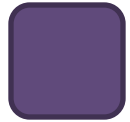




Bloomberg



# DISSECTING SOCIAL MEDIA

AND ITS IMPACT ON FINANCIAL MARKETS

Machine Learning & Natural Language Processing at Bloomberg

GRAB

# 1<Go> for Company Information

Bloomberg Social Velocity - Sep 13 2013 09:21:02

A BSV alert is triggered by an unusually high number of social media postings on a company.

1) INTEL CORP

Actual 21

Expected 2.2

Range Sep 13 09:00-09:30

## Number of Postings (30-Minute Increments)



## 10) Representative Postings

- 11) david peltier: Analysts' Actions: \$AET \$FINL \$INTC \$TEVA \$UA thestreet.com/s... TWT 09/13
- 12) BenzingaWire: \$INTC PreMarket Info Recap for September 13: Friday the 13th E... TWT 09/13
- 13) Earnings Impact: \$INTC Check out the details about \$300 Bay Trail clam shell... STK 09/13
- 14) Almostatrader: Watching: \$NQ \$SWY \$INTC \$GSVC... also names from earlier in... STK 09/13
- 15) John Voorheis: Upgrades \$WEN \$D \$INTC \$VZ \$LUL \$SWY \$MGM Downgrades \$UA... TWT 09/13
- 16) Charles Rankin: #stocks MARKET PULSE-Facebook, Intel, Pactera, Ulta Beauty,... TWT 09/13

20) View more postings on Bloomberg

21) View more postings on Twitter.com

# Bloomberg

**Impactful**  
The world is watching

Aug 13, 2013 14:21:29 [TWT] ☆ ✉ 🔍  
Carl Icahn: We currently have a large position in APPLE. We believe the company to be extremely...

We currently have a large position in APPLE. We believe the company to be extremely undervalued. Spoke to Tim Cook today. More to come. 🐦

Carl Icahn @Carl\_C\_Icahn

Sent With: web

[Click here to see tl](#)

1) APPLE INC Actual 635 Expected 48.2 Range Aug 13 14:00-14:30  
Number of Postings (30-Minute Increments)



## 10 Representative Postings

- 11) Linette Lopez: RT @clusterstock: Carl Icahn: I Have A Large Position In Apple And I... TWT 08/13
  - 12) Kooler Aid: \$AAPL Icahn's my favorite gran daddy TWT 08/13
  - 13) Briefing Trader: \$XLK & \$QQQ tick to a new almost 13-Yr. high led by \$AAPL +3.5% TWT 08/13
  - 14) Briefing Trader: \$XLK & \$QQQ tick to a new almost 13-Yr. high led by \$AAPL +3.5% TWT 08/13
  - 15) Shawn Langlois: Ellison praises the NSA, slams Google, baits the Apple camp, surpris... TWT 08/13
  - 16) Lev Naginsky: If you're wondering why Apple just spiked. \$AAPL <http://t.co/lADyl64...> TWT 08/13
- 20) [View more postings on Bloomberg](#) 21) [View more postings on Twitter.com](#)



Bloomberg



# Goal

Use Data and Knowledge about the real world in order to infer patterns and make predictions. Leverage this ability to enhance the tools our clients need.

Sentiment Analysis

Novelty Detection    Topic-based Clustering

Market Impact Analysis    Social Velocity    Language Detection

Event Detection & Extraction    Machine Translation

Social Graph Analysis    Tokenization

User Behaviour Analysis

Parsing



Bloomberg



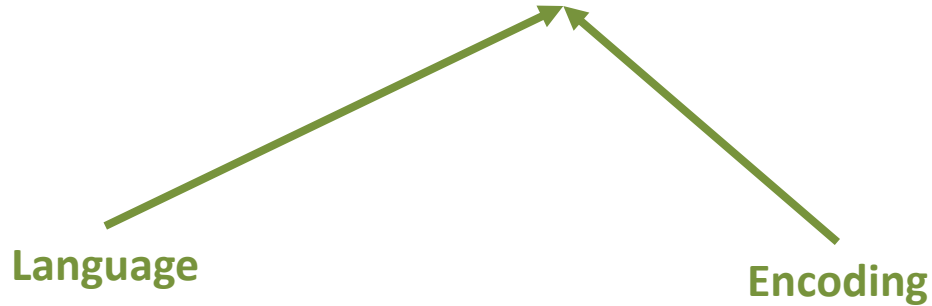
Start with some text.



# Conforming Content

Natural Language Processing

“James loves Machine Learning when the objective is well-specified.”





Bloomberg

# Natural Language Processing

## Sentence and Token Boundaries



Mark the boundaries of tokens and sentences. Maximize the probability of each segment.



[" James loves Machine Learning when the objective is well-specified . "]



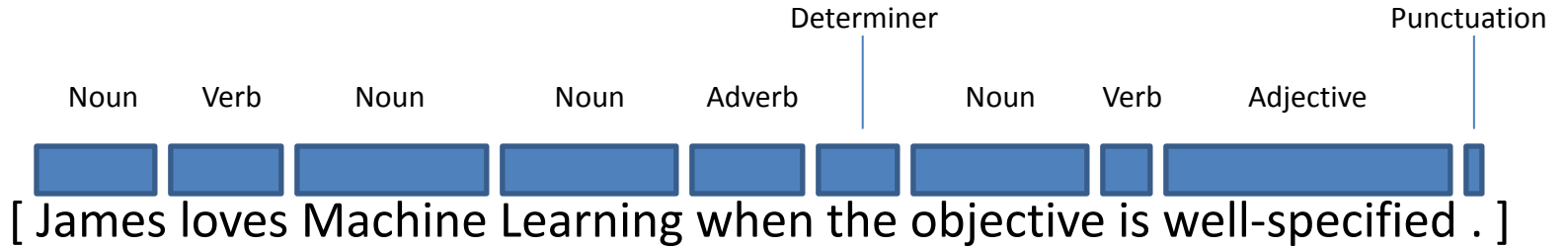
Bloomberg

# Natural Language Processing

## Parts-of-Speech Tagging



Identify the grammatical categorization of each token:





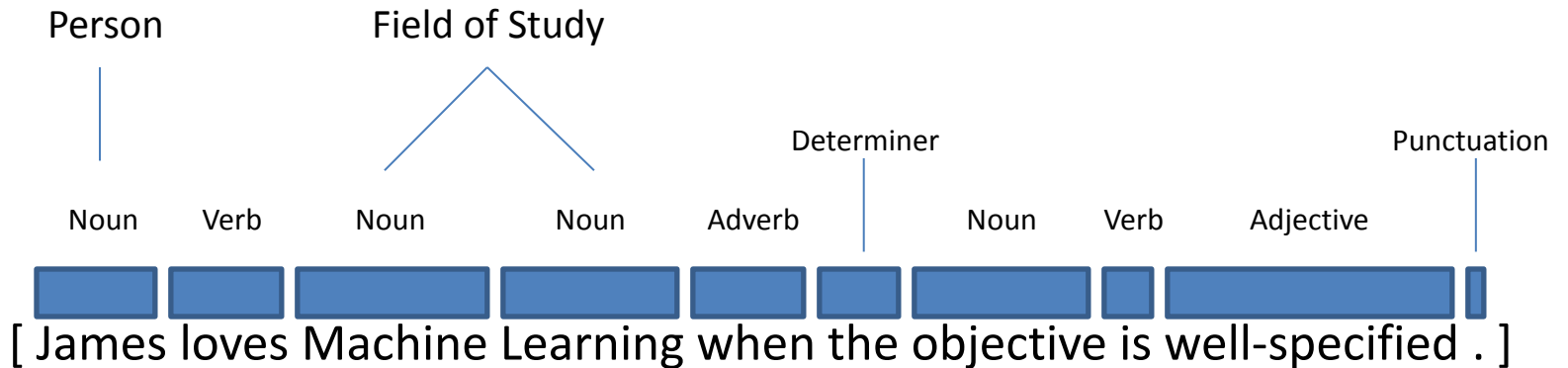
Bloomberg

# Natural Language Processing

## Named Entity Recognition



Noun groupings and named-entity recognition provides identification of logically connected entities. Named-entity resolution ties those entities to an ontology (knowledge-base) to differentiate them:



