Predictability of User Behavior in Social Media

Bottom-Up vs. Top-Down Modeling

AAAI Symposium: Social Networks and Social Contagion
15 November 2013

Jared Sylvester
David Darmon
William Rand
Michelle Girvan

CCB
Center for
Complexity in Business

ROBERT H. SMITH
SCHOOL OF BUSINESS
AT THE UNIVERSITY OF MARYLAND

UNIVERSITY OF MARYLAND
18 56

Predictability of User Behavior in Social Media

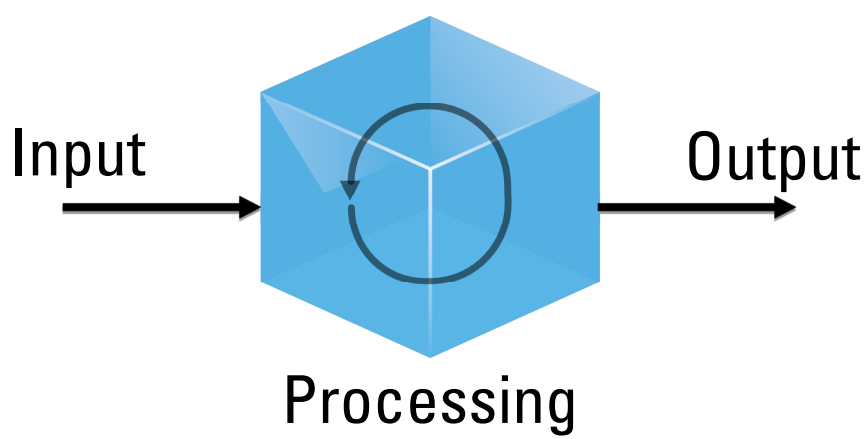
The Individual as a Computational Unit

Unprecedented access to people's behaviors

- Millions of users
- Second-level resolution.

Treat user as information processing unit

- Claude Shannon, 1948: Information Theory and Channels
- Simon DeDeo, 2012: Markov Models of Wikipedia Activity



Data Collection and Processing

Predictability of User Behavior in Social Media

The Dataset: Twitter Users

3k most garrulous users from a 15k follower network

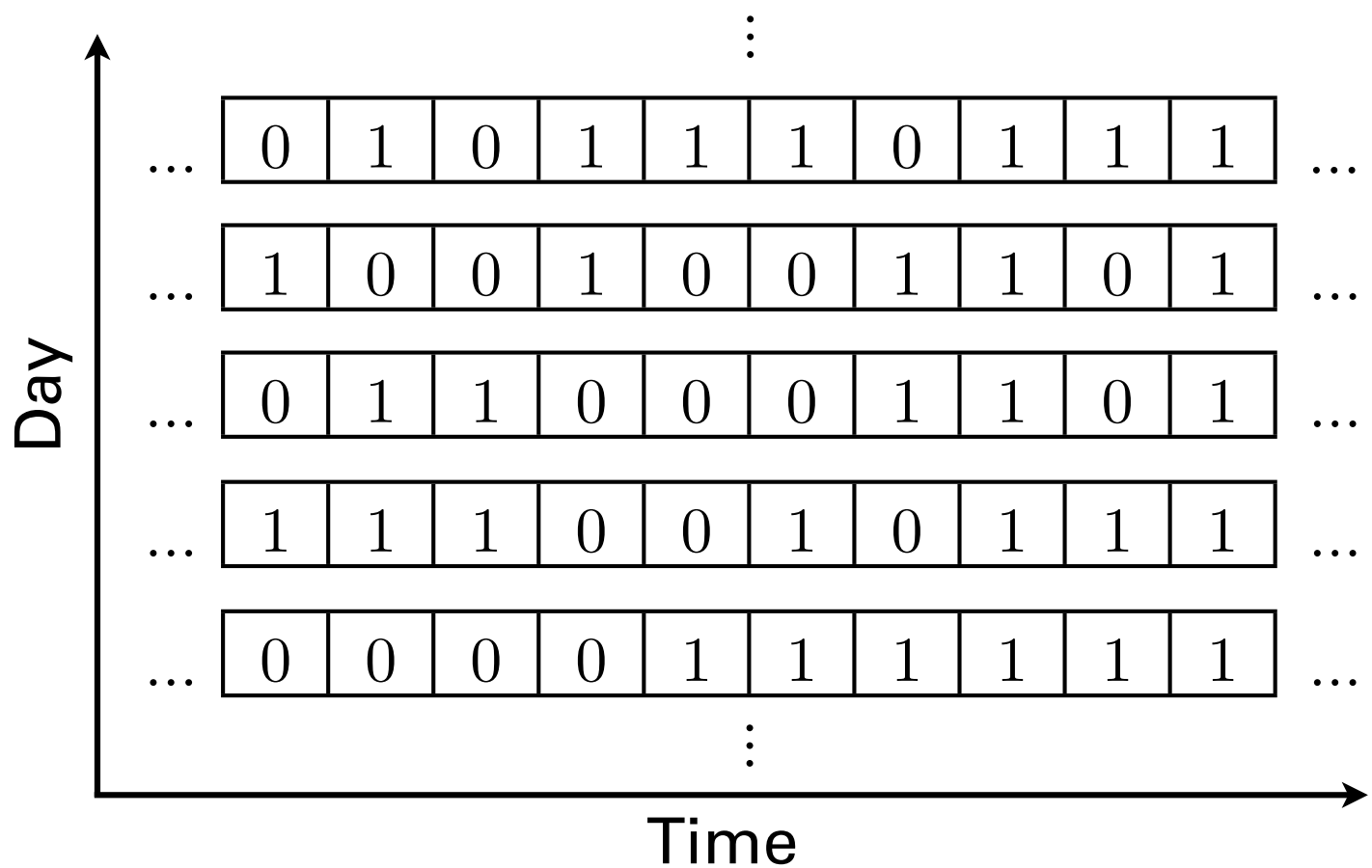
Build discrete time series for each user

$$X(v,t) = 1$$

user v tweets

$$X(v,t) = 0$$

user v doesn't tweet



Predictability of User Behavior in Social Media

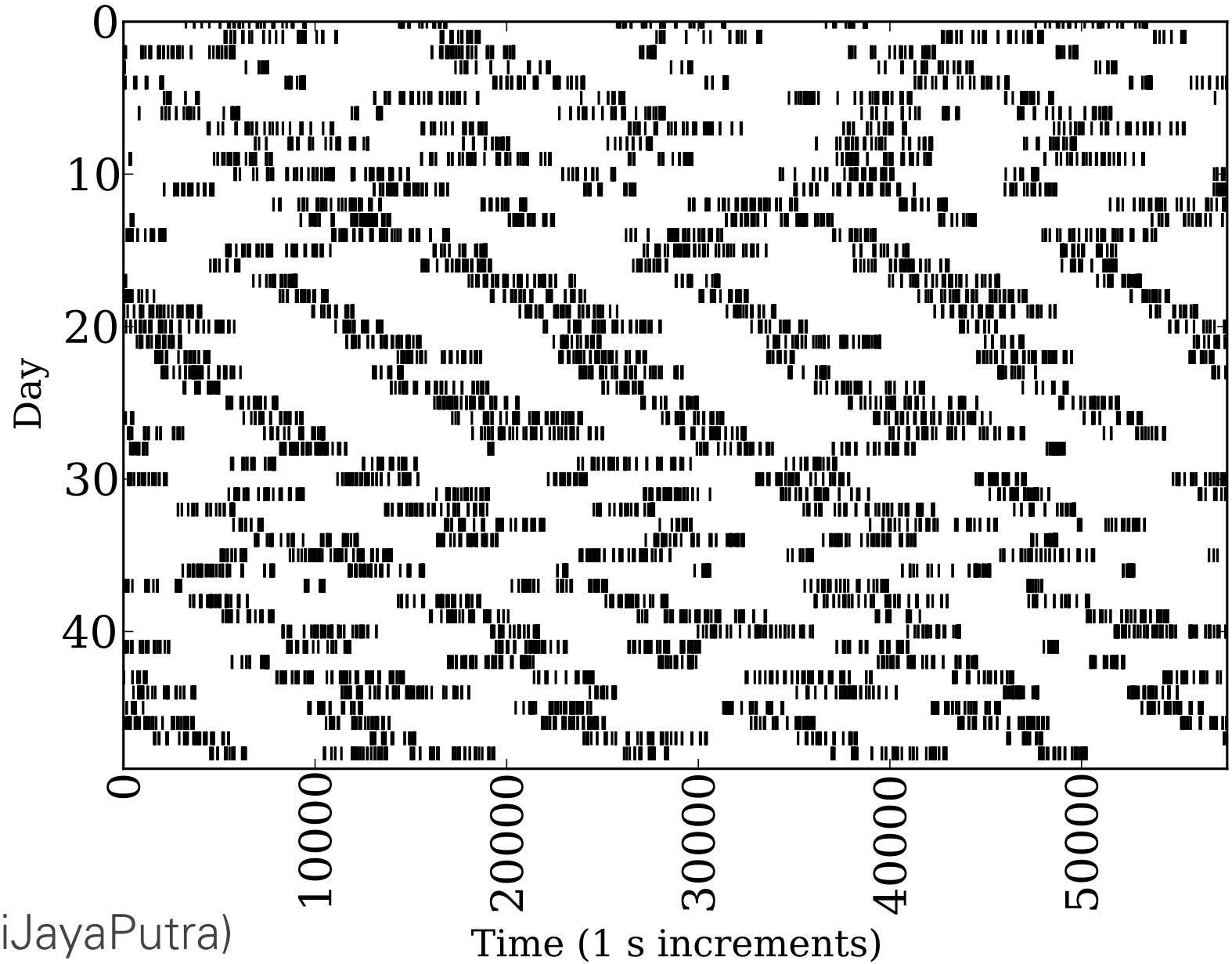
The Dataset

Timestamp	Status Text
2013-08-22 12:54:06	Is Your Gmail Social? How to Use Gmai
2013-08-22 13:11:22	Facebook's Embedded Posts Now Availab
2013-08-22 13:14:06	The Credible Hulk http://t.co/q17VrcSdBs
2013-08-22 13:29:02	25 Things You Didn't Know About Ninjas
2013-08-22 13:32:59	Twitter Users: Revoke and Reestablish Th
2013-08-22 13:48:46	10 Brilliant Facebook Marketing Tactics to
2013-08-22 14:17:11	Google Now Adds Cards for NCAA Footba
2013-08-22 15:18:03	What is the NSA Really Up To? [COMIC] h
2013-08-22 15:39:04	6 Things Every Good Business Blog MUST

