A Theory of Misinformation Spread on Social Networks

Chin-Chia Hsu Amir Ajorlou Ali Jadbabaie

Laboratory for Information and Decision Systems (LIDS)
Institute for Data, Systems, and Society (IDSS)
Massachusetts Institute of Technology (MIT)

Network Dynamics in the Social, Economic, and Financial Sciences
DISMA-Polytecnico Turino, November 4-8, 2019

Massachusetts Institute of Technology

IDSS
MIT INSTITUTE FOR DATA, SYSTEMS, AND SOCIETY

Spread of misinformation is a major societal problem

Misinformation spread has already created unrest in Burma, India, Indonesia, ... not a new problem, but social media have significantly amplified it.

- Online social networks facilitate the peer-to-peer interactions and serve as an amplification mechanism.
- People are passively fed information that their friends and the algorithms of the platforms subjectively select.
- Social media and its engagement-driven monetization scheme amplify misinformation and create polarization.

Vosoughi, S., Roy, D., and Aral, S.

Science, cover article, March 2018

Questions: Why do people forward information? Is it to inform? express their views? identify their tribe? Persuade others?

This talk

We develop a micro-founded model for forwarding decision of individuals whose main motive is persuasion.

Related literature:

Vosoughi et al. (2018) show that users on Twitter are more likely to retweet the false rumors than the truth. They hypothesize that novelty of fake news might be a major factor.

Pennycook and Rand (2018)'s experiments indicate that people are susceptible to fake news because they are lazy to think.

Swire et al. (2017) suggest that "people use political figures as a heuristic to guide evaluation of what is true or false."

Our goal

Aim to study:

- Rationalize sharing
- Uncover the mechanism underlying the news spread on social media

Our approach:

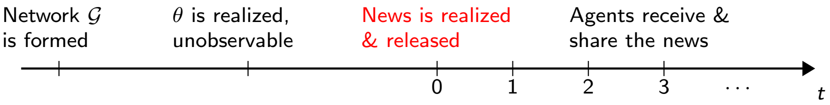
- Framework of perspectives and opinions in *Sethi and Yildiz (2016)*
- People's tendency to make their friends think like themselves
- Persuasion by changing the information set
- Sharing is costly

Outline

1. Micro-level
   - Model & Information structure
   - Utility function & decision rule of sharing news
2. Macro-level
   - Spread dynamics & condition of news cascade
   - Credibility of the news and the probability of cascade
3. Fully strategic agents who interpret "no news" as news
4. Concluding Remarks

The timeline of our model

1. There is an underlying state of the world θ. n agents in the society have individual beliefs on θ with n large.
2. At time $t = 0$, a noisy signal, henceforth called news x, is realized and released to a small number of agents.
3. When an agent receives the news, she knows the source of the news and current clickthrough rate, updates her belief on θ using Bayes rule, and decides whether to broadcast/forward the copies to her followers at a universal cost C.
4. As $t \to \infty$, in steady state we observe the size of the news cascade.

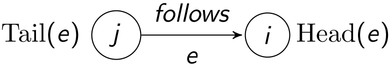


Outline

1. Micro-level
   - Model & Information structure
   - Utility function & decision rule of sharing news
2. Macro-level
   - Spread dynamics & condition of news cascade
   - Credibility of the news and the probability of cascade
3. Fully strategic agents who interpret “no news” as news
4. Concluding Remarks

Model Settings

- n agents form a directed network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, agent j follows agent i if and only if $e = (j, i) \in \mathcal{E}$.

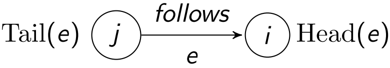


- Each agent has a Gaussian prior on the unobservable state $\theta \in \mathbb{R}$: Agents priors have different means but the same variance.

$$\theta \sim_i \mathcal{N}(\underbrace{\mu_i}_{\text{perspective of agent } i}, \underbrace{\sigma_\theta^2}_{\text{uncertainty of beliefs}}), \quad \forall i$$

Model Settings

- n agents form a directed network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, agent j follows agent i if and only if $e = (j, i) \in \mathcal{E}$.



