

# Seeking More Alpha

Dissecting M&A news blurbs  
utilizing unsupervised learning techniques

# Dimensions of a Deal

## Speculation



☐ Source (Reuters; *NY Post*; *WSJ*)

☐ “Exploring strategic options”

☐ Legal/fincl. advisor engagement

## Announcement



☐ Buyer type

☐ Dilutive/accretive to EPS

☐ Payment type

## Update



☐ Regulatory issues

☐ Divestment/arbitration

☐ Approval/deal close

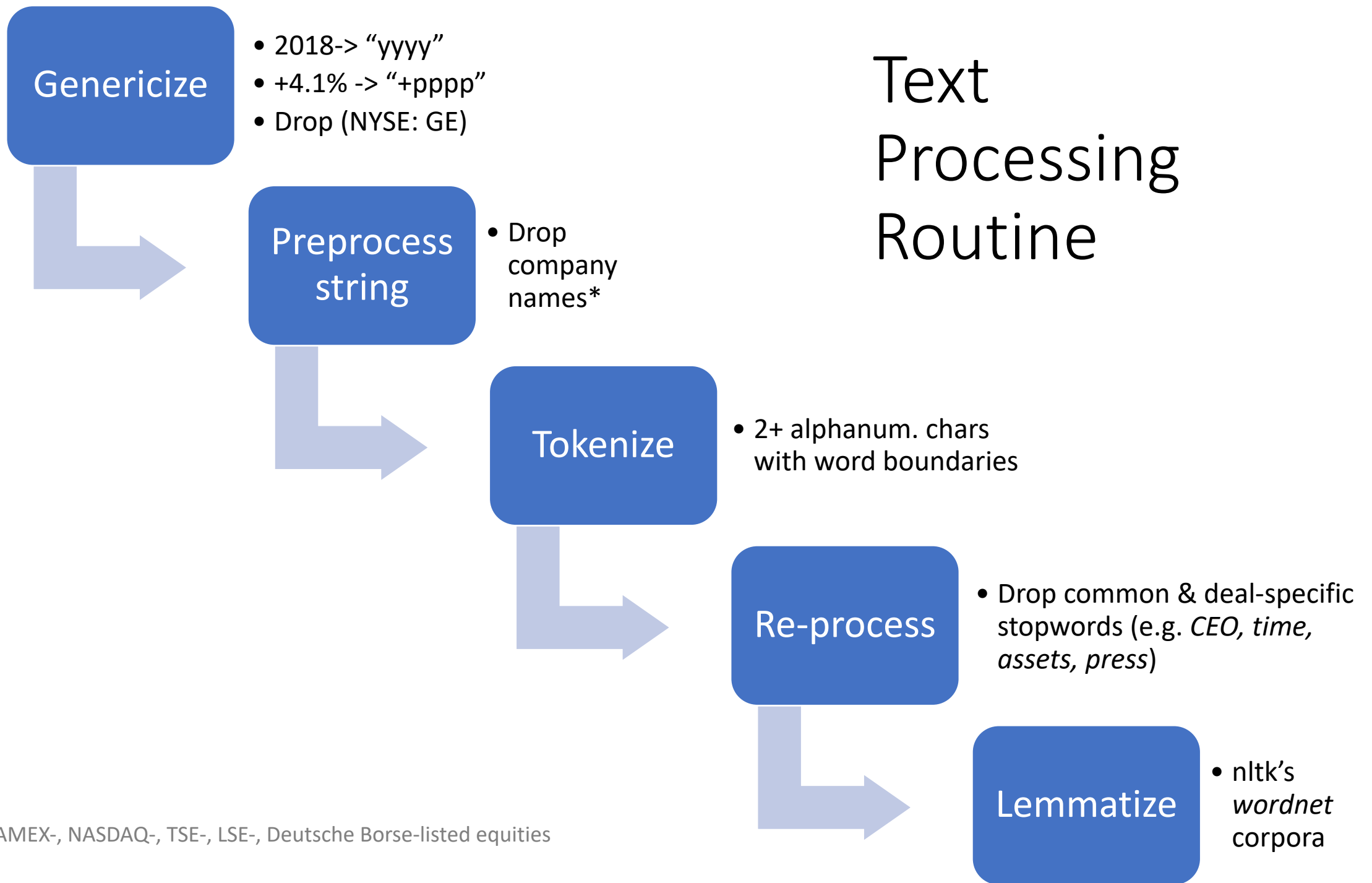
# Data

- SeekingAlpha.com (Financial news, crowd-sourced content)
- 18,061 news blurbs related to mergers & acquisitions
- April 2011 - Feb 2018

The logo for Seeking Alpha, featuring the text "Seeking Alpha" in white serif font and a stylized orange Greek letter alpha ( $\alpha$ ) to the right, all contained within a dark red rectangular background.

