2016 U.S. PRESIDENTIAL ELECTION

FAKE NEWS

OVERVIEW

- Introduction
- Main Paper
- Related Work and Limitation
- Proposed Solution
- Preliminary Result
- Conclusion and Future Work

TWITTER: A SOCIAL NETWORK AND A NEWS MEDIA

- Microblogging is a new form of communication in which users can describe their current status in short posts distributed by instant messages, mobile phones, email or the Web.
- Twitter, a popular microblogging tools has been a lot of growth since it launched in October, 2006.
- Twitter has become a new medium of information sharing.

TWITTER PLAYS A IMPORTANT ROLE IN ELECTION

- Both candidates used Twitter as social media platform for online campaigns.
- Millions of voters expressed their views and voting preferences through following and tweeting.
- Social media has been filled with fake news and rumor.
- Fake news and rumors on social media during the election has been criticized to have huge impacts on voters' decisions.

MAIN PAPER: RUMOR DETECTION ON TWITTER

- These are example of two sets of tweets about rumors that posted after the earthquake in Japan on March 11th.
- From the streaming API provided by twitter, there are 6 million tweets that were posted in March 2011.
- Extracting all 6 million tweets with keywords "server room" and "Cosmo Oil" for rumor1 and rumor2 separately, and then manually checked them to eliminate irrelevant tweets.
- As the result, there are 1,135 tweets for rumor1 and 2,042 tweets for rumor2.

Rumor1: Geek house Asakusa (Translated in English):

> When the earthquake occurred, I was in our server room. A rack collapsed on me now. My stomach is pressed and I'm shedding blood. Painful. Somebody help me. Doors look damaged. An announcement tells that no one should move before certain information. There is no time. I can use only my arm. Hard to breath.

Rumor2: Cosmo Oil (Translated in English):

> "Request for spreading" People living in Chiba-city area must use an umbrella or rain coat to avoid touching rain. Due to the explosion in Cosmo Oil Factory, harmful material will fall with rain.

CLUES FOR DETECTING RUMORS

- Burst: For rumors, they will burst suddenly in number of tweets and then it will go down when user posts a correction tweets.
- Retweet ratio: While the general retweet ratio is 8.03%, the retweet ratio for rumor1 is 96.2% and 70.7% for rumor2.
- Difference of word distribution: Extracting a list of content word from tweets by applying morphological analysis and give score. Finding out words with score higher than 1.0.

$$\text{score}(w) = \frac{\text{num_in_correction}(w)}{\text{num_in_rumor}(w)}$$

Rumor1		Rumor2	
word	score	word	score
デマ (false rumor)	630.00	確認 (check)	5690.00
		虚偽 (lie)	2910.00
		消防 (fire fighting)	2630.00
		送る (send)	2580.00
		否定 (negative)	1650.00
		事実 (truth)	1400.00
		:	:
		デマ (false rumor)	259.00

RUMOR CONTENT ANALYSIS

- To distinguish rumors from gossip or campaigns, it need content analysis. It may need deep semantic recognition from text, and it requires background knowledge to judge the tweets.
- Summarization by tag cloud consists of weighted keyword list can help. TF-IDF is a well known term weighting method, that works to extract keyword to summarize a set of tweets.
- These keywords help companies to recognize that a burst is a rumor or the result of their campaign.

word	TFIDF	*Frequency*
コスモ石油 (Cosmo Oil)	17248.70	2116
有害 (harmful)	10977.46	1405
雨 (rain)	10898.68	1847
物質 (material)	8777.04	1385
爆発 (explosion)	6862.35	1129
降る (fall)	6852.59	1102
付着 (attached)	5668.96	646
傘 (umbrella)	5415.50	703
外出 (go out)	4962.68	681
カッパ (rain coat)	4861.93	552

EXPERIMENT TO DETECT RUMORS

- Named Entity Extraction: Extracting named entities which occur more than 30 times at least in one day. 1,976 unique named entities were obtained. They were used as target source for the following steps.
- Filtering by Retweet Ratio: the system calculate the retweet ratio on each day in March for every targets. After filtering with high retweet ratio, 1,015 targets remained to be selected. Most of them were topics that widely spread, however, not all the topic were rumors.
- Filtering by Clue Keyword: Using keyword "false rumor" to detect rumors from the candidates. This keyword appears in 19.0% (in rumor1) and 12.5% (in rumor2) of tweets. So the system use 10% as the threshold to detect the topic is a rumor or not.