- Each agent has a Gaussian prior on the unobservable state $\theta \in \mathbb{R}$: Agents priors have different means but the same variance.

$$\theta \sim_i \mathcal{N}(\underbrace{\mu_i}_{\text{perspective of agent } i}, \underbrace{\sigma_\theta^2}_{\text{uncertainty of beliefs}}), \quad \forall i$$

- μ_i is private information of agent i, but $\mu_i, i \in \mathcal{V}$ are independently distributed as

$$\mu_i \sim \mathcal{N}(\underbrace{\bar{\mu}}_{\text{aggregate perspective}}, \underbrace{\sigma_\mu^2}_{\text{diversity of perspective}}), \quad \forall i$$

News as noisy signal

The news x is an uncertain observation of θ that is corrupted with an additive, independent from the θ, Gaussian noise ϵ.

$$x = \theta + \epsilon,$$
$$\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$$

Distribution of the noise of the source is common knowledge.

x is then released to a (randomly and uniformly chosen) finite number of the agents.

Reminder: Bayes rule in a Gaussian world

1. Linear combinations of normally distributed random variables is still normally distributed.

2. Given two independent normal random variables:

$$\theta \sim \mathcal{N}(\mu, \sigma_\theta^2)$$
$$\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$$

   and $X = \theta + \epsilon$. Given $X = x$, using Bayes rule we can write

$$\theta | X = x \sim \mathcal{N}\left(\frac{\sigma_\epsilon^2}{\sigma_\theta^2 + \sigma_\epsilon^2}\mu + \frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_\epsilon^2}x, \frac{\sigma_\theta^2 \sigma_\epsilon^2}{\sigma_\theta^2 + \sigma_\epsilon^2}\right)$$

Belief update

Agents receiving the news form their posterior according to Bayes rule as

$$\theta \sim_i \mathcal{N}\big((1-\beta)\mu_i + \beta x, (1-\beta)\sigma_\theta^2\big), \tag{1}$$

where $\beta = \frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_\epsilon^2}$.

Agents not receiving the news keep their prior belief on θ.

To brief,

$$\theta \sim_i \begin{cases} \mathcal{N}\big((1-\beta)\mu_i + \beta x, (1-\beta)\sigma_\theta^2\big), & \text{if agent } i \text{ has ever received the news} \\ \mathcal{N}\big(\mu_i, \sigma_\theta^2\big), & \text{if agent } i \text{ has never received the news} \end{cases}$$

Outline

1. Micro-level
   - Model & Information structure
   - Utility function & decision rule of sharing news
2. Macro-level
   - Spread dynamics & condition of news cascade
   - Credibility of the news and the probability of cascade
3. Fully strategic agents who interpret "no news" as news
4. Concluding Remarks

Distance between the beliefs

Agents derive their utility from proximity of their followers' beliefs to their own's.

$\Rightarrow$ need a distance measure on the normal beliefs

Distance between the beliefs

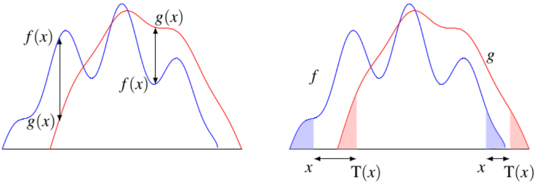
Agents derive their utility from proximity of their followers' beliefs to their own's.

$\Rightarrow$ need a distance measure on the normal beliefs

For normally distributed beliefs $P_1 = \mathcal{N}(\mu_1, \sigma_1^2)$ and $P_2 = \mathcal{N}(\mu_2, \sigma_2^2)$, we use the L2-Wasserstein distance

$$W_2^2(P_1; P_2) = (\mu_1 - \mu_2)^2 + (\sigma_1 - \sigma_2)^2. \quad (2)$$

also known as the earth-movers' distance.



from *Optimal Transport* by Filippo Santambrogio

The utility function

Let $a_i^B \in \{0, 1\}$ denote the decision of sharing news.

Utility function:

$$
\begin{aligned}
&u\Big(a_i^B, P^i, \{P^k | k \in N^{\text{in}}(i)\}, C\Big) \\
&= -\frac{1}{|N^{\text{in}}(i)|} \sum_{k \in N^{\text{in}}(i)} W_2^2(P^i; P^k) - C \times \mathbf{1}\{a_i^B = 1\}, \qquad (3)
\end{aligned}
$$

where P^i (similarly P^k) denotes the belief of agent i (agent k) on θ, and $C \geq 0$ represents the cost associated with broadcasting/forwarding.