Verb, Present  
Indicative: to hold  
dear : CHERISH



The ontology may also be used to resolve definitions of ambiguous words:





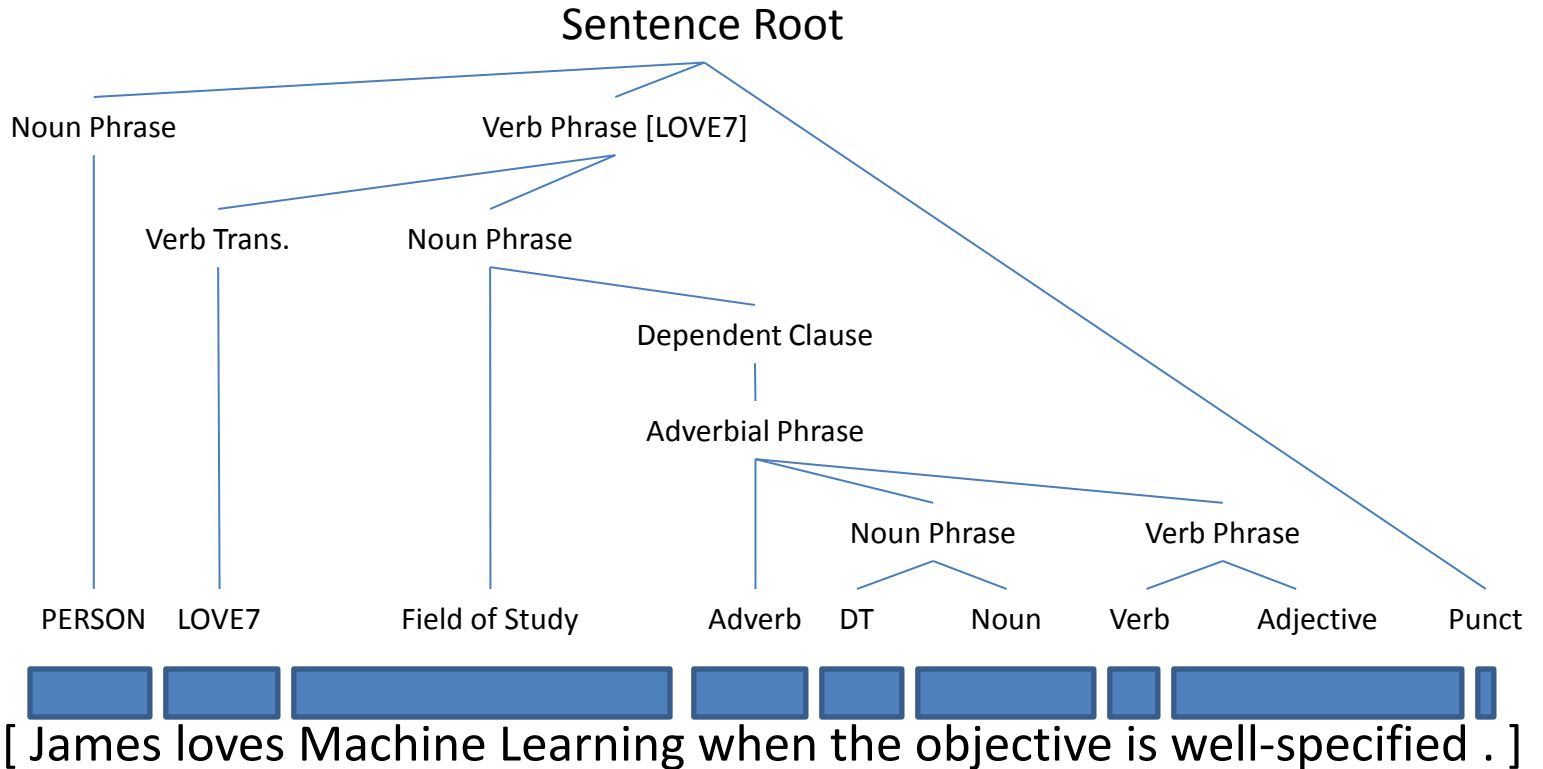
Bloomberg

# Natural Language Processing

## Syntactic Parsing



Build the dependency relationship diagram between constituents of the sentence:





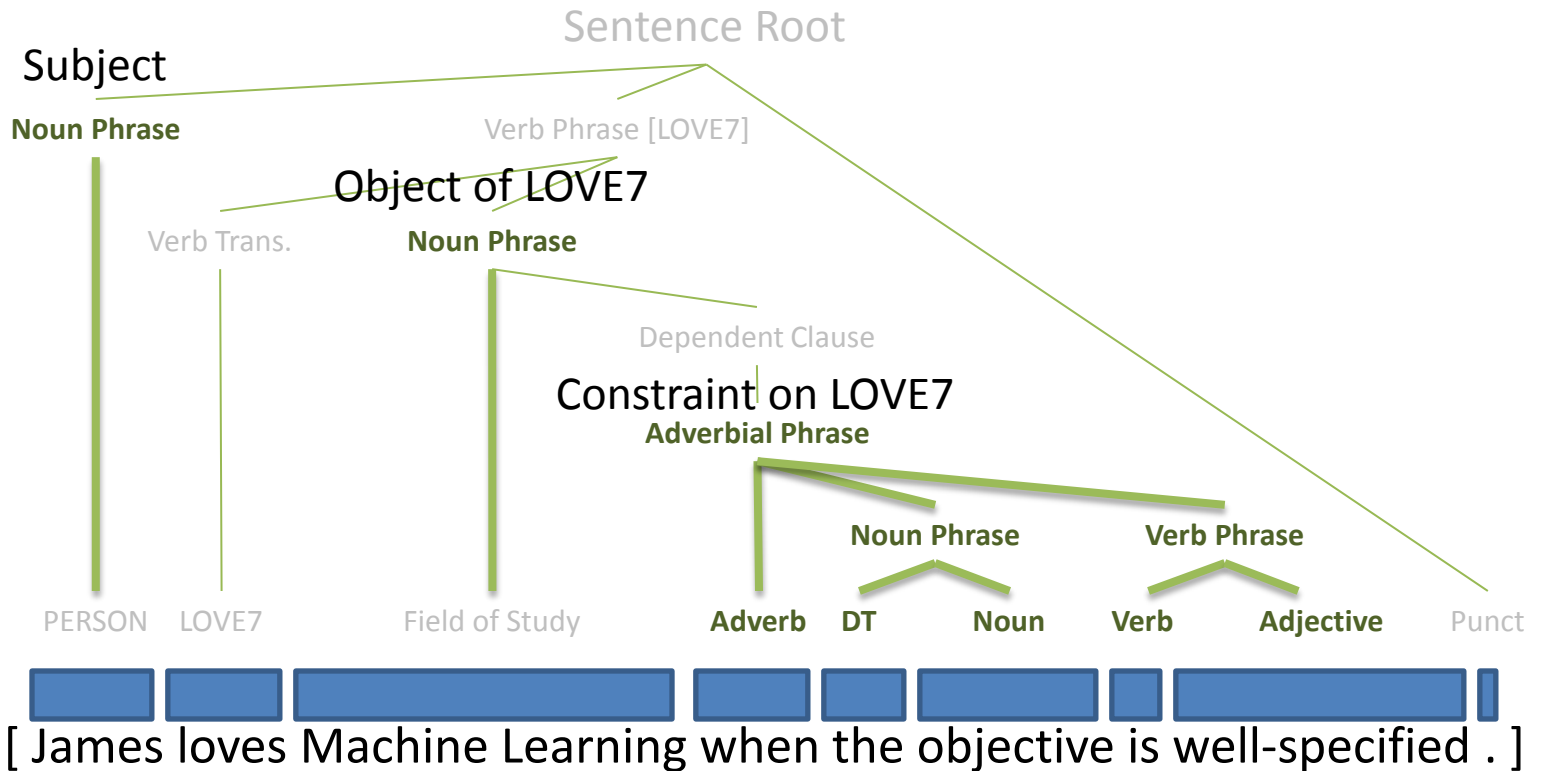
Bloomberg

# Natural Language Processing

## Semantic Role Labeling



What are the roles of different constituents in the sentence?