(@DanielZeevi)

Predictability of User Behavior in Social Media

The Dataset: Rastergram



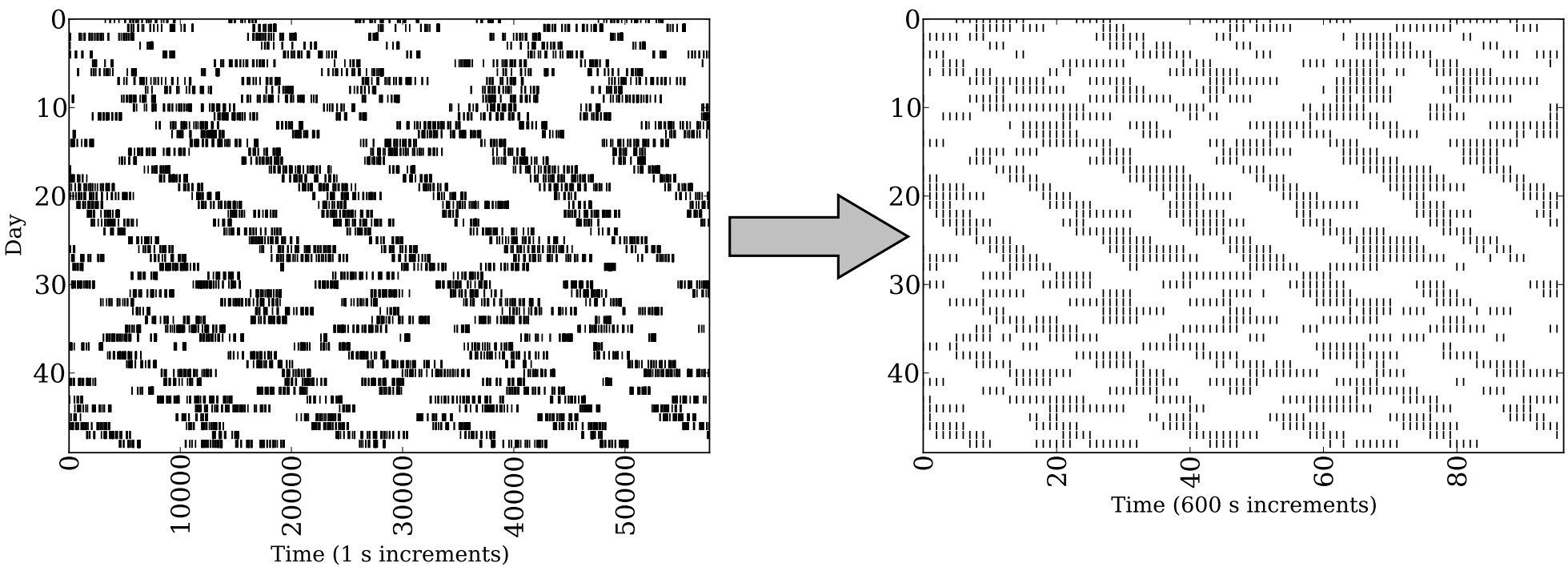
Predictability of User Behavior in Social Media

The Dataset: Coarsening

Looking L steps back in time:

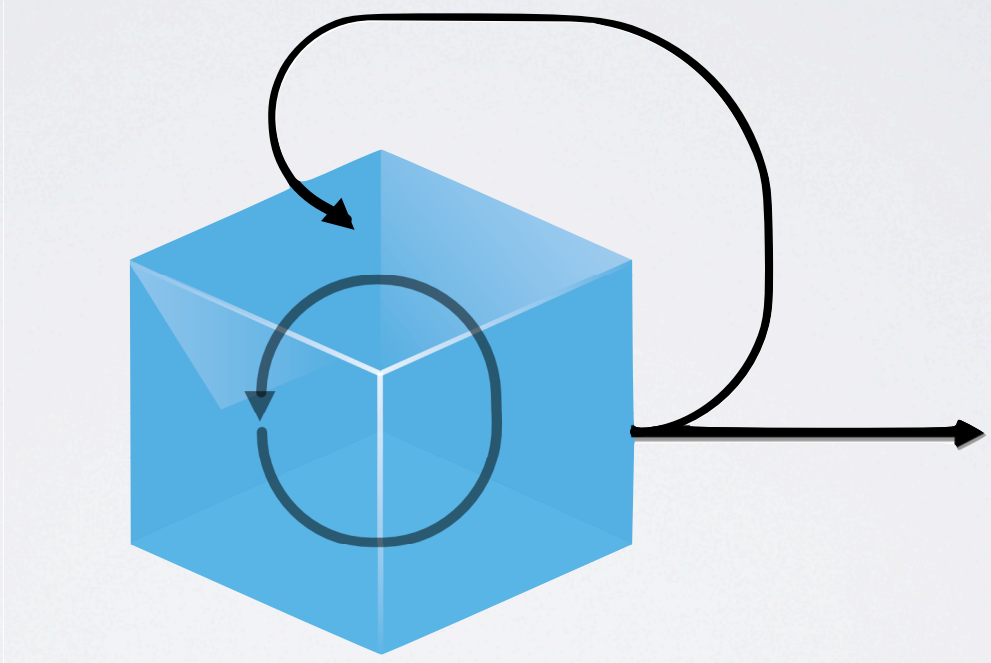
Dimensionality of predictive space is $\mathcal{O}(2^L)$

Need to coarsen



bin size = 600 s = 10 min

Models



Predictability of User Behavior in Social Media

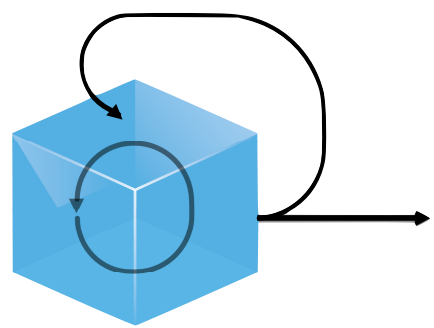
Modeling Framework: A Predictive View

A model that predicts well captures something about the computational capabilities of a user.

Necessary but not sufficient.

Begin by modeling a single user at a time

Building up to full networks



Observe: $X_{i-L}^{i-1} = (X_{i-L}, \ldots, X_{i-2}, X_{i-1})$

Predict: $\hat{X}_i = \underset{x \in \{0,1\}}{\arg\max}\ r(x; X_{i-L}^{i-1})$

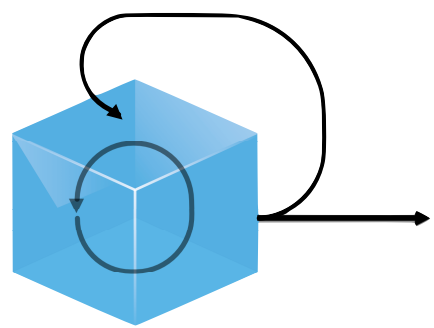
Predictability of User Behavior in Social Media

Modeling Framework: A Predictive View

Goal: Learn r

i.e. find function to map **past** histories to **future** behavior

i.e. autoregression



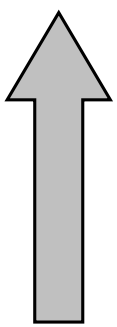
Observe: $X_{i-L}^{i-1} = (X_{i-L}, \ldots, X_{i-2}, X_{i-1})$

Predict: $\hat{X}_i = \underset{x \in \{0,1\}}{\arg\max}\ r(x; X_{i-L}^{i-1})$

Predictability of User Behavior in Social Media

Modeling Framework: A Predictive View

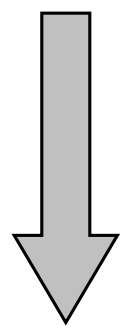
Two approaches to learning r



Computational Mechanics

Causal State Machines

"Bottom-Up"

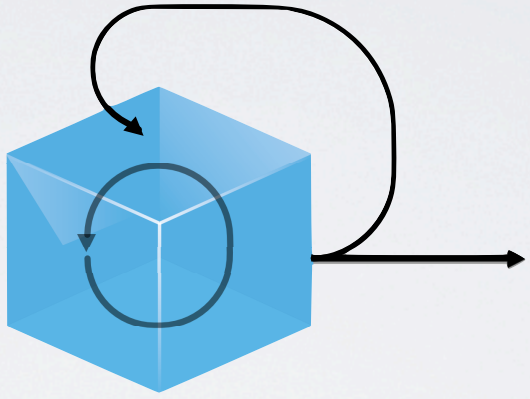


Neural Networks

Echo State Nets

"Top-Down"

