Science Volume 359(6380):1146-1151 March 9, 2018

The spread of true and false news online

Soroush Vosoughi, Deb Roy, Sinan Aral
Massachusetts Institute of Technology

Major Takeaways

- Investigates how falsity spreads differently from the truth and what factors of human judgment explain these differences?
  - In general, Falsehood spread farther, faster, deeper and more broadly than truth
- Humans, not robots are more likely to spread false news. Bots accelerated the spread of true and false news at the same rate

Major Takeaways

- User characteristics and network structure could not explain why truth and falsity diffuse or spread at different rates. It was the content

- False news were measurably more novel than true news and was also perceived as more “novel” than true news by humans

Setting - The spread of true and ~~fake~~ false news online

- News - Any story or claim on Twitter with an assertion in it, regardless of its institutional source

- True news Vs. False news
  - “Fake news” no longer associated with veracity[1] of a news story
  - Not making any claim regarding the intention with which it was spread

[1]conformity to the facts

Question

How do they know what is true and what is false in a post-truth world

- Investigated by six independent fact-checking orgs[1] with veracity of 95 to 98 percent agreement
- Fact checking agencies have their own biases. No single agency is trusted across all ideological spectrums

[1] snopes.com, politifact.com, factcheck.org, truthorfiction.com, hoax-slayer.com, urbanlegends.about.com

Reliable Sources of News

No correlation exists between the degree to which the American public finds a “source” reliable and the fraction of verified stories that are true

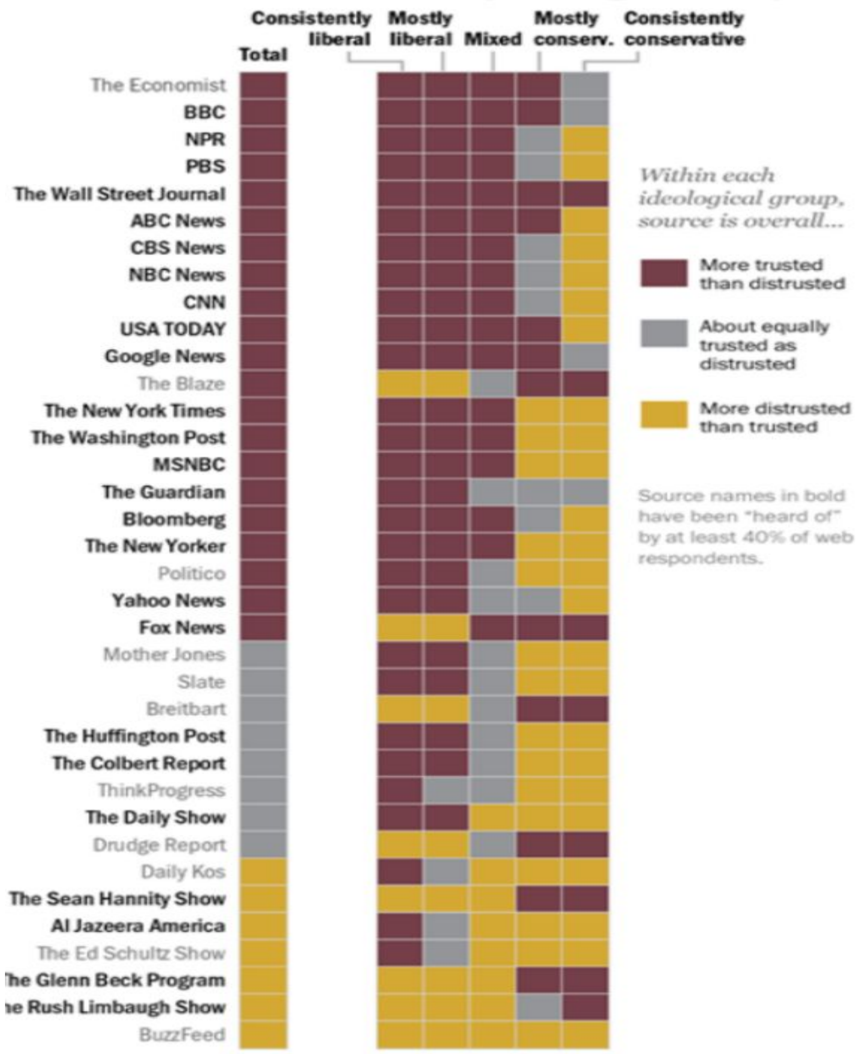


Fig 1 - Trust Levels of News Sources by Ideological Group by Pew Research Centre

News to Rumor

- Rumor - Social phenomena of a news story or claim spreading or diffusing through the Twitter network.
  - User makes an assertion on a topic
  - People can retweet or may also reply to the original tweet

- Rumor cascade - Exhibit an unbroken retweet chain with a common, singular origin(original tweet)

No. of cascades of a Rumor = No. of times the rumor was independently tweeted by a user(not retweeted)