Seeking Alpha  $\alpha$

# Text Processing Routine



\*NYSE-, AMEX-, NASDAQ-, TSE-, LSE-, Deutsche Borse-listed equities

# Text Processing Routine: An Example

Raw  
string

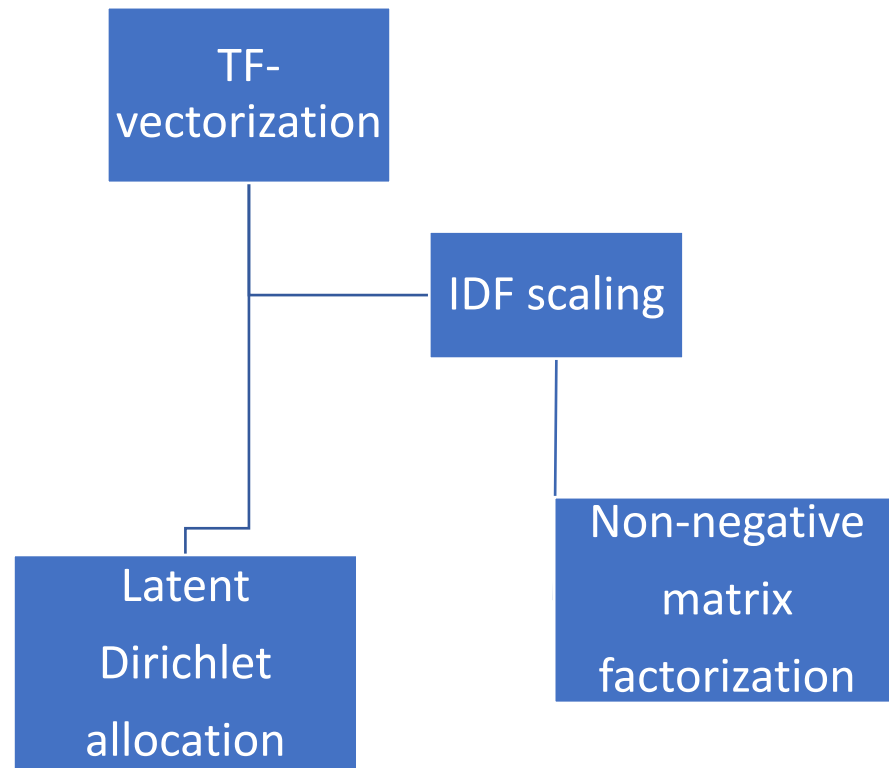
A source close to the matter says Wal-Mart (NYSE:WMT) isn't "actively considering" making an offer for Whole Foods (NASDAQ:WFM). WMT -0.9%, WFM -0.6% Previously: Bidding war on Whole Foods Market seen (June 23).

Fully  
processed

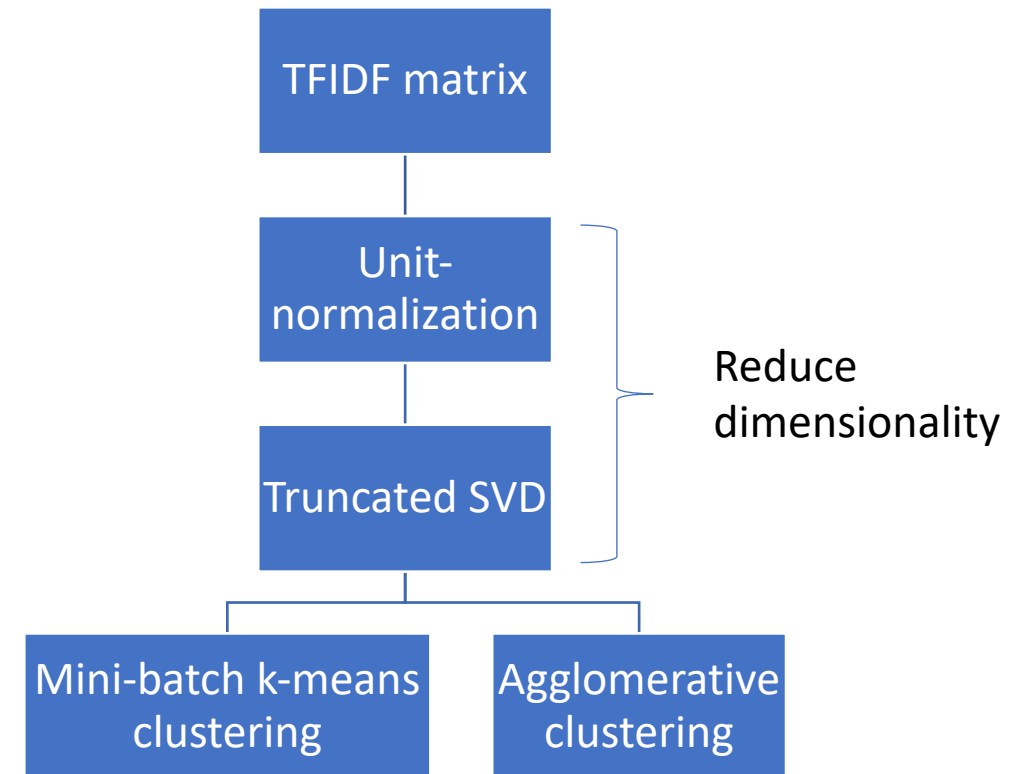
a source close to the matter says \_company\_ isn't "actively considering" making an offer for \_company\_ . \_company\_ \_pppp\_, \_company\_ \_pppp\_ previously: bidding war on \_company\_ seen

