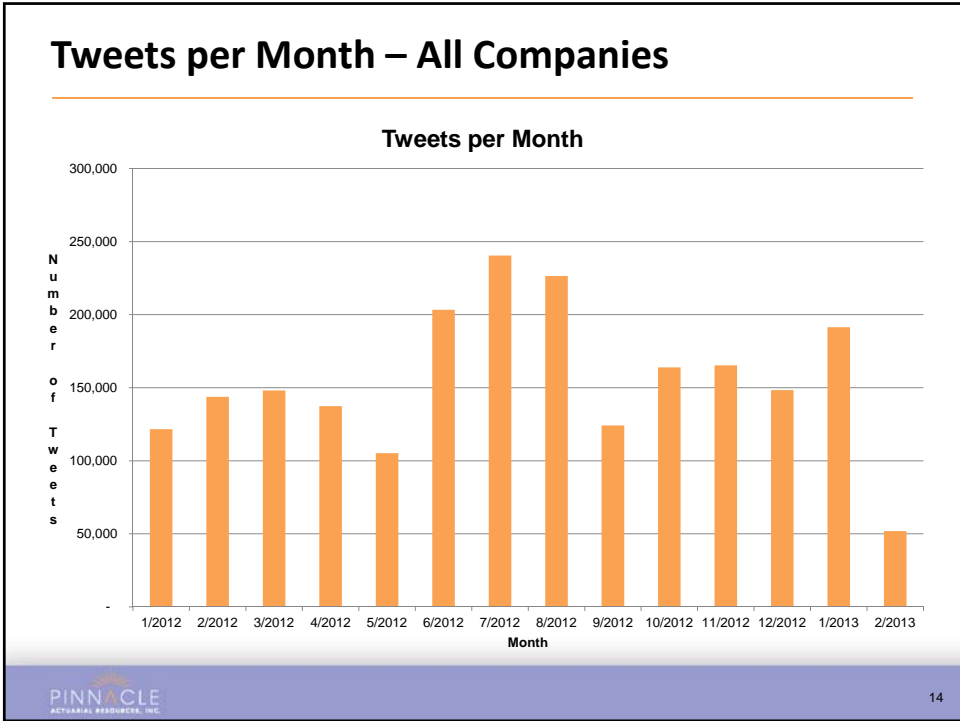
- Keyword searches for State Farm, Allstate, Geico, esurance, and #Progressive – 2.1 million tweets
- January 25, 2012 – February 6, 2013
- Tracked through hootsuite.com
- Data
 - **text**: content of the tweet
 - **to user id**: specific tweet recipient
 - **from user**: sender of the tweet
 - **iso language code**: language of tweet
 - **source**: where did the tweet originate?
 - **profile image**: picture of user
 - **geo**: latitude and longitude of the user
 - Date and time

Sources of Social Media Data

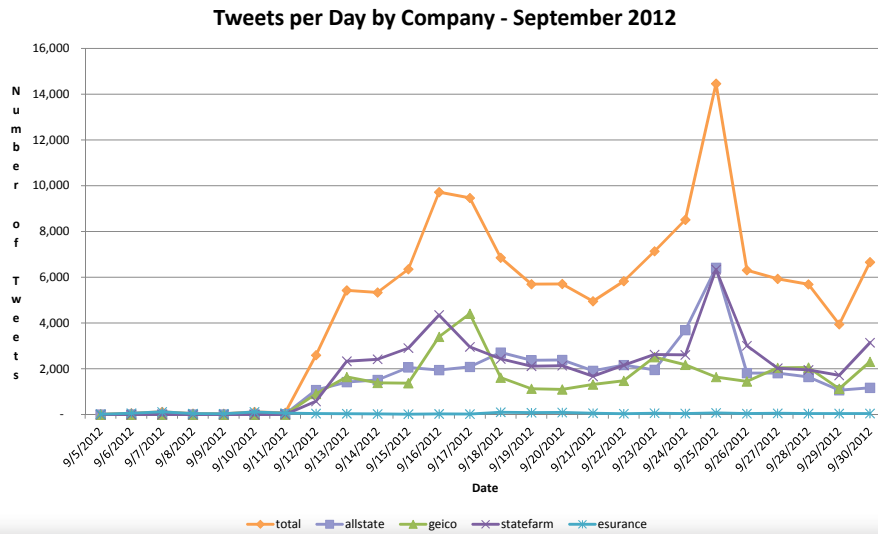
- Third party data aggregators (hootsuite, GNIP)
- API
- Company developers
- Screen scraping



