RUTGERS
THE STATE UNIVERSITY
OF NEW JERSEY

ICML
International Conference
On Machine Learning

Optimally Improving Cooperative Learning in a Social Setting

Shahrzad Haddadan, Cheng Xin, Jie Gao

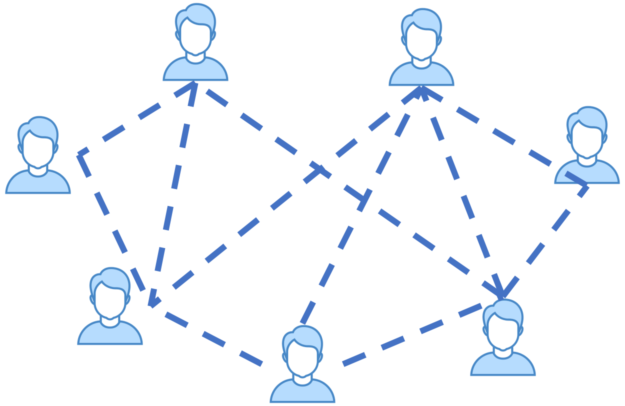
The 41st International Conference on Machine Learning (ICML 2024)

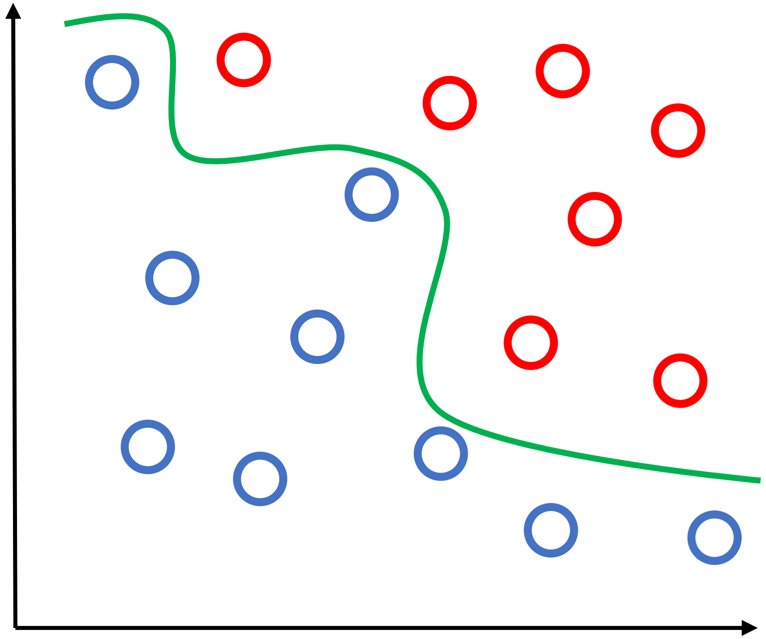
Motivation

- Consider a set of networked agents who solve a common classification problem by learning separate models

Examples:

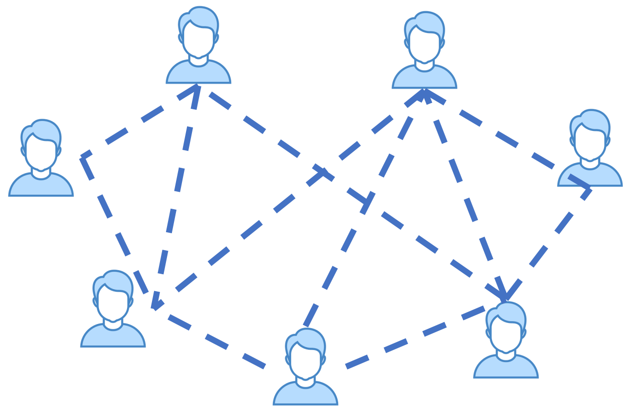
- whether a content in OSN is AI-generated
- whether a stock value is over-priced
- whether a security system is under attack