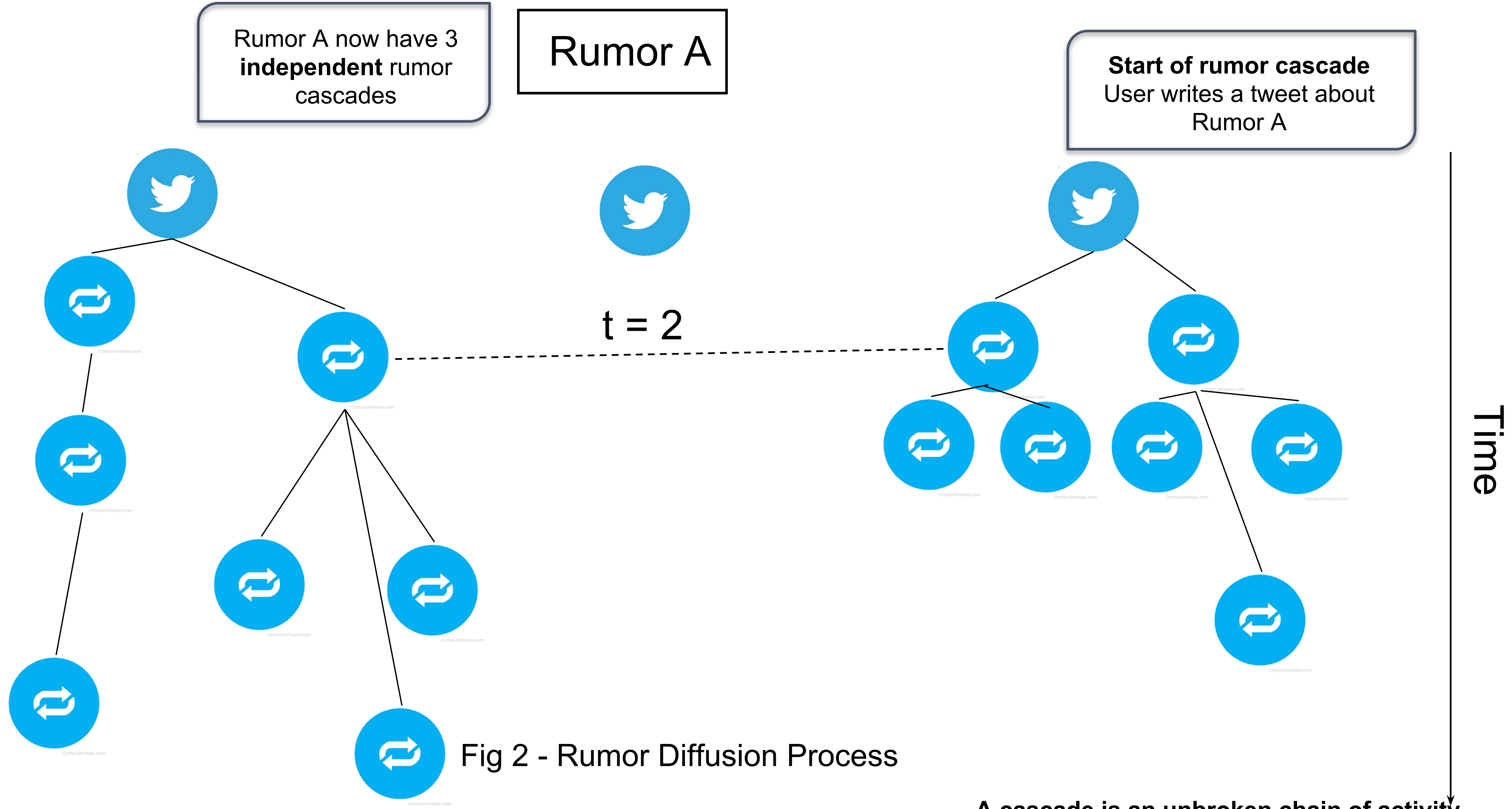


Fig 2 - Rumor Diffusion Process

A cascade is an unbroken chain of activity, centered around the tweet activity of an individual

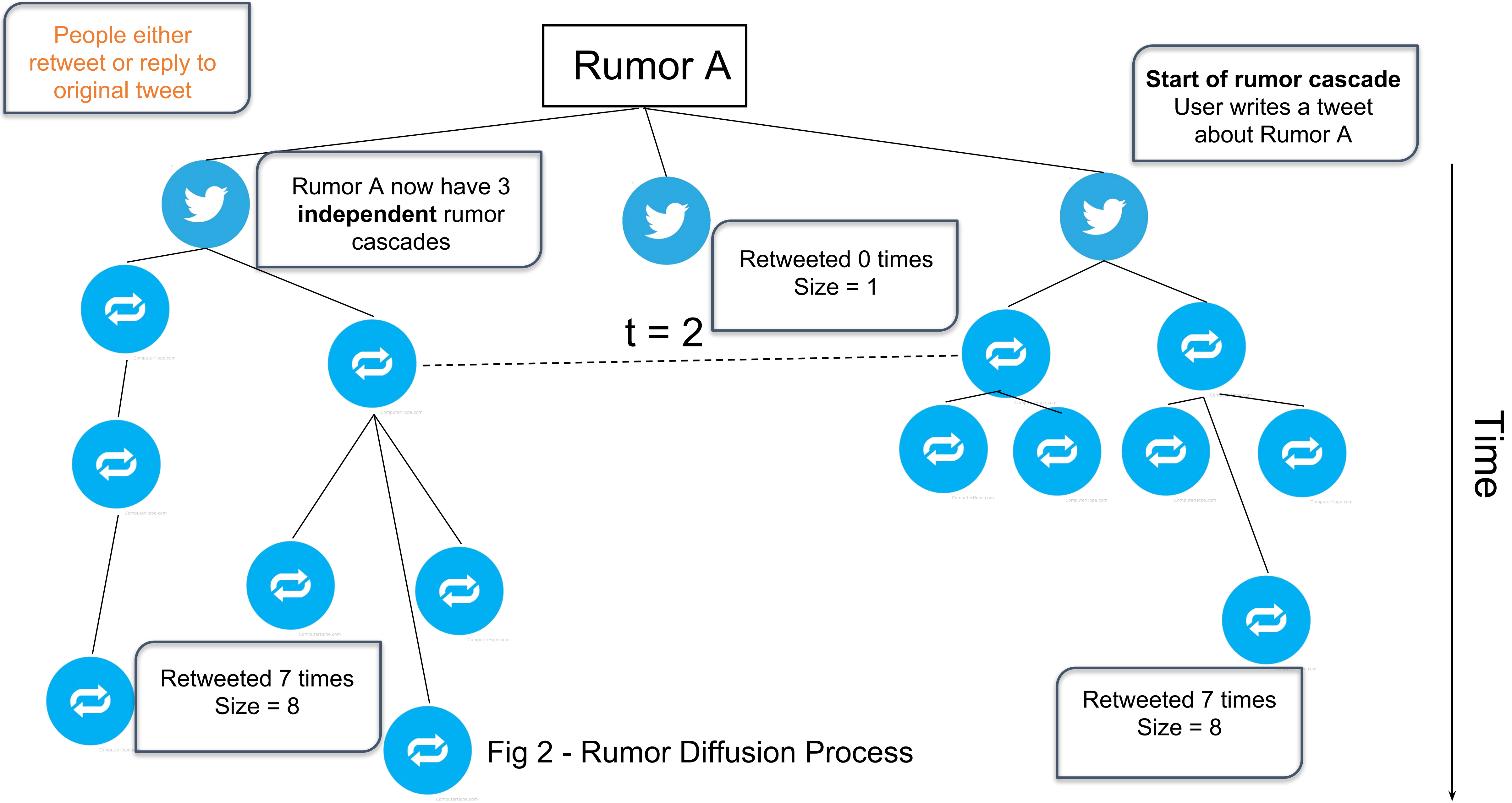


Fig 2 - Rumor Diffusion Process

Rumor Dataset

- Collected ~126,000 rumor cascades spread by ~3 million people on Twitter for 11 years

Verdict	Cascade Count	Rumour Count
False	82,605	1699
True	24,409	490
Mixed	19,287	259

500k tweets

Link to fact-checking orgs

Collect all English tweets that contain link to any website of the fact-checking orgs from Sept, 2006 to Dec, 2016

158k tweets

Replies to tweets

- For each reply tweet, extract the original tweet and extract all the retweets of the original tweet
- Removed original tweets directly linked with a fact-checking orgs headline

Canonicalization and Twitter Historical API

- Identifying images and external article links in the original tweets using Google reverse image search and OCR
- Extracted all English-tweets containing any urls or text

126,301 tweets

Last step - Removing bots

State-of-the-art bot detection algorithm by Varol et al. 13.2% of bot accounts were removed

Fact Check › Medical

Did the World Health Organization Declare Medical Marijuana Has No Health Risks?

A December announcement by a WHO committee stated that one specific component of marijuana is safe and should not require regulation.

CLAIM

The World Health Organization has declared that medical marijuana has no health risks.