LIMITATIONS

- Many of the threshold parameters are not well established such as those threshold that filter out small topic name entity and filter out low retweet ratio tweets.
- The difference of world distribution that used to extract keyword for rumor detection is limited within specific topic. Such keyword found in rumor2 is not suitable for rumor1. and thus produce a lot of mis- classification.

RELATED WORK

DECEPTION DETECTION WITH DEEP SYNTAX

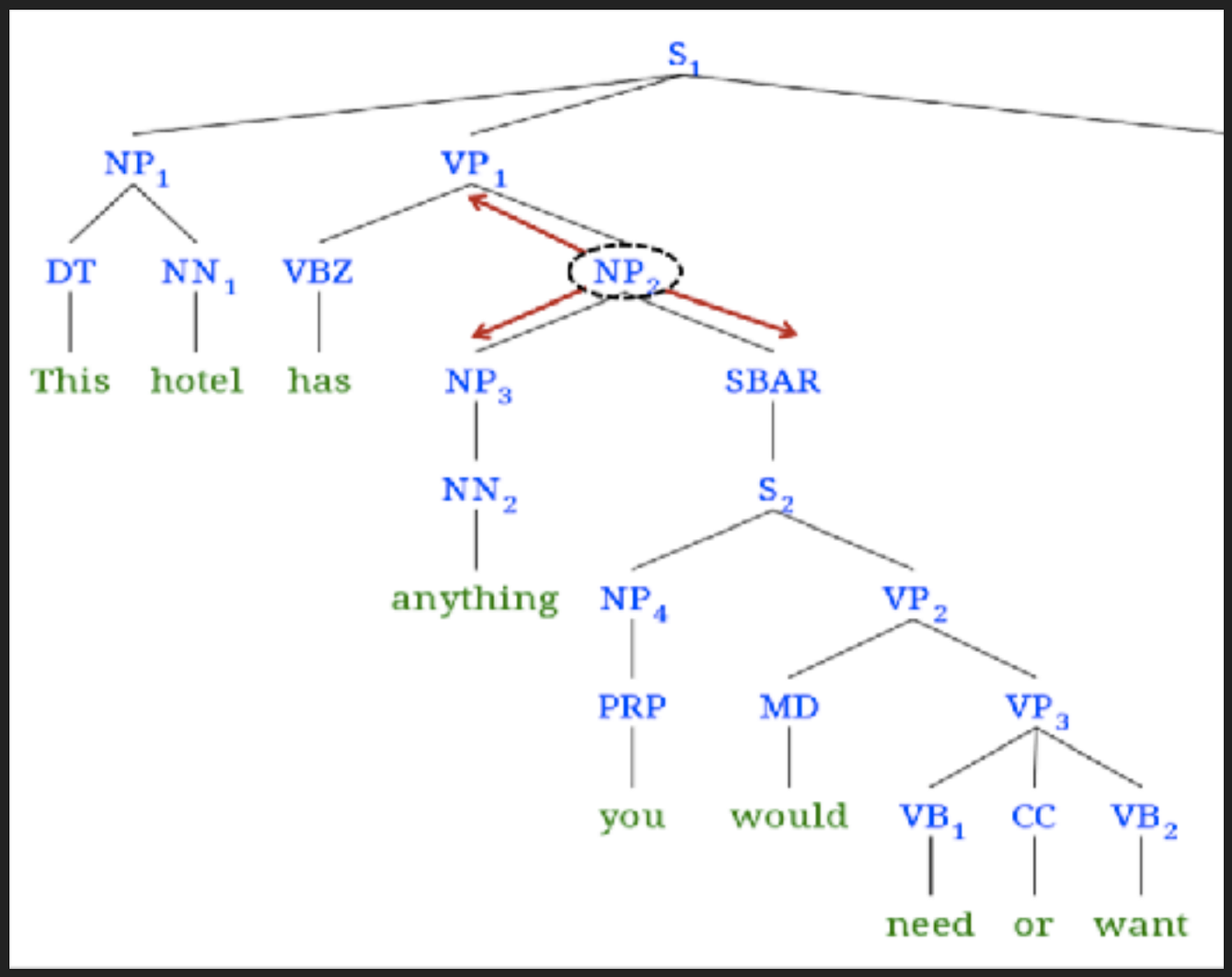
(Feng et.al) proposed a SVM classifier to detect deceptive reviews using features:

- simple unigram from bag-of-word
- rules driven by Probabilistic Context Free Grammars

Testing mainly on a TripAdvisor dataset that contains:

- 400 trustful reviews
- 400 deceptive reviews
- gathered by Amazon Mechanical Turk and evenly distributed cross 20 Chicago hotels
- achieved remarkable 91.2% accuracy

PCFG



ADJP	Adjective phrase
ADVP	Adverb phrase
NP	Noun phrase
PP	Prepositional phrase
S	Simple declarative clause
SBAR	Subordinate clause
SBARQ	Direct question introduced by *wh*-element
SINV	Declarative sentence with subject-aux inversion
SQ	Yes/no questions and subconstituent of SBARQ excluding *wh*-element
VP	Verb phrase
WHADVP	Wh-adverb phrase
WHNP	Wh-noun phrase
WHPP	Wh-prepositional phrase
X	Constituent of unknown or uncertain category
*	"Understood" subject of infinitive or imperative
0	Zero variant of *that* in subordinate clauses
T	Trace of wh-Constituent

TripAdvisor–Gold	
Deceptive	Truthful
NP^PP → DT NNP NNP NNP	S^ROOT → VP .
SBAR^NP → S	NP^NP → # CD
NP^VP → NP SBAR	PRN^NP → LRB NP RRB
NP^NP → PRP$ NN	NP^NP → NNS
NP^S → DT NNP NNP NNP	NP^S → NN
VP^S → VBG PP	NP^PP → DT NNP
NP^PP → PRP$ NN	NP^PP → CD NNS
VP^S → MD ADVP VP	NP^NP → NP PRN
VP^S → TO VP	PRN^NP → LRB PP RRB
ADJP^NP → RBS JJ	NP^NP → CD NNS

CONTENT-BASED MACHINE LEARNING

(Castillo et.al) proposed a supervised machine learning algorithm:

- automatically separates newsworthy topics from other type of conversations
- assess automatically the level of social media credibility of newsworthy topics

supervised process:

- Twitter Monitor extract burst topic keywords
- Mechanical Turk (real human being evaluators) judge whether a topic is a newsworthy topics or a personal comment or conversation topic. also used to judge the credibility of a topic (true, likely false, certain false)

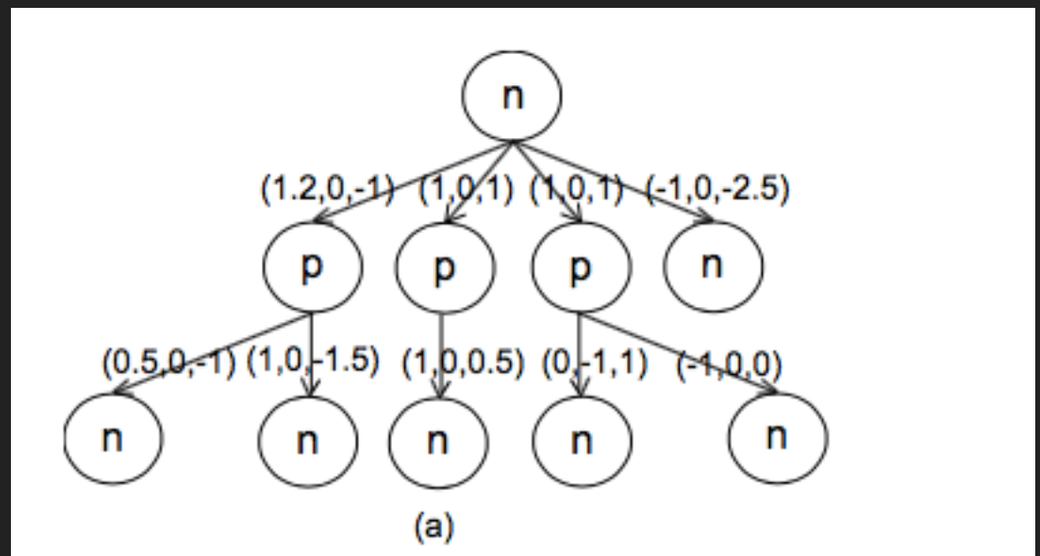
PROPAGATION-PATTERN BASED MACHINE LEARNING

