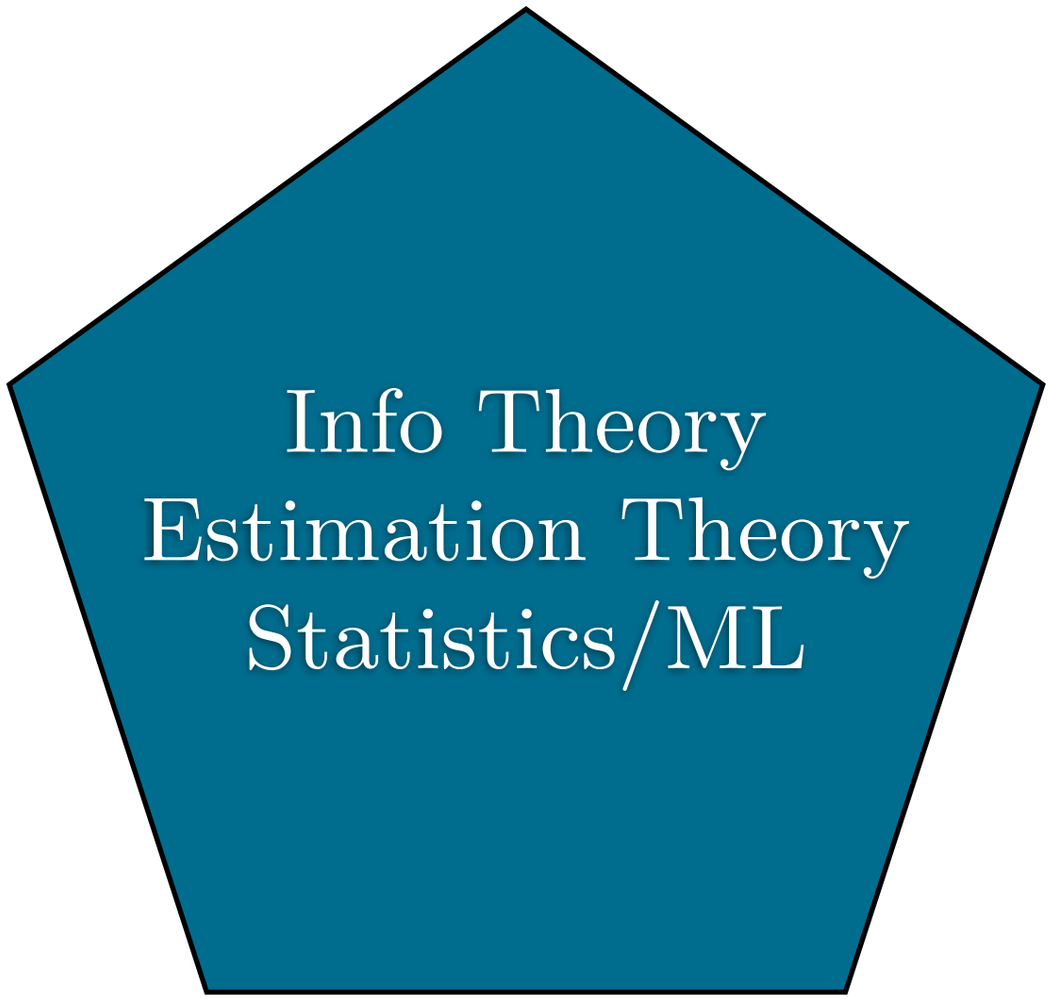
A Communication Network Perspective of Social Networks

Onkar Dabeer
School of Technology and Computer Science
Tata Institute of Fundamental Research
Mumbai, India