RATING

MOSTLY FALSE

- Normalized the verdicts into rating 1 to 5
  - False, Mostly false, Mixture, Mostly true, True (starting from 1)
- Tag rumors with one out of seven topics
  - Politics, Urban Legends, Business, Science and Technology, Terrorism and War, Entertainment , Natural Disasters
  - For missing and multiple or uncertain tags manually annotated by 3 annotators (Fleiss' kappa = 0.93)

Selection Bias

- Possible due to selection of 6 fact-checking orgs
  - Conducted robustness check by conducting study by considering human-identified stories
  - Analyzed separately 13,000 rumor cascades, which were fact-checked by 3 independent fact-checkers

- Towards stories with greater diffusion volume
  - May also be present in the robustness dataset
  - May under-sample stories that never diffused
  - Main sample is representative of verified stories
  - Robustness sample is representative of stories with a visible footprint on Twitter

Dataset Details

- Number of individuals spreading true news is more than the ones spreading false news
  - Greater fraction of false rumors between 1 and 1000 cascades
  - Greater fraction of true rumors more than 1000 cascades

	Count	False	True	Mixed
Political	44,095	27,600	9520	6,975
All others	82,206	55,005	14,889	12,312

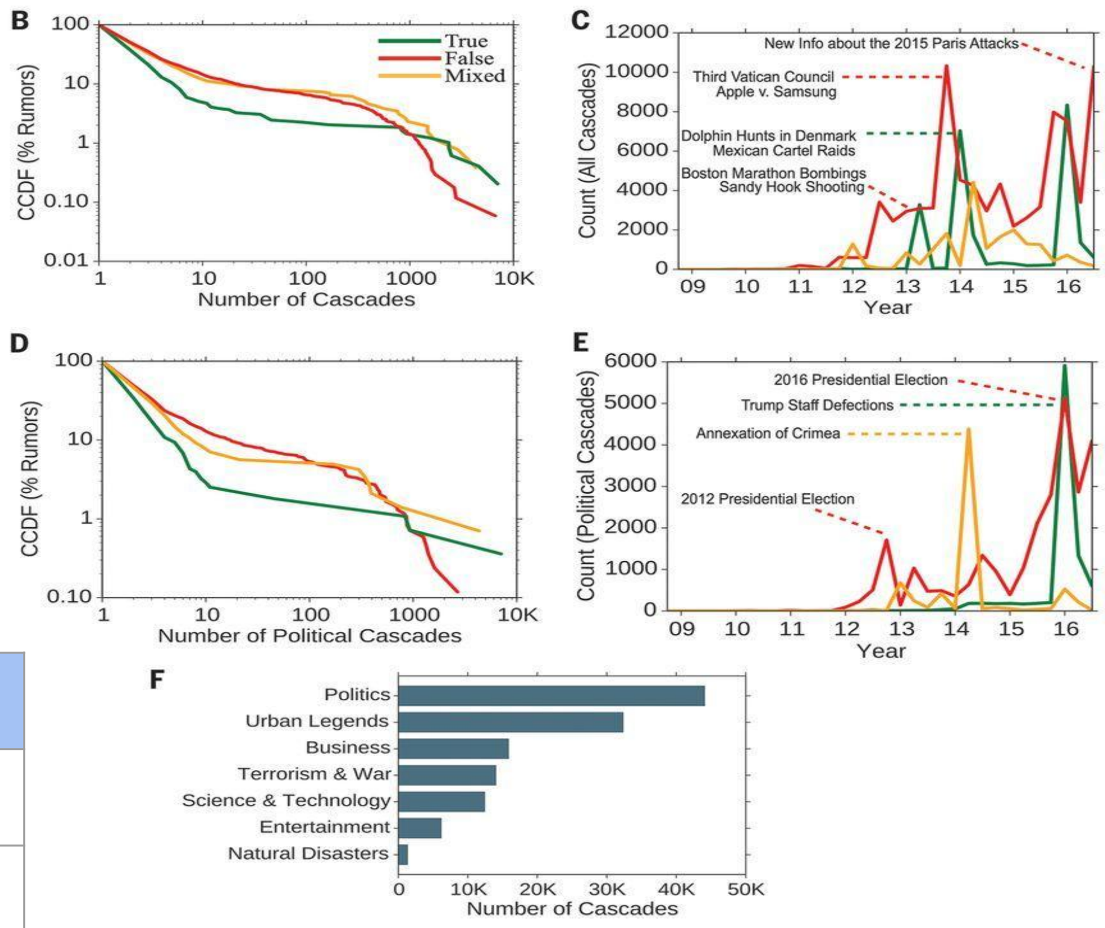
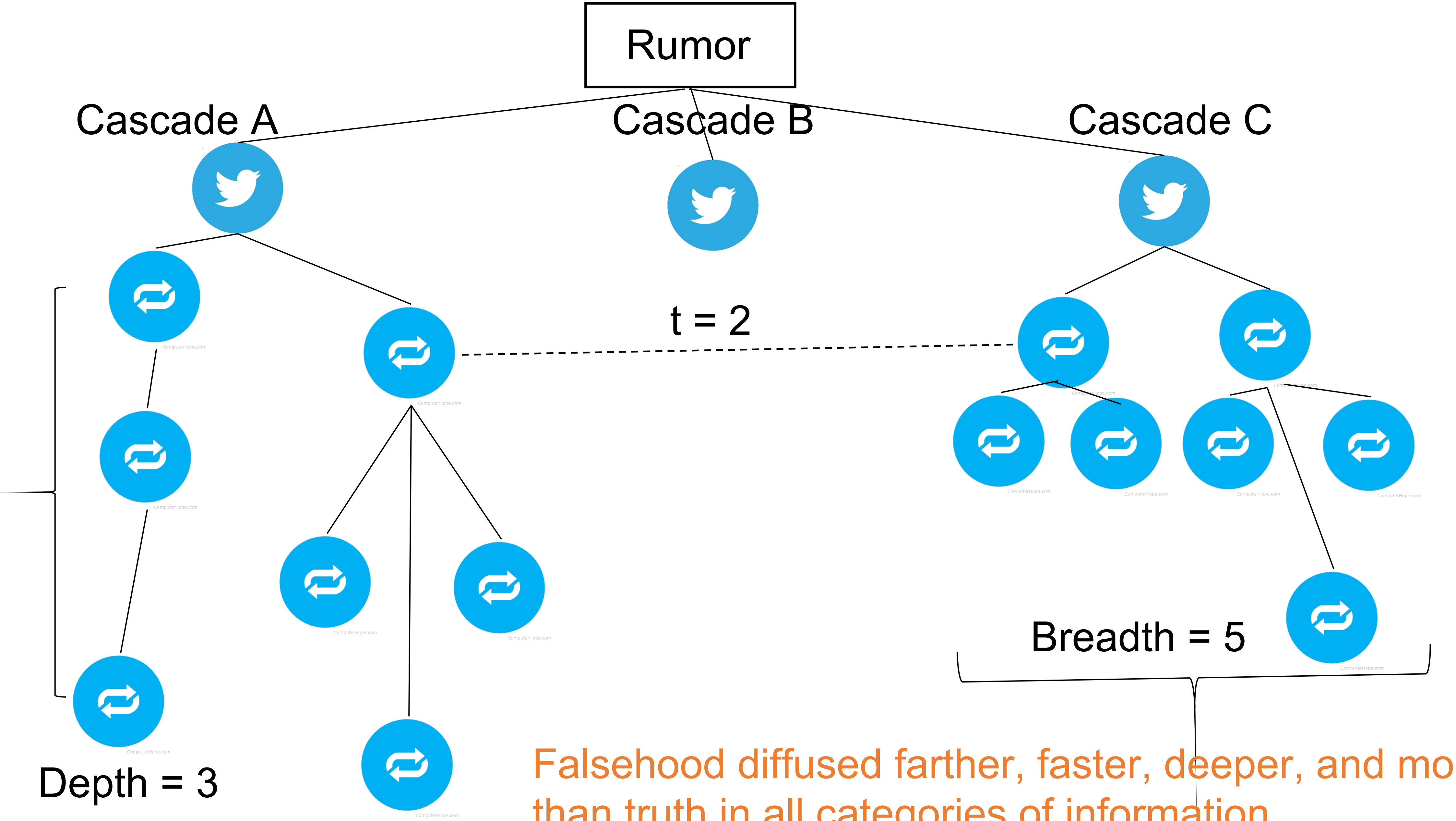


