Learning from Viral Content

Krishna Dasaratha   Kevin He

November 29, 2022

News Feed Virality Weight and Learning Accuracy

- In the past decade, viral content on social media platforms emerged as a prominent source of news for many people
- Key common feature: users see stories in **news feeds**, and stories that go into feeds depend on previous users' actions
  - E.g., Twitter homepage selects tweets based on likes and retweets and also displays a special section for trending stories
  - E.g., Reddit's front page is a list of highly upvoted posts
- Platform algorithm determines the weight given to **virality** (i.e., current popularity) when sampling stories for news feeds
  - Different iterations of Twitter design gave different levels of emphasis on trending / popular tweets
  - Reddit's ordering algorithm for posts evolved over a decade
- Today:
  - Equilibrium model where users interact with news stories
  - Study trade-offs for learning when platform shows more viral content (as opposed to random content) in news feeds
- This analysis relates to a recent policy debate: did news feeds that push viral content facilitate the spread of false info about public health, politics, etc?

Pros and Cons of a Viral News Feed

- Advantage: **conveys more info** with just a few stories
  - On Reddit, reading a few front-page stories gives you a lot more info than reading a few random stories
  - A positive story in a viral news feed carries more info than just the realization of a single signal: after accounting for selection, it also tells you about others' upvoting
  - This in turn lets you infer others' hidden (positive) info
- Disadvantage: can generate **bad steady states** where wrong stories dominate (even with rational users)
  - Most of the content on platform may be correct, but incorrect stories can still get shown more after weighting by popularity
  - Bad steady state self-perpetuates: see incorrect stories → believe state is negative → upvote incorrect stories → ...
  - Reddit's company blog in 2009: *"Once a comment gets a few early upvotes, it's moved to the top. The higher something is listed, the more likely it is to be read (and voted on), and the more votes the comment gets. It's a feedback loop."*
- Rest of talk: Formalize these intuition in an **equilibrium** model where agents form **Bayesian beliefs** after seeing the news feed and act rationally

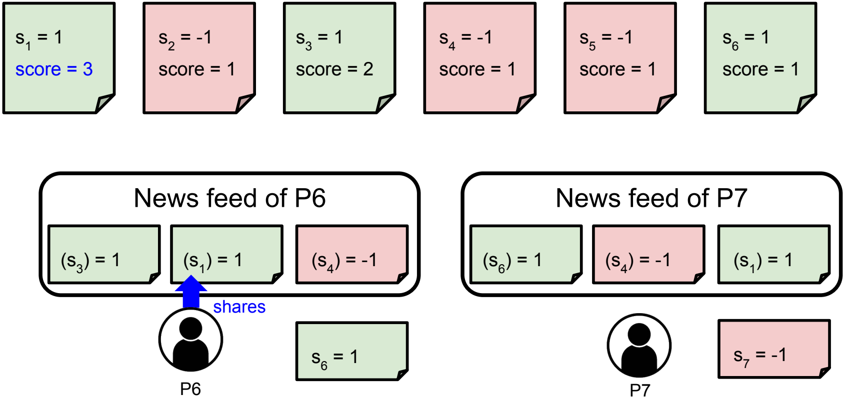
Outline

1. Model of equilibrium learning and sharing on social media
2. Steady states under a fixed (possibly non-equilibrium) strategy
3. When do misleading steady states exist in equilibrium?

A Model of Social Media News Feeds

- Unknown state of nature $\omega \in \{-1, 1\}$, equally likely
- Finite group of n agents in positions 1, ..., n, move in order
- Agents don't know own position, all n positions equally likely
- Agent in position i has a private signal, interpreted as a news **story** $s_i \in \{-1, 1\}$
- $\mathbb{P}[s_i = \omega \mid \omega] = q$ for some **story precision** $1/2 < q < 1$, independent across agents i
- Agent i posts their story s_i on the platform
- Each story i has a **popularity score**, and a newly posted story starts at score $+1$
- Agent i sees a **news feed** of K stories others posted, sampled based on stories' current popularity (more on this soon)
  - Does not see the popularity or arrival time of the stories
  - If $i < K$, then does not see others' stories
- Agent i then **shares** $C \leq K/2$ out of the K stories from their news feed, increasing their popularity score by 1
- Agent i gets utility $u > 0$ for each shared story that matches ω