This work was funded by India Science Lab, General Motors Global R&D

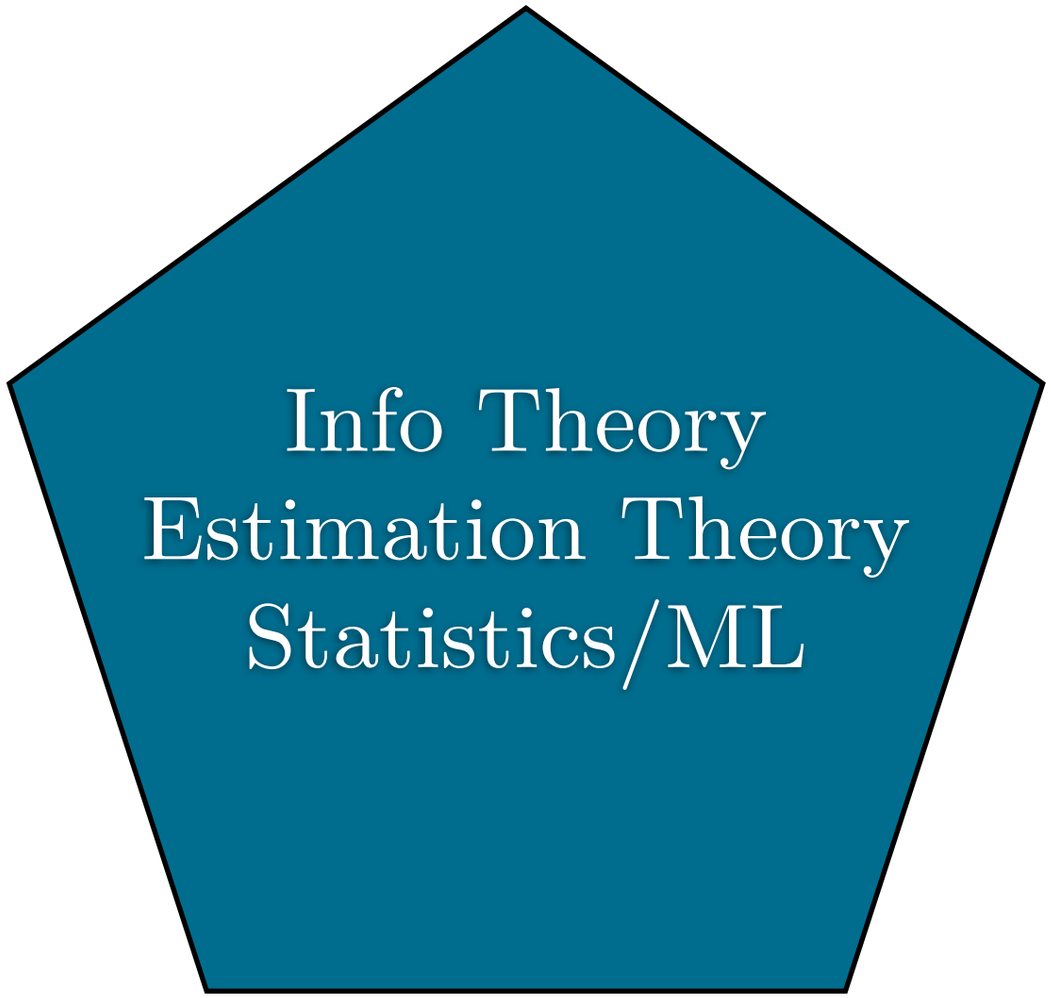
Recent Works



Recent Works

Collaborative Filtering

Social Networks



Haptics

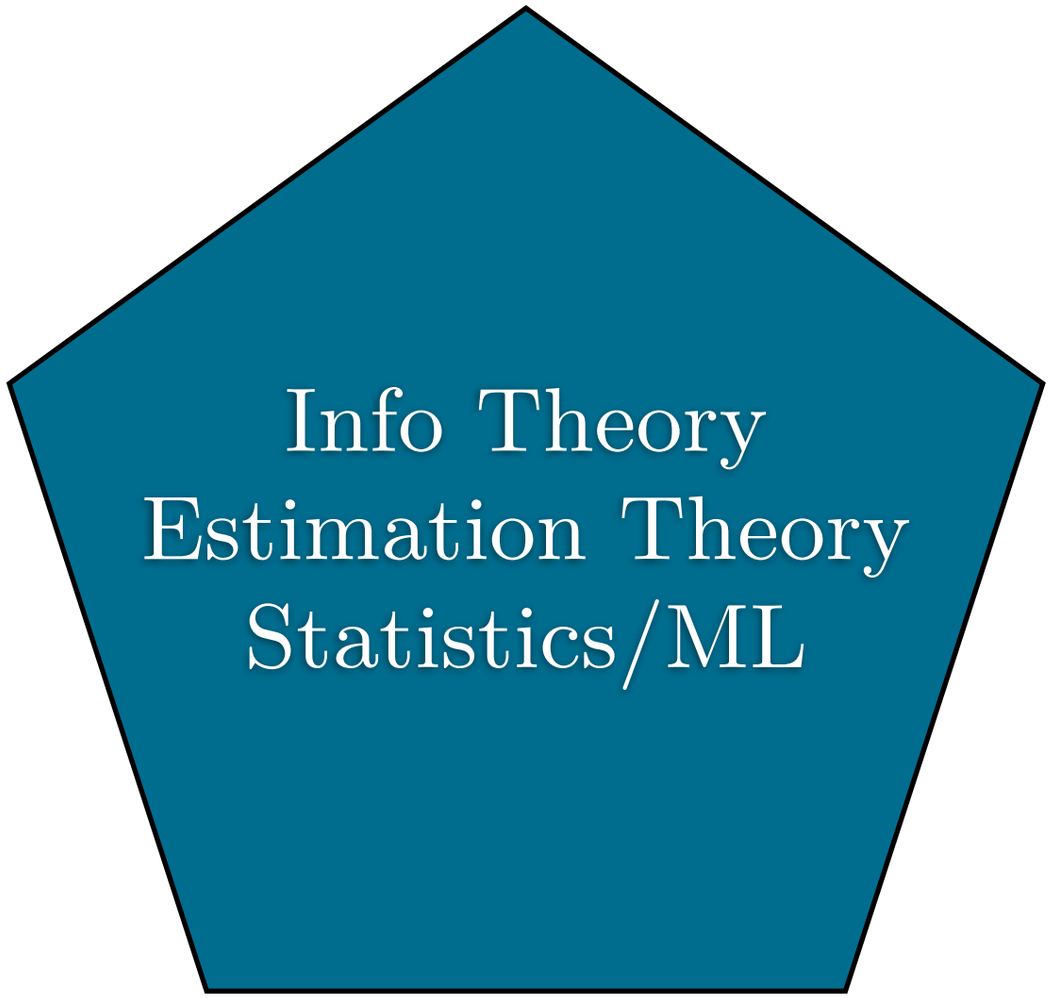
Inverse Lithography

Multi-Gbps Modems

Recent Works

Collaborative Filtering

Social Networks



Haptics

Inverse Lithography

Multi-Gbps Modems

Provably good schemes competitive in real world

Social Networks
(A Comm. Engineer's Perspective)

Social Networks
(A Comm. Engineer's Perspective)

- A lossy communication network
  - Overcrowded inbox, limited attention window

Social Networks
(A Comm. Engineer's Perspective)

- A lossy communication network
  - Overcrowded inbox, limited attention window
- A multi-objective system
  - Reach - how many people read the message?
  - Relevance - are people exposed to relevant content?
  - Age of messages

Social Networks
(A Comm. Engineer's Perspective)

- A lossy communication network
  - Overcrowded inbox, limited attention window
- A multi-objective system
  - Reach - how many people read the message?
  - Relevance - are people exposed to relevant content?
  - Age of messages
- Controls - message display, connection recommendations, etc.

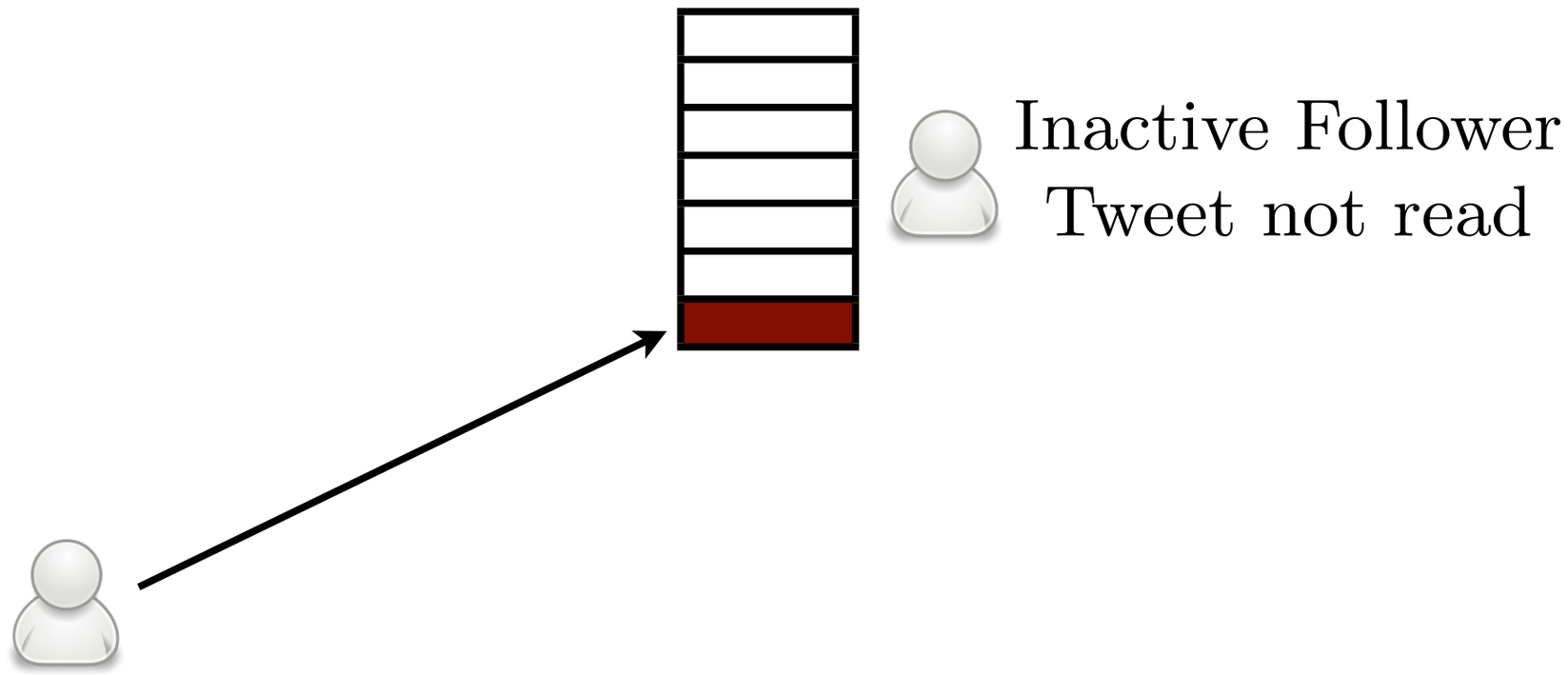
Social Networks
(A Comm. Engineer's Perspective)