Fig. 1 Rumor cascades.



Falsehood diffused farther, faster, deeper, and more broadly than truth in all categories of information

Observations

- Falsehood reached far more people than the truth
- Many more people retweeted falsehood than they did the truth
- Aided by structural virality, not just broadcast dynamics

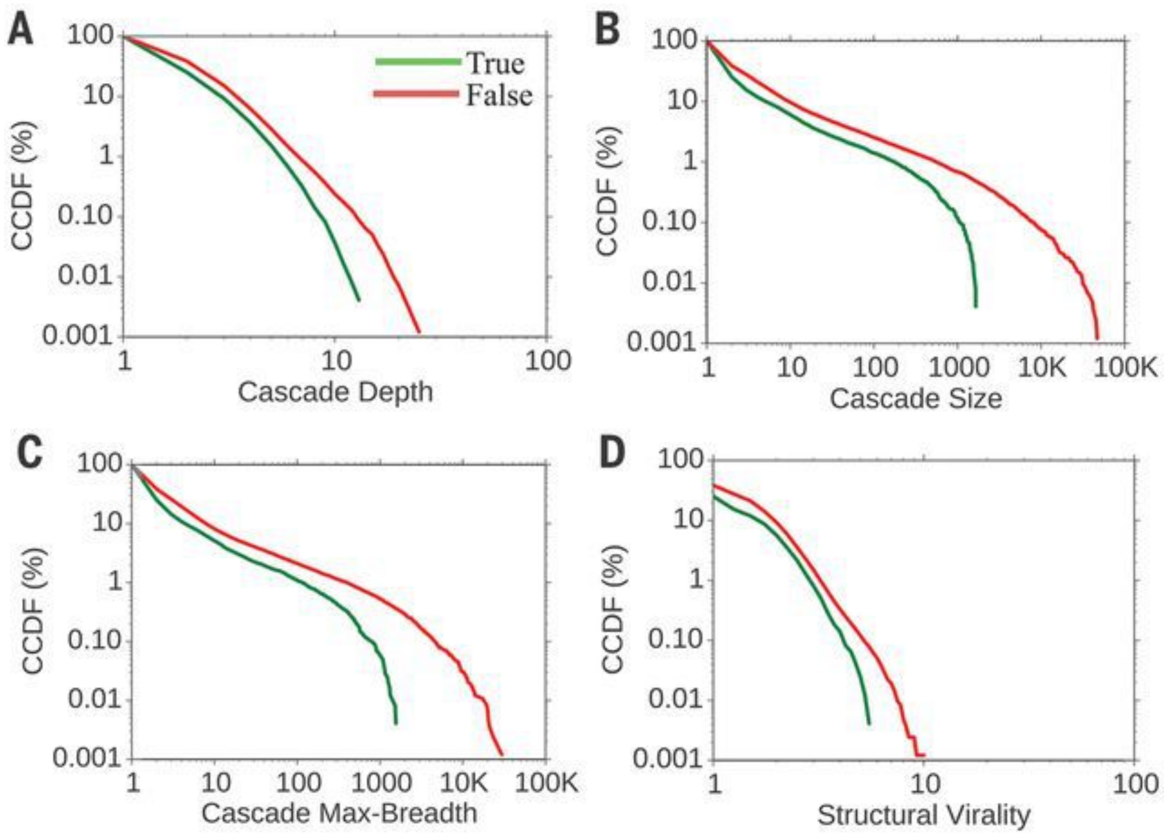


Fig. 2 Complementary cumulative distribution functions[1] of true and false cascades

[1] $F_x(x) = P(X > x)$

Rate of spread

- Truth took six times as long as falsehood to reach 1500 people and 20 times to reach a depth of 10

- Falsehood diffused more broadly and was retweeted by more unique users than the truth at every depth