Bloomberg

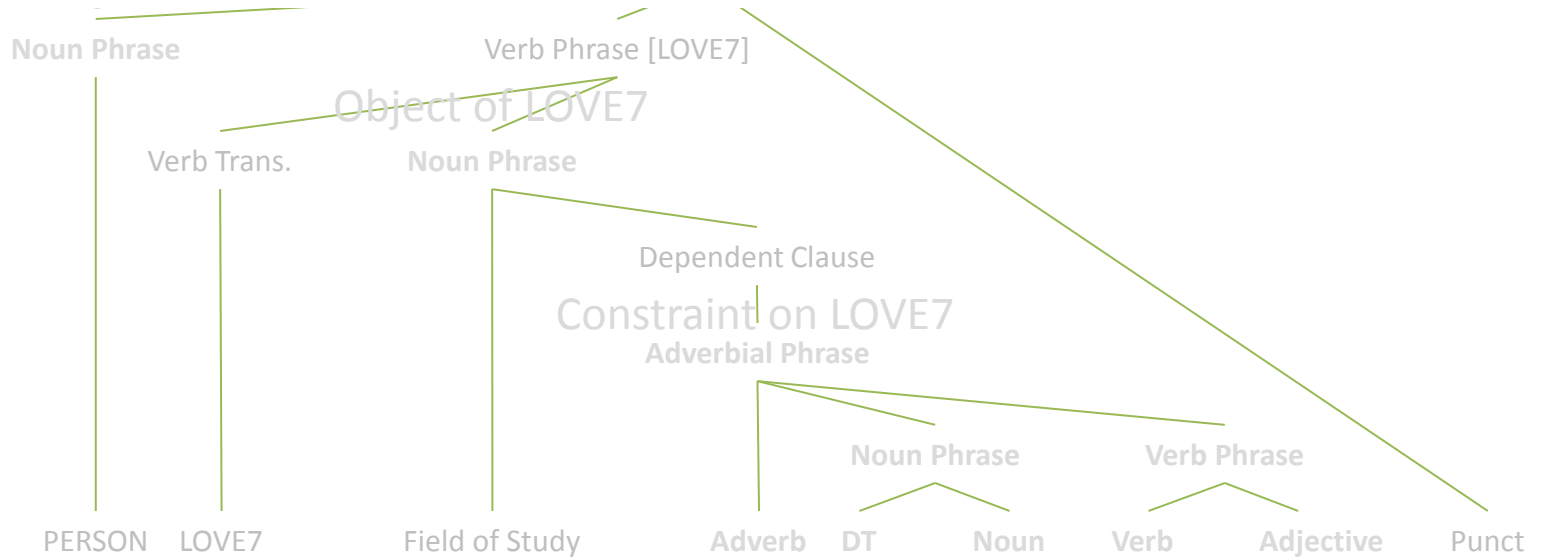
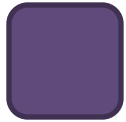
# Natural Language Processing Representation



If we get this far, we can make our findings portable:



$\exists x \in X, \exists y \in Y : \text{hasObjective}(x,y) \wedge \text{wellSpecified}(y) \wedge \text{loves}(\text{James},x)$



[ James loves Machine Learning when the objective is well-specified . ]



Bloomberg

# Natural Language Processing

Easy...?



What are the roles of different constituents in the sentence?

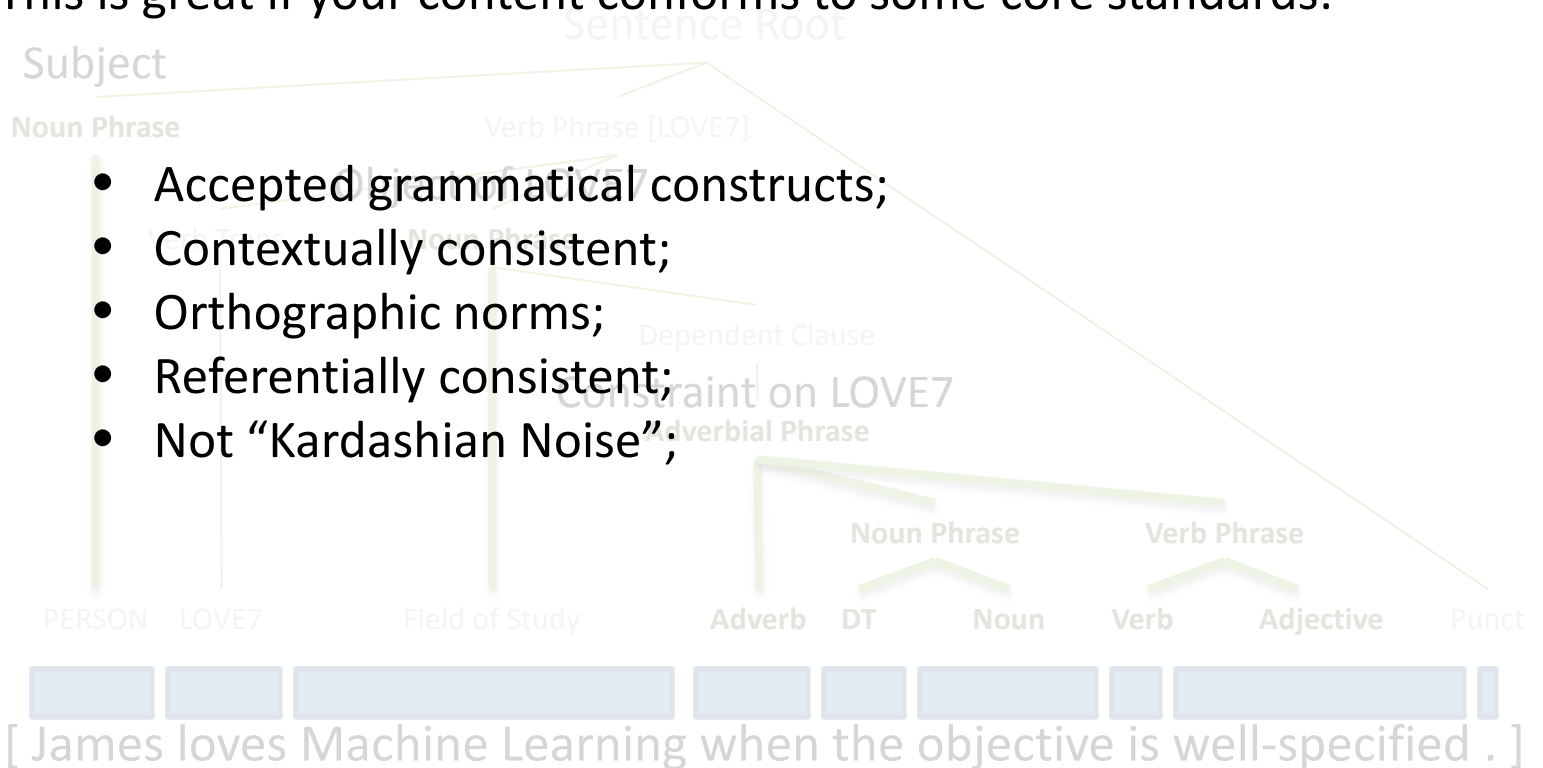


This is great if your content conforms to some core standards:

Subject

Noun Phrase

- Accepted grammatical constructs;
- Contextually consistent;
- Orthographic norms;
- Referentially consistent;
- Not “Kardashian Noise”;



# Bloomberg

# Social Publishing

A new kind of content

Shared conversations: Stocktwits, Twitter, Facebook, Bit.ly, LinkedIn, more...