An Illustration with $K = 3,\ C = 1$



A Model of Social Media News Feeds

- Platform's **virality weight** $\lambda \in [0, 1]$ determines how the K stories are sampled to generate the news feed
- For each slot in the news feed:
  - With prob λ, sample story with prob proportional to popularity
  - With prob $1 - \lambda$, sample a story uniformly at random
  - For simplicity, sample with replacement — not important when pool of stories is large
- Intermediate λ interpolates between the two cases of uniform sampling (ignoring popularity score) and proportional sampling (fully based on popularity score)

Strategies and Steady States

- A (mixed) **strategy** $\sigma(s,k) \in \Delta\{0,1,...,C\}$ gives distribution over number of positive stories shared when agent privately discovers story s and sees k out of K positive stories in feed
- State-symmetric strat = treat pos/neg stories symmetrically
- If everyone uses (not necessarily optimal) state-symmetric strategy σ, what happens on the platform in the long run?
- Given state ω, **viral accuracy** $x(t)$ after t acts is the fraction of the total popularity score on state-matching stories
- In an **infinite** society, if all agents $i \geq K+1$ use σ, get stochastic process $(x(t))_{t=1}^{\infty}$ of viral accuracy

Definition

A point x^* such that $x(t) \to x^*$ with positive probability is a **steady state** of σ.

- When $x(t) \to x^*$, fraction of total popularity score on state-matching stories remains close to x^* as fresh stories get added and agents use σ to share stories each period

Convergence to a Steady State

Proposition (viral accuracy convergence)

Given any σ, there is a finite set of steady states $X^* \subseteq (0,1)$ so that almost surely $x(t) \to x^*$ for some $x^* \in X^*$.

- Almost surely, viral accuracy converges (but the limit can be random and depend on randomness in signals and sampling)
- How do we characterize X^*, the steady states under σ?
- **Inflow accuracy function** at x under σ, denoted $\phi_\sigma(x)$
  - Say today's viral accuracy is x, and exactly q fraction of the stories match the state
  - New agent increases total popularity score by $C + 1$
  - $\phi_\sigma(x) =$ expected fraction of the $C + 1$ score allocated to state-matching stories (wrt randomness in story and sampling)
  - Fixed points of $\phi_\sigma(x)$ natural candidates for steady states ▶

Steady States and Fixed Points of ϕ_σ

- Much easier to study fixed points of ϕ_σ (polynomial function) than steady states (limit of a stochastic process)
- Next result uses **stochastic approximation** tools to establish the equivalence, provided the ϕ_σ fixed point is not "unstable from both sides"

Theorem 1

If $\phi_\sigma(x^*) = x^*$ and there is some $\epsilon > 0$ so that either

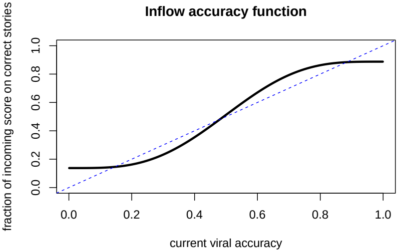
- $\phi_\sigma(x) < x$ *for all* $x \in (x^*, x^* + \epsilon)$, *or*
- $\phi_\sigma(x) > x$ *for all* $x \in (x^* - \epsilon, x^*)$,

then x^* is a steady state.
Conversely, if x^* is a steady state, then $\phi_\sigma(x^*) = x^*$.