Models



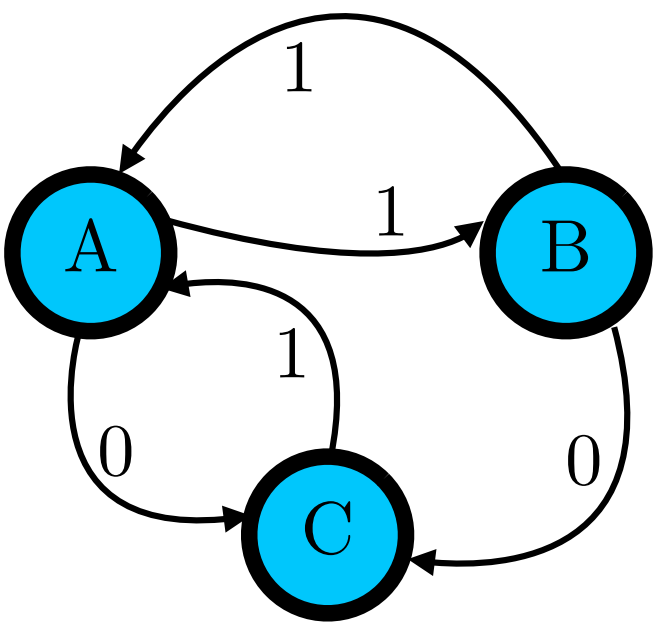
Computational Mechanics

Predictability of User Behavior in Social Media

Computational Mechanics

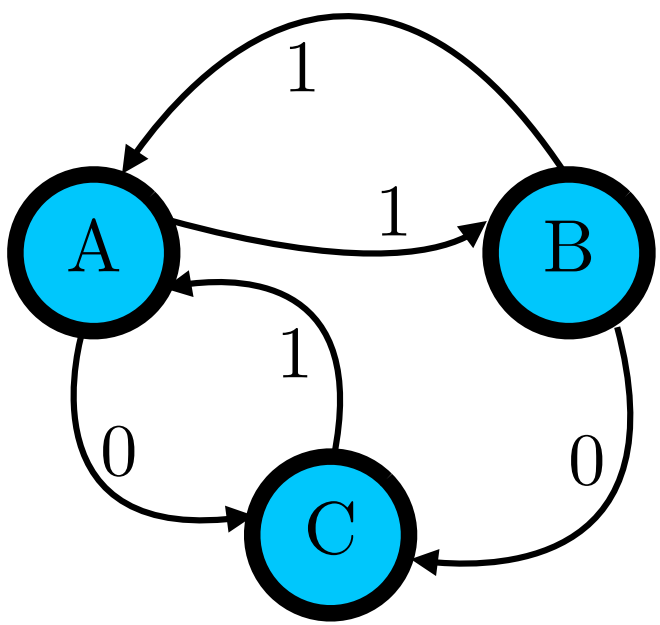
Assume $\{X_i\}_{i=1}^{N}$ was generated by a *conditionally stationary* stochastic process.

Explicitly learn the predictive distribution $P(X_i|X_{i-L}^{i-1} = x)$ by ***grouping together pasts*** x ***that give equivalent predictions***.



Predictability of User Behavior in Social Media

Computational Mechanics



Causal State Model (CSM)

built for each user using

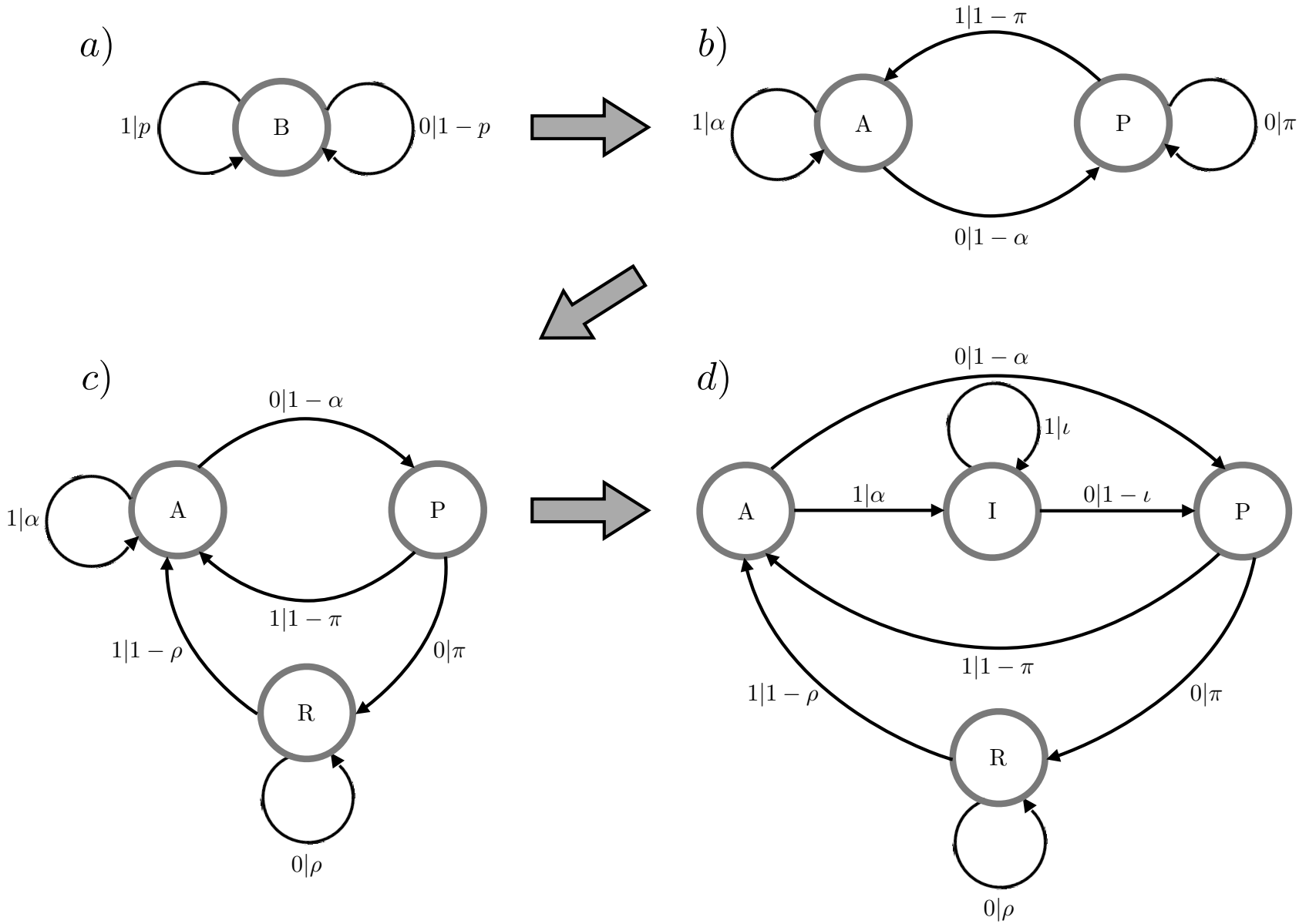
Causal State Splitting Reconstruction (CSSR).

Begin w/ only 1 state;
divide as necessary.

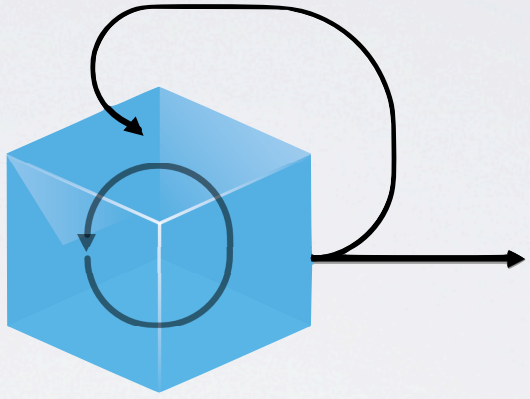
Like Markov model, only need to know which hidden state, not entire history to make predictions.