The screenshot displays the Bloomberg terminal interface. At the top, there are navigation tabs: 96) New Search, 97) Actions, 98) Settings, and Feedback. The main display area shows a search result for 'VALUE, Dividend Gainers Financial [4], INDU, DIVIDEND G...'. Below this, there is a list of 20 items, each with a number, a text snippet, and a source/time stamp. The items are:

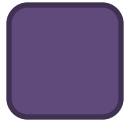
Rank	Text Snippet	Source	Time
1)	Clark: \$mmm	TWT	23:46
2)	Jeffrey-Ryan B.: Apple taking preorders for iPhone 5C cbsn.ws/1dhmja3	TWT	23:46
3)	Jeffrey-Ryan B.: Should you fear Apple's fingerprint scanner? cbsn.ws...	TWT	23:46
4)	Nathan: \$BA Lufthansa to buy Boeing jets http://stks.co/sSUT	STK	23:45
5)	Mila D. Aguilar: Has Apple Invented an Unbreakable iPhone Display? w...	TWT	23:45
6)	Option Alert: Commented on StockTwits: \$X Jan 17-20 put spread was...	TWT	23:45
7)	Jason Bond: "@landonsc1: I wish Apple would hurry up with an iWatch...	TWT	23:44
8)	Ali Madad: It would be very interesting to see a \$MSFT design-fiction v...	TWT	23:44
9)	BV Services: \$NFLX Netflix Makes History With Two Primetime Creative...	STK	23:43
10)	BV Services: \$NFLX Netflix Makes History With Two Primetime Creative...	TWT	23:43
11)	Flightglobal: Latest: Boeing delivers last USAF C-17 http://t.co/GroYw...	TWT	23:43
12)	auto clock: #microsoft claims to pull thier unfunny #apple ads but run...	TWT	23:42
13)	Vishal Mehra: A closer look at the private jet flights of Google execs,...	TWT	23:42
14)	Patrick Seitz: Netflix CEO Reed Hastings was wrong about financial cri...	TWT	23:41
15)	SynergyBuzzMarketing: Google founders Larry Page and Sergey Brin...	TWT	23:41
16)	Glenn Fleishman: ICYMI: I wrote how the iPhone 5c+5s are part of Ap...	TWT	23:40
17)	Ross D Pratt: Microsoft offers \$200 for used iPadsMicrosoft is trying to...	TWT	23:39
18)	Josh Noble: RT @M_C_Klein: A tumblr of Calvin and Hobbes cartoons wi...	TWT	23:38
19)	Lichartin Chartist: \$ES_F 1700 all time high coming, \$CMG should be f...	STK	23:37
20)	Justin M. Overdorff: did \$MSFT double down on these Surface commer...	TWT	23:37



Bloomberg



Opinion Leaders



# Social Publishing

A new kind of content

96) New Search    97) Actions    98) Settings    Feedback    Page 1 News

Topics << VALUE, Dividend Gainers Financial [4], INDU, DIVIDEND G...

My News < Narrow Search Sources Custom Lang My 09/15/13 Sort By Time

1) Clark: \$mmm	TWT 23:46
2) Jeffrey-Ryan B.: Apple taking preorders for iPhone 5C cbsn.ws/1dhmja3	TWT 23:46
3) Jeffrey-Ryan B.: Should you fear Apple's fingerprint scanner? cbsn.ws...	TWT 23:46
4) Nathan: \$BA Lufthansa to buy Boeing jets http://stks.co/sSUT	STK 23:45
5) Mila D. Aguilar: Has Apple Invented an Unbreakable iPhone Display? w...	TWT 23:45
6) Option Alert: Commented on StockTwits: \$X Jan 17-20 put spread was...	TWT 23:45
7) Jason Bond: "@landonsc1: I wish Apple would hurry up with an iWatch...	TWT 23:44
8) Ali Madad: It would be very interesting to see a \$MSFT design-fiction v...	TWT 23:44
9) BV Services: \$NFLX Netflix Makes History With Two Primetime Creative...	STK 23:43
10) BV Services: \$NFLX Netflix Makes History With Two Primetime Creative...	TWT 23:43
11) Flightglobal: Latest: Boeing delivers last USAF C-17 http://t.co/GroYw...	TWT 23:43
12) uto clock: #microsoft claims to pull thier unfunny #apple ads but run...	TWT 23:42
13) Vishal Mehra: A closer look at the private jet flights of Google execs,...	TWT 23:42
14) Patrick Seitz: Netflix CEO Reed Hastings was wrong about financial cri...	TWT 23:41
15) SynergyBuzzMarketing: Google founders Larry Page and Sergey Brin...	TWT 23:41
16) Glenn Fleishman: ICYMI: I wrote how the iPhone 5c+5s are part of Ap...	TWT 23:40
17) Ross D Pratt: Microsoft offers \$200 for used iPadsMicrosoft is trying to...	TWT 23:39
18) Josh Noble: RT @M_C_Klein: A tumblr of Calvin and Hobbes cartoons wi...	TWT 23:38
19) Lichartin Chartist: \$ES_F 1700 all time high coming, \$CMG should be f...	STK 23:37
20) Justin M. Overdorff: did \$MSFT double down on these Surface commer...	TWT 23:37



Bloomberg

# Social Publishing

A new kind of content



Different Grammatical Constructs



Opinion Leaders



The screenshot shows a Bloomberg terminal window with a search for "VALUE, Dividend Gainers Financial [4], INDU, DIVIDEND G...". The interface includes a navigation bar with "96) New Search", "97) Actions", "98) Settings", and "Feedback". The main content area displays a list of 20 items, each with a number, a title, and a timestamp. A red arrow points from the "Different Grammatical Constructs" text to the list, highlighting various sentence structures and user-generated content.

