Truth Learning in a Social Network

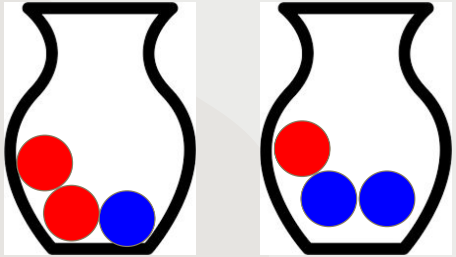
Roger Chen, John Wu, Eli Yablon
Professor Jie Gao
Daniel Baumgartner (PhD student)

Example of Herding/Information Cascade

Majority Red/Blue Vase

- There is an opaque vase with 3 balls
  - The vase is equally likely to be **majority red** or **majority blue**
- 50 rational people line up to *privately* sample a ball and *publicly* announce whether they think the vase is majority blue or majority red

50% **50%**



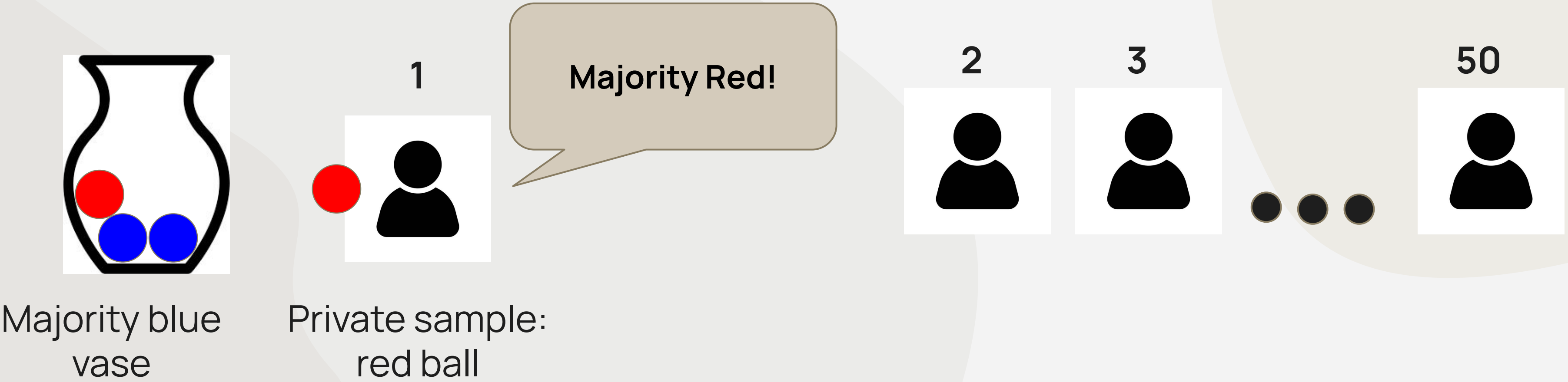