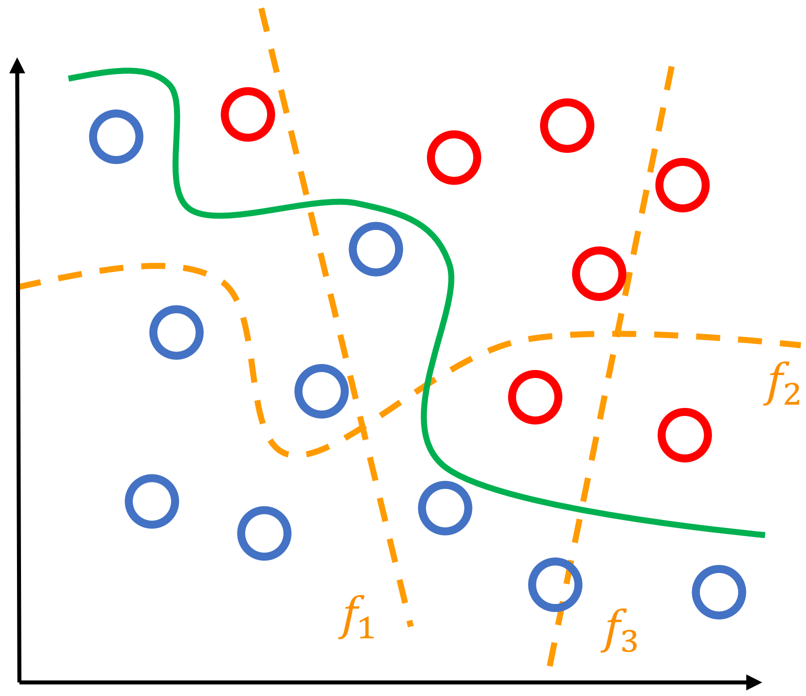




Modeling

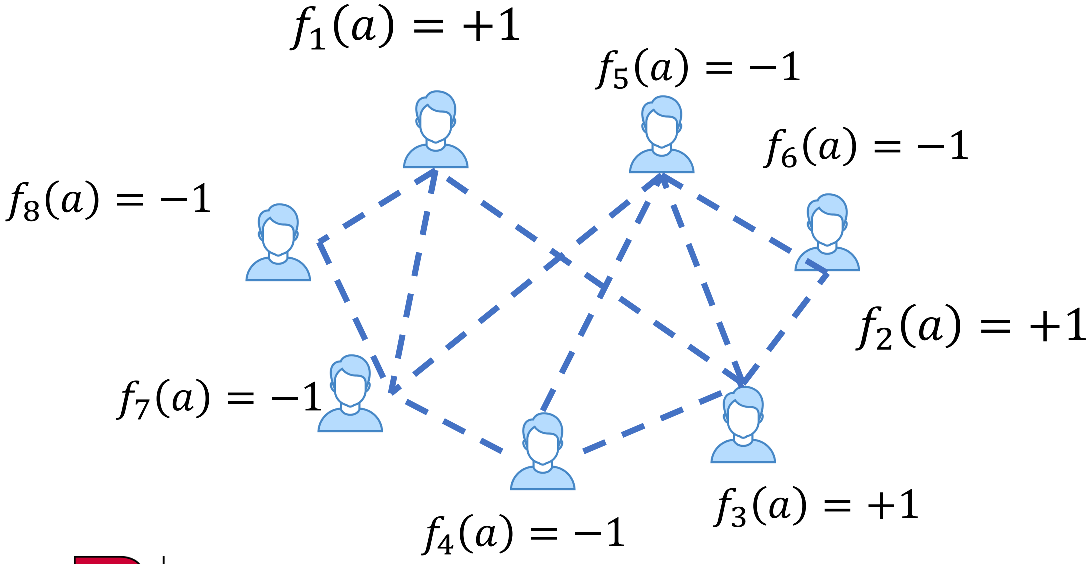
- Assume $f: \Omega \to \{-1, +1\}$ and a set of networked agents V
- Each agent $v_i \in V$ owns a classifier $f_i: \Omega \to \{-1, +1\}$
- Nature samples $a \in \Omega$