# Modeling processes

## Topic modeling



## LSA + Clustering



# Resulting n-grams

TF Matrix: 18061x3908

Type	Frequency
Unigrams	2727
Bigrams	965
Trigrams	183
Quadgrams	33

## Quadgrams

- “antitrust chief makan delrahim”
- “financial term weren disclosed”
- “valued mmmm including debt”
- “shareholder receive mmmm cash”
- “expected immediately accretive eps”

## Trigrams

- “advanced merger talk”
- “chief randall stephenson”
- “expected\_to\_close qqqq yyyy”

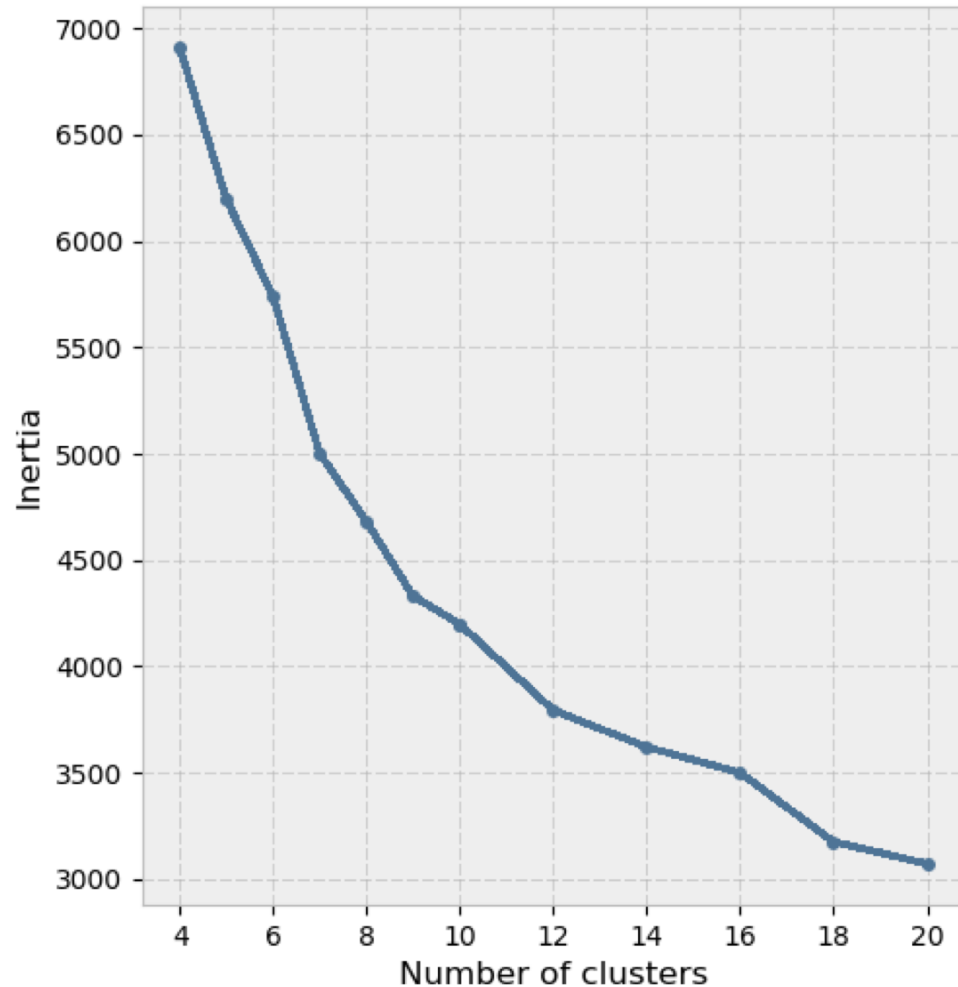
# LDA Topic Matrix: Word Associations

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8
Word 1	mmmm share	yyyy	mmmm mmmm	energy	stake	report	united_states	mobile
Word 2	cash	announce_acquisition	qqqq	announce_acquisition	shareholder	offer	approval	price
Word 3	stock	service	revenue	oil	group	bid	regulator	analyst
Word 4	price	software	medium	gas	board	talk	regulatory	market
Word 5	premium	cash	guidance	unit	capital	bloomberg	proposed	target
Word 6	closing	revenue	intel	partner	pppp stake	sale	antitrust	network
Word 7	pppp premium	based	result	united_states	private	mmmm share	government	yahoo
Word 8	announce_acquisition	provider	chip	production	equity	reuters	takeover	google
Word 9	close	technology	eps	largest	investment	reportedly	charter	buy
Word 10	mmmm cash	product	bayer	natural	investor	week	cable	make
	Announcement	Tech sector	[noise]	Energy sector	Private equity buyer	Speculation	Regulation	Telecom sector