Item	Author/Source	Content	Platform	Time
1)	Clark: \$mmm		TWT	23:46
2)	Jeffrey-Ryan B.:	Apple taking preorders for iPhone 5C cbsn.ws/1dhmja3	TWT	23:46
3)	Jeffrey-Ryan B.:	Should you fear Apple's fingerprint scanner? cbsn.ws...	TWT	23:46
4)	Nathan: \$BA	Lufthansa to buy Boeing jets http://stks.co/sSUT	STK	23:45
5)	Mila D. Aguilar:	Has Apple Invented an Unbreakable iPhone Display? w...	TWT	23:45
6)	Option Alert:	Commented on StockTwits: \$X Jan 17-20 put spread was...	TWT	23:45
7)	Jason Bond:	"@landonsc1: I wish Apple would hurry up with an iWatch...	TWT	23:44
8)	Ali Madad:	It would be very interesting to see a \$MSFT design-fiction v...	TWT	23:44
9)	BV Services:	\$NFLX Netflix Makes History With Two Primetime Creative...	STK	23:43
10)	BV Services:	\$NFLX Netflix Makes History With Two Primetime Creative...	TWT	23:43
11)	Flightglobal:	Latest: Boeing delivers last USAF C-17 http://t.co/GroYw...	TWT	23:43
12)	auto clock:	#microsoft claims to pull thier unfunny #apple ads but run...	TWT	23:42
13)	Vishal Mehra:	A closer look at the private jet flights of Google execs,...	TWT	23:42
14)	Patrick Seitz:	Netflix CEO Reed Hastings was wrong about financial cri...	TWT	23:41
15)	SynergyBuzzMarketing:	Google founders Larry Page and Sergey Brin...	TWT	23:41
16)	Glenn Fleishman:	ICYMI: I wrote how the iPhone 5c+5s are part of Ap...	TWT	23:40
17)	Ross D Pratt:	Microsoft offers \$200 for used iPadsMicrosoft is trying to...	TWT	23:39
18)	Josh Noble:	RT @M_C_Klein: A tumblr of Calvin and Hobbes cartoons wi...	TWT	23:38
19)	Lichartin Chartist:	\$ES_F 1700 all time high coming, \$CMG should be f...	STK	23:37
20)	Justin M. Overdorff:	did \$MSFT double down on these Surface commer...	TWT	23:37



Bloomberg

# Social Publishing

A new kind of content



Different Grammatical Constructs

Inconsistent Entity References



Opinion Leaders



Rank	Author	Snippet	Source	Time
1)	Clark: \$mm			TWT 23:46
2)	Jeffrey-Ryan B.	Apple taking preorders for iPhone 5C cbsn.ws/1dhmja3		TWT 23:46
3)	Jeffrey-Ryan B.	Should you fear Apple's fingerprint scanner? cbsn.ws...		TWT 23:46
4)	Nathan: \$BA	Lufthansa to buy Boeing jets http://stks.co/sSUT		STK 23:45
5)	Mila D. Aguilar:	Has Apple Invented an Unbreakable iPhone Display? w...		TWT 23:45
6)	Option Alert:	Commented on StockTwits: \$X Jan 17-20 put spread was...		TWT 23:45
7)	Jason Bond:	"@andonsc1: I wish Apple would hurry up with an iWatch...		TWT 23:44
8)	Ali Madad:	It would be very interesting to see a \$MSFT design-fiction v...		TWT 23:44
9)	BV Services: \$NFLX	Netflix Makes History With Two Primetime Creative...		STK 23:43
10)	BV Services: \$NFLX	Netflix Makes History With Two Primetime Creative...		TWT 23:43
11)	Flightglobal:	Latest: Boeing delivers last USAF C-17 http://t.co/GroYw...		TWT 23:43
12)	auto clock:	#microsoft claims to pull thier unfunny #apple ads but run...		TWT 23:42
13)	Vishal Mehra:	A closer look at the private jet flights of Google execs,...		TWT 23:42
14)	Patrick Seitz:	Netflix CEO Reed Hastings was wrong about financial cri...		TWT 23:41
15)	SynergyBuzzMarketing:	Google founders Larry Page and Sergey Brin...		TWT 23:41
16)	Glenn Fleishman:	ICYM: I wrote how the iPhone 5c+5s are part of Ap...		TWT 23:40
17)	Ross D Pratt:	Microsoft offers \$200 for used iPadsMicrosoft is trying to...		TWT 23:39
18)	Josh Noble:	RT @M_C_Klein: A tumblr of Calvin and Hobbes cartoons wi...		TWT 23:38
19)	Lichartin Chartist:	\$ES_F 700 all time high coming, \$CMG should be f...		STK 23:37
20)	Justin M. Overdorff:	did \$MSFT double down on these Surface commer...		TWT 23:37





Bloomberg

# Social Publishing

A new kind of content



Different Grammatical Constructs

Contextually Ambiguous

Inconsistent Entity References

Opinion Leaders



96) New Search 97) Actions 98) Settings Feedback Page 1 News

Topics << VALUE, Dividend Gainers Financial [4], INDU, DIVIDEND G...

My News <Narrow Search> Sources Custom Lang My 09/15/13 Sort By Time

1) Clark \$mmm	TWT 23:46
2) Jeffrey-Ryan B.: Apple taking preorders for iPhone 5C cbsn.ws/1dhmja3	TWT 23:46
3) Jeffrey-Ryan B.: Should you fear Apple's fingerprint scanner? cbsn.ws...	TWT 23:46
4) Nathan: \$BA Lufthansa to buy Boeing jets http://stks.co/sSUT	STK 23:45
5) Mila D. Aguilar: Has Apple Invented an Unbreakable iPhone Display? w...	TWT 23:45
6) Option Alert: Commented on StockTwits: \$X Jan 17-20 put spread was...	TWT 23:45
7) Jason Bond: "@landonsc1: I wish Apple would hurry up with an iWatch...	TWT 23:44
8) Ali Madad: It would be very interesting to see a \$MSFT design-fiction v...	TWT 23:44
9) BV Services: \$NFLX Netflix Makes History With Two Primetime Creative...	STK 23:43
10) BV Services: \$NFLX Netflix Makes History With Two Primetime Creative...	TWT 23:43
11) Flightglobal: Latest: Boeing delivers last USAF C-17 http://t.co/GroYw...	TWT 23:43
12) auto clock: #microsoft claims to pull thier unfunny #apple ads but run...	TWT 23:42
13) Vishal Mehra: A closer look at the private jet flights of Google execs,...	TWT 23:42
14) Patrick Seitz: Netflix CEO Reed Hastings was wrong about financial cri...	TWT 23:41
15) SynergyBuzzMarketing: Google founders Larry Page and Sergey Brin...	TWT 23:41
16) Glenn Fleishman: ICYMI: I wrote how the iPhone 5c+5s are part of Ap...	TWT 23:40
17) Ross D Pratt: Microsoft offers \$200 for used iPadsMicrosoft is trying to...	TWT 23:39
18) Josh Noble: RT @M_C_Klein: A tumblr of Calvin and Hobbes cartoons wi...	TWT 23:38
19) Lichartin Chartist: \$ES_F 1700 all time high coming, \$CMG should be f...	STK 23:37
20) Justin M. Overdorff: did \$MSFT double down on these Surface commer...	TWT 23:37