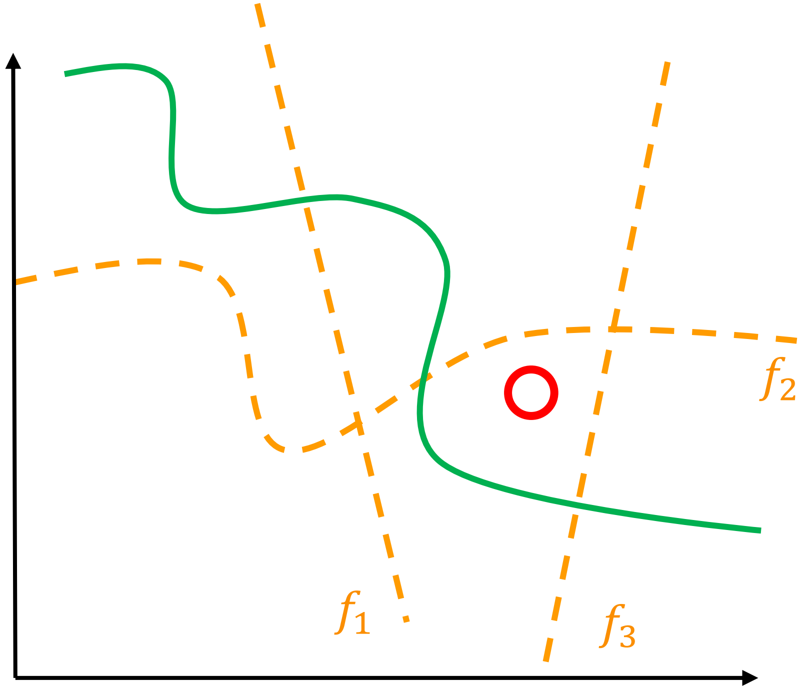




Modeling

- Assume $f : \Omega \to \{-1, +1\}$ and a set of networked agents V
- Each agent $v_i \in V$ owns a classifier $f_i : \Omega \to \{-1, +1\}$
- Nature samples $a \in \Omega$