Predictability of User Behavior in Social Media

Bottom-Up Complexity



Models



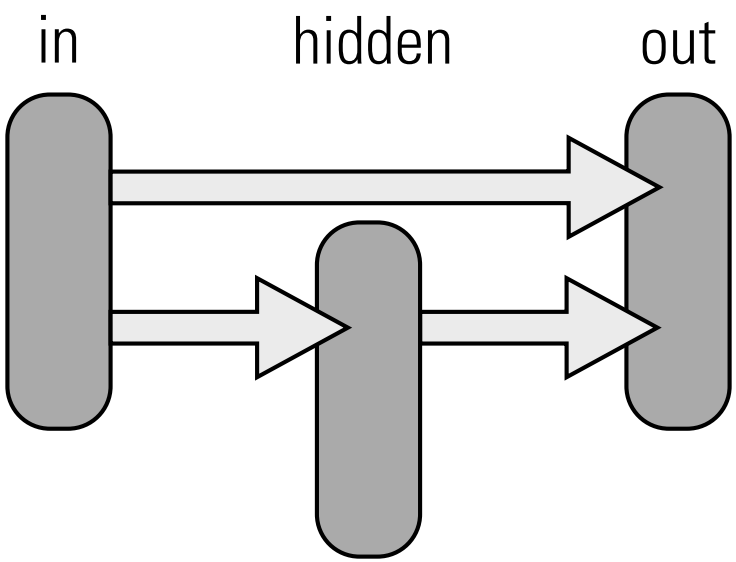
Echo State Networks

Predictability of User Behavior in Social Media

Echo State Networks

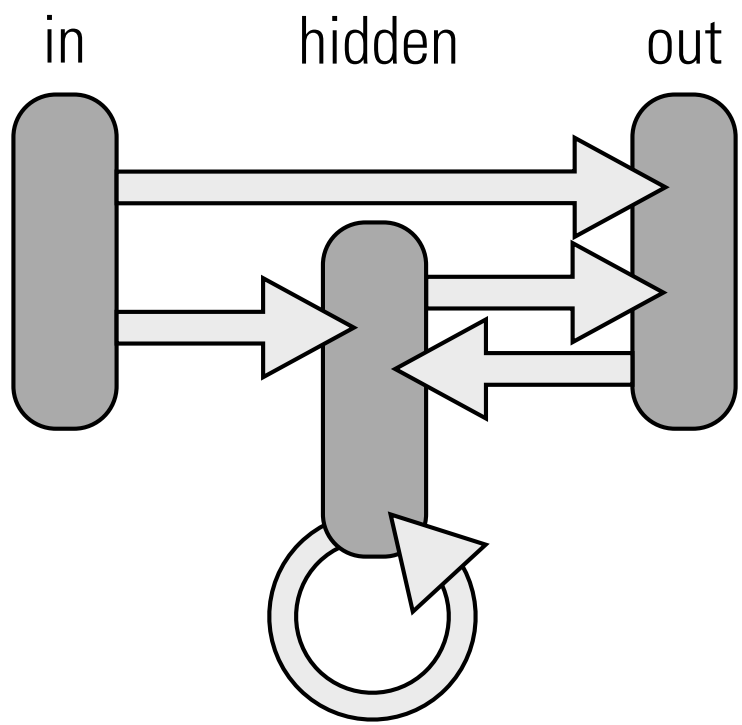
Feedforward Nets

Easy learning rules



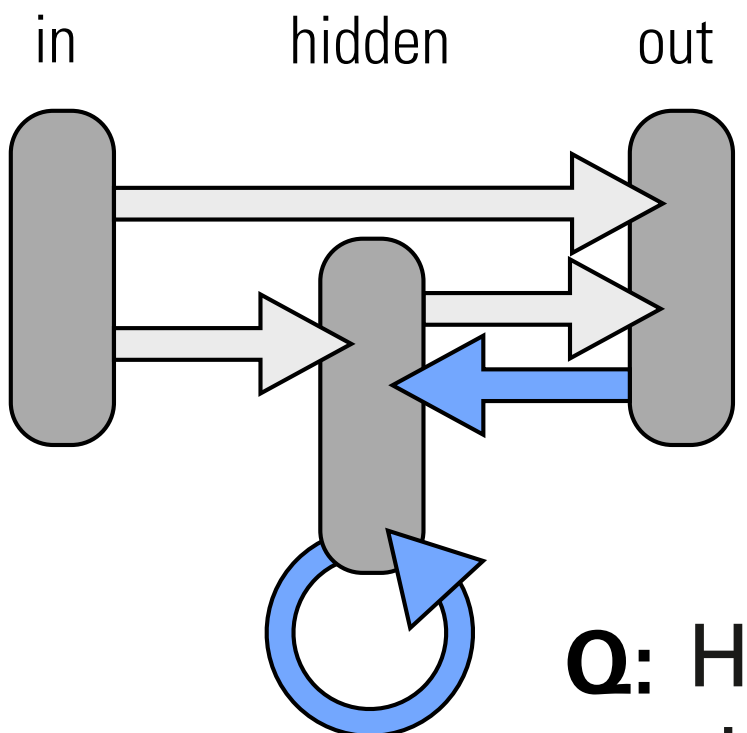
Recurrent Nets

Good with sequences



Predictability of User Behavior in Social Media

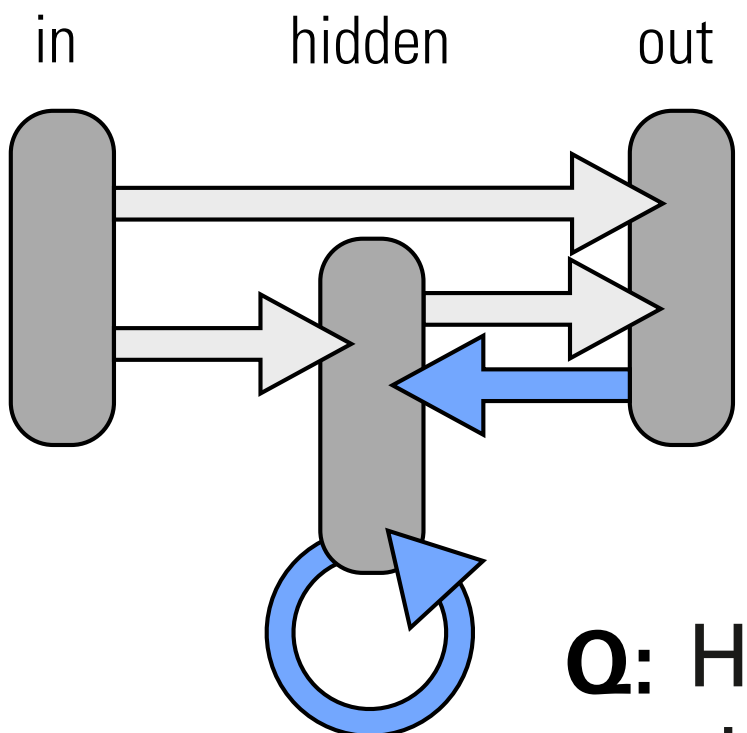
Echo State Networks



Q: How do we learn these weights?

Predictability of User Behavior in Social Media

Echo State Networks

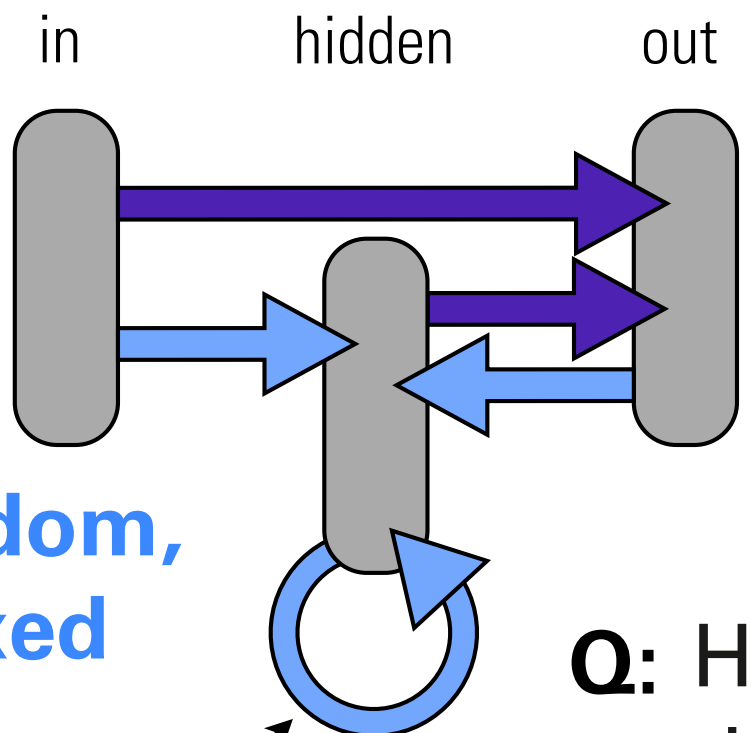


Q: How do we learn these weights?

A: ***We don't.***

Predictability of User Behavior in Social Media

Echo State Networks



Trained

Random, Fixed

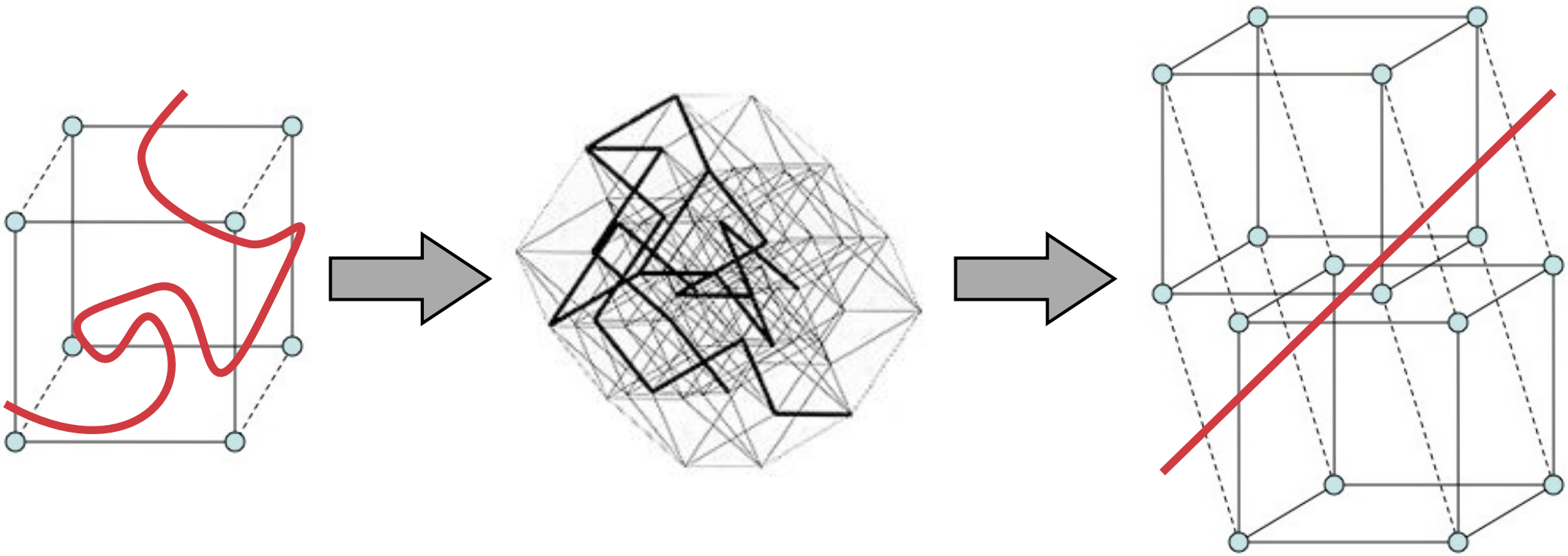
(sparse)