One majority-red or majority-blue vase

1 2 3 50

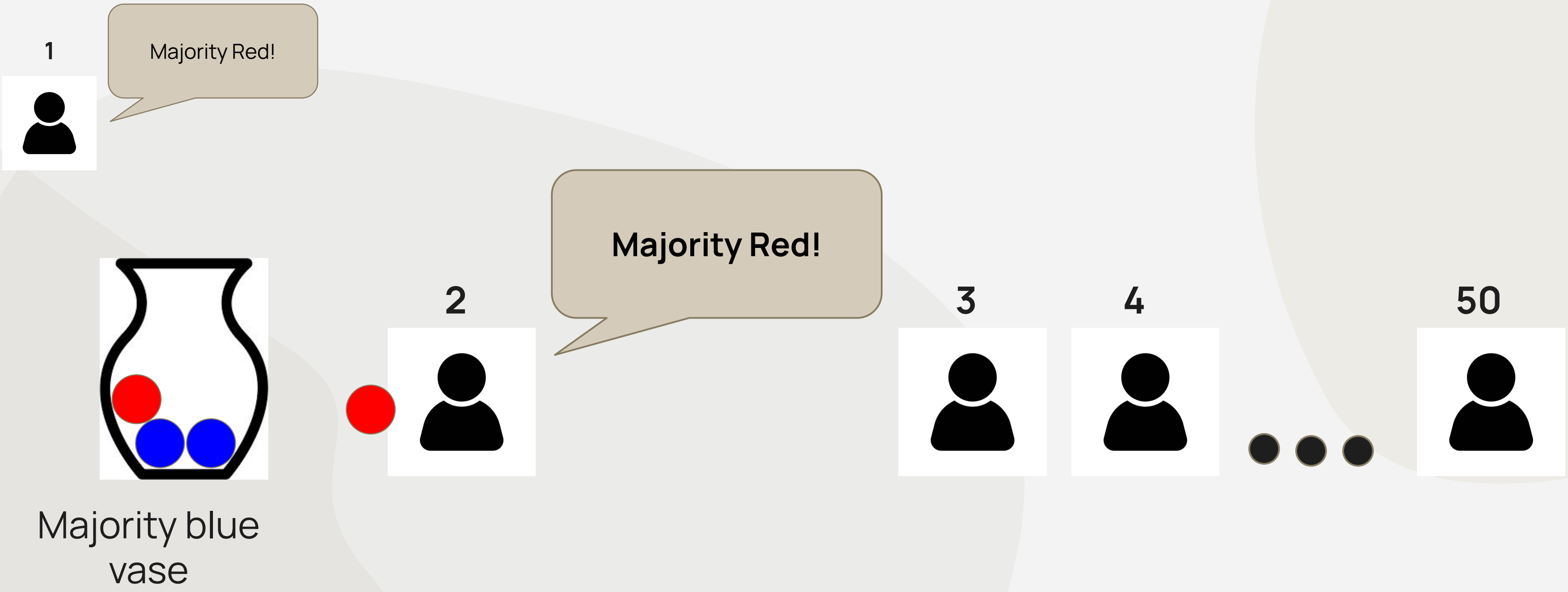
50 people in a line

- There is an opaque vase with 3 balls
  - The vase is equally likely to be majority red or majority blue
- 50 rational people line up to *privately* sample a ball and *publicly* announce whether they think the vase is majority blue or majority red

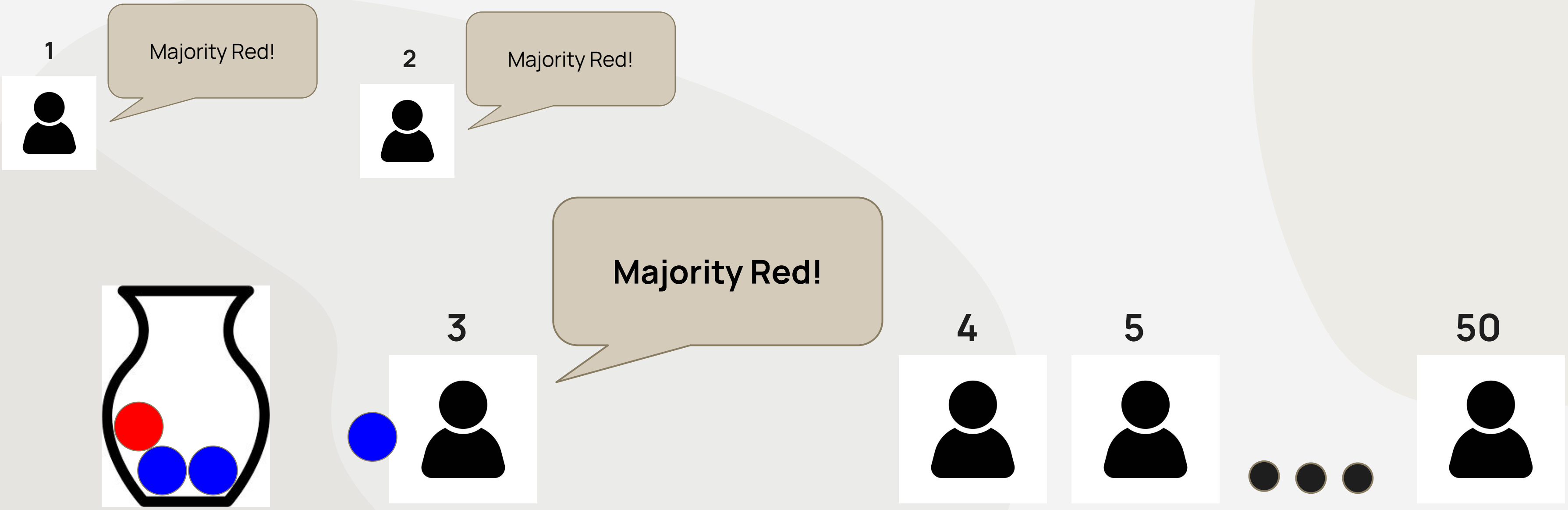
- **Consider a scenario where the vase is majority blue**
- **Suppose the first person samples a red ball. He will predict majority red.**