- A lossy communication network
  - Overcrowded inbox, limited attention window
- A multi-objective system
  - Reach - how many people read the message?
  - Relevance - are people exposed to relevant content?
  - Age of messages
- Controls - message display, connection recommendations, etc.
- **Goal:** Learn models for network, design optimal control, test and refine

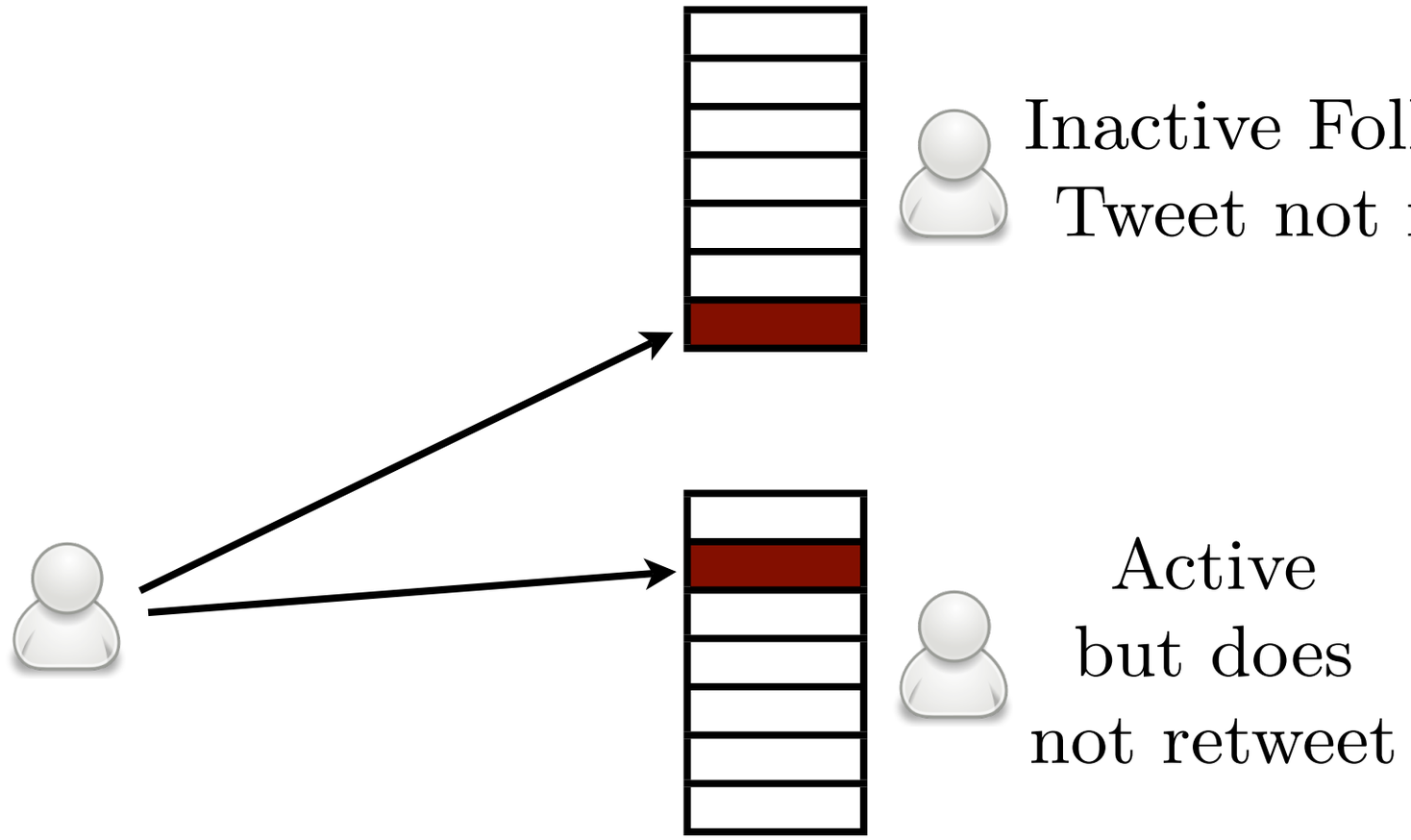
Campaigns Over Twitter

The Problem

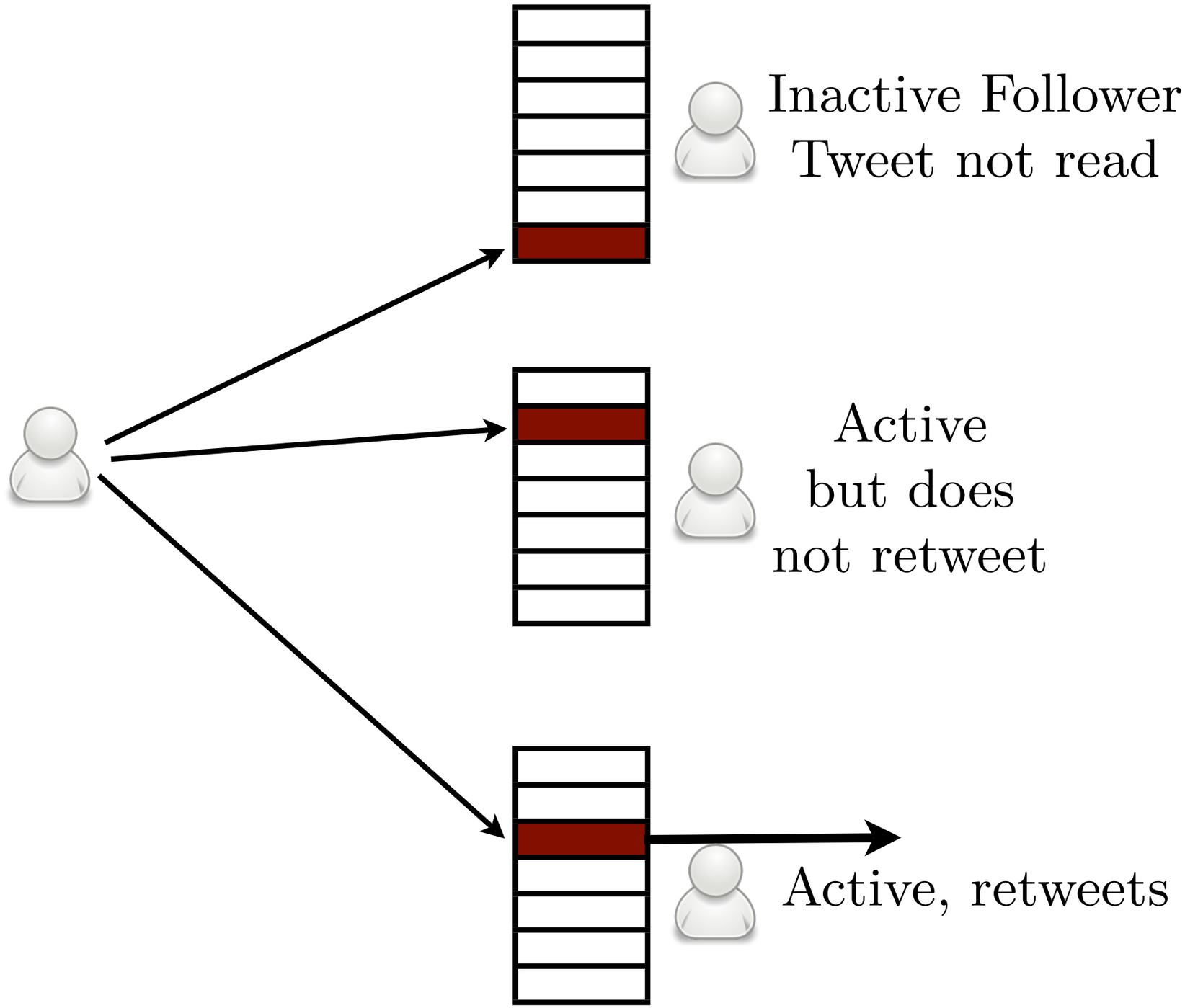
The Problem



The Problem

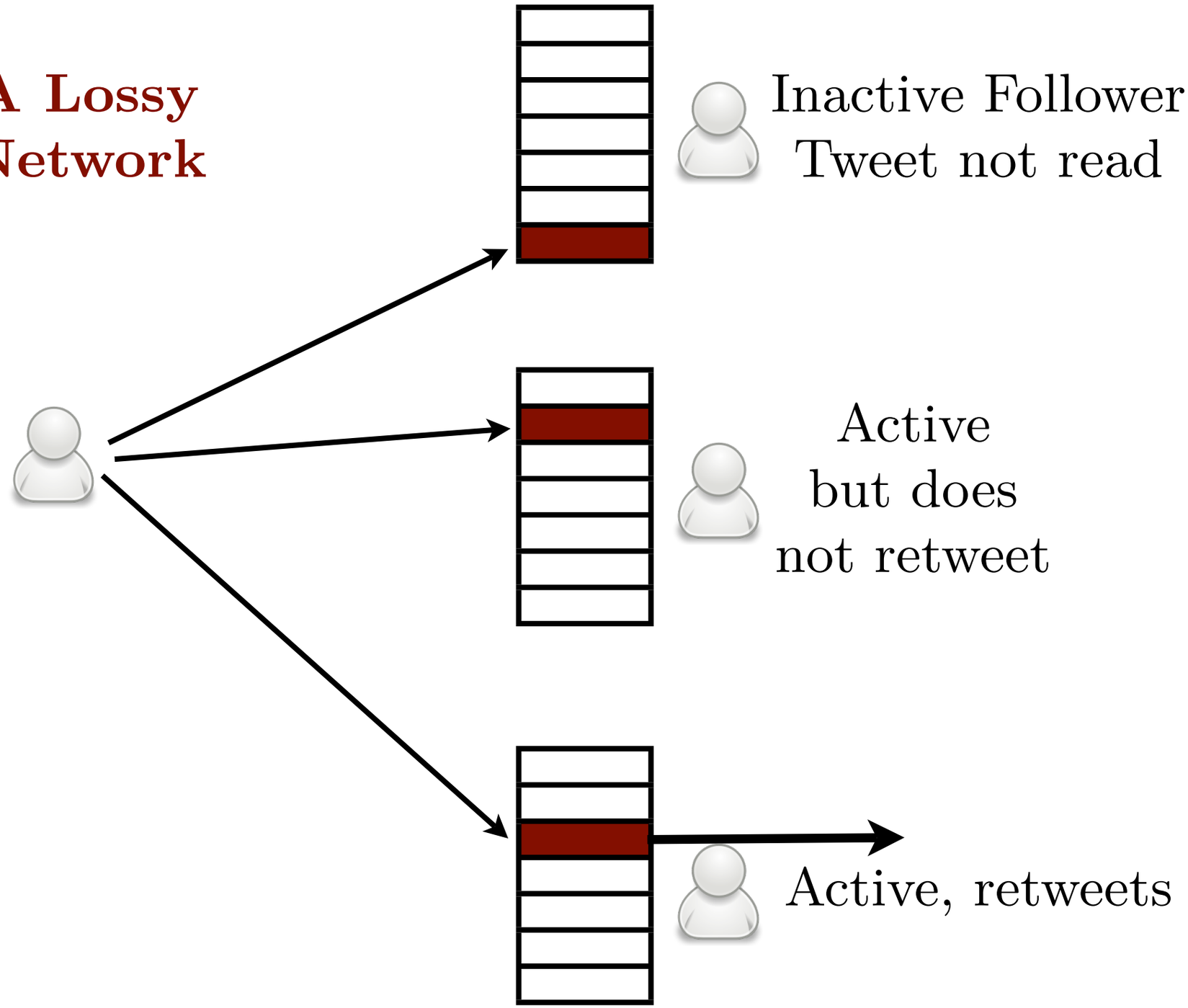

Inactive Follower
Tweet not read
Active
but does
not retweet

The Problem


Inactive Follower
Tweet not read
Active
but does
not retweet
Active, retweets

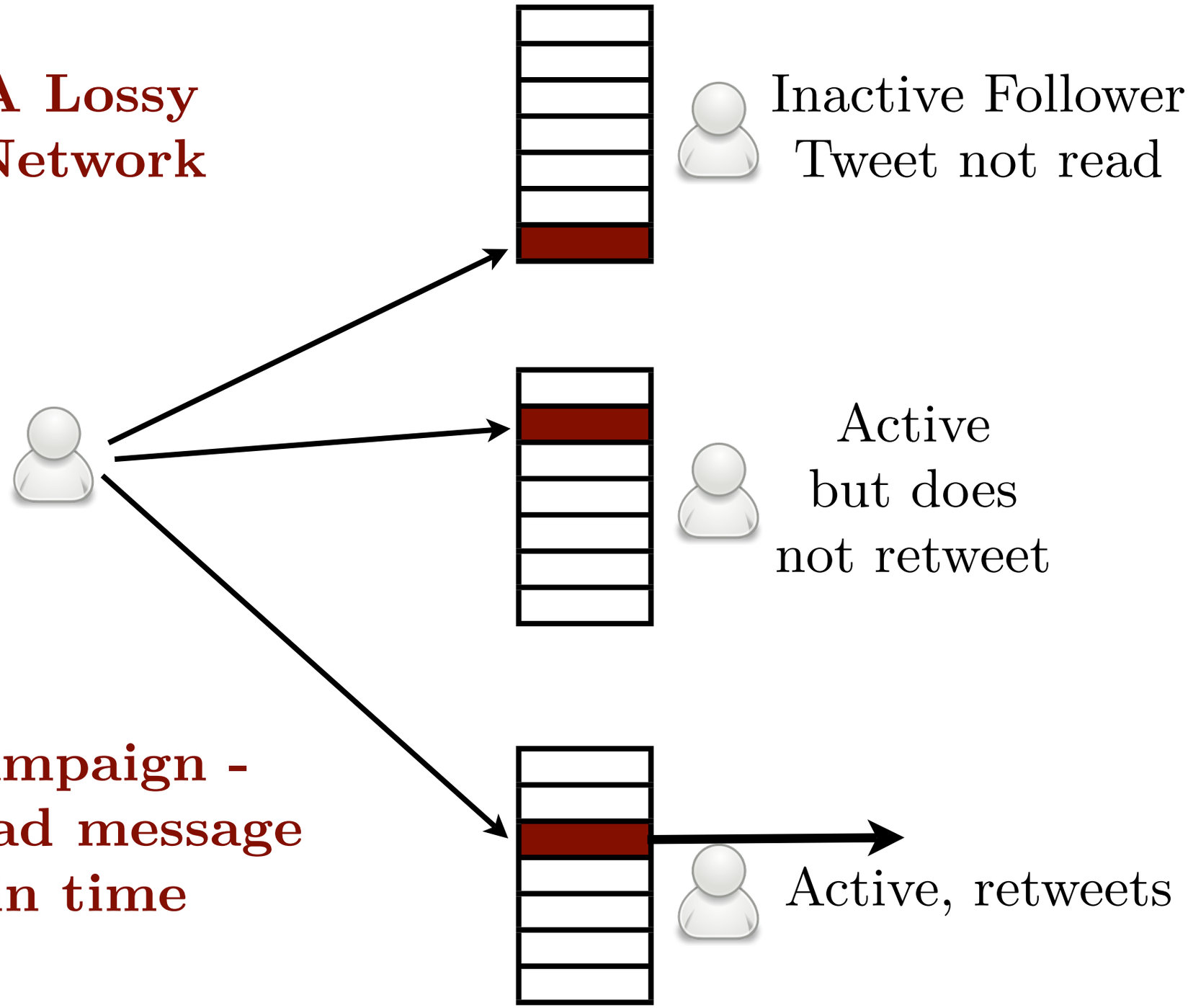
The Problem

A Lossy Network



The Problem

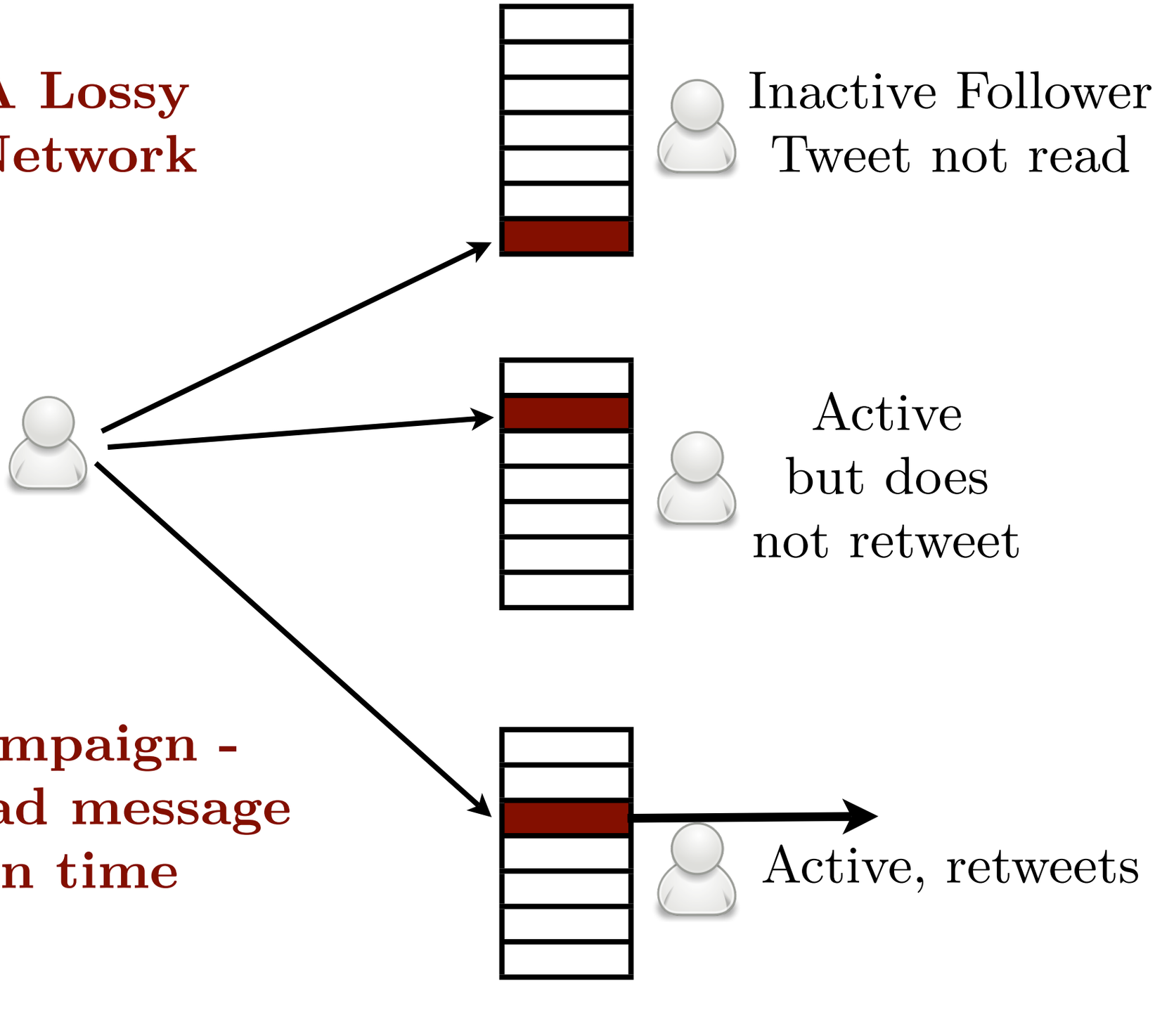
A Lossy Network



Campaign - spread message in time

