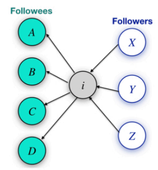
TOWARDS DECENTRALIZED SOCIAL REINFORCEMENT LEARNING VIA EGO-NETWORK EXTRAPOLATION

Mahak Goindani, Jennifer Neville, Department of Computer Science, Purdue University

Incentive Driven Policy Learning in Networks

- Directed relations between users, e.g., Followee-Follower
  - One-directional information flow — Follower can observe Followees
  - Information does not flow in opposite direction, unless Followee also follows the Follower
- Partially Observable Ego-Network
  - User i
    - Followees: A, B, C, D
      - Observed
    - Followers: X, Y, Z
      - Unobserved



- Individual Rewards. Eg. *visibility* among Followers
  - Number of followers exposed
  - Rank of user's posts in Followers' feeds
  - Amount of time for which the user's posts stay at top
- Depends on user's activities as well as activities of related users in local neighborhood
- Different *local reward* for each user based on her peers and local network structure

Social RL Objective

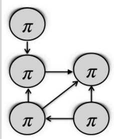
- *Local Observation* of user i $\mathbf{o}_i$
  - Activities of Followees of user i
- *Local Reward* of user i $R_i \in \mathbb{R}$
  - Correlation between exposures to Fake and True news among Followers of user i
  - Penalty or cost for a user to post more
- Objective of user i: Learn Policy $\pi_i : \mathbf{s}_{t,i} \rightarrow \mathbf{a}_{t,i}$ such that her total expected discounted local reward $\sum_{t=1}^{T} \gamma^{t-1} \mathbb{E}[R_{t,i}]$ is maximized

Challenges for Partially Observed Networks

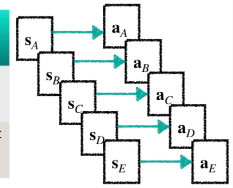
- Individual policies need to account for dependencies throughout the network
- Centralized learning and execution — improve sample efficiency per user
  - Different local reward, observation of each user — infeasible
- Decentralized learning
  - Does not scale for large N
  - Insufficient samples per user—sparse interaction data—large errors due to variance
- Directed nature of user interactions
  - Strong Partial Observability
  - Relevant state information cannot be utilized as history by the user
  - Storing complete network trajectory information for large N — space-prohibitive

Decentralized Ego-Network Policy Learning

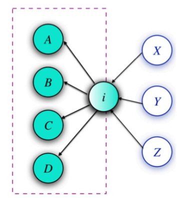
- *Main Approach:* Partially Centralized Learning and Decentralized Execution
  - Single policy function
  - Parameter Sharing to learn this function across users
    - Only share model parameters sequentially — Overcome sparsity
    - No sharing of samples/trajectories — privacy-aware (limit data sharing)



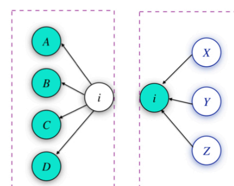
	Centralized	Decentralized
Learning	**Same** policy function for **all** agents	
Execution		Actions for an agent conditioned on her **individual state**



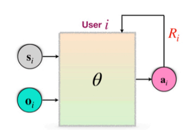
- A user i has two roles
  - Follower
  - Followee
- Learn dependency between Followees and Followers
- *Key Idea:* Ego-network extrapolation:
  - Learn a function to estimate user i's (Follower) state from her Followees' states
  - Use the learned function to extrapolate the state of user i's Followers from user i's (Followee) state

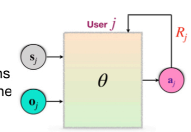


- Challenge:
  - User i's Followees' states → User i's state
    - Many-to-one
  - User i's state → User i's Followers' states
    - One-to-many
  - Less accurate estimates
- Insight:
  - Reciprocity — Retweets, Likes
    - Many-to-many mapping — better estimates
  - Dynamic peer-influence — Attention between activities of each Followee-Follower pair

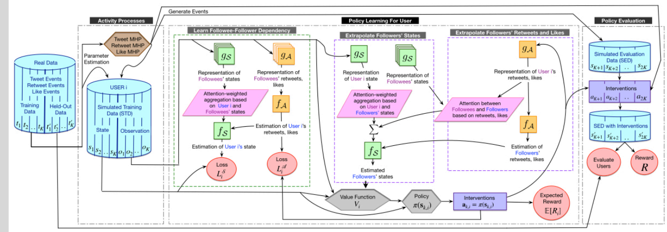


- Sequential Parameter Sharing
  - Common neural network, and agents access the network in a sequence
  - At a given iteration, only a single agent learns and updates the shared parameters based only on her state, observation and reward
- DENPL: 6 NN, 3 MHPs — First MARL approach to utilize user relations in a partially observable social network — transfer knowledge from one set of users to another — estimate hidden state



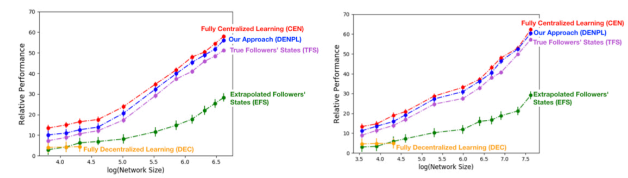


DENPL Framework and Evaluation



Real-world Twitter Datasets - Tweet, Retweet and Like Events. Multivariate Hawkes Proces to characterize user activities.
Policy Learning via Ego-network Extrapolation and transferring the knowledge from the Followees to Followers
Evaluation:
Performance: Reward along with *fraction of Followers exposed to fake news that become exposed to true news*
Relative Performance: Difference between performance after applying the learned policy and that without applying a policy

Experiments and Results



Twitter 2016 Twitter 2015
Relative Performance on Twitter Datasets

- DENPL achieves similar performance as Fully Centralized learning, along with overcoming the limitations of Fully Centralized Learning
- Sequential Parameter Sharing
  - Increased effective number of samples per user — Overcome sparsity
  - No samples, only parameters shared — privacy-aware (limit data sharing)
- Ego-Network Extrapolation
  - Effectively extrapolate dependencies learned from Followees to Followers
  - Pairwise user interactions, peer-influence as attention
  - Learn policies equivalent to centralized learning — without sharing trajectory information — for partially observable environments

Additional Publications

- Goindani, M., & Neville, J. Social Reinforcement Learning to Combat Fake News Spread. UAI 2019
- Goindani, M., & Neville, J. Cluster-based Social Reinforcement Learning. AAMAS 2020
- Goindani, M. Social Reinforcement Learning. Ph.D. Thesis. December 2020

mgoindani@apple.com, neville@purdue.edu