(Ke Wu et al.) proposed a Support vector machine algorithm that feed:

- 2,601 original rumor message from Sina Weibo community rumor measurement center, mixed with other 2,536 normal message.
- Propagation tree for all this 2,601 rumor.
- 23 features, 15 features from Castillo, and extra 8 features such as topic type, user type(verification) and other sentiment relative features.

PROPAGATION-PATTERN BASED MACHINE LEARNING

- N node is regular user, P node is opinion leader who has more follower than following and at least 1,000 follower.
- The value above each node is
  - approve score
  - doubt score
  - sentiment score



LIMITATION OF RELATIVE WORK

- All required human interference to flag a rumor tweets to construct the training dataset.
- Computational expensive to drive the features. The lack of full plot of features can result in bias in the classification process.
- Most of the algorithm focus the detection of a fake news topic, not focus who is spreading the fake information, thus hard to mitigating or regulating such behavior.

PROPOSED SOLUTION

TF-IDF & COSINE SIMILARITY

I like music, but I don't like movie.
I don't like music, and I don't like movie.

↓

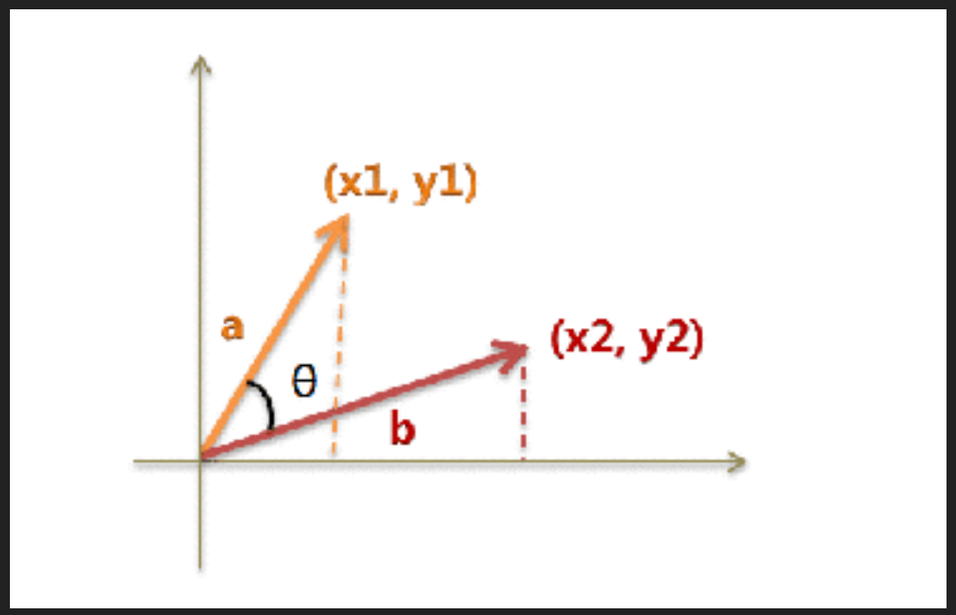
[I, don't, like, music, movie, and, but]

↓

I (2), don't (1), like (2), music (1), movie (1), but (1)
I (2), don't (2), like (2), music (1), movie (2), and (1)