Q: How do we learn these weights?

A: *We don't.*

Predictability of User Behavior in Social Media

Top-Down Complexity

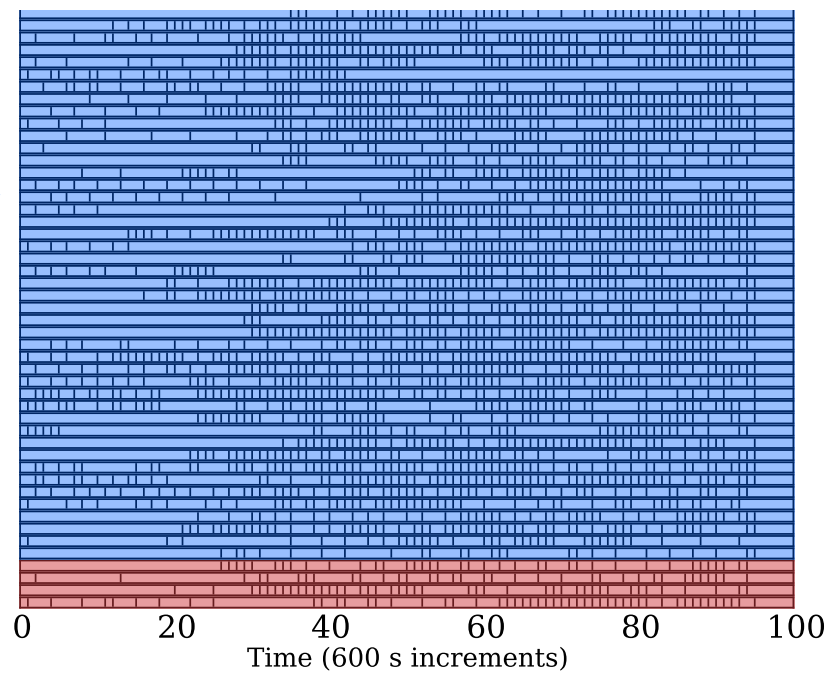


Results

Predictability of User Behavior in Social Media

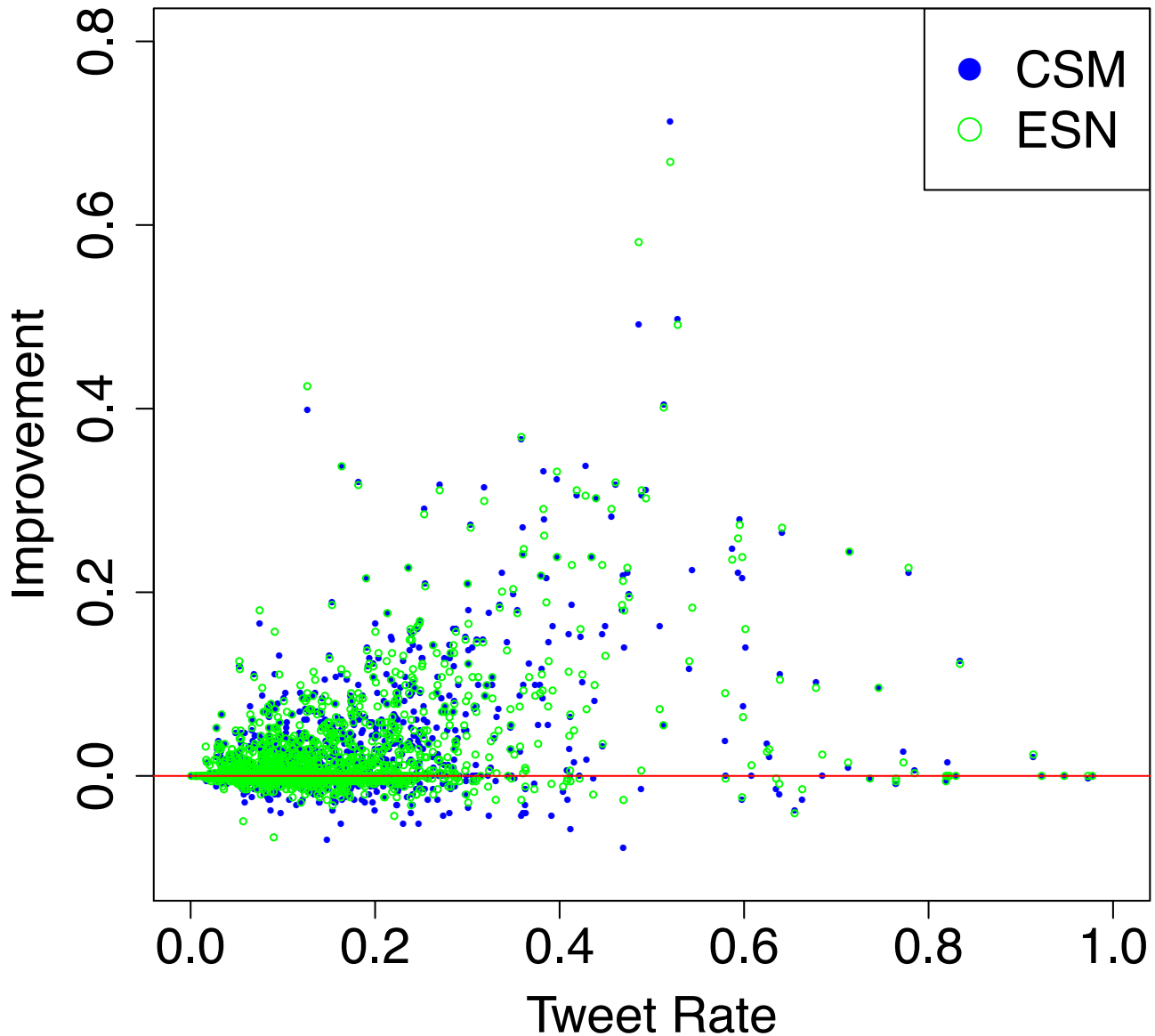
Testing Procedure

- Build model for each user separately
- Training: 45 days
- Testing: 4 days
- Look back 10 steps
- Predict ahead 1 step
- 0-1 Loss
- Compare to “majority vote” baseline