The Problem

A Lossy Network



The Problem (Contd.)

The Problem (Contd.)

- Incessant tweeting
  - If you believe tweeting does not cause any irritation

The Problem (Contd.)

- Incessant tweeting
  - If you believe tweeting does not cause any irritation
- Few tweets
  - Tweet in slots with high expected # of active users
  - Account for response probabilities

The Problem (Contd.)

- Incessant tweeting
  - If you believe tweeting does not cause any irritation
- Few tweets
  - Tweet in slots with high expected # of active users
  - Account for response probabilities
- Can we measure 'irritation'?

The Problem (Contd.)

- Incessant tweeting
  - If you believe tweeting does not cause any irritation
- Few tweets
  - Tweet in slots with high expected # of active users
  - Account for response probabilities
- Can we measure 'irritation'?
- Tradeoff between reach and irritation?
  - Structure of optimal strategies?

A Model

Graph, Activity, Response

Graph, Activity, Response

- **Graph:** A root node with several followers
  - Aim is to penetrate through the ring of followers
  - Leads to strategies based on 'local' data

Graph, Activity, Response

- **Graph:** A root node with several followers
  - Aim is to penetrate through the ring of followers
  - Leads to strategies based on 'local' data
- **Activity:**
  - Weekly periodicity
  - Activity probabilities for 24 x 7 slots known
  - Independence across followers

Graph, Activity, Response

- **Graph:** A root node with several followers
  - Aim is to penetrate through the ring of followers
  - Leads to strategies based on 'local' data
- **Activity:**
  - Weekly periodicity
  - Activity probabilities for 24 x 7 slots known
  - Independence across followers
- **Response:**
  - Response probability of each follower known