Decision rule for broadcasting (I)

Agent i with perspective μ_i and opinion θ_i decides to broadcast x iff

$$\begin{aligned}&\mathbb{E}_i\left[W_2^2(P^i;P^k)|k\in N^{\text{in}}(i) \text{ and } a_i^B=1\right]+C\\ \leq&\mathbb{E}_i\left[W_2^2(P^i;P^k)|k\in N^{\text{in}}(i) \text{ and } a_i^B=0\right]\end{aligned} \tag{4}$$

Decision rule for broadcasting (I)

Agent i with perspective μ_i and opinion θ_i decides to broadcast x iff

$$\begin{aligned}&\mathbb{E}_i\left[W_2^2(P^i;P^k)|k\in N^{\text{in}}(i)\text{ and }a_i^B=1\right]+C\\ \leq&\mathbb{E}_i\left[W_2^2(P^i;P^k)|k\in N^{\text{in}}(i)\text{ and }a_i^B=0\right]\end{aligned}\tag{4}$$

Broadcast: An agent knows that her followers will receive the news.

No broadcast: She needs to estimate the fraction of followers q who have already received the news.

Decision rule for broadcasting (I)

Agent i with perspective μ_i and opinion θ_i decides to broadcast x iff

$$\begin{aligned}&\mathbb{E}_i\left[W_2^2(P^i;P^k)|k \in N^{\text{in}}(i) \text{ and } a_i^B = 1\right] + C \\ \leq&\mathbb{E}_i\left[W_2^2(P^i;P^k)|k \in N^{\text{in}}(i) \text{ and } a_i^B = 0\right] \end{aligned} \tag{4}$$

Broadcast: An agent knows that her followers will receive the news.

No broadcast: She needs to estimate the fraction of followers q who have already received the news.

q can be computed from public information of the network metrics and the current scale of the news cascade.

Decision rule for broadcasting (II)

Lemma 1: Decision rule of broadcasting

Let $\beta = \frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_\epsilon^2}$, $\bar{q} = 1-q$. Agent i broadcasts the news x if and only if

$$\frac{C}{1-q} - (1-(1-\beta)^2)\sigma_\mu^2 - (1-\sqrt{1-\beta})^2\sigma_\theta^2$$
$$\leq \Big(((1-\beta)\mu_i + \beta x) - \bar{\mu}\Big)^2 - \Big(((1-\beta)\mu_i + \beta x) - ((1-\beta)\bar{\mu} + \beta x)\Big)^2.$$

- $(1-\beta)\mu_i + \beta x$: posterior perspective of agent i
- $\bar{\mu}$: average prior perspectives of followers
- $(1-\beta)\bar{\mu} + \beta x$: average of the posterior perspectives of followers'

Surprise and affirmation motives (I)

We re-write Lemma 1 to inspect the motives in the decision making:

$$2\beta(1-\beta)(\mu_i-\bar{\mu})(x-\bar{\mu})+\beta^2(x-\bar{\mu})^2 \geq \frac{C}{\bar{q}}-(1-(1-\beta)^2)\sigma_\mu^2-(1-\sqrt{1-\beta})^2\sigma_\theta^2.$$

Components:

1. $|x-\bar{\mu}|$: *surprise* of news relative to the aggregate perspective
2. $\mu_i-\bar{\mu}$: deviation of agent i's perspective from the aggregate perspective $(\mu_i-\bar{\mu})(x-\bar{\mu})$ shows the *affirmation* effect.
3. σ_μ^2: diversity of prior perspectives
4. σ_θ^2: uncertainty of the prior beliefs
5. The threshold on RHS increases in q.