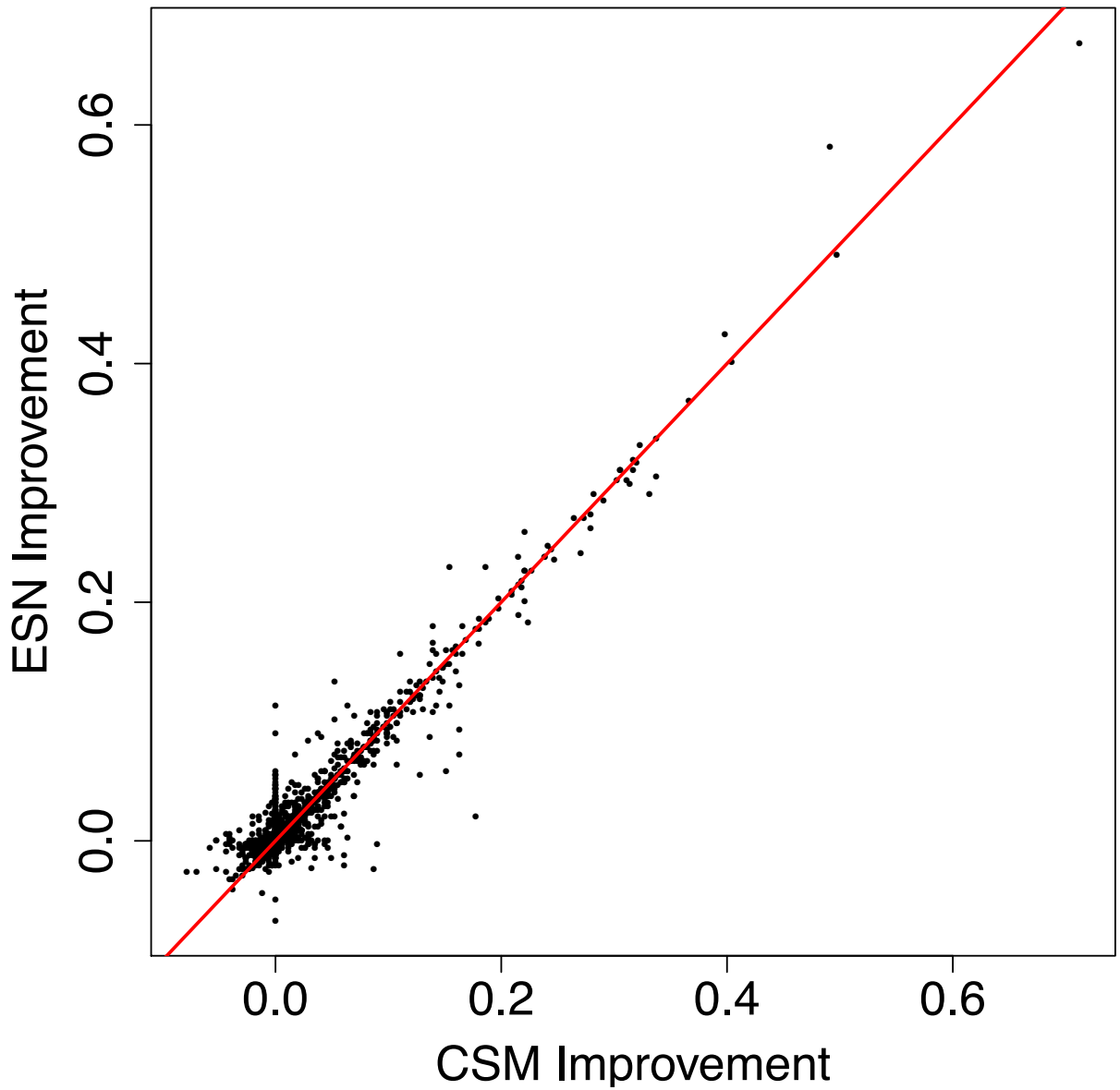
Predictability of User Behavior in Social Media

Results: Accuracy Improvement for CSM & ESN



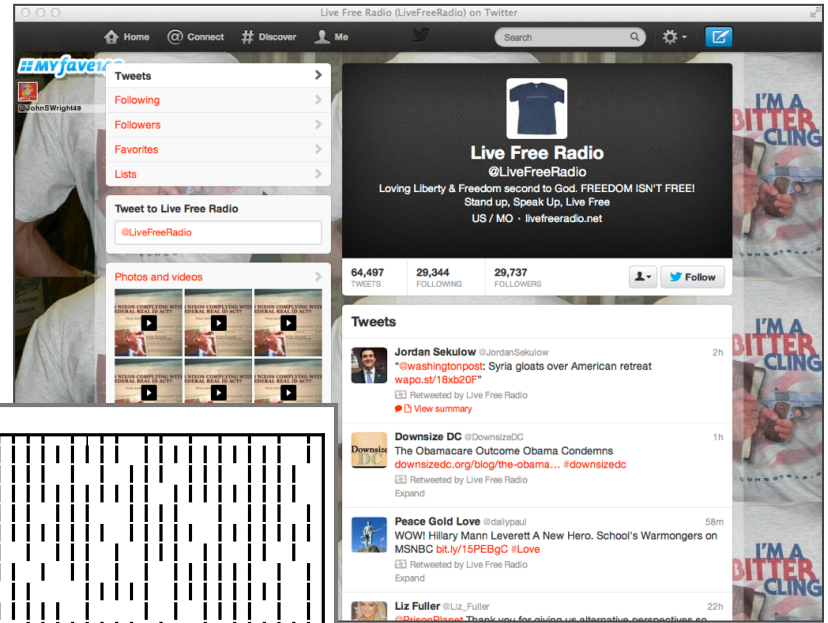
Predictability of User Behavior in Social Media

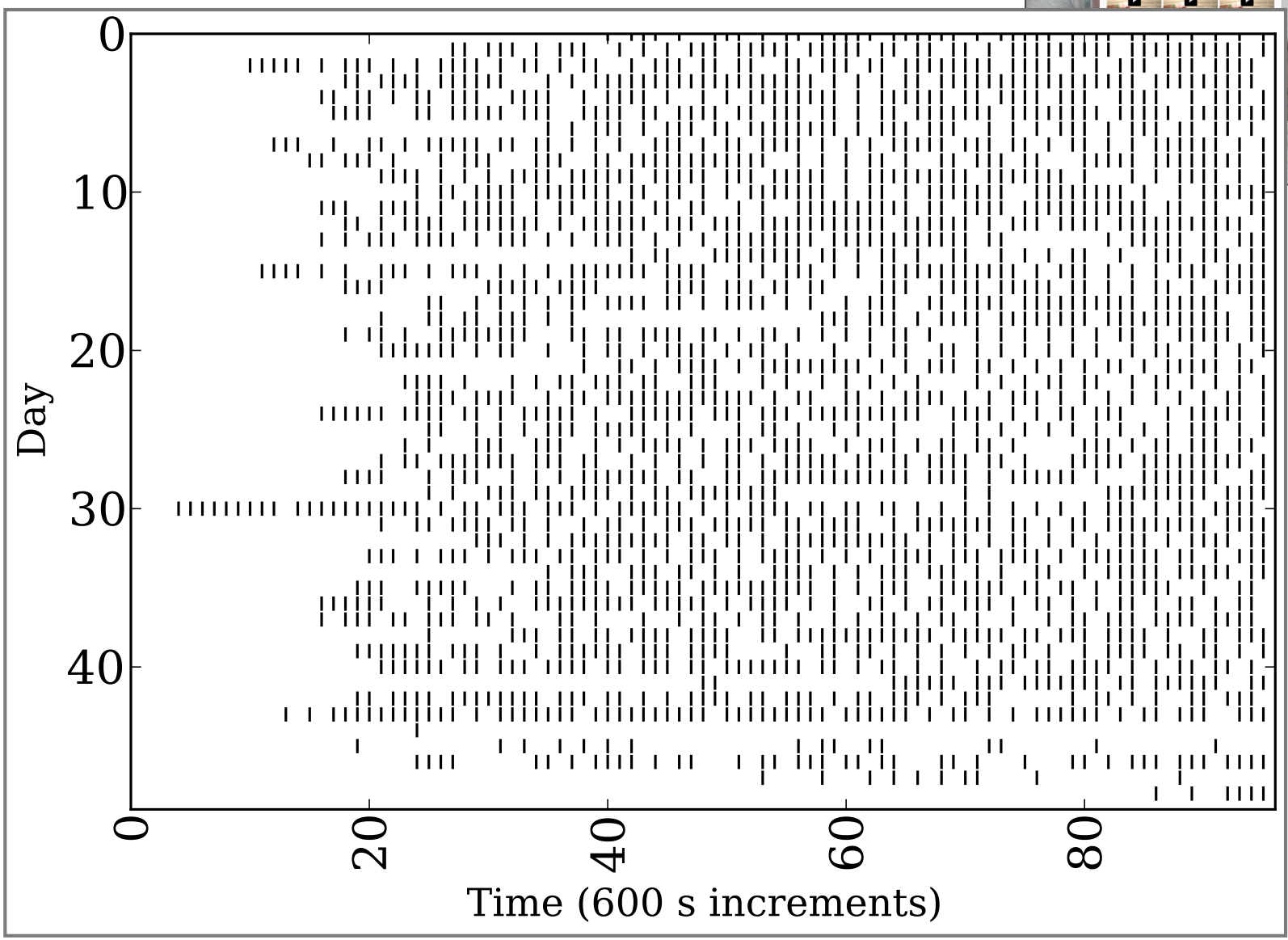
Results: Improvement of CSM vs. ESN



@LiveFreeRadio

Base Rate:	0.2122
CSM Rate:	0.7035
ESN Rate:	0.7936





Conclusions and Future Directions

Predictability of User Behavior in Social Media

Conclusions

Many users on Twitter are well-modeled as processes with self-feedback.

Didn't need social information.