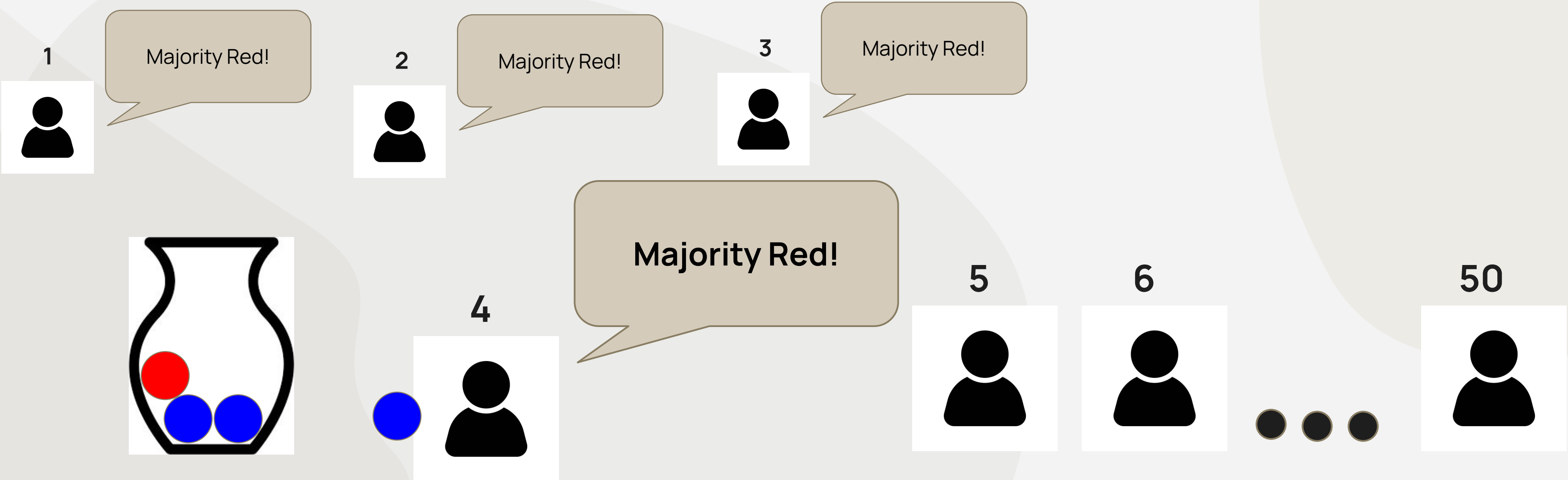
- Consider a scenario where the vase is majority blue
- Suppose the first person samples a red ball. He will predict majority red.
- **Suppose the second person also samples a red ball. He will also predict majority red.**