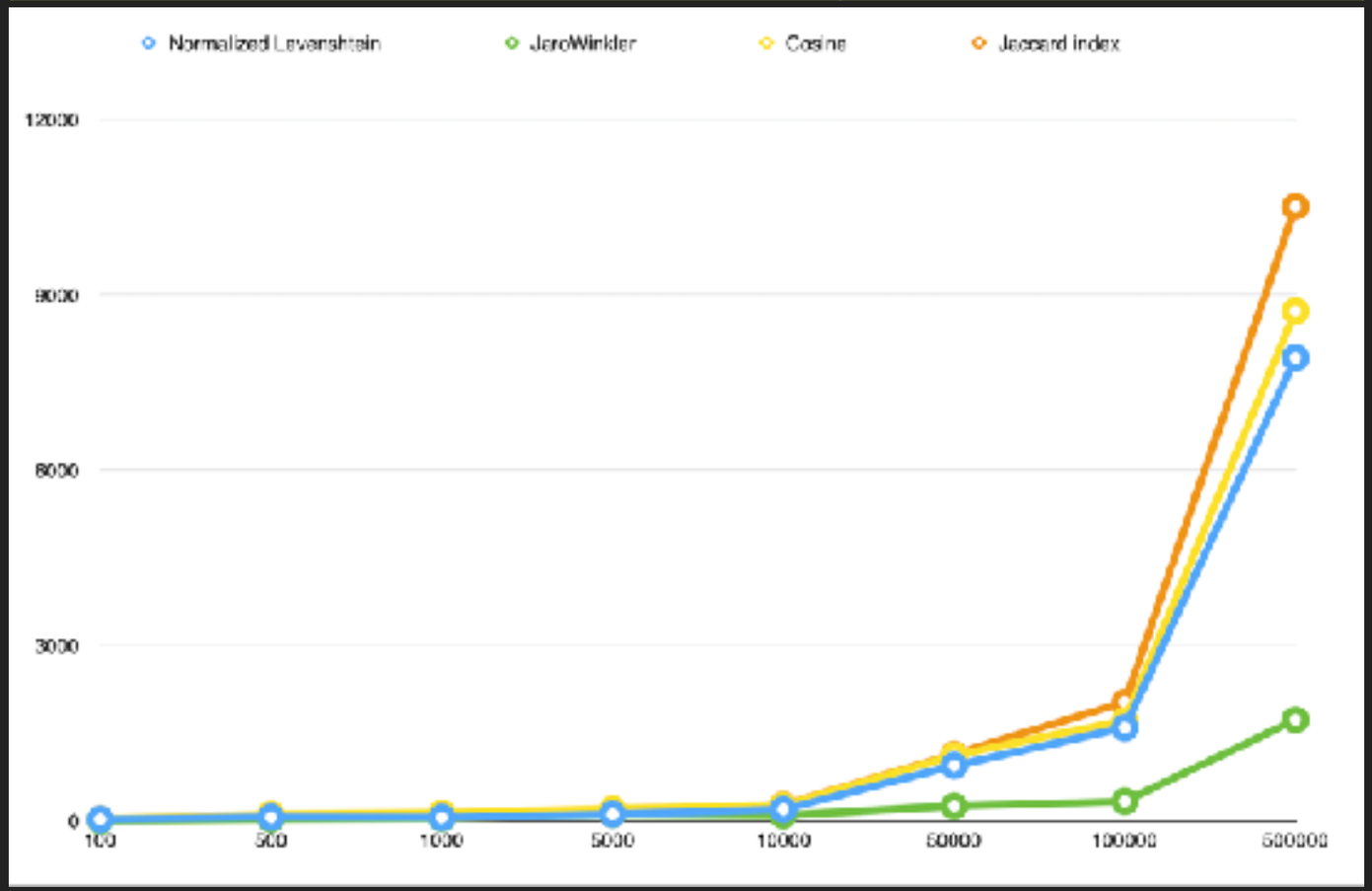
↓

X: [2, 1, 2, 1, 1, 1]
Y: [2, 2, 2, 1, 2, 1]



DISTANCE & SIMILARITY

Normalized Levenshtein	O(m*n)
Jaro-Winkler	O(m*n)
Cosine Similarity	O(m+n)
Jaccard Index	O(m+n)



Donald Trump's Intelligence Quotient

A chart purportedly showing Donald Trump's IQ ranking compared to that of other presidents is based on incomplete and inaccurate information.

CLAIM

Donald Trump's IQ, at 156, is comparable to that of the smartest U.S. presidents.