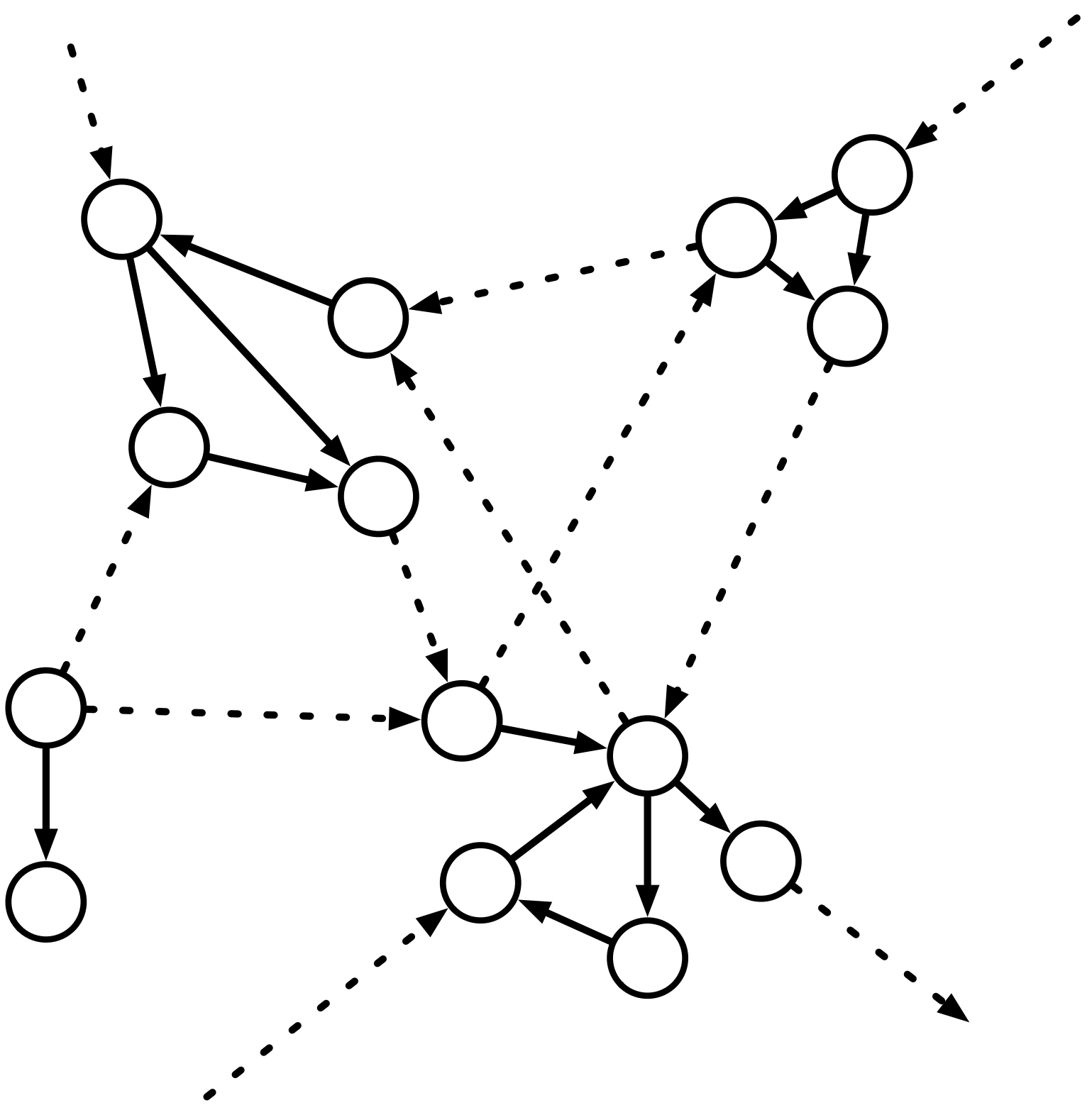
CSMs and ESNs performed similarly on a large proportion of users.

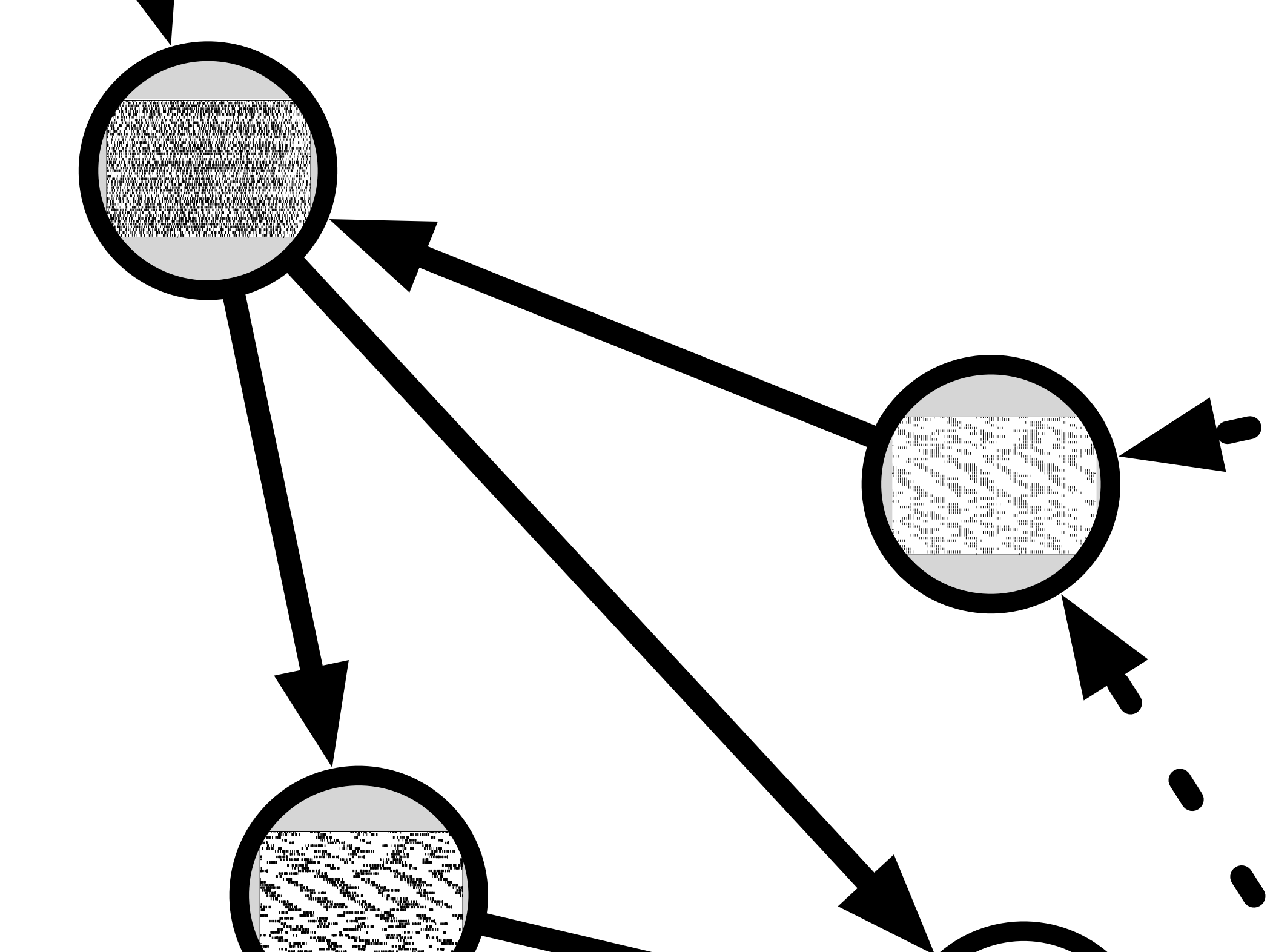
Despite very dissimilar modeling paradigms.

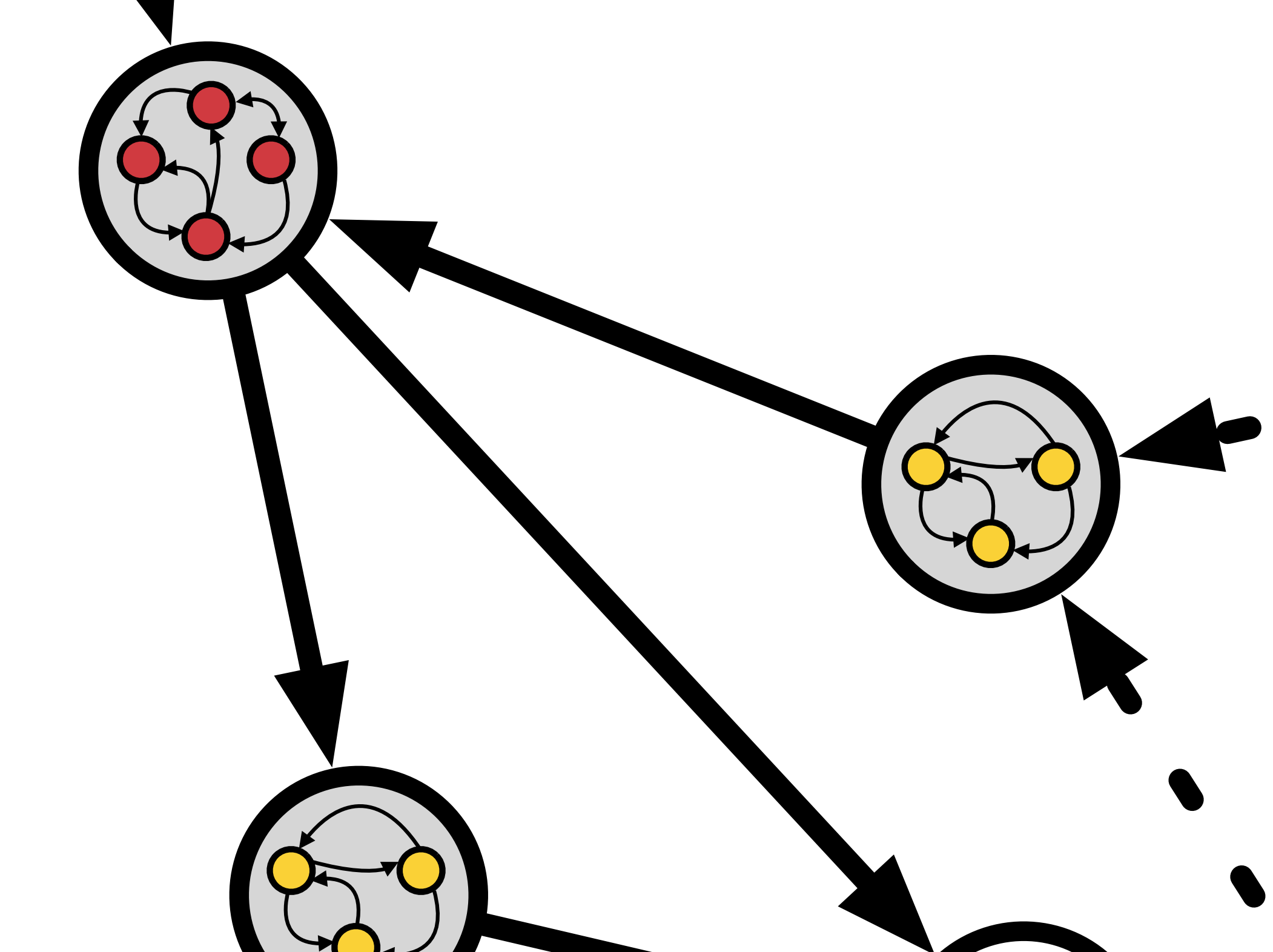
Predictability of User Behavior in Social Media