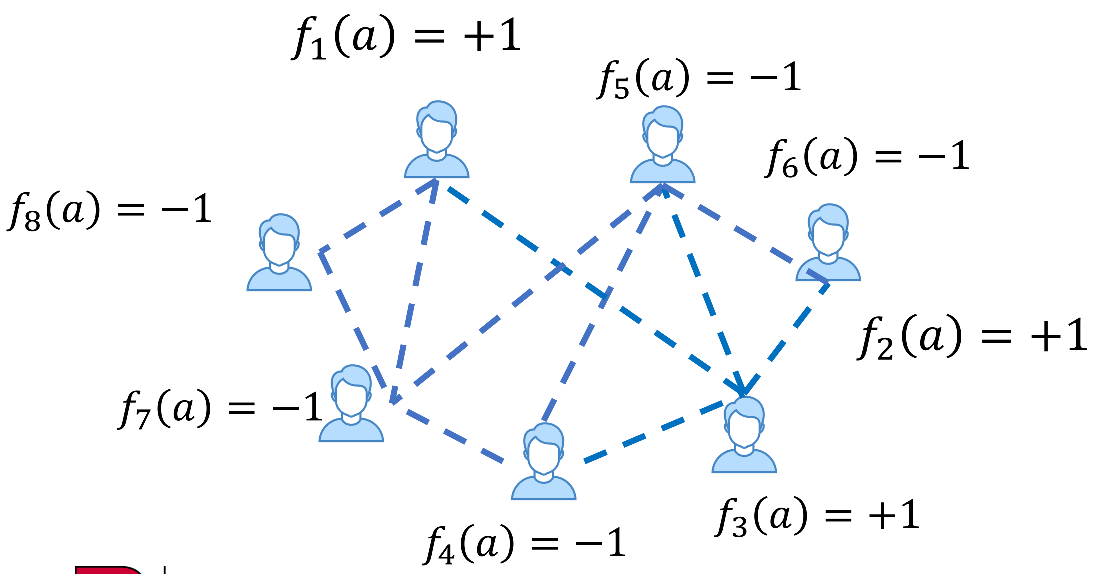




RUTGERS

Modeling

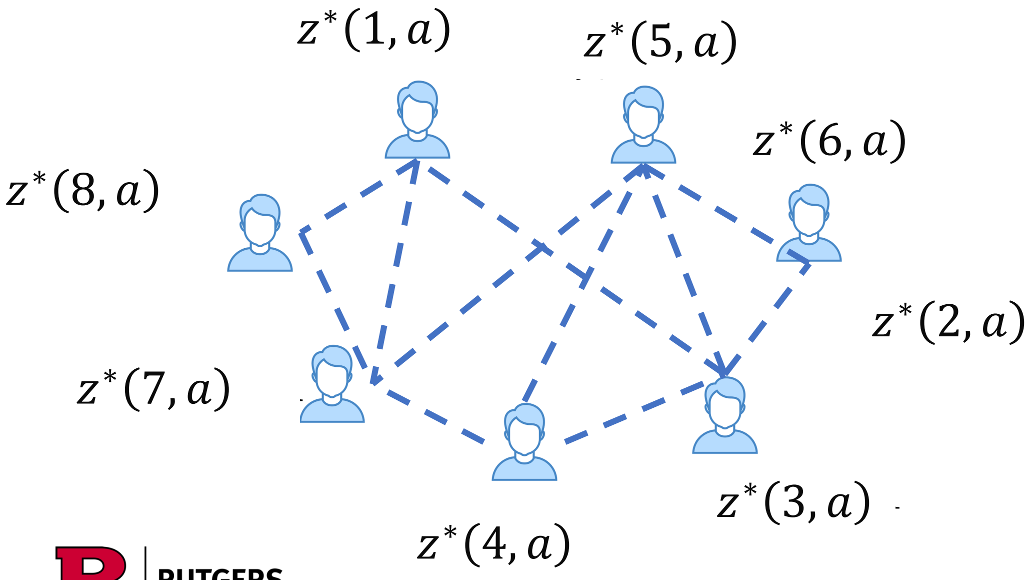
- Assume $f: \Omega \to \{-1, +1\}$ and a set of networked agents V
- Each agent $v_i \in V$ owns a classifier $f_i: \Omega \to \{-1, +1\}$
- Nature samples $a \in \Omega$
- Through the network agents exchange predictions $f_i(a)$ and will update it to:



$$z^*(i, a) = \sum_{v_j \in V} W_{ij} f_j(a)$$

Modeling

- Assume $f: \Omega \rightarrow \{-1, +1\}$ and a set of networked agents V
- Each agent $v_i \in V$ owns a classifier $f_i: \Omega \rightarrow \{-1, +1\}$
- Nature samples $a \in \Omega$
- Through the network agents exchange predictions $f_i(a)$ and will update it to:



$$z^*(i, a) = \sum_{v_j \in V} W_{ij} f_j(a)$$

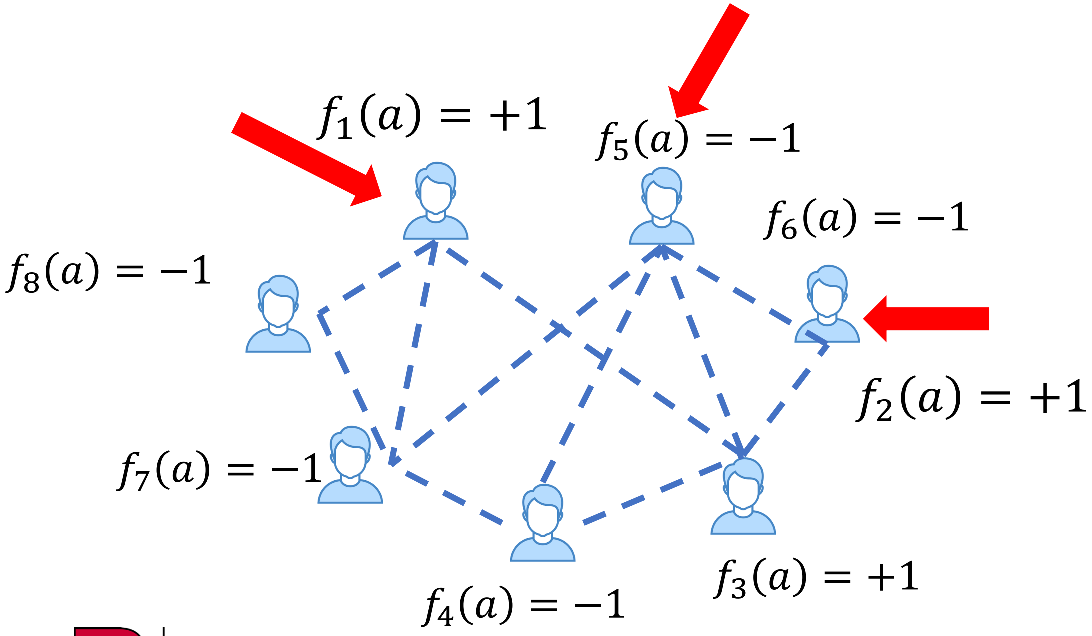
Accuracy measure:

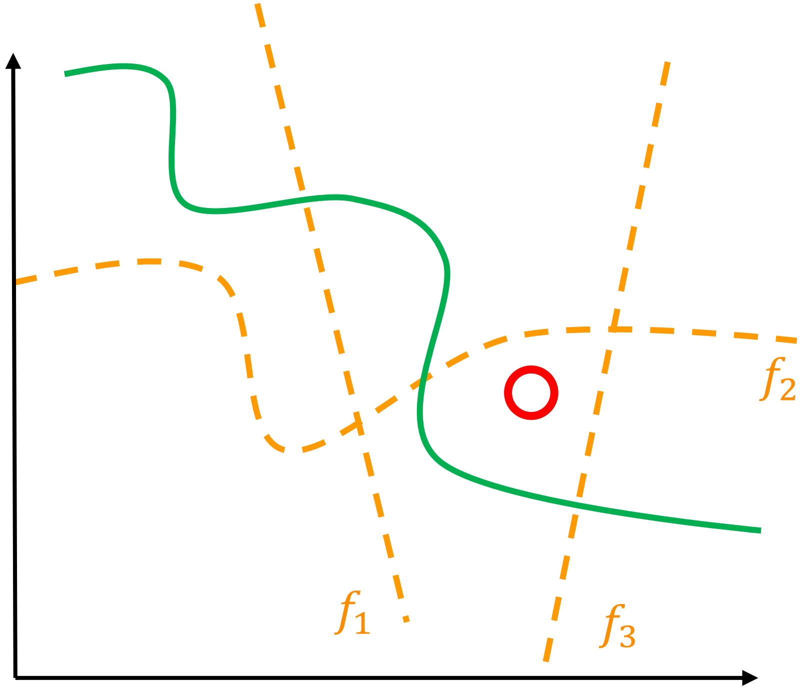
$$Z(i, a) \doteq f(a) \cdot z^*(i, a) \in [-1,1]$$

Algorithms

- A social planner who knows $f(a)$, selects $S \subseteq V$ and improves their predictions as:

$$\forall v_i \in S, \qquad \boldsymbol{f_i(a) = (1-\phi)f_i(a) + \phi f(a)}$$





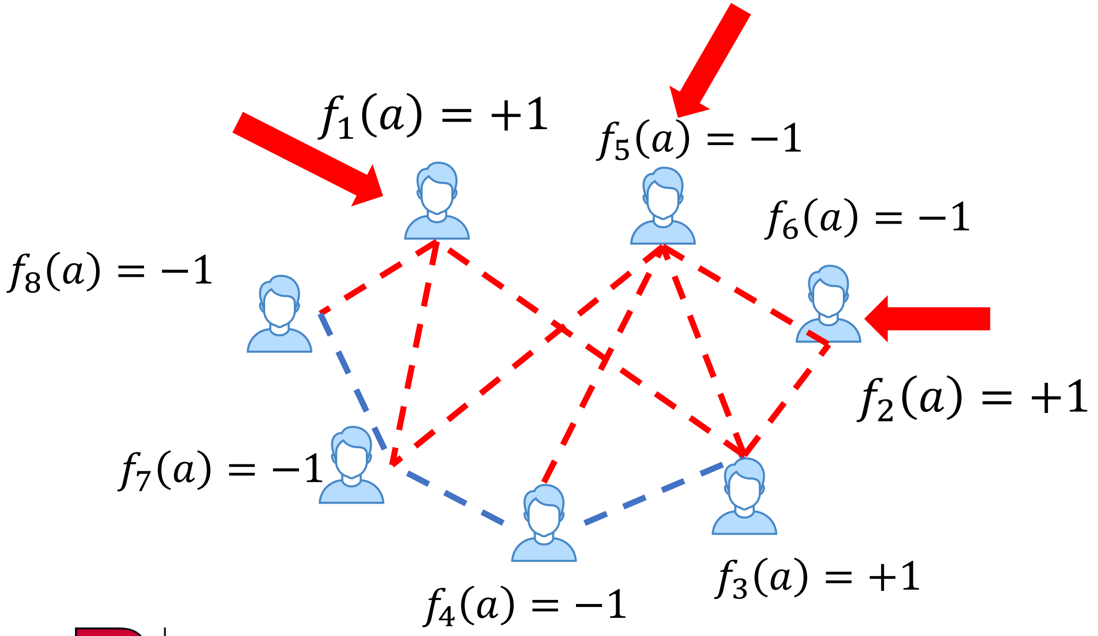
RUTGERS

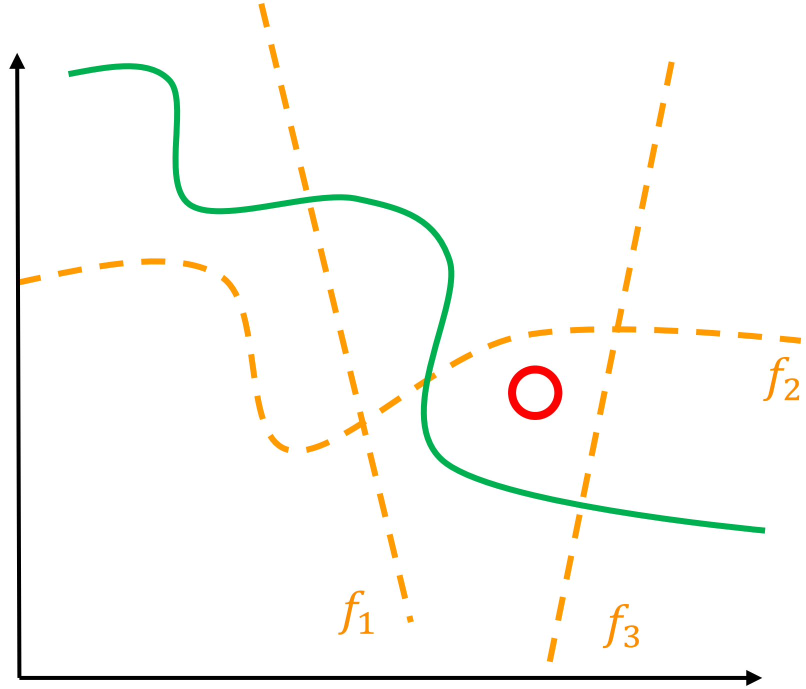
Algorithms

- A social planner who knows $f(a)$, selects $S \subseteq V$ and improves their predictions as:

$$\forall v_i \in S, \qquad f_i(a) = (1-\phi)f_i(a) + \phi f(a)$$

$$\forall\, v_i \in V \qquad Z(i,a) \Rightarrow Z_{new}(i,a)$$





RUTGERS

Summary of results: hardness

- **Aggregate improvement** $G^{(\mathrm{agg})}(S) \triangleq \mathbb{E}_{a\sim\Omega}\left[\sum_{i=1}^{n} Z_{\mathrm{new}}(i,a) - Z(i,a)\right]$

  Optimizing Aggregate improvement in EASY

- **Egalitarian improvement**

$$G^{(\mathrm{egal})}(S) \triangleq \mathbb{E}_{a\sim\Omega}\left[\sum_{i=1}^{n} \mathbf{1}\left(Z(i,a) < 0 \wedge Z(i,a) < Z_{\mathrm{new}}(i,a)\right)\right]$$

  Optimizing Egalitarian improvement in HARD

Summary of results: approximation algorithms for egalitarian improvement

- **EgalAlg**:
  - Assumption: full access to the joint probability distribution of classifiers.
  - Runtime: $\Theta(|\Omega|n^2k)$ Approximation ratio: $(1-1/e)$
- **EgalAlg(appx):**
  - Assumption: access to pairwise independence of agents' prediction & error rates
  - Runtime: $\Theta(n^3k)$ Approximation ratio: $(1-1/e)-\Delta_{\text{ind}}$