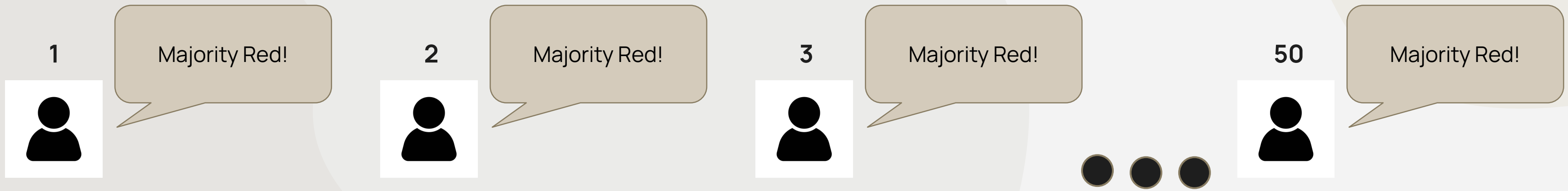
- Consider a scenario where the vase is majority blue
- Suppose the first person samples a red ball. He will predict majority red.
- Suppose the second person also samples a red ball. He will also predict majority red.
- **The third person will predict majority red regardless of what he samples**
  - **Under Bayesian inference, the third person can deduce that the first two people sampled red, so even if he samples blue, majority red is more likely**



- Consider a scenario where the vase is majority blue
- Suppose the first person samples a red ball. He will predict majority red.
- Suppose the second person also samples a red ball. He will also predict majority red.
- The third person will predict majority red regardless of what he samples
  - The third person can deduce that the first two people sampled red, so even if he samples blue, majority red is more likely
- **The fourth person is in the same position as the third person**



- Consider a scenario where the vase is majority blue
- Suppose the first person samples a red ball. He will predict majority red.
- Suppose the second person also samples a red ball. He will also predict majority red.
- The third person will predict majority red regardless of what he samples
    - The third person can deduce that the first two people sampled red, so even if he samples blue, majority red is more likely
- The fourth person is in the same position as the third person
- ...
- **An information cascade occurred since everyone from the third person onwards ignores their private signal and follows the first two people**
- **There is a 1/9th chance that the first two people predict majority red and causes everyone else to copy this incorrect decision!**



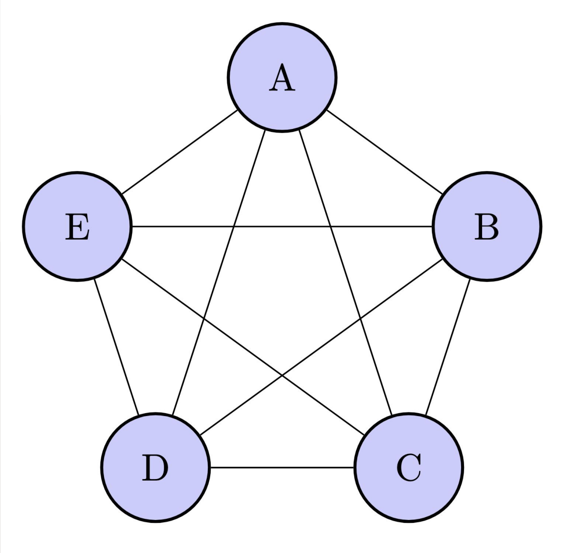
Background/Motivation

Background/Motivation

- Everyone’s beliefs are informed by their own experiences and attitudes, as well as the beliefs of others.
- Network learning naturally models social settings or multi-agent systems where
  - A group of agents need to make a decision
  - Each agent has their own private information
  - A network determines which agents can directly observe each other
  - Agents can observe the *decisions* of other agents but do not have access to their *private information*
- It is important to study in which settings information cascades occur, and more broadly which settings allow rational agents to learn the truth

Networks

Example Order:
ABCDE



- Networks can be modeled with a graph
- Vertices are agents
- Vertices should be ordered to determine the order in which agents make decisions
- Edges depict which agents can observe each other's decisions
- In a complete graph, each agent can observe the decision made by all other previous agents, just like in the majority red/blue vase case.
- As we saw previously, complete graph networks are generally not conducive to network learning due to information cascades

Past Work

DIMACS 2023 - Overall Objective

- Asymptotic Truth Learning - in a network of n agents, almost all agents make a correct prediction with probability approaching 1 as n goes to infinity

$$\frac{1}{n} \cdot \sum_{v} \mathrm{Prob}\{a_v = \theta\} \to 1$$

DIMACS 2023 - Erdös Rényi Graphs

- Sufficient condition: there exists vertex v with set of neighbors S (ISI = w(1)) that are independent of each other
    - Removing S will disconnect at most o(n) vertices from v
- Unless graph is too sparse or too dense, there exists valid decision ordering

Characterizing Truth Learning in an Erdös Rényi graph

$G(n,p)$: n nodes and each edge appears w/ prob $p \in [0,1]$.