Normalized variables:

$$\hat{C} \triangleq \frac{C}{\sigma_\mu^2}, \quad \hat{\mu}_i \triangleq \frac{\mu_i - \bar{\mu}}{\sigma_\mu}, \quad \hat{x} \triangleq \frac{x - \bar{\mu}}{\sigma_\mu}, \quad \hat{\sigma}_\theta \triangleq \frac{\sigma_\theta}{\sigma_\mu}$$

$$\begin{aligned}
&\mathbb{P}\left(a_i^B = 1 \mid \text{agent } i \text{ received the news}\right) \\
=&\mathbb{P}\left(\hat{\mu}_i \geq \underbrace{\frac{\hat{C}}{2\bar{q}\beta(1-\beta)\hat{x}}}_{\frac{K(\hat{x})}{\bar{q}}} - \underbrace{\left(\frac{2-\beta}{2(1-\beta)} + \frac{1-\sqrt{1-\beta}}{2(1-\beta)(1+\sqrt{1-\beta})}\hat{\sigma}_\theta^2\right)\frac{1}{\hat{x}} - \frac{\beta}{2(1-\beta)}\hat{x}}_{\eta(\hat{x})}\right) \\
=&\Phi\left(\eta(\hat{x}) - \frac{K(\hat{x})}{\bar{q}}\right) \qquad (5)
\end{aligned}$$

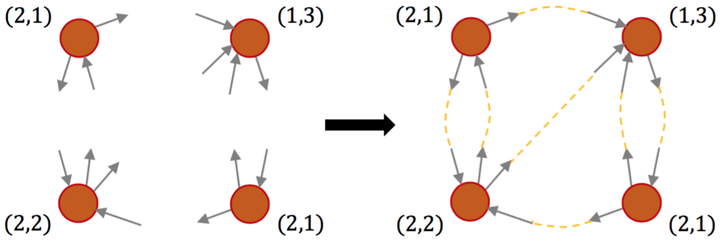
Outline

1. Micro-level
   - Model & Information structure
   - Utility function & decision rule of sharing news
2. Macro-level
   - Spread dynamics & condition of news cascade
   - Credibility of the news and the probability of cascade
3. Fully strategic agents who interpret “no news” as news
4. Concluding Remarks

Network formation and metrics

Network metrics are *public information* to all of n agents.

- Agent i has out-degree d_i and in-degree ℓ_i.
- $P(d, \ell)$ is the joint degree distribution. $P^{\text{out}}(\cdot)$ and $P^{\text{in}}(\cdot)$ stand for the marginal distributions.
- Network by configuration model: each outgoing link is independently connected to one of the agents' incoming links.



Analysis of spread dynamics

- Discrete time $t \in \mathbb{N} \cup \{0\}$, the news x is released at $t = 0$.
- Let $r(t)$ be the set of agents who **receive** the news at time t and $b(t) \subset r(t)$ be the set of those who **broadcast** at time t.
- $R(t) = \cup_{\tau=0}^{t} r(\tau)$. Similarly, $B(t) = \cup_{\tau=0}^{t} b(\tau)$.
- $|\cdot|$ means the normalized fraction.
- We need $\bar{q}(t)$ for individual's decision of broadcasting.

$$\bar{q}(t) = \mathbb{P}\Big(\mathrm{Tail}(e) \notin R(t) \mid \mathrm{Head}(e) \in r(t), e \in \mathcal{E}\Big) \tag{6}$$

- Note that the cascade dynamics is endogenous: agents don't just mechanically forward with some given probability!

Equations that govern the dynamics of the news spread

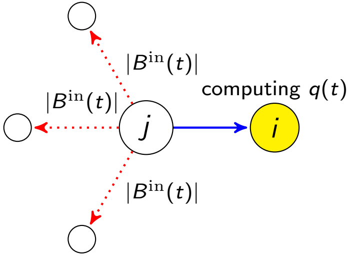
To compute $\bar{q}(t)$, need to figure out what fraction of network edges spread the news:

$$|B^{\text{in}}(t)| = \mathbb{P}\Big(\text{Head}(e) \in B(t) \mid e \in \mathcal{E}\Big) \tag{7}$$

Given the news x,

$$|R(t)| = 1 - \underbrace{\sum_{d=0}^{\infty} (1 - |B^{\text{in}}(t-1)|)^d P^{\text{out}}(d)}_{\text{fraction of agents who haven't received by time } t}$$

$$\bar{q}(t) = \frac{1}{\mathbb{E}[d]} \sum_{d=1}^{\infty} d(1 - |B^{\text{in}}(t-1)|)^{d-1} P^{\text{out}}(d)$$



Size of the news cascade in the steady state

Theorem 1

Denote with $|B^{\text{in}}_{\infty}| = |B^{\text{in}}(\infty)|$ the fraction of links through which the news has passed at the steady state.