- Surprising: a fixed point of ϕ_σ stable from only one side (a "**touchpoint**") is also reached by $x(t)$ with positive prob
- Distribution over steady states changes discontinuously in σ and parameters q, λ

Example: Steady States under a Simple Strategy

- A strategy that will be important later: σ^{maj} "**majority rule**"
  - Share C stories from the majority side of the news feed
  - If K even and $K/2$ stories on each side, break ties with s_i



- Above: inflow accuracy function for $K = 7$, $C = 3$, $q = 0.55$, $\lambda = 1$, σ^{maj} (i.e., share 3 of the majority stories)
- By Theorem 1, with positive prob learning converges to a steady state with viral accuracy < 0.5
  - Virality of false stories becomes **self-sustaining**
  - Most people see a news feed of false stories (even though most stories are true)
  - σ^{maj} shares these false stories and increases their popularity

Equilibrium

- So far have considered steady states induced by arbitrary strategies, but what do steady states look like when agents use **equilibrium** strategies?
- Consider player-symmetric and state-symmetric Bayesian Nash equilibrium of the n-player game, abbreviated "**equilibrium**"
- Mainly interested in limits of equilibria when n grows large

Definition

For fixed parameters q, K, C, λ, say σ^* is a **social equilibrium** if it is the limit of a sequence of equilibria $(\sigma^{(j)})_{j=1}^{\infty}$ for societies with n_j agents, where $n_j \to \infty$.

Informative and Misleading Steady States

Definition

Steady state x is **informative** if $\lambda x + (1-\lambda)q > 1/2$. Steady state x is **misleading** if $\lambda x + (1-\lambda)q < 1/2$.

- Informative / misleading denotes whether **sampling accuracy** is above or below 50%
  - Given the distribution of popularity scores on the platform, is it more likely that each story in the news feed is true or false?
- When do misleading steady states arise in equilibrium?
- Answer depends on whether λ is above or below a critical level, defined in terms of when the inflow accuracy function **of a particular strategy** first admits a misleading steady state

Critical Virality Weight

Definition

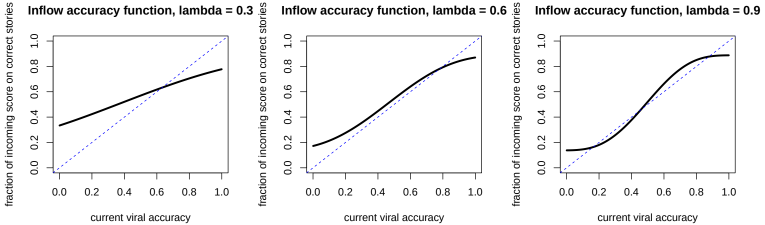
The **critical virality weight** λ^* is

$$\lambda^* := \inf\{\lambda \in [0,1] : \phi_{\sigma^{\mathsf{maj}}}(x^*) = x^* \text{ for some } x^* \in [0, 1/2)\}$$

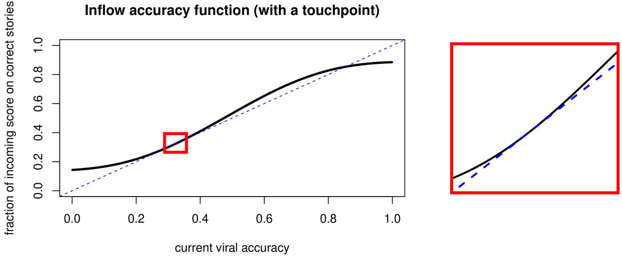
provided this set is non-empty. Otherwise, let $\lambda^* = 1$.

- λ^* = least virality weight s.t. σ^{maj} has a misleading steady state
- Next theorem: whether **any** social eqm admits a misleading steady state depends on if strategy σ^{maj} admits one
    - Equilibrium depends on distribution over steady states (no closed form characterization), while analyzing σ^{maj} is simple
    - Can use σ^{maj} to describe structure of equilibrium steady states for different λ (even when σ^{maj} is not an equilibrium)

An Example of Critical Virality Weight



- $K = 7,\ C = 3,\ q = 0.55,\ \lambda \in \{0.3, 0.6, 0.9\},\ \sigma^{\text{maj}}$
- Critical virality weight $0.6 < \lambda^* < 0.9$, in fact $\lambda^* \approx 0.76$