Bloomberg

# Social Publishing

A new kind of content



Different Grammatical Constructs

Contextually Ambiguous

Inconsistent Entity References

Orthographically Challenging



Opinion Leaders



96) New Search		97) Actions		98) Settings		Feedback		Page 1 News	
Topics	<<	VALUE, Dividend Gainers Financial [4], INDU, DIVIDEND G...							
My News	<>	<Narrow Search>	Sources	Custom	Lang	M	09/15/13	Sort By	Time
Search		1)	Clark: \$mmm					TWT	23:46
( (TICKERLIST		2)	Jeffrey-Ryan B.: Apple taking preorders for iPhone 5C cbsn.ws/1dhmja3					TWT	23:46
My Pages		3)	Jeffrey-Ryan B.: Should you fear Apple's fingerprint scanner? cbsn.ws...					TWT	23:46
My News Pag		4)	Nathan: \$BA Lufthansa to buy Boeing jets http://stks.co/sSUT					STK	23:45
+Customize		5)	Mila D. Aguilar: Has Apple Invented an Unbreakable iPhone Display? w...					TWT	23:45
Recommend		6)	Option Alert: Commented on StockTwits: \$X Jan 17-20 put spread was...					TWT	23:45
Bookmarks		7)	Jason Bond: "@landonsc1: I wish Apple would hurry up with an iWatch...					TWT	23:44
Recently Rea		8)	Ali Madad: It would be very interesting to see a \$MSFT design-fiction v...					TWT	23:44
Filters, Alerts		9)	BV Services: \$NFLX Netflix Makes History With Two Primetime Creative...					STK	23:43
GLOBAL GOLD		10)	BV Services: \$NFLX Netflix Makes History With Two Primetime Creative...					TWT	23:43
JAPANESE		11)	Flightglobal: Latest: Boeing delivers last USAF C-17 http://t.co/GroYw...					TWT	23:43
TENDER		12)	auto clock: #microsoft claims to pull thier unfunny #apple ads but run...					TWT	23:42
NYTSEARCH		13)	Vishal Mehra: A closer look at the private jet flights of Google execs,...					TWT	23:42
+Customize		14)	Patrick Seitz: Netflix CEO Reed Hastings was wrong about financial cri...					TWT	23:41
Suggested Filt		15)	SynergyBuzzMarketing: Google founders Larry Page and Sergey Brin...					TWT	23:41
Securities		16)	Glenn Fleishman: ICYMI: I wrote how the iPhone 5s+5s are part of Ap...					TWT	23:40
People		17)	Ross D Pratt: Microsoft offers \$200 for used iPadsMicrosoft is trying to...					TWT	23:39
Sources		18)	Josh Noble: RT @M_C_Klein: A tumblr of Calvin and Hobbes cartoons wi...					TWT	23:38
		19)	Lichartin Chartist: \$ES_F 1700 all time high coming, \$CMG should be f...					STK	23:37
		20)	Justin M. Overdorff: did \$MSFT double down on these Surface commer...					TWT	23:37



Bloomberg

# Social Publishing

A new kind of content



Different Grammatical Constructs

Contextually Ambiguous

Inconsistent Entity References

Orthographically Challenging

Opinion Leaders

Frequent Spelling Inconsistencies



96) New Search		97) Actions		98) Settings		Feedback		Page 1 News	
Topics	<<	VALUE, Dividend Gainers Financial [4], INDU, DIVIDEND G...							
My News	<	<Narrow Search>	Sources	Custom	Lang	My	09/15/13	Sort By	Time
Search		1)	Clark: \$mmm					TWT	23:46
( (TICKERLIST		2)	Jeffrey-Ryan B.: Apple taking preorders for iPhone 5C cbsn.ws/1dhmja3					TWT	23:46
My Pages		3)	Jeffrey-Ryan B.: Should you fear Apple's fingerprint scanner? cbsn.ws...					TWT	23:46
My News Pag		4)	Nathan: \$BA Lufthansa to buy Boeing jets http://stks.co/sSUT					STK	23:45
+Customize		5)	Mila D. Aguilar: Has Apple Invented an Unbreakable iPhone Display? w...					TWT	23:45
Recommend		6)	Option Alert: Commented on StockTwits: \$X Jan 17-20 put spread was...					TWT	23:45
Bookmarks		7)	Jason Bond: "@landonsc1: I wish Apple would hurry up with an iWatch...					TWT	23:44
Recently Rea		8)	Ali Madad: It would be very interesting to see a \$MSFT design-fiction v...					TWT	23:44
Filters, Alerts		9)	BV Services: \$NFLX Netflix Makes History With Two Primetime Creative...					STK	23:43
GLOBAL GOLD		10)	BV Services: \$NFLX Netflix Makes History With Two Primetime Creative...					TWT	23:43
JAPANESE		11)	Flightglobal: Latest: Boeing delivers last USAF C-17 http://t.co/GroYw...					TWT	23:43
TENDER		12)	auto clock: #microsoft claims to put thier unfunny #apple ads but run...					TWT	23:42
NYTSEARCH		13)	Vishal Mehra: A closer look at the private jet flights of Google execs,...					TWT	23:42
+Customize		14)	Patrick Seitz: Netflix CEO Reed Hastings was wrong about financial cri...					TWT	23:41
Suggested Filt		15)	SynergyBuzzMarketing: Google founders Larry Page and Sergey Brin...					TWT	23:41
Securities		16)	Glenn Fleishman: ICYMI: I wrote how the iPhone 5c+5s are part of Ap...					TWT	23:40
People		17)	Ross D Pratt: Microsoft offers \$200 for used iPadsMicrosoft is trying to...					TWT	23:39
Sources		18)	Josh Noble: RT @M_C_Klein: A tumblr of Calvin and Hobbes cartoons wi...					TWT	23:38
		19)	Lichartin Chartist: \$ES_F 1700 all time high coming, \$CMG should be f...					STK	23:37
		20)	Justin M. Overdorff: did \$MSFT double down on these Surface commer...					TWT	23:37



Bloomberg

# Social Publishing

A new kind of content



In order to answer the same questions as we answer for editorial content, we must first address several new questions:

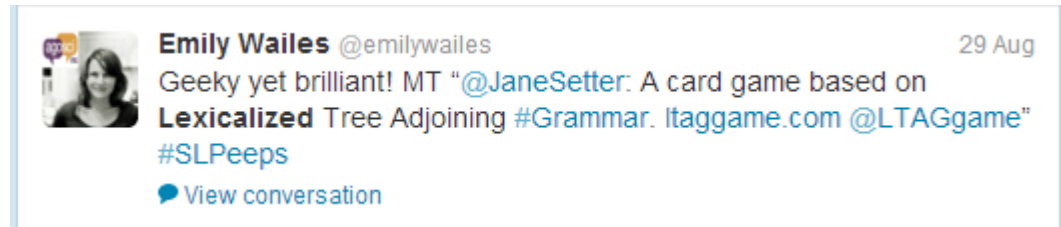


- **How trustworthy is this message?**
  - AP News Twitter account hacked: “Breaking: Two Explosions in the White House and Barack Obama is injured”
  - The false alarm sent the Dow plunging 145 points;
- **How authoritative is the source?**
  - Do people listen to what is being published?
  - Is it timely, novel, impactful?
- **What is the social context for this message?**
  - What has already been said?
  - Is this a response?
  - How many people are listening?
  - Is it spreading?



## What does the content from this source usually look like?

- Normalize the input text to reduce dimensionality:



- [Geeky, yet, brilliant], [!], [@modified@, @conversation@, A, card, game, based, on, lexicalized, tree-adjointing, grammar, @url@, @conversation@], [@topic@], [@topics@, grammar], [@topics@, SLPeeps];
- Build a semi-lexicalized, feature-rich channel-based language model (LM);
- Estimate the likelihood that new content published is genuinely from this channel, by using a modified Query Likelihood Model:
  - $P(\textit{tweet}|\textit{channel}) = \sum_{k=0}^n \log(P(K|Q))$   
Where  $k$  represents the semi-lexicalized feature,  $K$  the set of  $n$ -gram features rooted at  $k$ , and  $Q$  the semi-lexicalized channel LM.



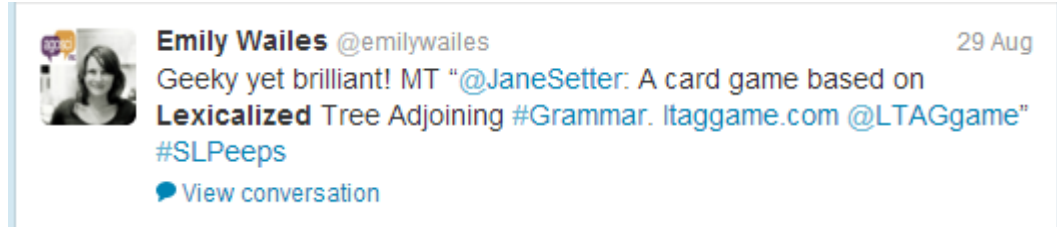
Bloomberg

# Social Publishing

Is it trustworthy?