RATING

FALSE

ORIGIN

150 MILLION TWEETS

121 days, July 13 ~ Nov 10

Dev

STRING MATCHING

0.957173

CLAIM

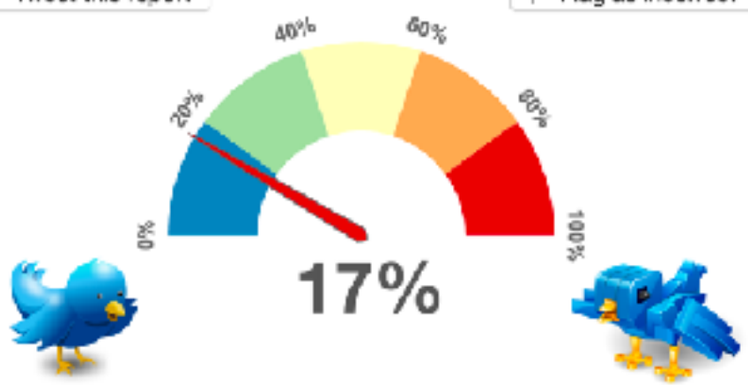
Donald Trump's IQ, at 156, is comparable to that of the smartest U.S. presidents.

BOTORNOT

Classification

Tweet this report

Flag as incorrect



Friend 23%

User 14%

Content 30%

Temporal 14%

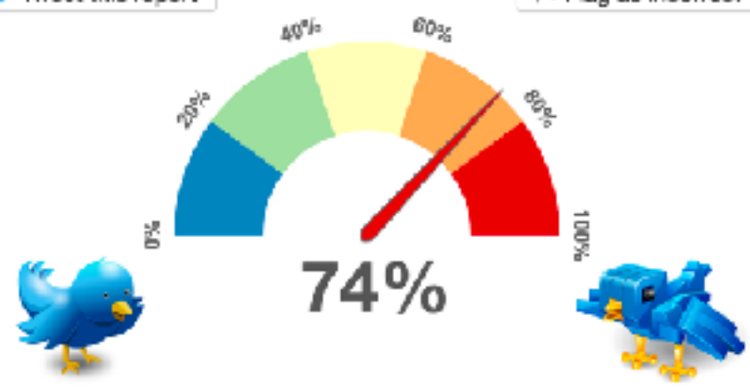
Sentiment 13%

Network 40%

Classification

Tweet this report

Flag as incorrect



Friend 79%

User 99%

Content 78%

Temporal 74%

Sentiment 61%

Network 96%

Mon Oct 17 16:25:26 EDT 2016	Pence: "Michelle Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/kE7WeowsEt https://t.co/TwHaVsWzsU	@DeplorableJeff	12730	830	741
Mon Oct 17 16:25:26 EDT 2016	Pence: "Michelle Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/uY1AqtbcSI #News https://t.co/fbECYMBQra	@NationPolitic	51219	1966	1675
Mon Oct 17 16:25:26 EDT 2016	Pence: "Michelle Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/4D5Qp4hy80 https://t.co/mRcbGxVA6a	@Trump_Rvt	4629	4257	3270
Mon Oct 17 16:25:25 EDT 2016	#News Pence: "Michele Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/1JoDqEMm3U https://t.co/m4CIEIRL3I	@Trumppridetv	81671	3700	3386
Mon Oct 17 16:25:25 EDT 2016	#News Pence: "Michele Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/bCl0J710hA https://t.co/Tj1PWtXP8Y	@Trumpleecs	50661	1079	1524
Mon Oct 17 16:25:25 EDT 2016	#News Pence: "Michele Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/nO5MeP4Uj3 https://t.co/2Na8JDpMp	@AILEENTRUMP	45844	1555	1231
Mon Oct 17 16:25:25 EDT 2016	Pence: "Michelle Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/UnXrgRgrN1 https://t.co/63DeYYC5uY	@Vinzle4Trump	29852	1089	3762
Mon Oct 17 16:25:30 EDT 2016	Pence: "Michelle Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/xJRxHF3R5n https://t.co/STdTSjR2h2	@nationalnewsb	3133	734	1624
Mon Oct 17 16:25:30 EDT 2016	Pence: "Michelle Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/RmUjmRmaHH https://t.co/AOKjH5H2mW	@NewsTodayUS	14583	7362	7386
Mon Oct 17 16:25:30 EDT 2016	Pence: "Michelle Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/45z2IwvE5a https://t.co/cR66cv8Glu	@TPolitcmanager	113005	18788	12707
Mon Oct 17 16:25:31 EDT 2016	Pence: "Michelle Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/5xXmoFR1t https://t.co/1RJ9OcGWdV	@TheRiseOfTrump	19171	1010	4766
Mon Oct 17 16:25:31 EDT 2016	Pence: "Michelle Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/CK2wcoWUEE https://t.co/aWZOW1d43V	@VictoriaCorleto	29808	5044	5158
Mon Oct 17 16:25:33 EDT 2016	#News Pence: "Michele Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/YYDO4ZdmhD https://t.co/aJzbF4lbG4	@BreakitheNewsUSA	53902	92	135
Mon Oct 17 16:25:33 EDT 2016	Pence: "Michelle Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/1WJTKQgLHW https://t.co/AsiejfFFIV	@InsiderToday	11831	24	20
Mon Oct 17 16:25:33 EDT 2016	Pence: "Michelle Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/CQmWOz07aW https://t.co/WZX3wYDL1G	@Theycallicim	49258	5750	4687