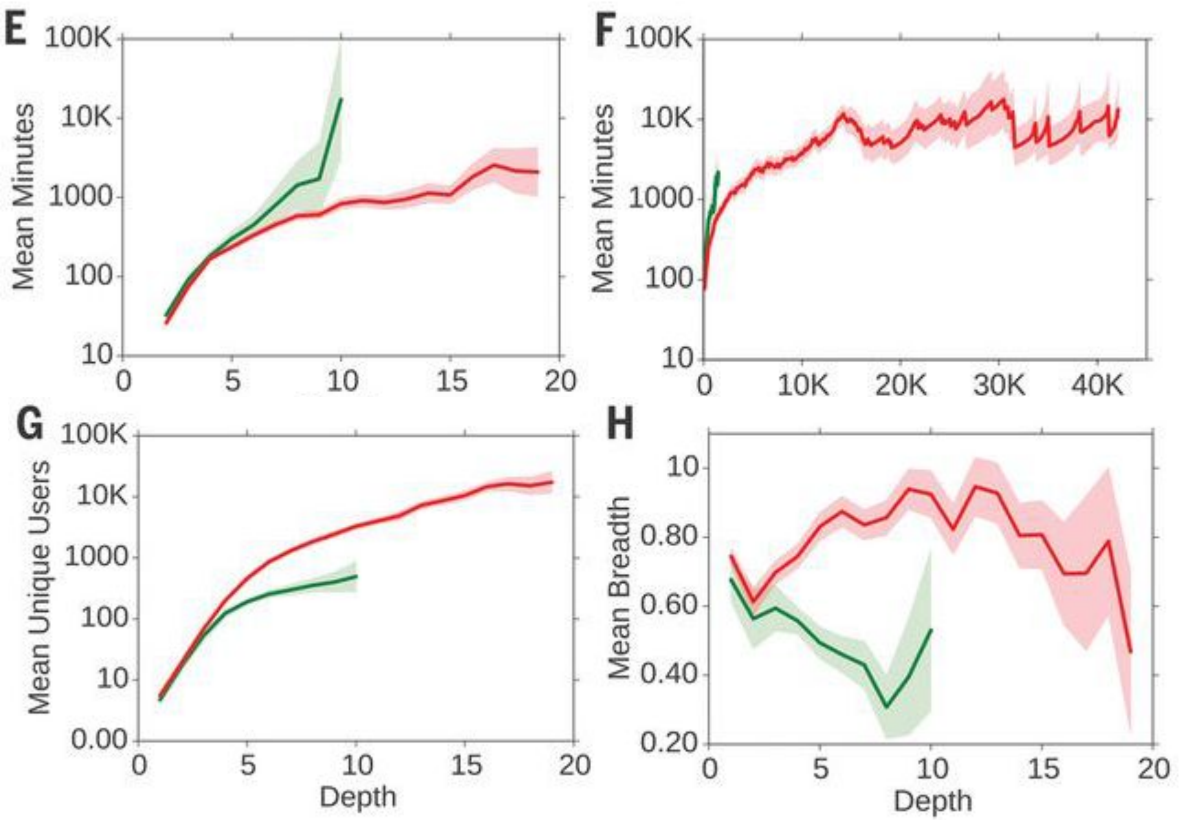


Fig. 2 Complementary cumulative distribution functions of true and false cascades

Political News

- False political news vs All other
  - Deeper
  - More broadly
  - Reached more people
  - More viral
  - Other categories travelled more broadly at shallower depths
  - Political news travelled more broadly at greater depths
  - Reached >20k people ~3 times faster than all other types of false news reached 10k people

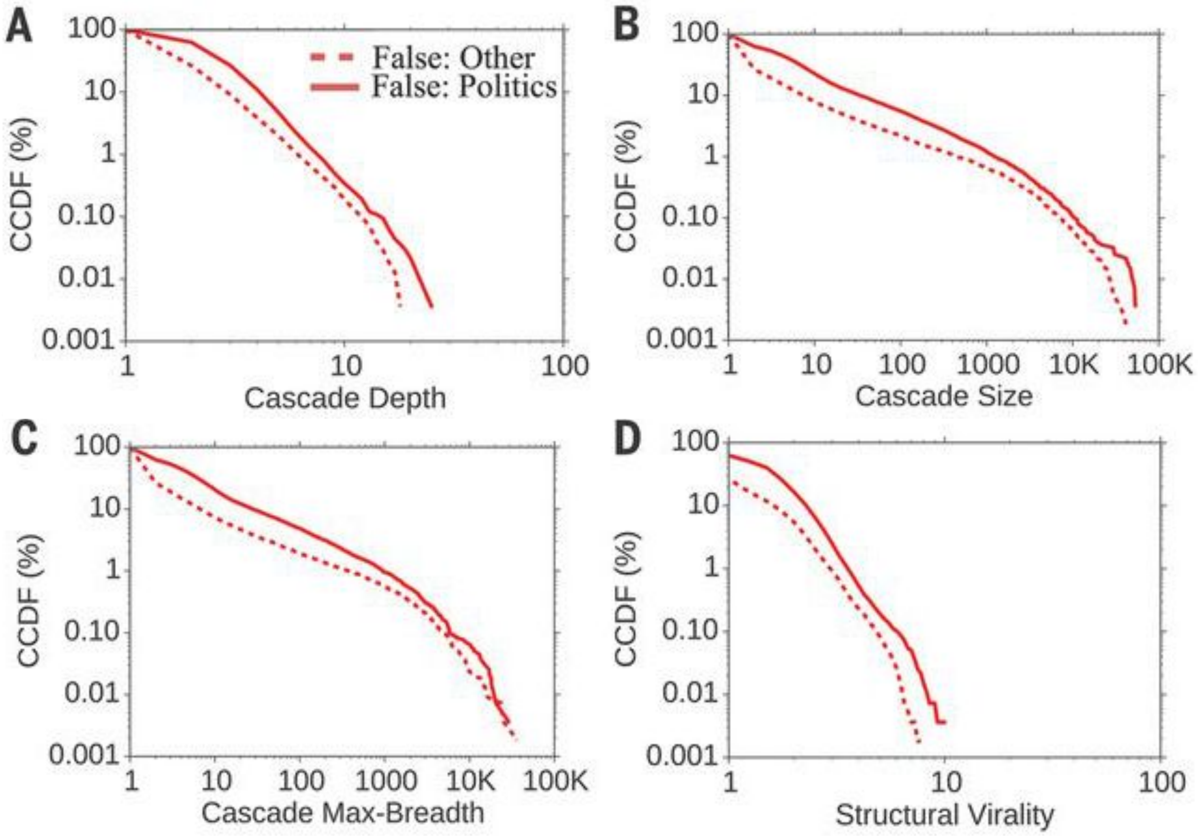


Fig. 2 Complementary cumulative distribution functions of false political and other types of rumor cascades

Question

- What causes falsity to travel faster than the truth

- Investigating whether it is the structural elements of the network or individual characteristics of spreaders

Observations

- Users spreading false news have -
  - Fewer followers
  - Followed by fewer people
  - Less active on Twitter
  - Verified less often
  - Short period of Twitter use

	median		mean		mean (log)		stdv (log)		ks-test
	false	true	false	true	false	true	false	true	
followers	410	466	2234	5240	2.62	2.68	0.69	0.88	D=0.104, p∼0.0
followees	383	509	1002	1707	2.59	2.72	0.85	0.96	D=0.136, p∼0.0
verified	0	0	0.002	0.006	nd	nd	nd	nd	D=0.005, p<0.001
engagement	9.52	9.54	19.70	24.65	0.91	0.90	0.65	0.76	D=0.054, p∼0.0
account age	982	1214	1072	1269	2.90	2.97	0.39	0.42	D=0.125, p∼0.0

	coef	odds ratio	std err	z	P>|z|	[95.0% Conf.	Int.]
account age	0.0002	1.000160	2.07e-05	7.759	0.000	0.000	0.000
engagement	0.0066	1.006648	0.000	18.019	0.000	0.006	0.007
falsehood	0.5350	1.707489	0.084	6.366	0.000	0.370	0.700
followees	-1.639e-05	0.999984	8.73e-06	-1.877	0.060	-3.35e-05	7.22e-07
followers	5.192e-05	1.000052	7.77e-06	6.682	0.000	3.67e-05	6.72e-05
intercept	-2.3941	0.091257	0.072	-33.437	0.000	-2.534	-2.254
verified	1.4261	4.162467	0.090	15.915	0.000	1.250	1.602

Table 1 - Model estimating correlates of news diffusion

Are Bots Responsible?

- None of the main conclusions changes. False news still spread farther, faster, deeper and more broadly than truth in all categories of information
  - Bot detection algorithm by Davis et al. (2016) used to identify and remove all bots
  - Removed all tweet cascades started by bots,including human retweets of original bot tweets
  - Varied the algorithm's sensitivity with another bot-detection to verify the robustness of analysis

- Inclusion of bot traffic, accelerated the spread of both true and false news roughly equally. Humans, and not robots are more likely to spread false news than the truth

Are bots responsible

- Role of bots should not be entirely discarded. It may be that -
  - Over the entire timeline, bots do not favor spread of false news
  - In some of the recent cases, bots are reported to be used to spread false news
  - Did not study the change in the role of bots with time

Question

- Whether falsity was more novel than the truth
- Whether Twitter users were more likely to retweet information that is more novel

What makes humans to act

Novelty in the content

Novelty attracts Human attention!

a. Novelty arouses surprise
b. In a Bayesian Sense, surprise quantifies how data affects an observer, by measuring the difference between posterior and prior beliefs of the observer.