- $p = O(1/n)$: too sparse, × with any ordering.
- $p = o(n)/n$, $p = \omega(1)/n$: ✓ with a good ordering
- $p = \Theta(1)$: × with a random ordering – herding happens
- $p = 1 - \omega(n^{\varepsilon})/n$: ✓ with a good ordering
- $p = 1 - O(n^{\varepsilon})/n$: × with any ordering – herding happens.

DIMACS 2024 - NP Hardness

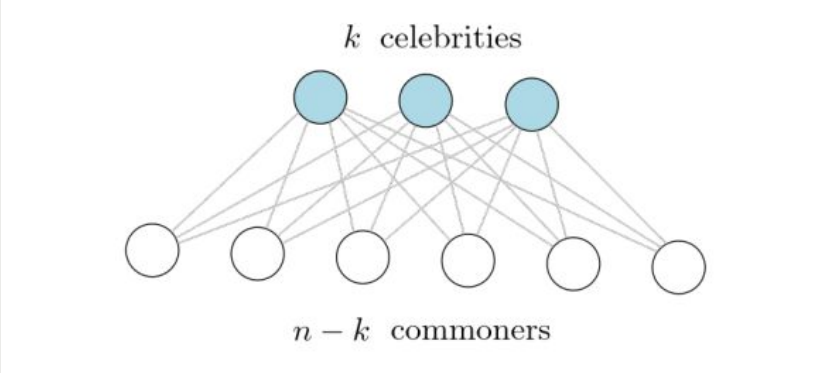
- Decision Problem: for a given network, does there exist an (optimal) decision ordering where the learning rate exceeds some threshold?
- Learning rate: expected proportion of agents to guess correctly

Theorem 3.1. *Network Learning with the Bayesian learning rule $\mu = \mu^B$ is NP-hard.*

Theorem 4.1. *Network Learning with the majority vote rule $\mu = \mu^M$ is NP-hard.*

DIMACS 2025 - Random Orderings

- Analyze networks under random-order learning and not fixed orderings.
- Construct algorithm to alter network to support random-order learning
- Depends on the learning oracle, which is an oracle that answers whether a vertex's learning rate exceeds some threshold under random orderings
- Conjectured to be NP-Hard



The celebrity graph, an example of a graph that learns under random orderings

Research Goals

Idea 1 - NP Hardness

Motivation: Extend the results of the 2025 and 2024 papers

- Is determining the learning rate of a network under *random* orderings an NP-hard problem?
- Is determining the learning rate of a specific vertex in a network (i.e. computing the learning oracle) an NP-hard problem? This is an open conjecture from 2025.
- If computing the learning oracle is indeed NP-hard, what are some other techniques that can transform a network into one that achieves random-order asymptotic truth learning that do not rely on a learning oracle?

CONJECTURE 4.10. *For general graphs G, $v \in V$, and $t \in [0,1]$, computing the output of a learning oracle $f(v,t)$ is NP-hard.*

Idea 2 - Majority Vote Context

Motivation: A lot of focus has been on Bayesian or majority vote protocols, but what about other aggregation methods?

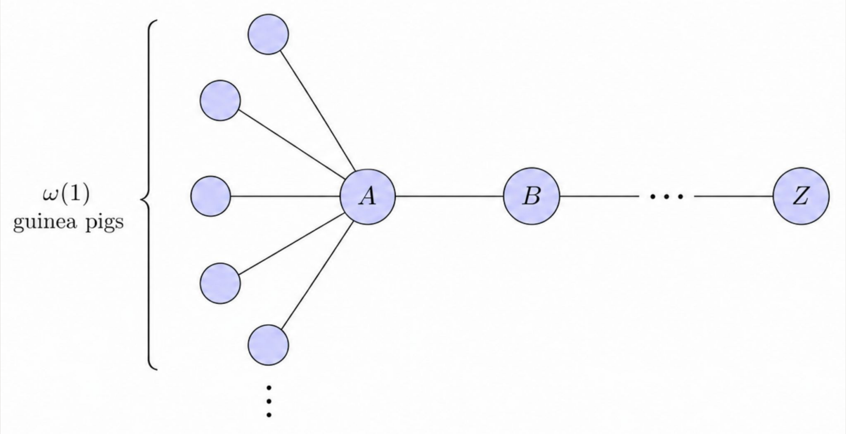
- What if we precompute how each node should weight their neighbors' predictions and then run a weighted majority vote aggregation based on these weights?
- What guarantees can we get? Are there strictly better heuristics than majority vote?
- How efficient can we make our precomputation?