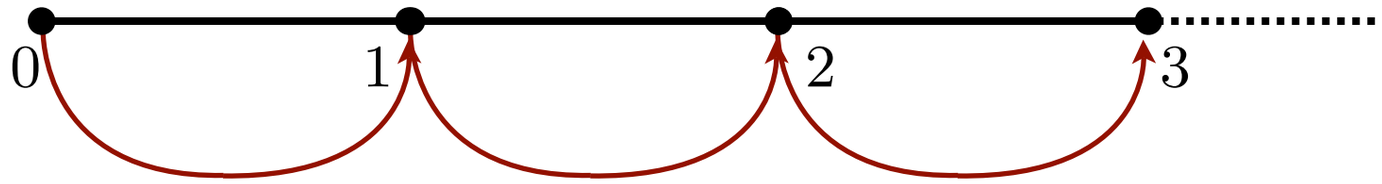
(Potential) Irritation

(Potential) Irritation

Irritation state of a follower

(Potential) Irritation

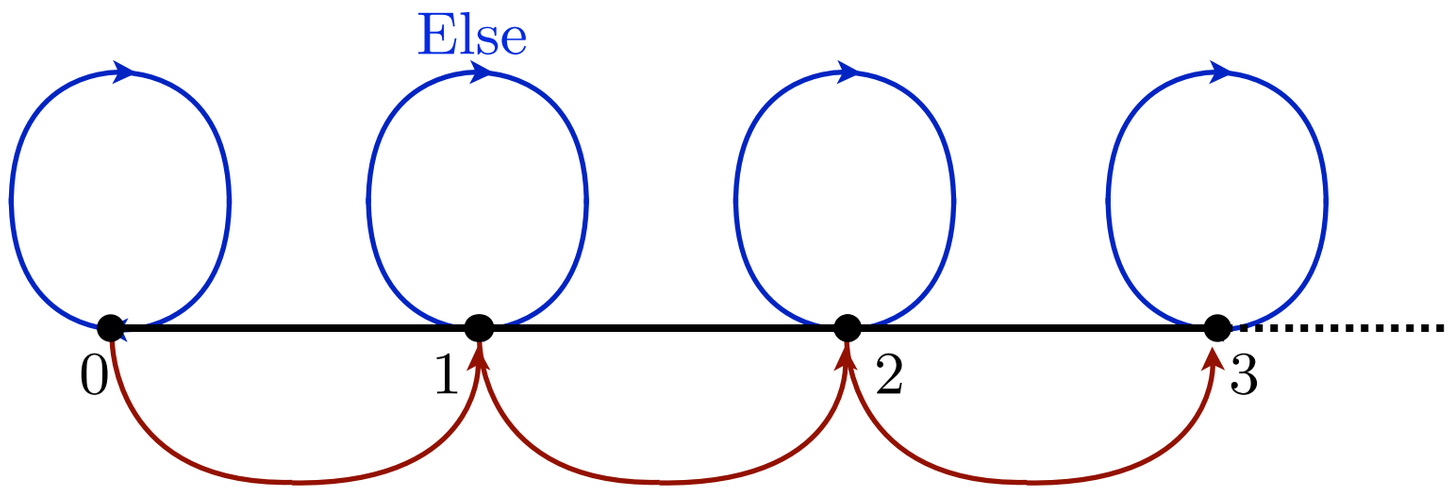
Irritation state of a follower



Root tweets, follower active,
but no response

(Potential) Irritation

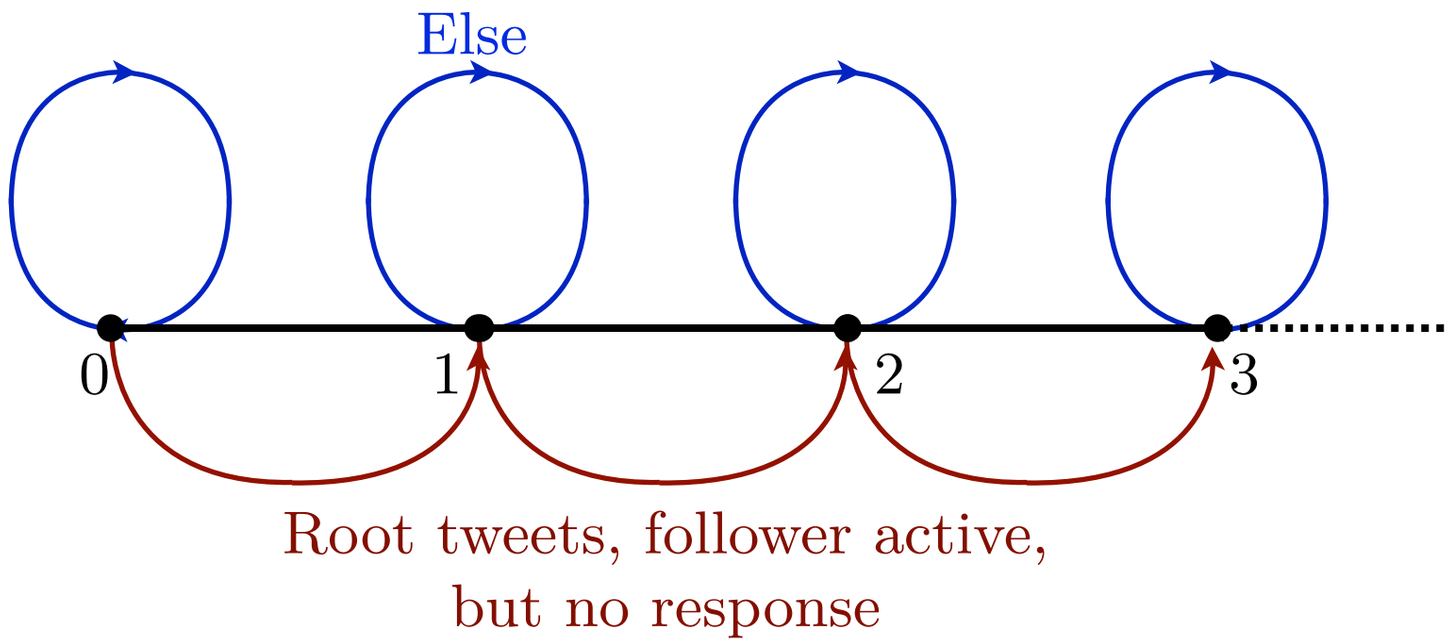
Irritation state of a follower



Root tweets, follower active,
but no response

(Potential) Irritation

Irritation state of a follower



Response probability may depend on state

- **Optimistic Case:** State independent response prob.
- **Pessimistic Case:** No response under irritation

Data Collection and Parameter Estimation

The Data

The Data

- 1 root node, 11916 followers, 6837 allow tweet collection

The Data

- 1 root node, 11916 followers, 6837 allow tweet collection
- 50+ accounts to get around rate limitation

The Data

- 1 root node, 11916 followers, 6837 allow tweet collection
- 50+ accounts to get around rate limitation
- Avg of about 800 tweets per user

The Data

- 1 root node, 11916 followers, 6837 allow tweet collection
- 50+ accounts to get around rate limitation
- Avg of about 800 tweets per user
- Few months to few years

The Data

- 1 root node, 11916 followers, 6837 allow tweet collection
- 50+ accounts to get around rate limitation
- Avg of about 800 tweets per user
- Few months to few years
- Estimation of activity probabilities:
  - 26 most recent weeks
  - 24 x 7 time slots
  - # weeks with tweet in the slot/# weeks

Estimating Response Probability

Estimating Response Probability

- Assume tweets from root are independent

Estimating Response Probability

- Assume tweets from root are independent
- Ratio of
  - # of 'responses' to root tweets, and,
  - # of tweets by root when follower is 'active'