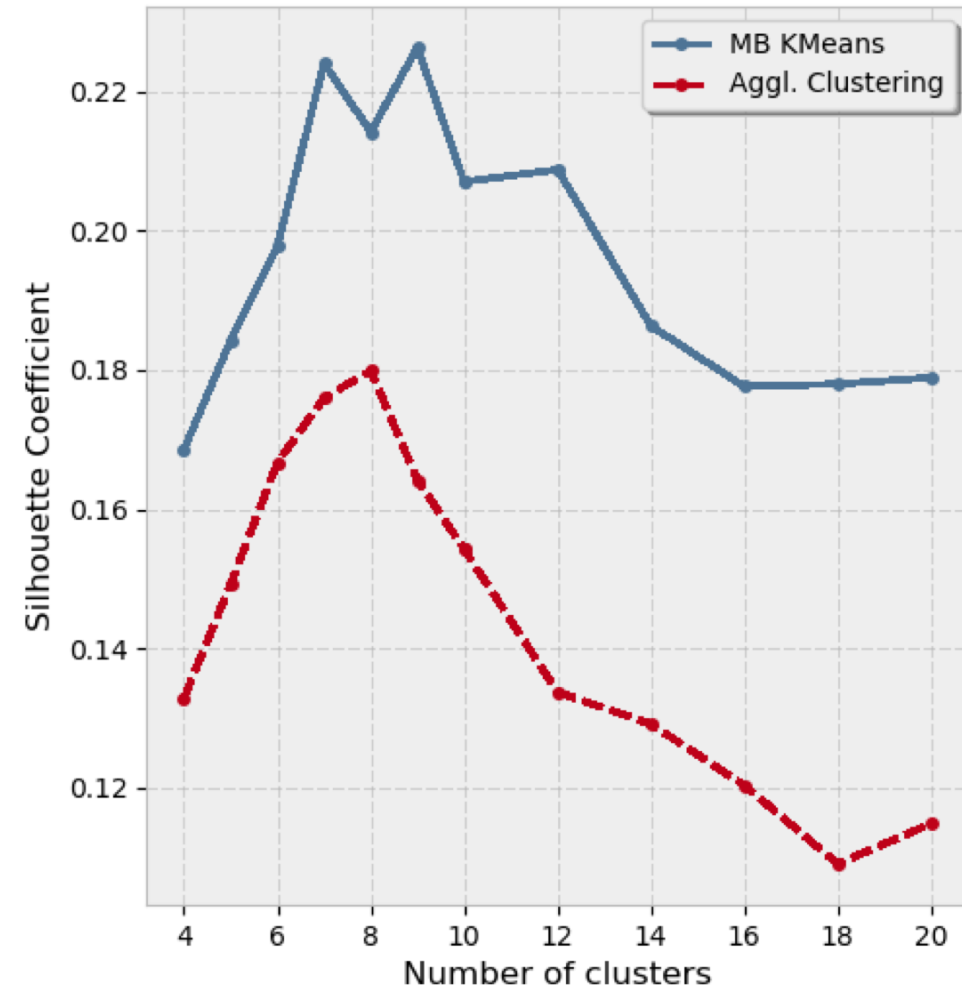


# Clustering metrics

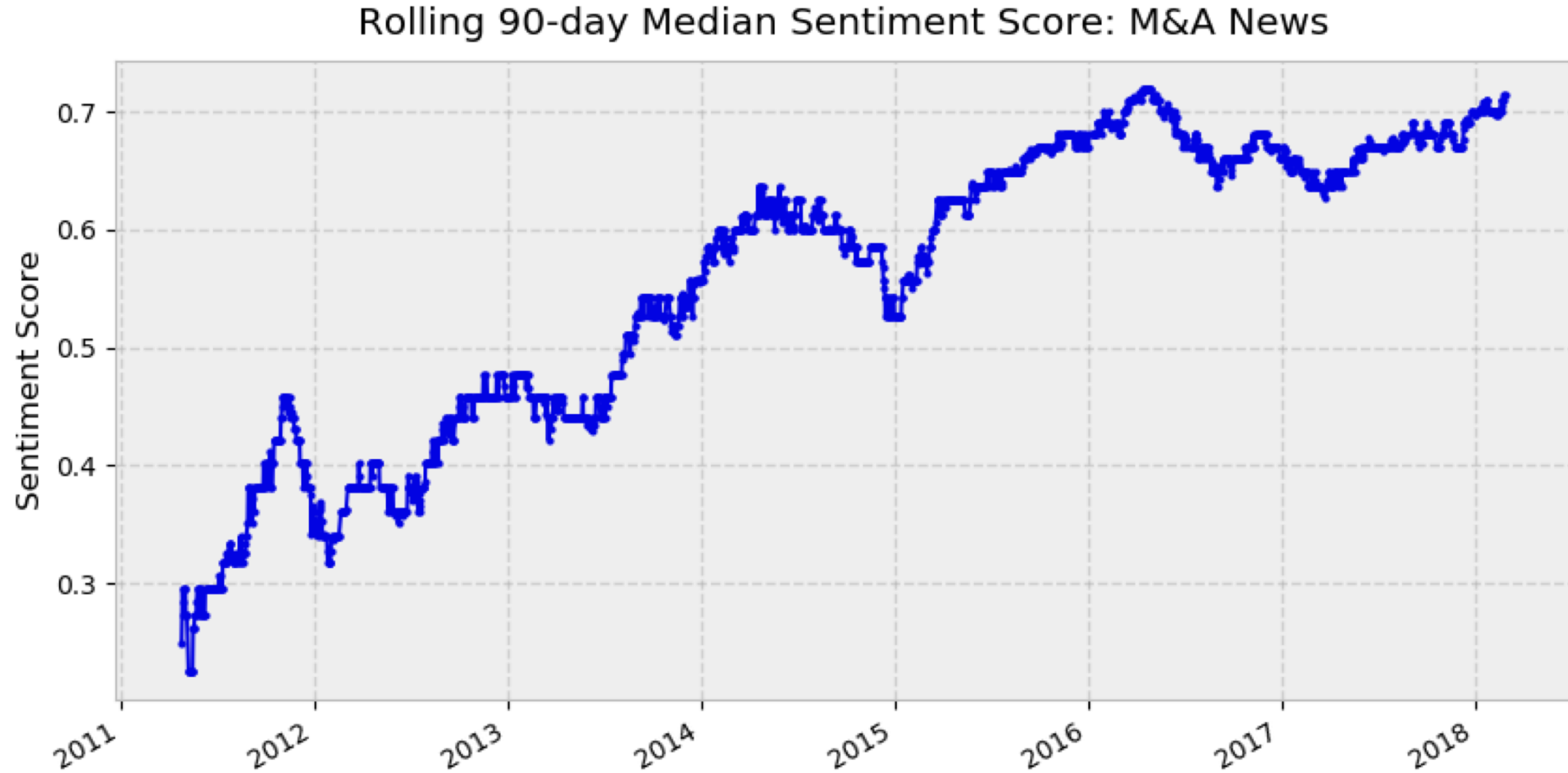
K-Means Inertia as a function of clusters



Silhouette Coefficient as a function of clusters



# Smoothed Sentiment Over Time



Normalized, Weighted Composite Score  
Source: VADER (Valence Aware Dictionary and sEntiment Reasoner)

# Next Steps

- Run clustering algorithms better-suited for uneven cluster size (affinity propagation, mean-shift) in distributed environment.
- Supervised learning: predict changes in arbitrage spreads
  - E.g.  $y$  = forward  $n$ -day arb spread change from article release
- Construct live data feed
  - Real-time aggregated stream from multiple sources (*NYT Dealbook*, *WSJ*)
  - Informational advantage from fast document clustering