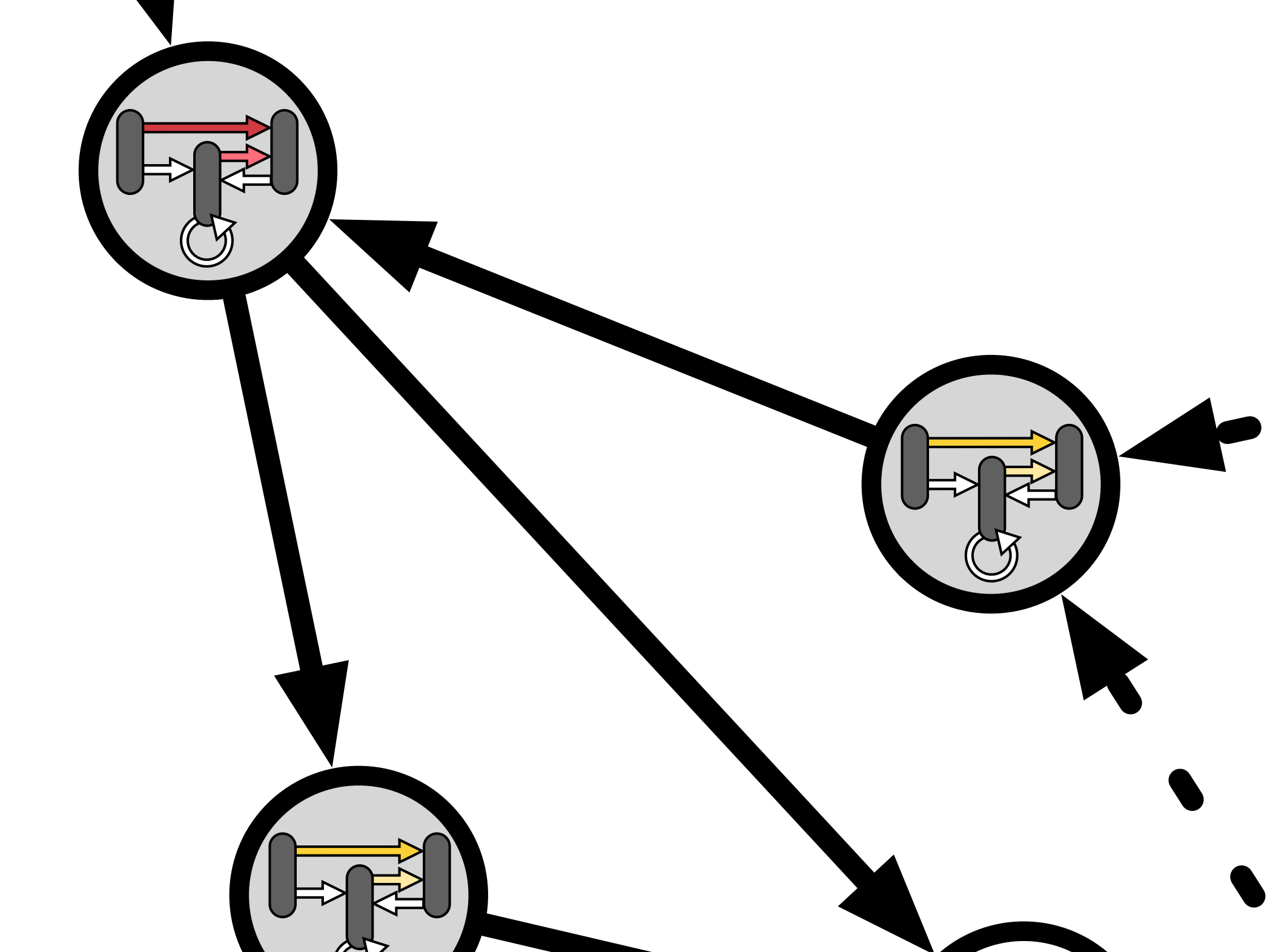
Future Work

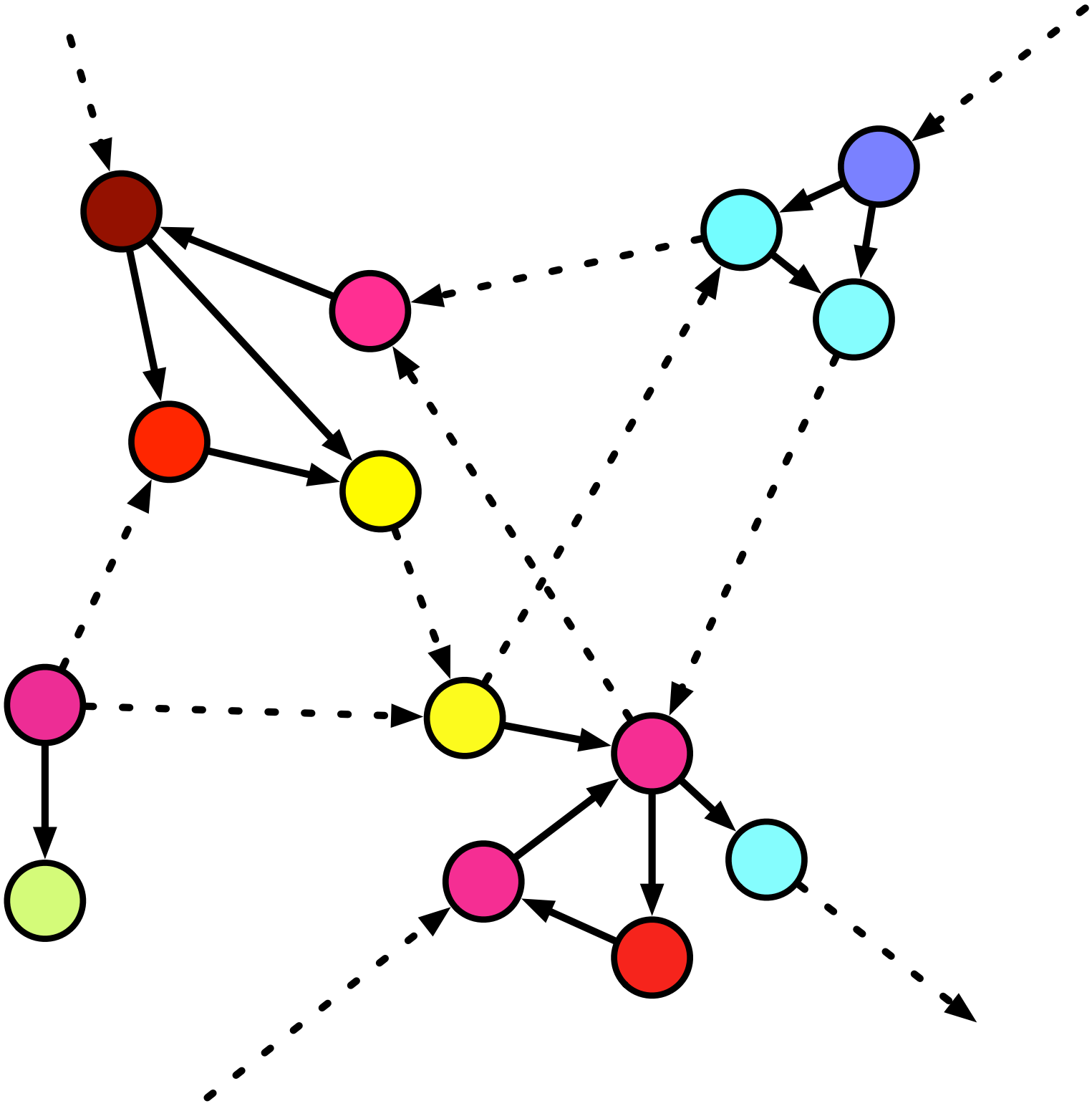
Content from tweets	Sentiment, hashtags, links, etc.
Longitudinal studies	Do users change over time?
Network effects	Explicitly model social dynamics
Multiple types of users	Group users into equivalence classes











Thanks!

Questions?

For details, please see:

Darmon, Sylvester, Girvan & Rand (2013). "Understanding the Predictive Power of Computational Mechanics and Echo State Networks in Social Media." *ASE Human Journal*, vol. 2(2), pp.13–25.

Darmon, Sylvester, Girvan & Rand (2013). "Predictability of User Behavior in Social Media: Bottom-Up v. Top-Down Modeling." ASE/IEEE Int'l Conf. on Social Computing, p. 102–107.

`www.cs.umd.edu/~jared/`