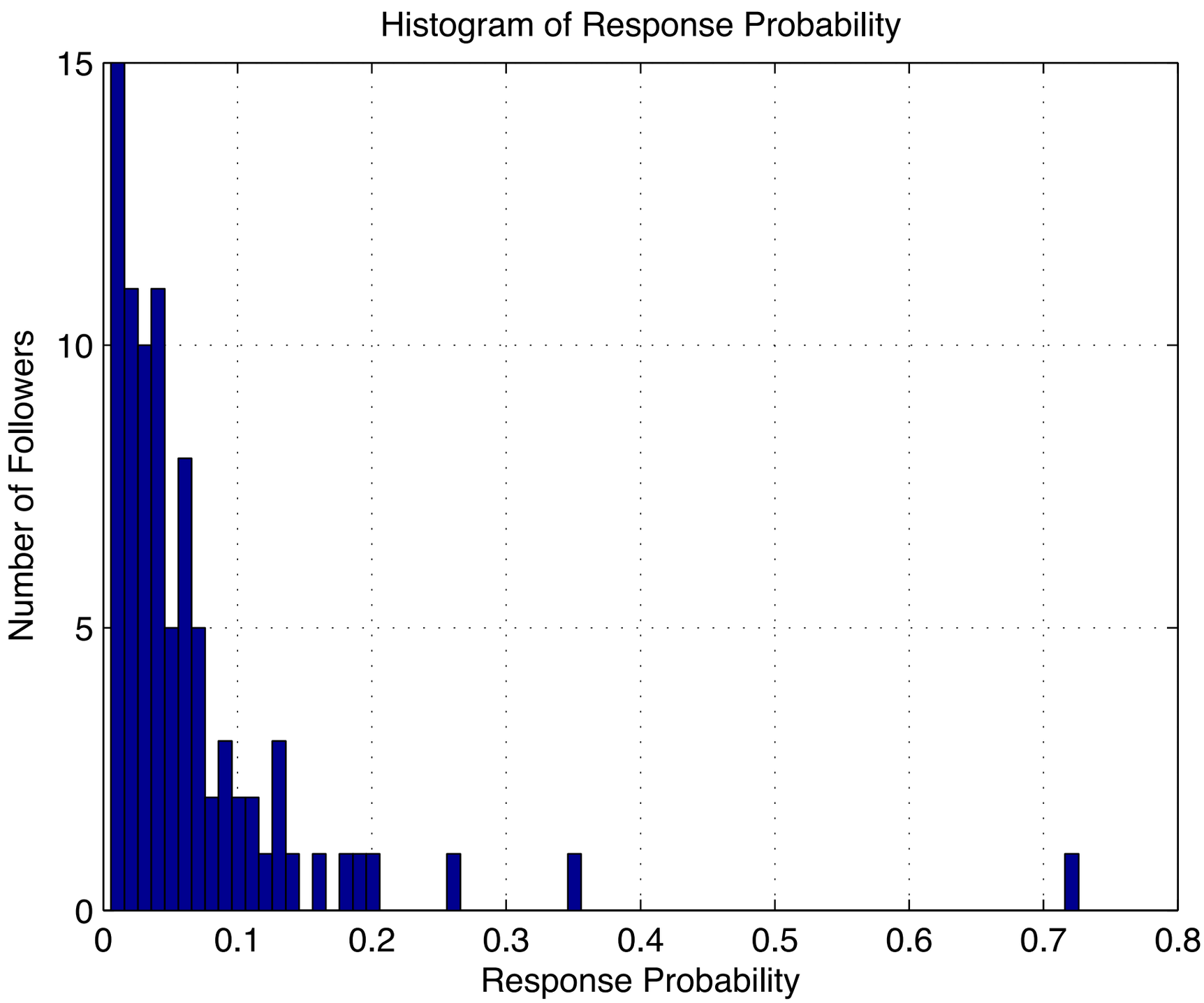
Estimating Response Probability

- Assume tweets from root are independent
- Ratio of
  - # of 'responses' to root tweets, and,
  - # of tweets by root when follower is 'active'
- Confidence interval: pick only those followers with good estimates

Estimating Response Probability

- Assume tweets from root are independent
- Ratio of
  - # of 'responses' to root tweets, and,
  - # of tweets by root when follower is 'active'
- Confidence interval: pick only those followers with good estimates
- Only 84 followers have non-zero response probability

Response Probability



The Markov Decision Process (Finite Horizon)

Markov Chain and Its Control

Markov Chain and Its Control

- **State:** Vector of irritation states of all users
  - Independent activity across followers
  - Activity independent of irritation state
  - Conditioned on current state, the response is independent of the past

Markov Chain and Its Control

- **State:** Vector of irritation states of all users
  - Independent activity across followers
  - Activity independent of irritation state
  - Conditioned on current state, the response is independent of the past
- **Control:** The decision to transmit in the slot or not
  - May depend on the state; root needs to follow followers
  - State independent for many policies we consider

Rewards

Rewards

- Reaching a wide audience
  - Reach = weighted sum of retweets by followers
  - Weight = out-degree
  - Could take into account authority scores, sentiment estimates, etc.

Rewards

- Reaching a wide audience
  - Reach = weighted sum of retweets by followers
  - Weight = out-degree
  - Could take into account authority scores, sentiment estimates, etc.
- Passive consumers
  - Many consume information but retweet rarely
  - # Active followers
  - A reward just to reach followers

Costs

Costs

- Net irritation in campaign
  - Campaign ends in a fixed time
  - Terminal cost = weighted sum of terminal state vector
    - Aggregate irritation at end of campaign
    - Weight = out-degree, but could be more involved

Costs

- Net irritation in campaign
  - Campaign ends in a fixed time
  - Terminal cost = weighted sum of terminal state vector
    - Aggregate irritation at end of campaign
    - Weight = out-degree, but could be more involved
- Human resource
  - # Tweets in campaign
  - Fewer tweets, lesser human effort in constructing novel tweets in campaign

Total Reward

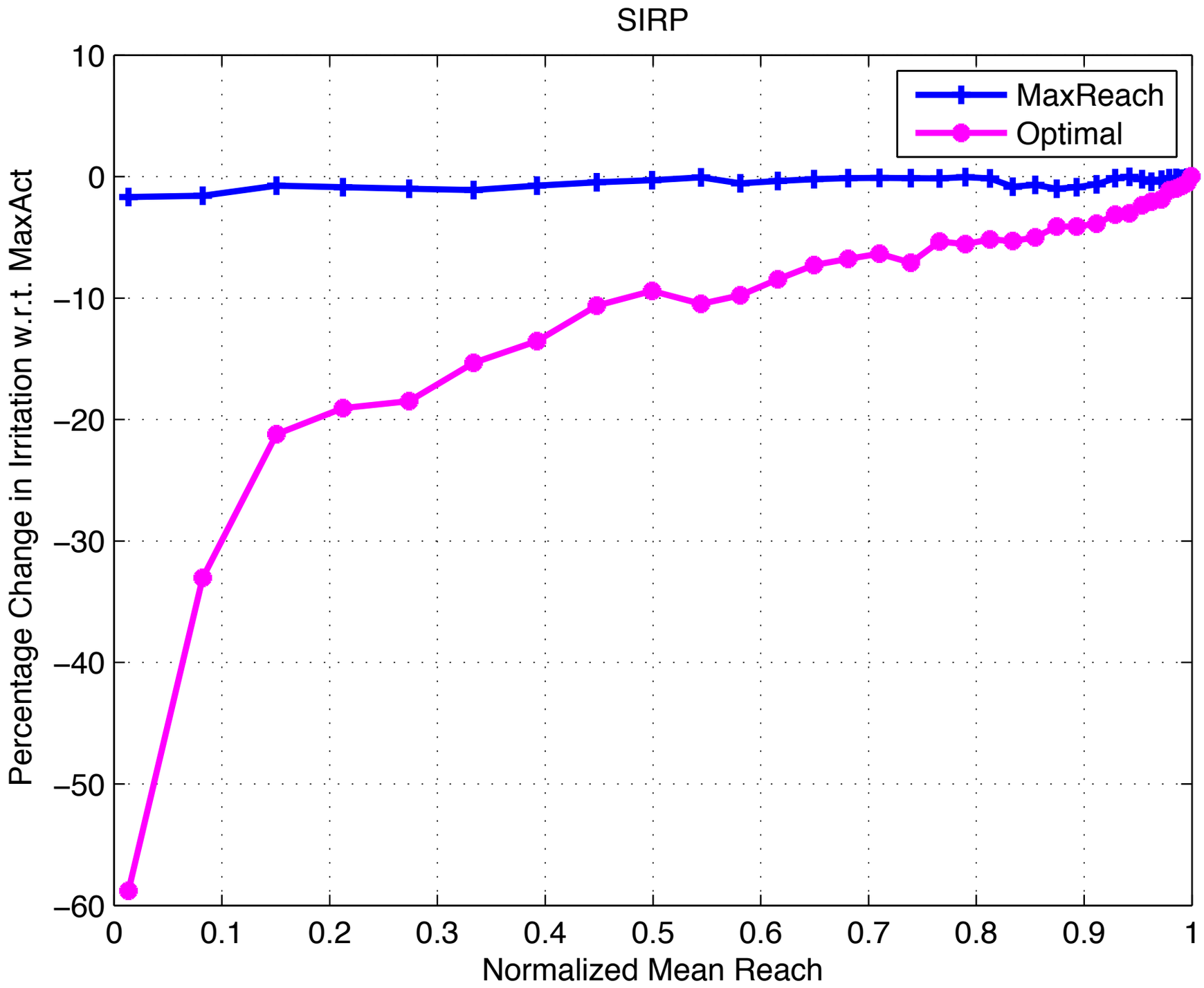
$$\lambda_1 \underbrace{\sum_{t=1}^{T}\sum_{n=1}^{N} w_n 1(n \text{ retweets at } t)}_{\textbf{Reach}} + \lambda_2 \underbrace{\sum_{t=1}^{T}\sum_{n=1}^{N} 1(n \text{ active at } t)}_{\textbf{\# Active}}$$