Approximately improve egalitarian

Our greedy algorithms iteratively optimizes some marginal gain gr(S):

$$\mathsf{gr}(S) \triangleq \underset{u \in V}{\operatorname{argmax}}\ \mathcal{G}^{(\mathrm{egal})}(S \cup \{u\}) - \mathcal{G}^{(\mathrm{egal})}(S)\ ,$$
$$= \underset{u \in V}{\operatorname{argmax}} \sum_{\substack{i=1:n \\ \bar{W}_{iu} \neq 0}} \Delta\mathcal{G}_i(S, u).$$

- EgalAlg:

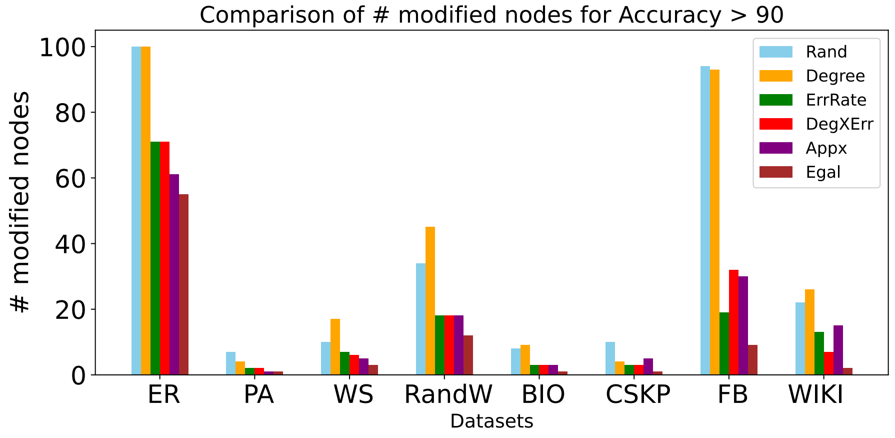
$$\Delta\mathcal{G}_i := \underset{a \sim \Omega}{\mathbb{P}} \left(\mathcal{Z}(i, a) \leq 0 \wedge \left(\bigwedge_{\substack{v_j \in S \\ \bar{W}_{ji} \neq 0}} y(a) = \hat{y}_j(a) \right) \wedge y(a) \neq \hat{y}_u(a) \right)$$

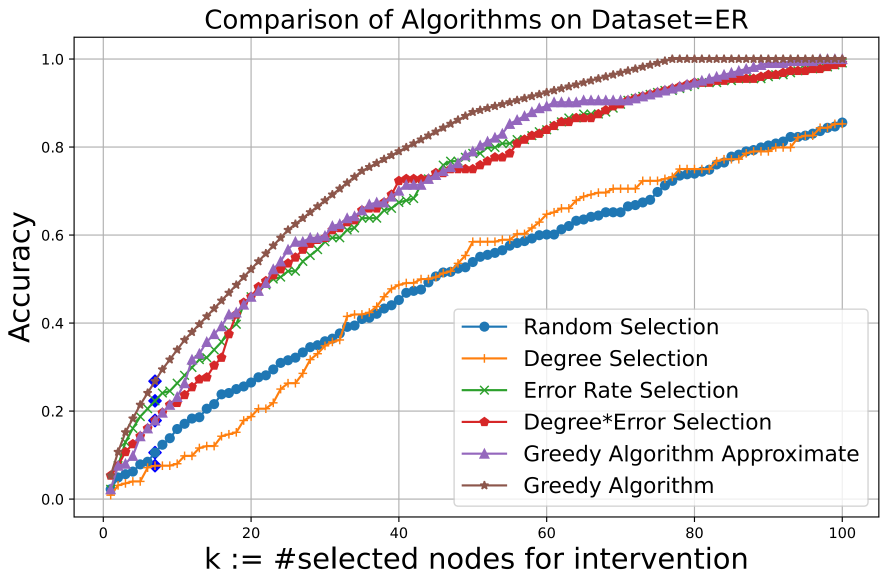
- EgalAlg(appx):