Virality Weight and Equilibrium Steady States

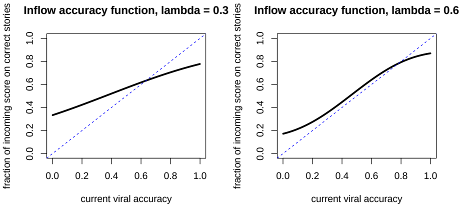
- Under equilibrium behavior, misleading steady states **discontinuously** emerge at the critical virality weight λ^*

Theorem 2

- *For* $0 < \lambda \leq \lambda^*$*, the only social equilibrium is* σ^{maj}
- *At every* $0 < \lambda < \lambda^*$*,* σ^{maj} *only has one equilibrium steady state, and it is higher than* q *(thus, informative)*
- *For* $\lambda = \lambda^*$*, provided* $\lambda^* < 1$*,* σ^{maj} *induces a misleading steady state*
- *For* $\lambda > \lambda^*$*, every social equilibrium induces at least one misleading steady state*

- Formalizes the message that platform algorithms that push viral content into news feeds generate misleading steady states

Steady State Accuracy Increases in λ if $\lambda < \lambda^*$



Proposition

For $\lambda < \lambda^*$, viral accuracy and sampling accuracy in the unique equilibrium steady state strictly increase in λ.

- Formalizes the message that platform algorithms that weigh virality more heavily helps aggregate more info
- When λ increases, news feed stories become more accurate indicators of the true state
- For $\lambda > \lambda^*$, steady states "move away" from $1/2$ with λ
- **Trade-off** for increasing λ: aggregates **more info**, but can create **systemic risk** that most people hold wrong beliefs

Comparative Statics of the Critical Virality Weight

- Results about eqm steady states hinge on the value of λ^*
- Higher λ^* means platform less susceptible to misleading steady states

Proposition

Let $\lambda^*(q, K, C)$ be the critical virality weight for q, K, C.

- $\lambda^*(q', K, C) \geq \lambda^*(q, K, C)$ *if* $q' > q$
- $\lambda^*(q, K, C') \geq \lambda^*(q, K, C)$ *if* $C' < C$
- $\lambda^*(q, K-2, C) \geq \lambda^*(q, K, C)$

All inequalities are strict whenever $\lambda^*(q, K, C) < 1$.

- Tend to get misleading steady states when people consume and interact with **large amount of social info**, relative to amount of private info (small q, large K, C)
- Private info untainted by relative popularity of different stories

Optimal News-Feed Design

- Can use characterization of eqm steady states to ask about optimal platform design (in large but finite societies)
- Suppose designer wants to maximize users' equilibrium utility from sharing stories
  - Likely affects people's decision to continue using the platform
- (Result also goes through for other objectives like expected viral accuracy or degree of consensus in news feeds)

Proposition

For any sequence (λ_n) *where each* λ_n *maximizes* $W_n(\lambda)$, *we have*

$$\liminf_{n \to \infty} \lambda_n \geq \lambda^*.$$

- For large n, the designer must either:
  - choose $\lambda > \lambda^*$ and accepts misleading steady states
  - or engage in "brinkmanship" and chooses λ just below a threshold that causes a discontinuous drop in user welfare

Related Literature: Learning from Shared Signals

- Bloch, Demange, and Kranton (2018), Papanastasiou (2020), Hsu, Ajorlou, and Jadbabaie (2021), Bowen, Dmitriev, and Galperti (2022), Kranton and McAdams (2022), Acemoglu, Ozdaglar, and Siderius (2022), etc.
- Either dissemination of a single signal, or signals are shared once with network neighbors and not re-shared
- In our model:
  - multiple signals about same state interact (multiple stories that corroborate each other increase the prob that each is shared)
  - signals shared and re-shared through a central algorithm
- Combination of these two features generates the social version of confirmation bias
- Our focus is also different: effect of showing more or less viral content, instead of the effect of the social network or the fact-checking technology