Bayesian Surprise Attracts Human Attention. Laurent Itti, Pierre Baldi. NIPS 2005

What makes humans to act

1. Novelty encourages information sharing -
   a. Content that evokes high-arousal positive (awe) is more viral

J. Berger, K. L. Milkman. , What makes online content viral? J. Mark. Res. 49, 192–205 (2012). doi:10.1509/jmr.10.0353

What makes humans to act

1. Novelty encourages information sharing -
   a. Content that evokes high-arousal positive (awe)
   b. or negative (anger or anxiety) emotions is more viral

J. Berger, K. L. Milkman. , What makes online content viral? J. Mark. Res. 49, 192–205 (2012). doi:10.1509/jmr.10.0353

What makes humans to act

1. Novelty encourages information sharing -
   a. Content that evokes high-arousal positive (awe)
   b. or negative (anger or anxiety) emotions is more viral
   c. Convey social status on one that is “in the know” or has access to unique “inside” information

J. Berger, K. L. Milkman. , What makes online content viral? J. Mark. Res. 49, 192–205 (2012). doi:10.1509/jmr.10.0353

What makes humans to act

1. Novelty encourages information sharing -
   a. Content that evokes high-arousal positive (awe)
   b. or negative (anger or anxiety) emotions is more viral
   c. Convey social status on one that is “in the know” or has access to unique “inside” information

J. Berger, K. L. Milkman. , What makes online content viral? J. Mark. Res. 49, 192–205 (2012). doi:10.1509/jmr.10.0353

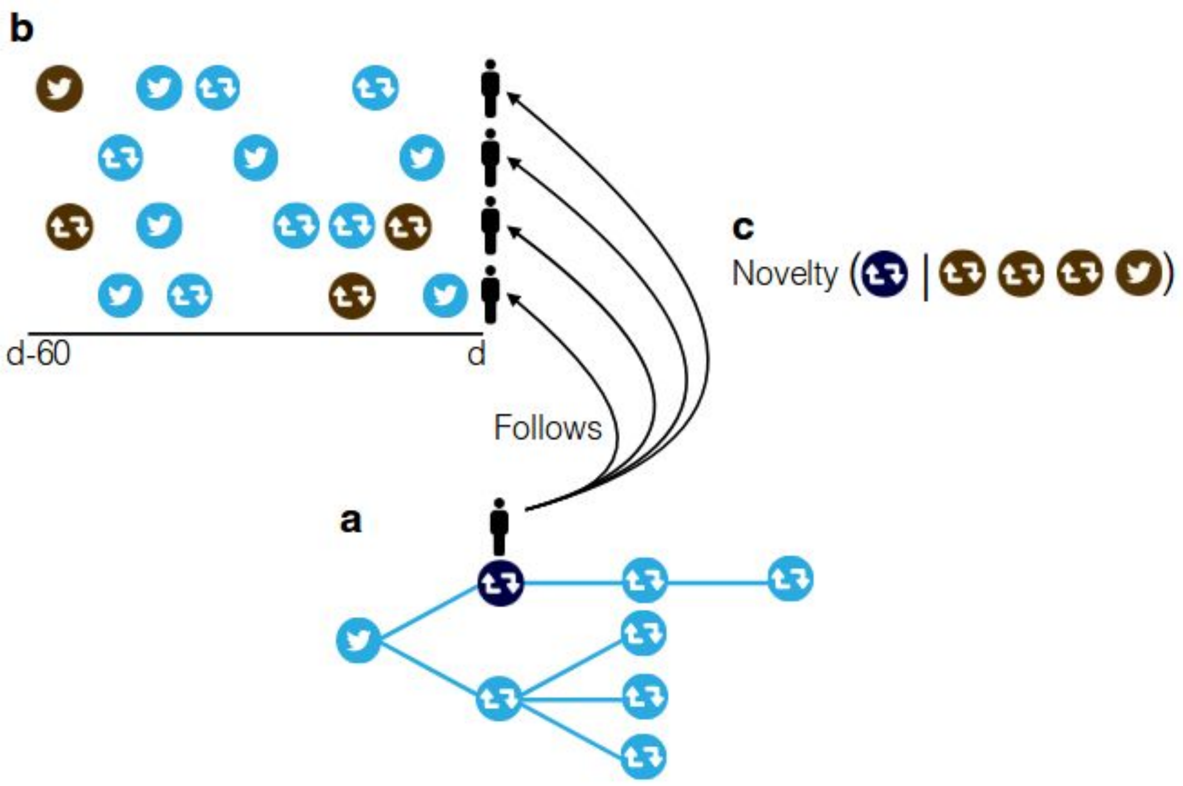
What makes humans to act

Novelty

1. Novelty contributes to productive decision-making
   a. It provides the greatest aid to decision-making

S. Aral, M. Van Alstyne. ,The diversity-bandwidth trade-off. Am. J. Sociol. 117, 90–171 (2011). doi:10.1086/661238

Measuring novelty of a tweet in the network



A. **Rumour set** - randomly selected ~5000 users who propagated true and false rumors and
B. **Exposure Set** - extracted a random sample of ~25,000 tweets that they were exposed to in the 60 days prior to their decision to retweet a rumor.
C. Measure novelty of retweeted tweet w.r.t. tweets they were exposed to.

Measuring novelty of a tweet in the network

- Topic Modelling - LDA (Specific model for Twitter)
  - Trained on 10M english tweets
  - 200 topics
- For a tweet t_r in Rumour Set , obtain a topic distribution for the tweet
- Compare it with every *background tweet of t_r*
  - those tweets tweeted/retweeted by the user's (who tweeted t_r) followees, i.e. a tweet from Exposure set, which the user got exposed to. A subset of Exposure set.
  - KL-Divergence (KL), Information Uniqueness (IU), Bhattacharyya Distance (BD)

Γ_r Tweet in the rumour set

Γ_b Background tweet.