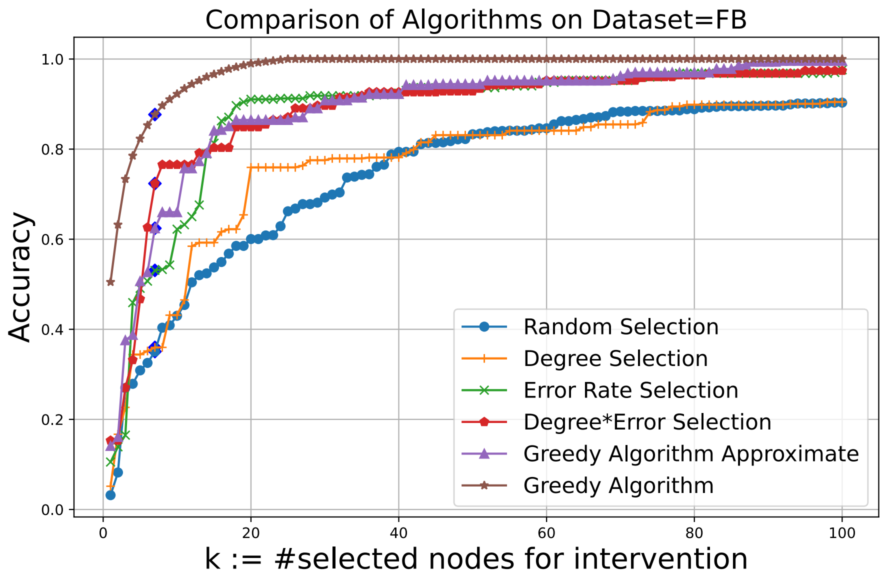
$$\widehat{\Delta\mathcal{G}}_i(S, u) := \mathbf{1}\left(\Psi_i(S, u) < 0\right) \mathsf{err}(u) \prod_{\substack{v_j \in S \\ \bar{W}_{ij} \neq 0}} (1 - \mathsf{err}(v_j))$$

Experiments

- Compare to baselines with heuristic or random marginals:
  $\Delta\mathcal{G}_i \leftarrow$ node degree / error rate / random
- Results of algorithms can be categorized into four tiers:
  Tier 1 (EgalAlg) >> Tier 2 (EgalAlg(appx)) ≥ Tier 3 (heuristics
  >> Tier 4 (random)
- High accuracy achieved with only log(n) modified nodes







RUTGERS

Experiments

		Datasets							
Score	**Method**	ER (128)	PA (128)	WS (128)	RandW (128)	BIO (297)	CSPK (39)	FB (620)	WIKI (890)
Acc@ k=log(n)	Rand	0.11	0.88	0.53	0.18	0.80	0.63	0.35	0.48
	Degree	0.08	0.96	0.42	0.12	0.78	0.84	0.36	0.49
	ErrRate	0.22	**1.00**	0.76	0.47	0.96	0.94	0.53	0.54
	DegXErr	0.18	**1.00**	0.89	0.37	0.96	**1.00**	0.72	0.78
	Appx	0.18	**1.00**	0.87	0.41	0.94	0.84	0.62	0.64
	Egal	**0.27**	**1.00**	**1.00**	**0.58**	**1.00**	**1.00**	**0.88**	**0.96**
#k @ Acc>90%	Rand	>100	7	10	34	8	10	94	22
	Degree	>100	4	17	45	9	4	93	26
	ErrRate	71	2	7	18	3	3	19	13
	DegXErr	71	2	6	18	3	3	32	7
	Appx	61	**1**	5	18	3	5	30	15
	Egal	**55**	**1**	**3**	**12**	**1**	**1**	**9**	**2**
#k @ Acc>75%	Rand	83	8	16	61	15	14	37	55
	Degree	83	5	28	64	14	6	20	54
	ErrRate	46	3	10	31	5	4	14	26
	DegXErr	51	3	8	36	4	3	8	16
	Appx	47	**2**	9	35	6	7	11	39
	Egal	**36**	**2**	**4**	**19**	**2**	**2**	**4**	**3**

Conclusion and Future Work

- We introduce a new model in which networked agents help each other to improve the accuracy of their prediction using distinct classifiers and by solely **exchanging predictions.**
- Our theoretical analyses and the experiments on real and synthetic networks show that **both model parameters** play a critical role in the study of this model and development of algorithms.
- In **future work**, we may expand this work in several directions:
  - Considering networks with negative edge weights (signed graphs)
  - Considering different improvement formulations (agent based)
  - Extending binary classification to more general learning algorithms.