General Descriptive Statistics



Tweets per Day by Company – September, 2012



September 25, 2012 – NFL Replacement Referees

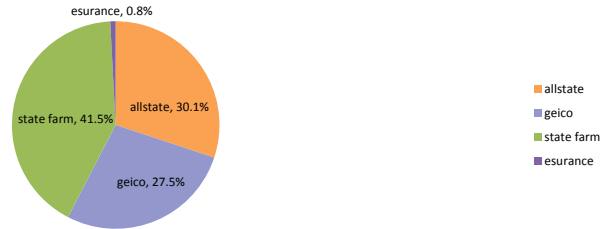
If only Aaron Rodgers had Allstate instead of State Farm. He'd be protected from mayhem like this. #MondayNightFootball #Robbery



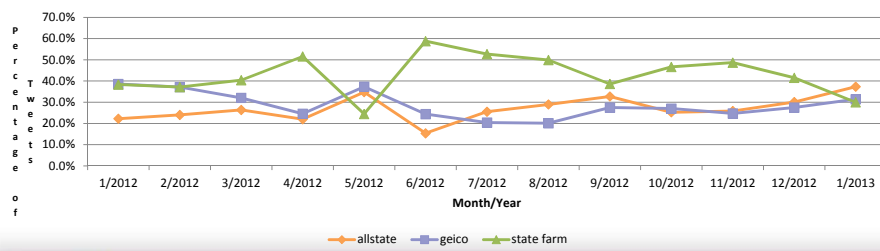
BREAKING NFL NEWS: the Packers will switch to Allstate so they will be protected from "mayhem" (like tonight) #NFL

Distribution of Tweets by Company

Tweets by Company - December 2012



Percentage of Tweets by Company



Data Processing Steps

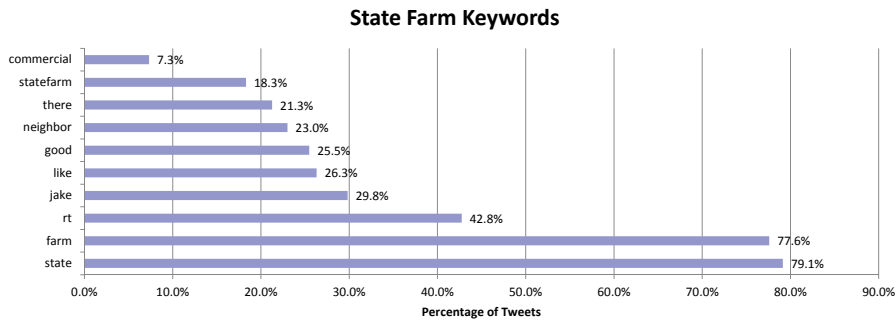
Data Processing Steps

- Remove punctuation and symbols (retain @ and #)
- Parse the tweet (35 words worked for Twitter – will need many more for other sources)
- Change table structures from tweets in rows to tweets in columns – keep indicator of order
- Correct spelling errors
- Add word indicators

Tweet ID	User	Tweet	Word1	Word2	...	Word35
1	@mosley	Text of tweet	W1	W2	...	W35

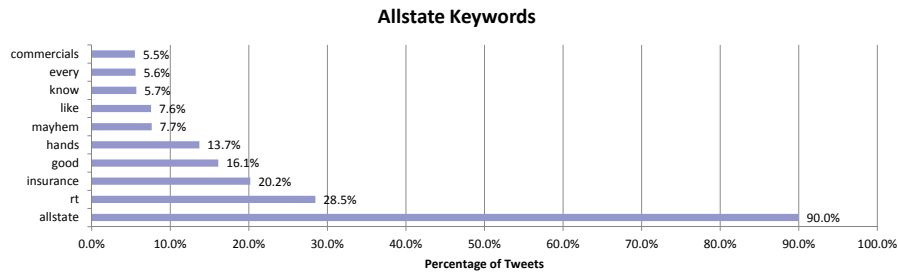
Tweet ID	Word Order	Word
1	1	Word1
1	2	Word2
...
1	35	Word35