Related Literature: Herding

- Banerjee (1992), Bikhchandani, Hirshleifer, and Welch (1992), and a large subsequent literature on observational social learning
- As in our model, wrong initial signals can lead to persistent wrong beliefs, but differ in mechanism
- In herding, obstruction to learning is that agents observe others' actions that coarsely reflect these people's private signals
- In our world, agents observe and learn from others' signals, and the obstruction is that these observed signals depend on endogenous **selection**
- Important for herding results is the stark "information cascade" where later agents' signals become completely lost
- In our world, misleading steady state persists even though new private info keeps arriving and influencing later agents' beliefs

Conclusion

- An equilibrium model of a social media platform where:
  - Users exogenously **discover** stories and post them on platform
  - Users also read a small selection of past stories in a **news feed**
  - They **form beliefs** and **share** some of the stories they read, maximizing expected number of accurate stories they share
  - Technical contribution: apply stochastic approximation tools in an **equilibrium** setting, where behavior optimal given the induced distribution of stories
- Study how the sampling rule that populates the news feed affects long-run learning in equilibrium
  - When virality weight low enough, increasing this weight creates **more informative** equilibrium steady states with no downside
  - But high enough virality weight ($\lambda \geq \lambda^*$) generates **misleading** equilibrium steady states
  - At the threshold λ^*, misleading steady states emerge **discontinuously**

Thank you!

Inflow Accuracy Function ϕ_σ

Definition

The inflow accuracy function is

$$\phi_\sigma(x) = \frac{q + \sum_{k=0}^{K} P_k(x,\lambda) \cdot [q \cdot \mathbb{E}[\sigma(1,k)] + (1-q) \cdot \mathbb{E}[\sigma(-1,k)]]}{1+C}$$

where $P_k(x,\lambda) := \mathbb{P}[\mathrm{Binom}(K, \lambda x + (1-\lambda)q) = k]$.

◀ back

Proof Outline of Theorem 2

For $\lambda < \lambda^*$, need to show any social equilibrium σ^* is σ^{maj}

- Idea: use eqm behavior to deduce steady states under σ^*, and vice versa
- By optimality of eqm, $\mathbb{E}[\sigma^*(1, k)] \geq \mathbb{E}[\sigma^*(-1, k)]$ for each k
- Can show if σ^* has a misleading steady state, so does σ^{maj} — intuitively, majority rule makes it easier to perpetuate misleading steady state relative to any other strategy that's monotonic wrt own story
- σ^* can't have misleading steady states (else σ^{maj} has a misleading steady state at $\lambda < \lambda^*$, contradicting defn of λ^*)
- Applying optimality of eqm again, $\sigma^*(1, k)$ shares C positive stories for any $k \geq K/2$
- Can show this implies ϕ_{σ^*} only has fixed points in $(q, 1]$
- This means news-feed stories more precise than privately discovered stories, so must have $\sigma^* = \sigma^{\text{maj}}$ by optimality

Proof Outline of Theorem 2

For $\lambda > \lambda^*$, need to show any social equilibrium σ^* has a misleading steady state

- If not, then by the same arguments as before, $\sigma^* = \sigma^{\mathsf{maj}}$
- But we know σ^{maj} has a misleading steady state at λ^*
- Can show σ^{maj} continues to have misleading steady states at all higher $\lambda > \lambda^*$, contradiction

Comparative Statics of the Critical Virality Weight

- Next result shows to what extent K, C can affect λ^*

Proposition

For any q, K, C, we have $\lambda^*(q, K, C) > 1 - 1/2q$. But, for any fixed $1/2 < q < 1$ and any $\overline{\lambda} > 1 - 1/2q$, there exist $\underline{K}, \underline{C}$ so that whenever $K \geq \underline{K}, C \geq \underline{C}$, we have $\lambda^*(q, K, C) \leq \overline{\lambda}$.

- Platforms that show users large enough news feeds and let them share many stories will have λ^* arbitrarily close to $1 - 1/2q$
- No matter how precise individual stories, every social eqm admits a misleading steady state if $\lambda \geq 1/2$, provided K, C big enough
- But if algorithm close enough to random sampling, impossible to get misleading steady states no matter how much social info users see and interact with