Both are 1x200 vectors, each showing a discrete distribution of topics

$$IU(\Gamma_r, \Gamma_b) = 1 - cos(\Gamma_r, \Gamma_b)$$

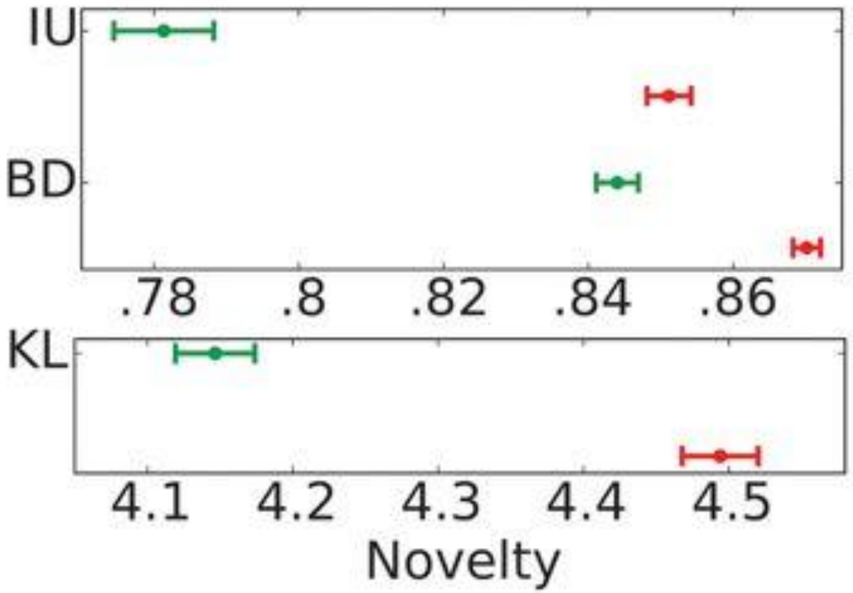
$$KL(\Gamma_r, \Gamma_b) = \frac{D_{KL}(\Gamma_r \parallel \Gamma_b) + D_{KL}(\Gamma_b \parallel \Gamma_r)}{2}$$

$$D_{KL}(P \parallel Q) = \sum_i P(i) log(\frac{P(i)}{Q(i)})$$

$$BD(\Gamma_r, \Gamma_b) = -ln(\sum_{x \in X} \sqrt{\Gamma_r(x)\Gamma_b(x)})$$

X is the topic set. x is the individual topic

aggregated and averaged the scores for false and true rumors to produce a mean and a standard error for each of the novelty metrics.



Differences in the information uniqueness (IU), scaled Bhattacharyya distance (BD), and K-L divergence (KL) of true (green) and false (red) rumor tweets compared to the corpus of prior tweets the user was exposed to in the 60 days before retweeting the rumor tweet.

	mean		variance	
	false	true	false	true
IU	0.85	0.78	0.0052	0.0072
KL	4.49	4.15	0.1618	0.0948
BD	0.87	0.84	0.0008	0.0008

Mean and variance of the IU, KL, and BD of true and false rumor tweets compared to the corpus of prior tweets the user has seen in the 60 days before seeing the rumor tweet

Evaluating LDA

- Convert a Tweet to a vector using Tweet2Vec
- (Hard) Label the tweet with its most prominent topic from LDA
- With every tweet as a node, construct a network
- The edge weights are vector similarity of the nodes using tweet2vec
- Run Louvain Clustering over the network
- Correlation between the clusters and the topic label obtained
- Pearson $r = 0.63$, p 0.001

Question

- Although false rumors are measurably more novel than true rumors, users may not perceive in the same manner
- Assessed users' perceptions of the information contained in true and false rumors by comparing emotional content of replies to true and false rumors

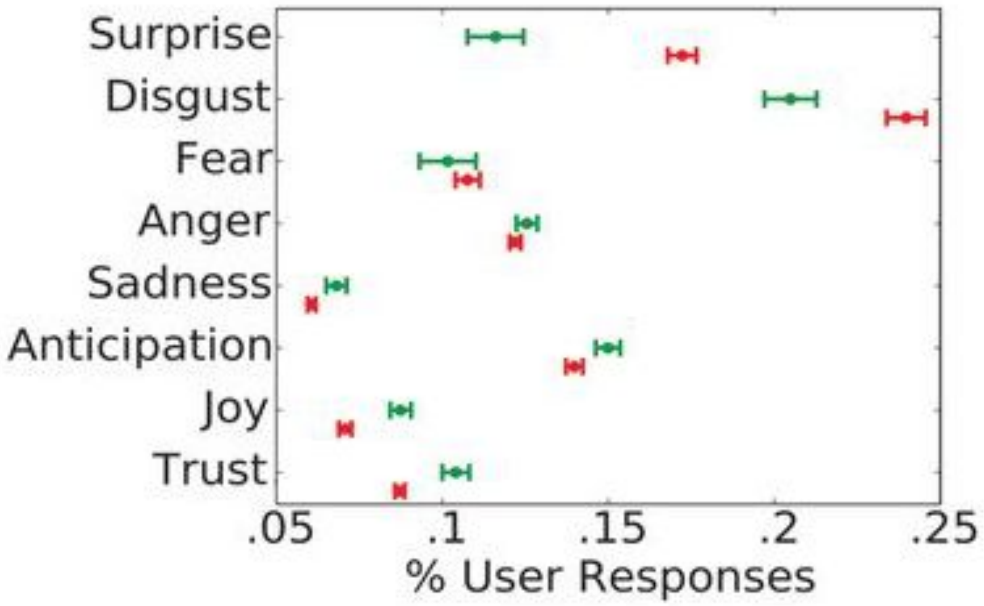
How users perceive

- Lexicon- based approach(NRC)
  - ~140,000 English words and association with the above emotions
  - ~32,000 Twitter hashtags and their weighted association with the above emotions
  - Calculated the fraction of words in the tweets related to each of the 8 emotions - anger, fear, anticipation, trust, surprise, sadness, joy, disgust

False rumors	True rumors
Greater surprise	Greater sadness
Support novelty with evidence	Anticipation
Greater disgust	Joy
	Trust

Fig 4 - Models estimating the emotional content of replies to news

How users perceive



	mean		variance		ks-test
	false—true		false—true		
surprise	0.172	0.116	0.0167	0.0072	D=0.205, p∼0.0
disgust	0.240	0.205	0.0260	0.0227	D=0.102, p∼0.0
fear	0.108	0.102	0.0120	0.0095	D=0.021, p∼0.164
anger	0.122	0.126	0.0074	0.0111	D=0.023, p∼0.078
sadness	0.061	0.068	0.0038	0.0065	D=0.037, p∼0.0
anticipation	0.140	0.150	0.0093	0.0154	D=0.038, p∼0.0
joy	0.071	0.087	0.0054	0.0104	D=0.061, p∼0.0
trust	0.087	0.104	0.0058	0.0119	D=0.060, p∼0.0

Table 2 - Mean and variance of the emotional content of replies to true and false rumor tweets

Falsehood may have additional factors beyond novelty that inspire people to share false news

Conclusion

- Falsehood diffused farther, faster, deeper, and more broadly than truth in all categories of information
  - Humans, and not bots are more likely to spread false news to the truth
- Network structure and individual characteristics of spreaders ~~favor and promote false news~~
- Often found that false news is more novel and that novel information is more likely to be retweeted

Conclusion

- Cannot claim whether the insights hold true for other social networks

- Exact same study is possible with Facebook data
- Currently, working on developing interventions to stop the spread of false new
    - Approach 1 - Labelling
    - Approach 2 - Deincentivizing