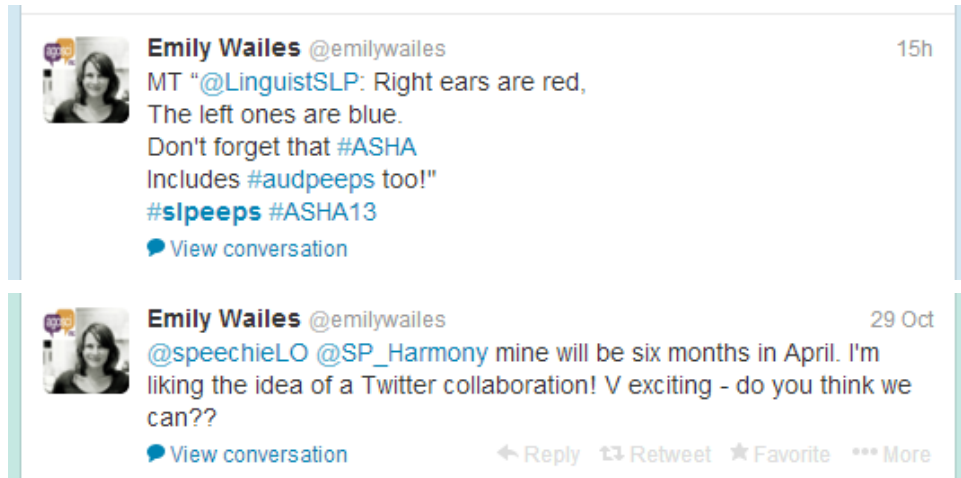


What does the content from this source usually look like?



- People tend to maintain a consistent set of styles across conversations.
  - Threshold of around 0.5 seems to separate genuine user content from spam behaviors;

MT: "Modified Tweet"  
 @Frequent  
 #Consistency  
 N-gram Overlap in LM





Bloomberg

# Social Publishing

Is it trustworthy?



Is this genuine user content?

- Auto-generated template-spam from market activity;



```

<< 30DC Inc, Twitter
Top All
<Narrow Search> Sources TWT Lang My 11/17/13 Sort By Time
1) goarmy123: STOCKGOODIES PLAYS Pintus: PHOT!!! Marijuana $GOL... TWT 11/04
2) goarmy123: $UNGS Nostradamus Stocks: All the $SE*K money will... TWT 11/04
3) goarmy123: STOCKGOODIES PLAYS ManicTrader: I like $BBRY as a... TWT 11/03
4) goarmy123: STOCKGOODIES PLAYS mrchipper: $JCP first off congr... TWT 11/03
5) goarmy123: STOCKGOODIES PLAYS mrchipper: $PAY ATTENTION : Y... TWT 11/03
6) goarmy123: STOCKGOODIES PLAYS ManicTrader: $YELP and i'm ridi... TWT 11/02
7) goarmy123: STOCKGOODIES PLAYS ManicTrader: $HOTR I'm holding... TWT 11/02
8) goarmy123: STOCKGOODIES PLAYS nobluff55: In Agreement on $HO... TWT 11/02
9) goarmy123: Winners Circle Pharaoh: $FNMA Monthly Chart: Pharao... TWT 11/02
10) goarmy123: STOCKGOODIES PLAYS mrchipper: $NFLX at 329 close... TWT 11/02
11) goarmy123: STOCKGOODIES PLAYS mrchipper: $SPY chart we poste... TWT 11/02
12) goarmy123: STOCKGOODIES PLAYS mrchipper: $HOTR continues to... TWT 11/02
13) goarmy123: STOCKGOODIES PLAYS mrchipper: $JCP just hit the 2... TWT 11/01
14) goarmy123: STOCKGOODIES PLAYS myopinion: **SHORT SQUEEZE (... TWT 11/01
15) goarmy123: $UNGS TechnoAnalyst: BLACK $GOLD BABY!!! $UNGS!:... TWT 11/01
16) goarmy123: $MDHI drKen: No lies in the $TWO WEEK trend.: drKen... TWT 11/01
17) goarmy123: STOCKGOODIES PLAYS myopinion: $GLUU 2nd Resistan... TWT 11/01
18) goarmy123: STOCKGOODIES PLAYS myopinion: $$ $MEET $$ **SOCI... TWT 11/01
19) goarmy123: STOCKGOODIES PLAYS l2 hunter: $BBRY GAPPER ON BU... TWT 11/01
20) goarmy123: STOCKGOODIES PLAYS mrchipper: you small cap folks... TWT 11/01

```





Bloomberg

# Social Publishing

Is it trustworthy?



## Are the assertions corroborated?



- The message syntax can provide us with clues;
  - Re-tweets imply agreement, propagation;
  - Citations from trusted sources (news);
- Have we seen this information before?
  - Simple n-gram overlap measures tend to perform poorly:



**ROL - RS/BRASIL** @RSBR\_ROL 4m  
UN confirms sarin gas used in **Syria** attack ( video) [aje.me/16xG7nR](http://aje.me/16xG7nR)  
via @AJEnglish  
[View summary](#) [Reply](#) [Retweet](#) [Favorite](#) [More](#)

**Caleb Little** @CalebLittle6 5m  
Shortly after the #UN confirmed #chemicalweapons were used in #Syria, #WashingtonShooting happens. Coincidence? Is there oil in Syria?  
[Expand](#) [Reply](#) [Retweet](#) [Favorite](#) [More](#)



- Need to move beyond n-grams to entity recognition and propositional structure;





Bloomberg

# Social Publishing

Is it trustworthy?



Are the assertions corroborated?



**ROL - RS/BRASIL** @RSBR\_ROL 4m  
UN confirms sarin gas used in Syria attack ( video) aje.me/16xG7nR  
via @AJEnglish  
View summary Reply Retweet Favorite More

**Caleb Little** @CalebLittle6 5m  
Shortly after the #UN confirmed #chemicalweapons were used in #Syria , #WashingtonShooting happens. Coincidence? Is there oil in Syria?  
Expand Reply Retweet Favorite More



- Dependency Parse of phrasal portions;
  - “UN Confirms” – X – “used”
- Super-sense mapping;
  - “sarin” >> WordNet.Noun.Substance
  - “chemical” >> WordNet.Noun.Substance
- FrameNet *Attack* frame;