Then, $|B^{\text{in}}_{\infty}|$ is the largest solution of $G(|B^{\text{in}}_{\infty}|) = 0$, where

$$G(|B^{\text{in}}_{\infty}|) = \int_0^{|B^{\text{in}}_{\infty}|} \Big(-1 + \frac{\Phi(\eta - \frac{K}{\bar{q}(y)})}{\mathbb{E}[d]} \sum_{d=0}^{\infty} d\mathbb{E}[\ell|d](1-y)^{d-1} P^{\text{out}}(d) \Big) dy,$$

if $G(1) < 0$; $|B^{\text{in}}_{\infty}| = 1$ otherwise.

The steady-state size of the spread, denoted by $|R_{\infty}| = |R(\infty)|$, is

$$|R_{\infty}| = 1 - \sum_{d=0}^{\infty} (1 - |B^{\text{in}}_{\infty}|)^d P^{\text{out}}(d)$$

Condition for emergence of news cascade

A news cascade occurs if the news spreads to a **non-zero** fraction of the population in steady state, i.e., a giant component of size $\Theta(n)$ emerges.

Proposition 1: Condition of cascade

A news cascade happens if and only if

$$\frac{\mathbb{E}[\ell d]}{\mathbb{E}[d]} > \frac{1}{\Phi\big(\eta(\hat{x}) - K(\hat{x})\big)} > 1. \tag{8}$$

Second inequality is just the network condition for cascade when agents simply broadcast in the configuration model.

Condition for emergence of news cascade

A news cascade occurs if the news spreads to a **non-zero** fraction of the population in steady state, i.e., a giant component of size $\Theta(n)$ emerges.

Proposition 1: Condition of cascade

A news cascade happens if and only if

$$\frac{\mathbb{E}[\ell d]}{\mathbb{E}[d]} > \frac{1}{\Phi\big(\eta(\hat{x}) - K(\hat{x})\big)} > 1. \tag{8}$$

Second inequality is just the network condition for cascade when agents simply broadcast in the configuration model.

So far, our results are built on the realized news.

We move on to the *ex-ante* analysis before the news is realized $(t = 0^{-})$.

Outline

1. Micro-level
   - Model & Information structure
   - Utility function & decision rule of sharing news
2. Macro-level
   - Spread dynamics & condition of news cascade
   - Credibility of the news and the probability of cascade
3. Fully strategic agents who interpret "no news" as news
4. Concluding Remarks

Credibility of the news and probability of cascade

Given the network statistics, what is the optimal σ_ϵ^2, or $\beta = \frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_\epsilon^2} \in [0,1]$ ($\beta = 1$: truth), that maximizes the probability of cascade?

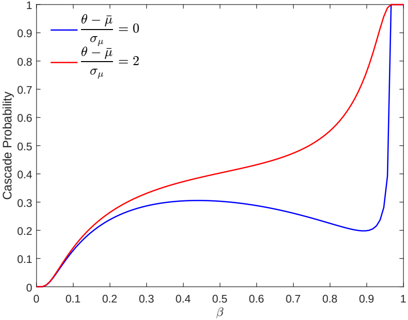
$$\max_{\beta\in[0,1]} \mathbb{P}_{\hat{x}}\Big(\frac{\mathbb{E}[\ell d]}{\mathbb{E}[d]} > \frac{1}{\Phi\big(\eta(\hat{x}) - K(\hat{x})\big)}\Big) \tag{9}$$

Since $\Phi(\cdot)$ is strictly increasing, we can rewrite it as

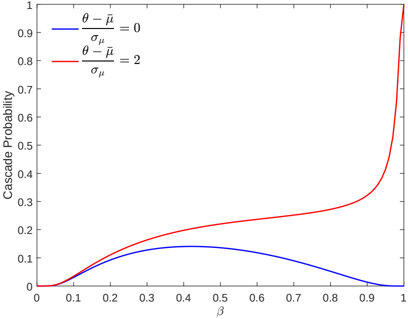
$$\max_{\beta\in[0,1]} \mathbb{P}_{\hat{x}}\Big(\text{a quadratic function in } |x-\bar{\mu}| > \Phi^{-1}(\frac{\mathbb{E}[d]}{\mathbb{E}[\ell d]})\Big)$$