Mon Oct 17 16:25:34 EDT 2016	Pence: "Michelle Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/Ioefmzrjet https://t.co/VBsel2Zjw5	@Bestamericanew	21448	1013	2532
Mon Oct 17 16:25:50 EDT 2016	Pence: "Michelle Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/AcJIG6pen9 https://t.co/W1Hg5AlOB	@usa_today_24	15753	1178	4793
Mon Oct 17 16:25:50 EDT 2016	Pence: "Michelle Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/95s7Jxo6F2 https://t.co/k7y9E3y9vB	@Icarus_Politic	10335	612	2203
Mon Oct 17 16:25:50 EDT 2016	Pence: "Michelle Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/YSvGQWAdC https://t.co/aVj4HsK1AW	@uc1NEWp	17089	1086	1676
Mon Oct 17 16:25:51 EDT 2016	Pence: "Michelle Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/HeEjYlYnah https://t.co/9rfSvaaXfL	@EBussines64Now	8749	788	2628
Mon Oct 17 16:25:51 EDT 2016	Pence: "Michelle Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/xHF1bx18Q https://t.co/OcMMdaHE9v	@newsolpolitic16	22603	589	2251
Mon Oct 17 16:25:52 EDT 2016	Pence: "Michelle Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/9rbiTv7A5U https://t.co/yhq42HHUDa	@trumpamericann	40607	4585	4691
Mon Oct 17 16:25:52 EDT 2016	Pence: "Michelle Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/AA5cDI6NbF https://t.co/ilTlvPZW7V	@TheNewsOfUSA	35806	161	378
Mon Oct 17 16:25:52 EDT 2016	Pence: "Michelle Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/1AI9E8Neh2 https://t.co/Ba45EZvNMo	@New_Trump	13129	2183	3662
Mon Oct 17 16:25:52 EDT 2016	Pence: "Michelle Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/xvTab6JNka https://t.co/KLzONIIMLC	@NoticiasErick	14546	790	2615
Mon Oct 17 16:25:53 EDT 2016	Pence: "Michelle Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/BazQTZVAC1 https://t.co/gIGMthR6vu	@AnnaAlvarez1992	24453	1491	1278
Mon Oct 17 16:25:53 EDT 2016	Pence: "Michelle Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/uxcsGJmabe https://t.co/8%LXaRbGf	@politic_report	8867	4892	6962
Mon Oct 17 16:25:54 EDT 2016	Pence: "Michelle Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/mCp1mwD630 https://t.co/0QVaEqE1Vk	@Politician_News	48958	5574	4595
Mon Oct 17 16:25:54 EDT 2016	Pence: "Michelle Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/hmcHXgBFz2 https://t.co/2E0rNovreT	@Politic2k	54171	4597	4726
Mon Oct 17 16:25:54 EDT 2016	Pence: "Michelle Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/deRBCP4cdO https://t.co/3qEEBCP93n	@worldsmichome158	17488	3080	3760
Mon Oct 17 16:25:54 EDT 2016	Pence: "Michelle Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/zPgmqK1scl https://t.co/BiUb20PDHW	@ThePrinceOfTrump	19178	1010	4766
Mon Oct 17 16:25:54 EDT 2016	Pence: "Michelle Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/6rj5C2JG7G https://t.co/CxoumxuGZe	@Fresh_News_24	61615	5345	5040
Mon Oct 17 16:25:55 EDT 2016	Pence: "Michelle Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/GrzU7YLySF https://t.co/pusNm3ogHZ	@POLITICS_USA2	34959	32	271