State Farm Top 10 Keywords



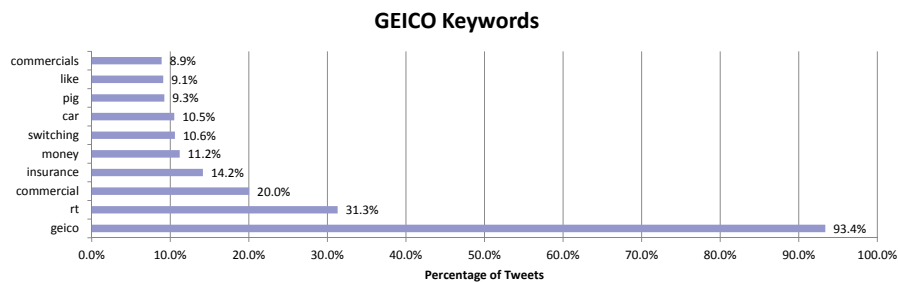
- Advertising (commercial, Jake)
- Brand recognition (“Like a good neighbor, State Farm is there”)
- Significant sharing – over 40% the tweets are retweets

Allstate Top 10 Keywords



- Advertising (Mayhem, commercials, ramen noodles “every” night budget, Mayhem “like” me)
- Brand recognition (you’re in good hands with Allstate)
- Sharing component not as significant (rt only 28%)

GEICO Top 10 Keywords



- Advertising (commercial, pig, commercials)
- Brand recognition (“Saved money by switching car insurance”)
- Sharing component not as significant (rt only 31%)



Analysis – Identifying Themes in the Data

Commitment Beyond Numbers

Clustering/Segmentation

- Unsupervised classification technique
- Groups data into set of discrete clusters or contiguous groups of cases
- Performs disjoint cluster analysis on the basis of Euclidean distances computed from one or more quantitative input variables and cluster seeds
- Data points are grouped based on the distances from the seed values
- Objects in each cluster tend to be similar, objects in different clusters tend to be dissimilar

Cluster Lift – State Farm

$$\text{Cluster Lift (word)} = \frac{\text{Percentage of tweets in a cluster that include word}}{\text{Percentage of all tweets that include word}}$$

Cluster	another_ind	anyone_ind	around_ind	business_ind	buy_ind	came_ind	care_ind	cause_ind
1	2.039	0.512	7.101	2.788	1.042	0.703	1.955	0.973
2	0.345	7.176	0.963	1.589	0.000	0.277	0.386	1.076
3	0.000	0.319	1.381	8.481	3.242	2.625	4.258	0.969
4	0.000	0.688	0.554	0.654	0.000	0.877	0.375	91.688
5	0.000	0.831	0.385	0.000	0.000	0.000	0.000	0.000
6	0.944	0.968	0.958	1.839	0.844	1.328	21.637	1.051
7	8.337	1.303	7.762	3.053	0.000	0.894	0.000	0.743
8	2.413	2.121	1.838	0.733	1.849	5.407	4.049	2.073
9	11.261	0.707	1.225	2.993	2.158	4.367	0.000	1.075
10	2.872	1.708	0.579	1.976	1.208	11.211	0.567	0.621
11	1.311	0.576	0.999	2.091	1.759	1.187	6.600	0.876
12	2.498	4.393	1.631	6.451	13.403	0.431	2.395	0.477
13	155.505	2.254	0.977	0.974	0.000	1.436	0.000	1.224