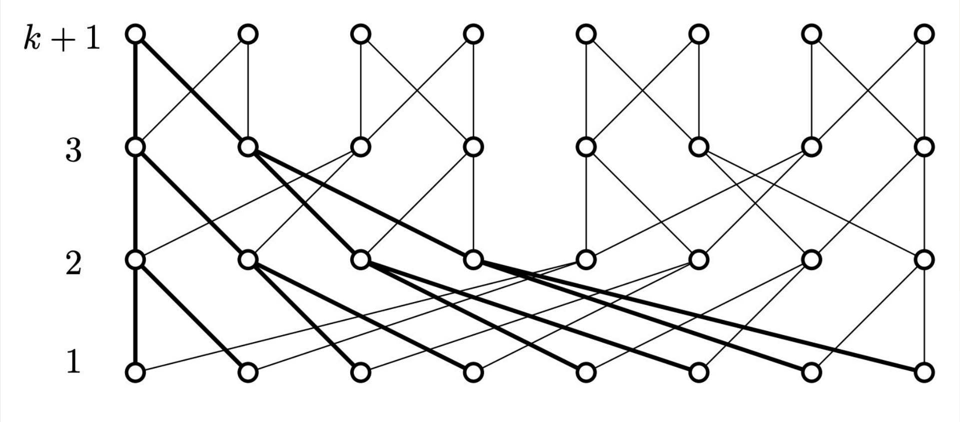
form. More concretely, fully rational agents predict via Bayesian inference, whereas agents with bounded rationality use a simpler heuristic, such as majority vote over all available signals.

Idea 3 - Other Networks

Motivation: A lot of truth learning graph families still rely on an information cascade.

- Are there other graph families that achieve asymptotic truth learning without this structure?
- What requirements characterize such graphs?
- What are the benefits of such networks? Are these networks more robust to adversarial agents?





Bibliography

- David Easley and Jon Kleinberg. 2010. *Networks, crowds, and markets : reasoning about a highly connected world*. Cambridge University Press, New York.
- Filip Úradník, Amanda Wang, and Jie Gao. 2025. Maximizing truth learning in a social network is NP-hard. Retrieved from https://arxiv.org/abs/2502.12704
- Jie Gao. 2024. *Enabling Asymptotic Truth Learning in a Social Network*. Retrieved June 1, 2026 from https://sites.rutgers.edu/jie-gao/wp-content/uploads/sites/375/2024/09/JSM24.pdf
- William Guo and Edward Xiong. 2025. *Truth Learning in a Social Network*. Retrieved June 1, 2026 from https://reu.dimacs.rutgers.edu/~wg270/DIMACS_Presentation_6_3_25.pdf
- William Guo, Edward Xiong, and Jie Gao. 2026. Robust sequential learning in random order networks. Retrieved from https://arxiv.org/abs/2602.08953
- Kevin Lu, Jordan Chong, Matt Lu, and Jie Gao. 2024. Enabling asymptotic truth learning in a social network. Retrieved from https://arxiv.org/abs/2410.04317

Thanks for listening!

This research is being conducted as part of the 2026 DIMACS REU program under the mentorship of Professor Jie Gao.

We like to thank the NSF for supporting the DIMACS REU program under grant CCF-2447342.

We would also like to thank ACTION (AI Institute for Agent-based Cyber Threat Intelligence and Operation), funded by NSF IIS-2229876