$$-\lambda_3 \underbrace{\sum_{n=1}^{N} w_n S_n(T)}_{\textbf{Net Irritation}} - \lambda_4 \underbrace{\sum_{t=1}^{T} 1(\text{Root tweets at } t)}_{\textbf{\# Tweets}}$$

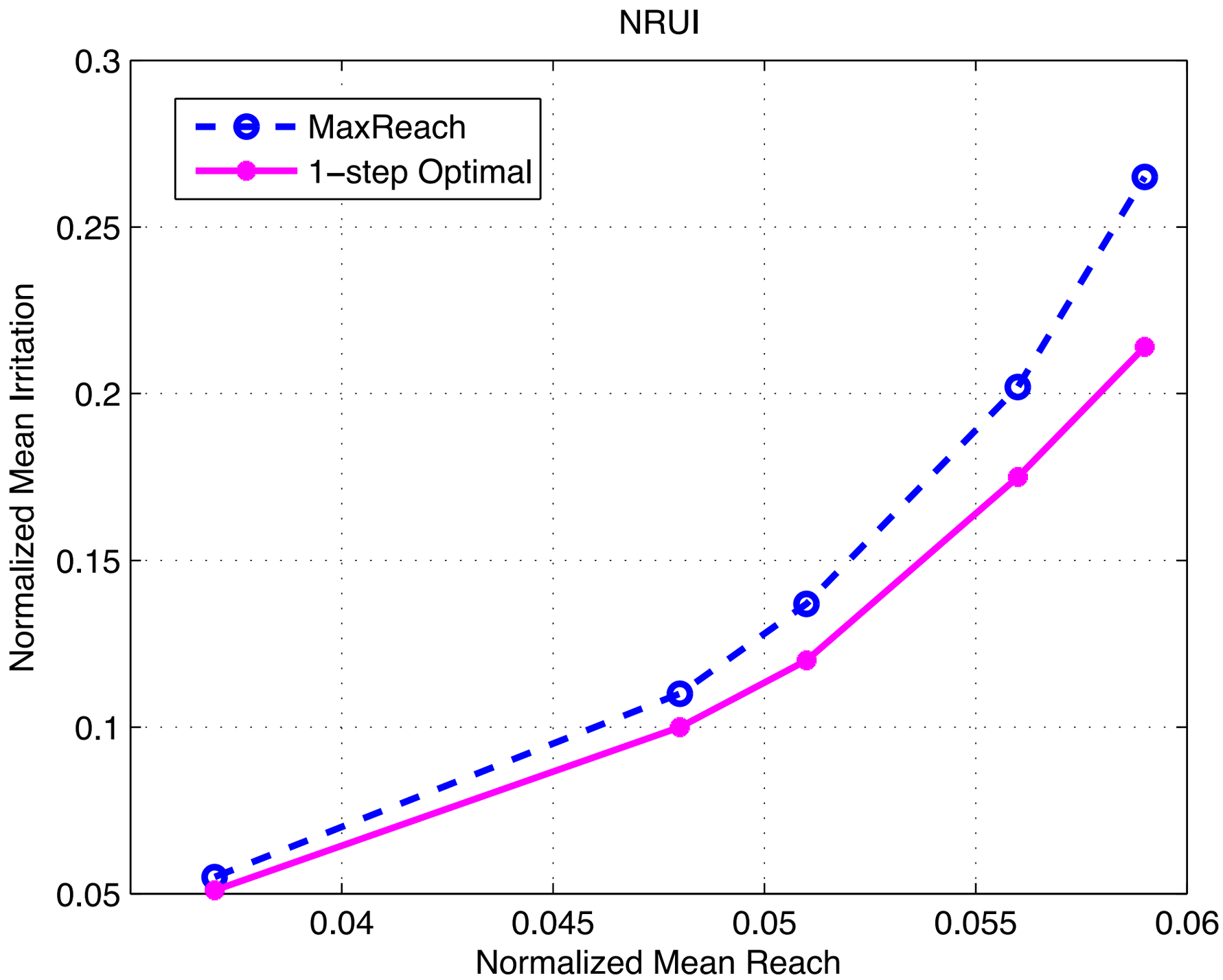
- Non-negative parameters
- **Goal:** Maximize expected reward
- In principle, optimal policy can be found by dynamic programming

Simulations

Optimistic Case



Pessimistic Case



Discussion

Conclusion

Conclusion

- Importance of tweet composition
  - Pessimistic case reach << optimistic case

Conclusion

- Importance of tweet composition
  - Pessimistic case reach << optimistic case
- MaxRatio
  - Different slots than MaxReach/MaxAct
  - But many more

Conclusion

- Importance of tweet composition
  - Pessimistic case reach << optimistic case
- MaxRatio
  - Different slots than MaxReach/MaxAct
  - But many more
- MaxReach
  - Within 10-20% of best policy we have tried

Conclusion

- Importance of tweet composition
  - Pessimistic case reach << optimistic case
- MaxRatio
  - Different slots than MaxReach/MaxAct
  - But many more
- MaxReach
  - Within 10-20% of best policy we have tried
- MaxAct
  - Reward for reaching passive followers

Much More Remains...

Much More Remains...

- Composition codes
  - Play role of FEC - spread message in time, but avoid repetition
  - Relation between 'irritation' & response?

Much More Remains...

- Composition codes
  - Play role of FEC - spread message in time, but avoid repetition
  - Relation between 'irritation' & response?
- Network performance
  - Considered a single node in isolation
  - What happens if a significant number of nodes employ analytics?

Much More Remains...

- Composition codes
  - Play role of FEC - spread message in time, but avoid repetition
  - Relation between 'irritation' & response?
- Network performance
  - Considered a single node in isolation
  - What happens if a significant number of nodes employ analytics?
- The general network with order of message display as control

Thanks!