Keywords – State Farm

Cluster #	Number of Tweets	Category	Keywords	Description
1	1,491	sponsorship	around hit miss play run	Home Run Derby
2	3,778	sponsorship	anyone college comes kids lebron	College Sports Sponsorships
3	1,198	advertising	business music win	Music Festivals, Music Download
6	2,762	competition	care cool funny much nationwide pretty	Comparisons with Nationwide
7	4,689	sponsorship	another around football looks lost packers week	Football sponsorships, Packers, Hish School Football
8	1,260	advertising	couple help jingle million nobody nothing remember	Do the jingle
9	1,080	advertising	another long named pretty talk wife woman	Jake from State Farm, woman singing "Waterfall"
10	5,142	advertising	aaron came check discount double giants lost maybe p	Discount Double Check, Packers
12	1,217	quotes	buy coverage insurance month quote start	State Farm Marketing, Insurance Quotes
15	471	advertising	discounts drive insurance less through	Discounts
19	1,000	sports	discounts giants hanging less more packers spent time	Packers & Giants
20	2,941	competition	change having same	Changing Insurance Companies
25	2,414	marketing	about care don florida insurance kids rates	Great rates
26	45	advertising	accident another anyone anything believe commercial	French model
27	1,152	claims	claims customers emergency help lost nationwide san	Hurricane Sandy

Cluster Results – Top 10 Words per Cluster

Rank										
Cluster	1	2	3	4	5	6	7	8	9	10
1	run	hit	miss	play	around	city	watching	today	kids	long
2	college	comes	kids	lebron	anyone	about	drop	think	football	never
3	music	win	business	part	song	happy	ready	hear	live	care
4	full	yes	cause	coverage	double	emergency	room	thats	favorite	house
5	atwasoverv	walks	brakes	parents	texting	awkward	security	wearing	caught	morning
6	nationwide	pretty	care	cool	funny	much	allstate	follow	okay	same
7	looks	football	week	lost	another	around	packers	drop	named	today
8	times	nobody	million	couple	help	remember	same	nothing	jingle	called
9	woman	long	wife	talk	pretty	another	named	fire	looks	dead
10	double	discount	check	maybe	packers	came	aaron	lost	rodgers	giants
11	very	sorry	happy	nice	thank	drop	glad	hear	much	customer
12	quote	buy	coverage	start	insurance	month	business	get	rates	getting
13	okay	though	another	statefarm	week	nothing	cool	think	chris	no
14	password	found	wtf	put	week	happy	wife	neighbors	tomorrow	car
15	discounts	drive	less	through	insurance	live	more	policy	coverage	discount

Key Themes – State Farm

Themes	Percentage of Tweets
advertising	8.7%
claims	1.2%
competition	2.8%
customer feedback	19.1%
customer service	7.2%
employment	0.7%
marketing	1.2%
quotes	0.6%
sponsorship	6.0%
sports	1.1%
tweets	1.7%
blank	49.9%

Association Analysis

- Background in market basket analysis
- Identification of items that occur together in the same record
- Produces event occurrence as well as confidence interval around the occurrence likelihood
- Can lead to sequence analysis as well, which considers timing and ordering of events

Association Analysis Formulas

$$\text{Support} = \frac{\text{Transactions that contain Items A \& B}}{\text{All transactions}}$$

$$\text{Confidence} = \frac{\text{Transactions that contain Items A \& B}}{\text{Transactions that contain Item A}}$$

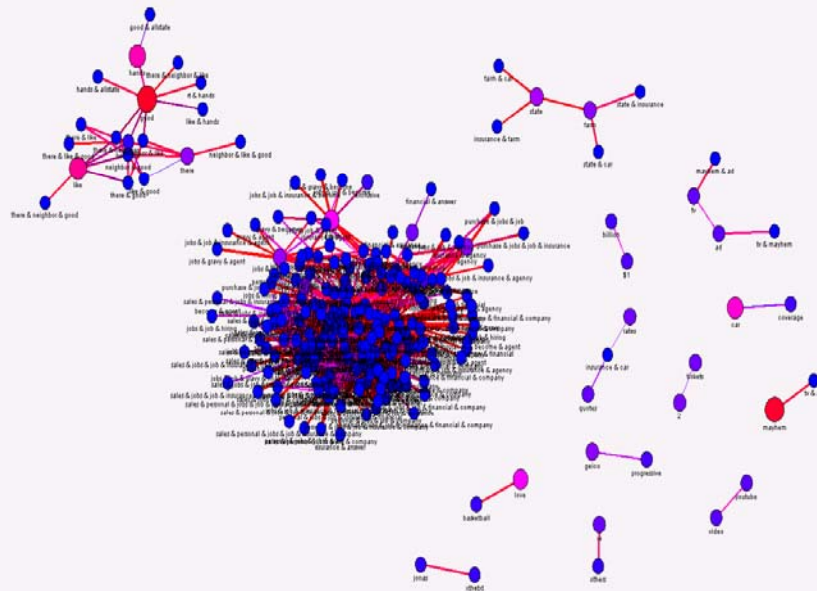
$$\text{Expected Confidence} = \frac{\text{Transactions that contain Item B}}{\text{All transactions}}$$