References

- L. J. Savage, The theory of statistical decision. J. Am. Stat. Assoc. 46, 55–67 (1951).
- H. A. Simon, The New Science of Management Decision (Harper & Brothers Publishers, New York, 1960).
- R. Wedgwood, The aim of belief. Noûs 36, 267–297 (2002).
- E. Fehr, U. Fischbacher, The nature of human altruism. Nature 425, 785–791 (2003).
- C. E. Shannon, A mathematical theory of communication. Bell Syst. Tech. J. 27, 379–423 (1948). doi:10.1002/j.1538-7305.1948.tb01338.
- S. Bikhchandani, D. Hirshleifer, I. Welch, A theory of fads, fashion, custom, and cultural change as informational cascades. J. Polit. Econ. 100, 992–1026 (1992).
- K. Rapoza, “Can ‘fake news’ impact the stock market?” Forbes, 26 February 2017; www.forbes.com/sites/kenrapoza/2017/02/26/can-fake-news-impact-the-stock-market/.
- M. Mendoza, B. Poblete, C. Castillo, “Twitter under crisis: Can we trust what we RT?” in Proceedings of the First Workshop on Social Media Analytics (Association for Computing Machinery, ACM, 2010), pp. 71–79.

References

- A. Gupta, H. Lamba, P. Kumaraguru, A. Joshi, “Faking Sandy: Characterizing and identifying fake images on Twitter during Hurricane Sandy,” in Proceedings of the 22nd International Conference on World Wide Web (ACM, 2010), pp. 729–736.
- K. Starbird, J. Maddock, M. Orand, P. Achterman, R. M. Mason, “Rumors, false flags, and digital vigilantes: Misinformation on Twitter after the 2013 Boston Marathon bombing,” in iConference 2014 Proceedings (iSchools, 2014).
- J. Gottfried, E. Shearer, “News use across social media platforms,” Pew Research Center, 26 May 2016;
- C. Silverman, “This analysis shows how viral fake election news stories outperformed real news on Facebook,” BuzzFeed News, 16 November 2016;
- M. De Domenico, A. Lima, P. Mougel, M. Musolesi, The anatomy of a scientific rumor. Sci. Rep. 3, 2980 (2013).
- O. Oh, K. H. Kwon, H. R. Rao, “An exploration of social media in extreme events: Rumor theory and Twitter during the Haiti earthquake 2010,” in Proceedings of the International Conference on Information Systems (International Conference on Information Systems, ICIS, paper 231, 2010).

References

- M. Tambuscio, G. Ruffo, A. Flammini, F. Menczer, “Fact-checking effect on viral hoaxes: A model of misinformation spread in social networks,” in Proceedings of the 24th International Conference on World Wide Web (ACM, 2015), pp. 977–982.
- Z. Zhao, P. Resnick, Q. Mei, “Enquiring minds: Early detection of rumors in social media from enquiry posts,” in Proceedings of the 24th International Conference on World Wide Web (ACM, 2015), pp. 1395–1405.
- M. Gupta, P. Zhao, J. Han, “Evaluating event credibility on Twitter,” in Proceedings of the 2012 Society for Industrial and Applied Mathematics International Conference on Data Mining (Society for Industrial and Applied Mathematics, SIAM, 2012), pp. 153–164.
- G. L. Ciampaglia, P. Shiralkar, L. M. Rocha, J. Bollen, F. Menczer, A. Flammini, Computational fact checking from knowledge networks. PLOS ONE 10, e0128193 (2015).
- A. Friggeri, L. A. Adamic, D. Eckles, J. Cheng, “Rumor cascades,” in Proceedings of the International Conference on Weblogs and Social Media (Association for the Advancement of Artificial Intelligence, AAAI, 2014)

References

- A. Bessi, M. Coletto, G. A. Davidescu, A. Scala, G. Caldarelli, W. Quattrociocchi, Science vs conspiracy: Collective narratives in the age of misinformation. PLOS ONE 10, e0118093 (2015).
- Friggeri et al. (19) do evaluate two metrics of diffusion: depth, which shows little difference between true and false rumors, and shares per rumor, which is higher for true rumors than it is for false rumors. Although these results are important, they are not definitive owing to the smaller sample size of the study; the early timing of the sample, which misses the rise of false news after 2013; and the fact that more shares per rumor do not necessarily equate to deeper, broader, or more rapid diffusion.
- S. Goel, A. Anderson, J. Hofman, D. J. Watts, The structural virality of online diffusion. Manage. Sci. 62, 180–196 (2015).
- L. Itti, P. Baldi, Bayesian surprise attracts human attention. Vision Res. 49, 1295–1306 (2009).
- S. Aral, M. Van Alstyne, The diversity-bandwidth trade-off. Am. J. Sociol. 117, 90–171 (2011).
- J. Berger, K. L. Milkman, What makes online content viral? J. Mark. Res. 49, 192–205 (2012).
- D. M. Blei, A. Y. Ng, M. I. Jordan, Latent Dirichlet allocation. J. Mach. Learn. Res. 3, 993–1022 (2003).

References

- T. M. Cover, J. A. Thomas, Elements of Information Theory (Wiley, 2012).
- T. Kailath, The divergence and Bhattacharyya distance measures in signal selection. IEEE Trans. Commun. Technol. 15, 52–60 (1967).
- R. Plutchik, The nature of emotions. Am. Sci. 89, 344–350 (2001).
- S. M. Mohammad, P. D. Turney, Crowdsourcing a word-emotion association lexicon. Comput. Intell. 29, 436–465 (2013).
- S. M. Mohammad, S. Kiritchenko, Using hashtags to capture fine emotion categories from tweets. Comput. Intell. 31, 301–326 (2015).
- S. Vosoughi, D. Roy, “A semi-automatic method for efficient detection of stories on social media,” in Proceedings of the 10th International AAAI Conference on Weblogs and Social Media (AAAI, 2016), pp. 707–710.
- C. A. Davis, O. Varol, E. Ferrara, A. Flammini, F. Menczer, “BotOrNot: A system to evaluate social bots,” in Proceedings of the 25th International Conference Companion on World Wide Web (ACM, 2016), pp. 273–274.

Thank you for your attention

Why fake news alarming today

- New social technologies facilitate rapid information sharing and large-scale information cascades
  - Affected stock prices [Forbes 2017]
  - Responses to natural disasters to terrorist attacks have been disrupted

- "Fake news" has been named the word of the year 2017 by Collins dictionary

Previous works

- Spread of rumors
  - Single rumors like Discovery of the Higgs boson by Domenico et al. [Sci. Rep 2013] and 2010 Haitian earthquake by Oh et al. [ICIS 2010]
  - Multiple rumors from single event like 2013 Boston Marathon bombing by Starbird et al. [iSchools 2014]
- Other works
  - Spread of scientific and conspiracy-theory stories by Del Vicario et al. [2016] and Bessi et al. [2015]
  - Friggeri et al. [AAAI 2014] analyzed how fact checking affects rumor propagation on Facebook

Kolmogorov-Smirnov test

- To test whether two samples are drawn from identical distributions

- Measures a distance between the empirical distribution functions(ECDF) of the two samples or between the ECDF of a sample and the CDF of a reference distribution