Bloomberg

# Social Publishing

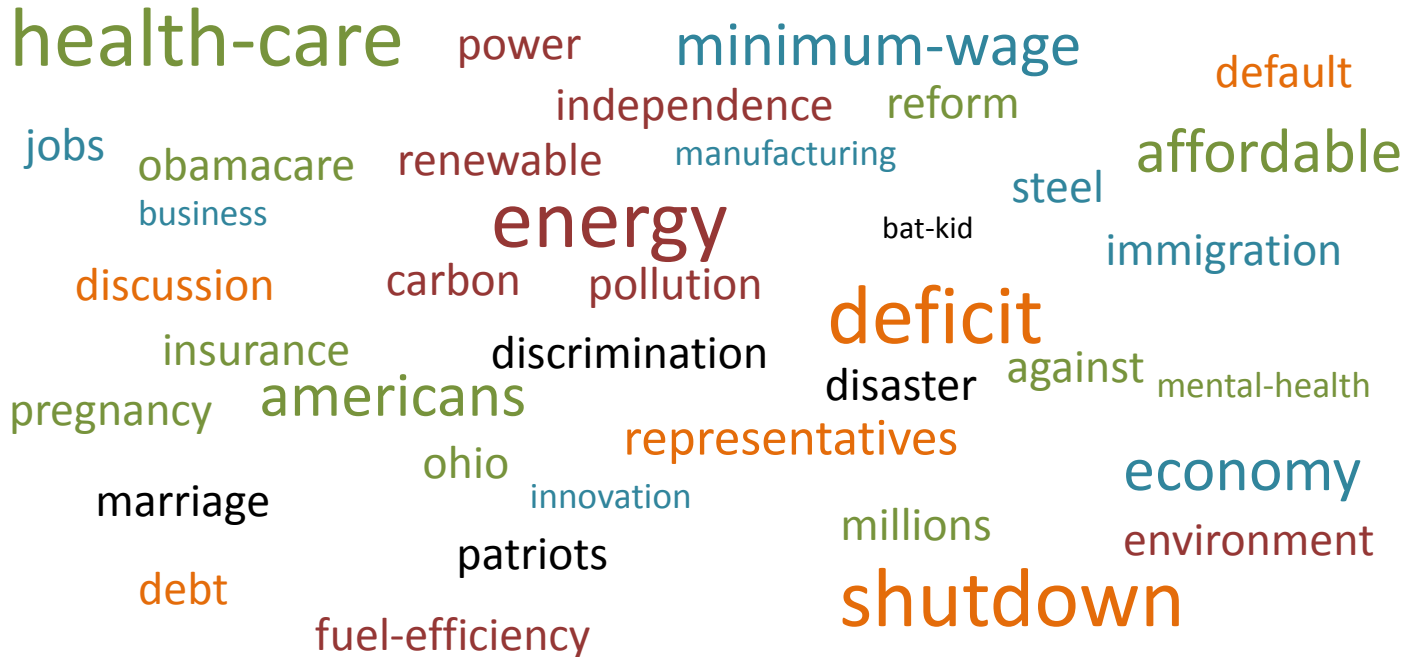
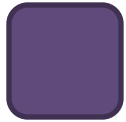
How authoritative?



Do people listen to what is being published?



- Take Obama as an example;
- 6 months of posts give the following (roughly scaled by probability of each topic):





Bloomberg

# Social Publishing

How authoritative?



Do people listen to what is being published?



- Next, associate each source to their most prolific topics;



- $\operatorname{argmax}_{\{t_1, \dots, t_n\} \in T} [\prod_{d \in D} (\sum_{t \in \{t_1, \dots, t_n\}} P(d|t))]$ ;



Obama





Bloomberg

# Social Publishing

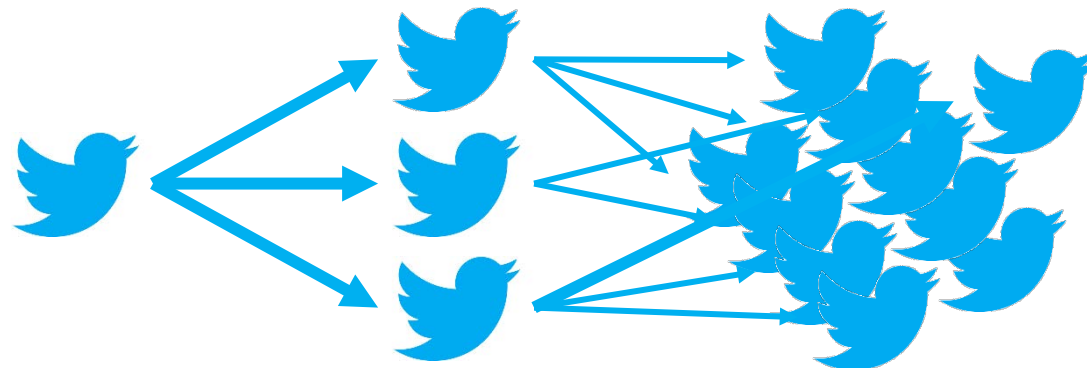
How authoritative?



Do people listen to what is being published?



- For each topic, and each document representative of that topic for the source in question, we want to understand the relationship between prolific topics and their audience:
  - Build a document feature vector based on followers and re-tweets at the time of publication, trends of related #topics, URL's, etc.
  - Use a PageRank-like algorithm:
    - Treat each user-topic as a node;
    - Each message is treated as an edge to the set of peers who interact with it directly or indirectly;
    - Relationship strength decays with distance and time;





Bloomberg

# Social Publishing

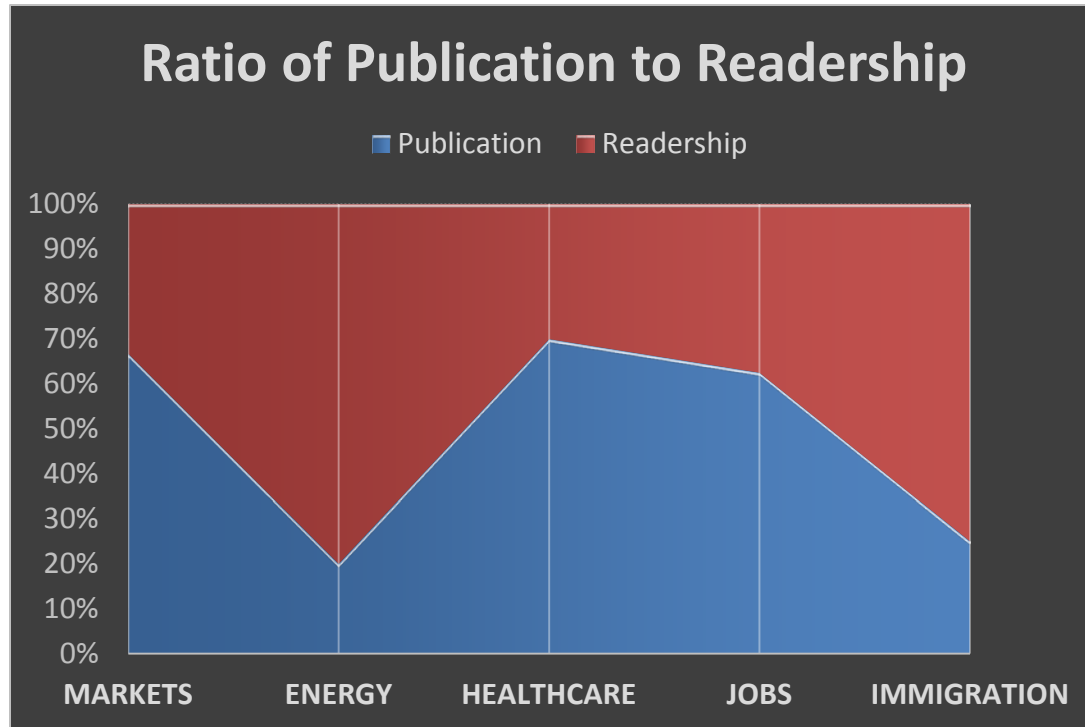
How authoritative?



Do people listen to what is being published?



- For Obama, the graph of topics vs. interaction profile—surprising result for healthcare, but potentially explained by temporal bias?



## What is the social context for this message?

- Track conversations through the network;
- Analyze the strength of a topic by how quickly it emerges/grows/disappears;
- Identify when new important information is being released;
  - High authority score;
  - Unseen statement;
- Predict impact (information propagation through social graph);
- Predict sentiment (positive, negative, neutral)
- Verify impact (market reaction);



Bloomberg

# Social Publishing

What is the impact?



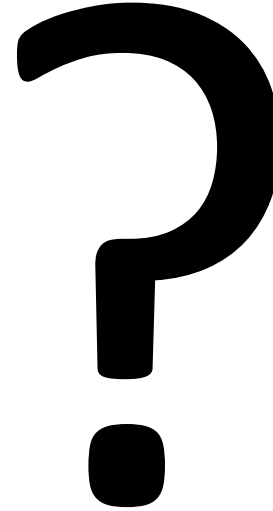




Bloomberg



# Questions:



**[jhodson2@bloomberg.net](mailto:jhodson2@bloomberg.net)**