Transactions = tweets
Items = words

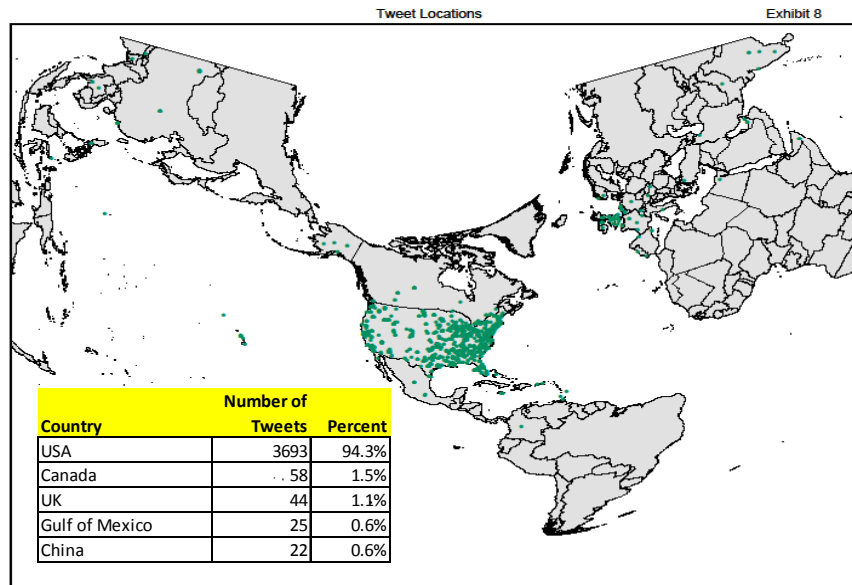
Association Rules – Dataset #1

Expected			Transaction				Rule		Rule
Confidence(%)	Confidence(%)	Support(%)	Lift	Count	Rule	Left Hand of Rule	Right Hand of Rule	Index	
0.632	73.391	0.618	116.070	422	neighbor ==> there & like & good	neighbor	there & like & good	73	
0.754	86.831	0.618	115.194	422	neighbor & like ==> there & good	neighbor & like	there & good	78	
0.754	81.043	0.682	107.516	466	neighbor ==> there & good	neighbor	there & good	110	
0.771	79.174	0.618	102.644	422	neighbor & good ==> there & like	neighbor & good	there & like	134	
0.888	91.053	0.760	102.486	519	jonas ==> xthetxt	jonas	xthetxt	135	
0.834	85.502	0.760	102.486	519	xthetxt ==> jonas	xthetxt	jonas	136	
0.771	75.130	0.632	97.402	432	neighbor ==> there & like	neighbor	there & like	177	
1.762	99.167	0.697	56.273	476	financial & answer ==> esurance	financial & answer	esurance	277	
0.921	50.911	0.777	55.299	531	tv ==> mayhem & ad	tv	mayhem & ad	281	
1.527	84.420	0.777	55.299	531	mayhem & ad ==> tv	mayhem & ad	tv	282	
0.700	37.679	0.618	53.855	422	like & good ==> there & neighbor	like & good	there & neighbor	295	
1.639	88.285	0.618	53.855	422	there & neighbor ==> like & good	there & neighbor	like & good	296	
1.041	52.679	0.777	50.620	531	ad ==> tv & mayhem	ad	tv & mayhem	311	
1.762	88.516	0.733	50.229	501	answer ==> esurance	answer	esurance	323	
1.639	81.913	0.689	49.968	471	neighbor ==> like & good	neighbor	like & good	331	
1.399	65.843	0.897	47.056	613	video ==> youtube	video	youtube	364	
2.122	87.639	0.809	41.294	553	state & car ==> farm	state & car	farm	530	
0.924	38.138	0.809	41.294	553	farm ==> state & car	farm	state & car	529	
2.122	80.923	0.950	38.130	649	state & insurance ==> farm	state & insurance	farm	588	
1.174	44.759	0.950	38.130	649	farm ==> state & insurance	farm	state & insurance	587	