An example sample case: probability of cascade versus β

Example: $\sigma_\mu = 1$, $\sigma_\theta = 3$, $\theta - \bar{\mu} \in \{0, 2\}$, and $\Phi^{-1}(\frac{\mathbb{E}[\ell d]}{\mathbb{E}[d]}) = -2$



$C = 7$: $C < \sigma_\mu^2 + \sigma_\theta^2$.



$C = 11$: $C = \sigma_\mu^2 + \sigma_\theta^2 + 1$.

$1 < (\theta - \bar{\mu})^2$ for $\theta - \bar{\mu} = 2$;
$1 > (\theta - \bar{\mu})^2$ for $\theta - \bar{\mu} = 0$.

Comparative statics: What noise level in the news maximizes cascade probability?

Theorem 2: Optimal credibility level

Let Ω^* be the set of values β maximizing the ex-ante likelihood of a cascade. Assume also that $\frac{\mathbb{E}[d]}{\mathbb{E}[\ell d]} < 1$, and let $\Delta = \Phi^{-1}\left(\frac{\mathbb{E}[d]}{\mathbb{E}[\ell d]}\right)$. Then,

i) If $C < \sigma_\mu^2 + \sigma_\theta^2$, then $\Omega^* = [\bar{\beta}^*, 1]$, where $\bar{\beta}^* < 1$ is the unique solution to $S(\bar{\beta}^*) = 0$, where

$$S(\beta) = C + (1-\beta)^2(\Delta^2+1)\sigma_\mu^2 - (1-\sqrt{1-\beta})^2{\sigma_\theta}^2 - \sigma_\mu^2$$

ii) If $\sigma_\mu^2 + \sigma_\theta^2 \leq C < (\theta - \bar{\mu})^2 + \sigma_\mu^2 + \sigma_\theta^2$, then $\Omega^* = \{1\}$.

iii) If $C \geq (\theta - \bar{\mu})^2 + \sigma_\mu^2 + \sigma_\theta^2$, then $1 \notin \Omega^*$.

Implications of Theorem 2

Theorem 2

Cascade probability is maximized with inaccurate/false news (nonzero noise) if and only if $C > (\theta - \bar{\mu})^2 + \sigma_\mu^2 + \sigma_\theta^2$.

- If the society is wise, i.e., aggregation of individual priors is concentrated around the truth, then the agents are not incentivized to share the truth. False news can result in surprise and is more likely to cause a cascade.
- If the society is unwise and aggregation of prior perspectives doesn't concentrate on truth then truth causes cascade. It happens if
  (1) Aggregate perspective is concentrated away from the truth.
  (2) Perspectives (mean of priors) are highly diverse.
  (3) There is large uncertainty of the priors.

Outline

1. Micro-level
   - Model & Information structure
   - Utility function & decision rule of sharing news
2. Macro-level
   - Spread dynamics & condition of news cascade
   - Credibility of the news and the probability of cascade
3. Fully strategic agents who interpret “no news” as news
4. Concluding Remarks

Concluding remarks

- A micro-foundation for news sharing among rational individuals who want to persuade their followers to think more like them
  - Discussed factors that affect broadcast decisions, including persuasion, surprise, and affirmation motives
- Macro-level
  - Described the endogenous spread dynamics and necessary and sufficient conditions for emergence of a cascade
  - Relationship between "wisdom of the crowd" and optimal precision levels that maximize the likelihood of a news cascade

References

Pennycook, G. and Rand, D. G. (2018). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*.

Sethi, R. and Yildiz, M. (2016). Communication With Unknown Perspectives. *Econometrica*, 84(6):2029–2069.

Swire, B., Berinsky, A. J., Lewandowsk, S., and Ecker, U. K. (2017). Processing political misinformation: comprehending the trump phenomenon. *Royal Society Open Science*, 4(3):160802.

Vosoughi, S., Roy, D., and Aral, S. (2018). News Online. *Science*, 359(March):1146–1151.