Mon Oct 17 16:25:56 EDT 2016	Pence: "Michelle Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/87FdLwrgbzw https://t.co/9PJlbc8lIu	@CNNManager	37041	1815	5091
Mon Oct 17 16:25:58 EDT 2016	Pence: "Michelle Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/ePwYpSTiAh https://t.co/pcl9CWIEN	@TheTrumpNews	46591	1872	3396
Mon Oct 17 16:26:16 EDT 2016	Pence: "Michelle Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/NKsUzbCfUv https://t.co/Q8zEsSdBq9	@thenews_paper	1252	422	1022
Mon Oct 17 16:26:16 EDT 2016	Pence: "Michelle Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/CCEISHbago https://t.co/Xdsr7Z0sUZ	@UsaNewsEnglish	3313	1119	3387
Mon Oct 17 16:26:16 EDT 2016	Pence: "Michelle Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/K7PwsGTJ0a https://t.co/DVzhmxlGu2	@PolitiNewsUSA	625	494	1730
Mon Oct 17 16:26:16 EDT 2016	Pence: "Michelle Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/pNuADYNFmW https://t.co/mSLN8FlwKo	@Politicsmessenger	15429	2410	4066
Mon Oct 17 16:26:16 EDT 2016	Pence: "Michelle Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/e26VFlkvjm https://t.co/1VdfYjNRkO	@PolicyArea	10675	817	5069
Mon Oct 17 16:26:16 EDT 2016	Pence: "Michelle Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/hmcHXgBFz2 https://t.co/5odMCMGcJb	@Politic2k	54177	4597	4726
Mon Oct 17 16:26:22 EDT 2016	Pence: "Michelle Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/ggFVxHbZ0I https://t.co/ech40rGmtv	@TrumpGo_2016	2086	1945	4276
Mon Oct 17 16:27:13 EDT 2016	Pence: "Michelle Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/ogKpUQBzQA https://t.co/jpNvFcpCJp	@TrumpTrainNews	3905	6693	9531
Mon Oct 17 16:27:15 EDT 2016	Pence: "Michelle Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/YmDO4vBIHC https://t.co/TaPsw3zzsO	@PolitidWorldONE	75	19	284
Mon Oct 17 16:27:15 EDT 2016	Pence: "Michelle Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/NMc5TDxCLT https://t.co/bU9Jnc3bx47	@alexiekelly2021	3804	212	330
Mon Oct 17 16:27:16 EDT 2016	Pence: "Michelle Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/IcevcwdK1 https://t.co/WVGuHNMX0r	@breaking_news	9503	773	2294
Mon Oct 17 16:27:16 EDT 2016	Pence: "Michelle Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/Eqxz40R0Hc https://t.co/vGHn1FWqT	@mywonlinews	9855	868	2626
Mon Oct 17 16:27:17 EDT 2016	Pence: "Michelle Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/D0uYaf1sm1 https://t.co/DBqRAWaZxx	@politics_today4	2115	1518	2080
Mon Oct 17 16:27:17 EDT 2016	Pence: "Michelle Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/azSSIADEDc https://t.co/7HFK3vDmCu	@julyrose299	435	468	765
Mon Oct 17 16:28:23 EDT 2016	Pence: "Michelle Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/x4bflKRHc https://t.co/626kAZvR9	@Foxnewsplus	4627	539	981
Mon Oct 17 16:28:33 EDT 2016	Pence: "Michelle Obama Is The Most Vulgar First Lady We've Ever Had" https://t.co/N23coI19eos https://t.co/Ga1cm2fSx1	@AmericanNews200	36284	4393	6696

CONCLUSION