Association Analysis Link Graph



Tweet Locations



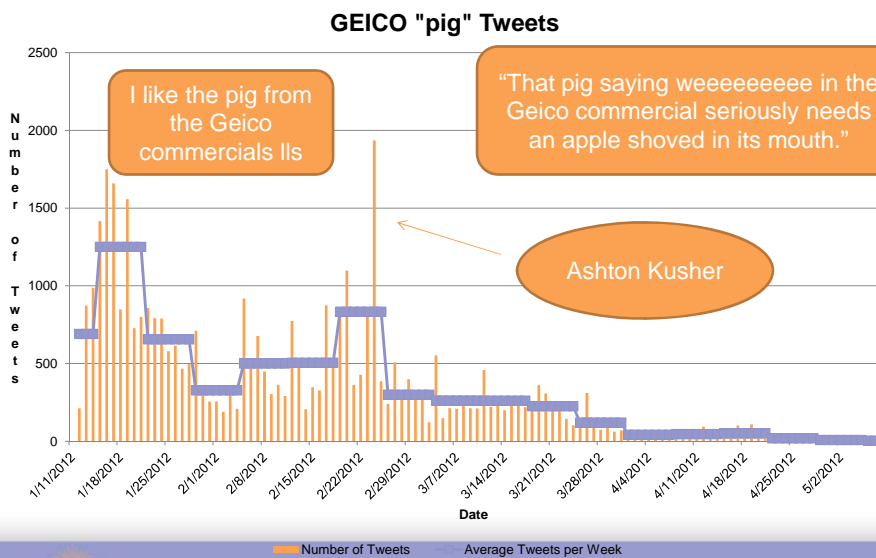
Analysis Challenges

- Accessing and collecting information
- Context
- Relevance
- Influence
- Sentiment
- Raw, unfiltered customer data
- Near real-time analysis needed

Applications of Social Media Analytics

- Customer service
- Understanding customer sentiment relating to company (advertising, etc.)
- Competitive intelligence
- Broad market trends

Commercial "Focus Groups"



Claims Applications

- Service questions
 - hi I was just looking at the site and I need assistance on faxing a claim
 - What is the Toll Free Telephone Number for the Claims Department for XXXX? <http://t.co/Y8ypSEkV>
- Feedback
 - Thank you for the fast windshield chip repair claim processing.
- Information
 - XXXX opens temporary claims sites to help AL storm victims.
 - Claims to XXXX from Dallas-Ft Worth tornado&hail on Apr 3 have now reached 16,000. We're working 7days/wk to help customers recover.
- Complaints
 - XXXX not paying your claim in a timely manner? Get a LAWYER here NOW: <http://t.co/IJL5YgAL> #RT

Quotes

- Many companies/agencies using for advertising
- Current and potential customers
 - I really wish i could get a quote on insurance without XXXX or XXXX hassling me for 3 weeks afterwards
 - Called my insurance company for new policy here in TX made my insurance jump \$204 dollars. Called XXXXfor a quote & they wanted \$300 more.
 - Does anyone use XXXX Auto Insurance? I was looking at quotes and they put me \$450+ below everyone else. Wondering what the deal is.

Thank You for Your Attention

Gary C. Wang, FCAS, MAAA

309-807-2331

gwang@pinnacleactuarials.com

