NEURAL INFORMATION PROCESSING SYSTEMS

Free-Rider and Conflict Aware Collaboration Formation for Cross-Silo Federated Learning

Mengmeng Chen[1], Xiaohu Wu[1], Xiaoli Tang[2], Tiantian He[3], Yew-Soon Ong[2,3], Qiqi Liu[4], Qicheng Lao[1], Han Yu[2]

[1]Beijing University of Posts and Telecommunications, [2]Nanyang Technological University, [3]Agency for Science, Technology and Research (A∗STAR), [4]Westlake University

Outline

- Background & Motivation
- Problem Description
- Solution
- Experiments

Federated Learning

Federated Learning (FL)

- Federated learning (FL) is a promising paradigm of distributed machine learning as it does not require sharing raw data between FL participants (FL-PTs), thereby upholding the privacy considerations.

General FL Training Process

- Multiple FL-PTs train a shared model locally with their own dataset, and upload their local model updates to a central server (CS), which then aggregates these model updates and distributes the model updates to each FL-PT.
- This iterative interplay between the CS and FL-PTs persists until the global model achieves convergence.

Application Domains in business sector

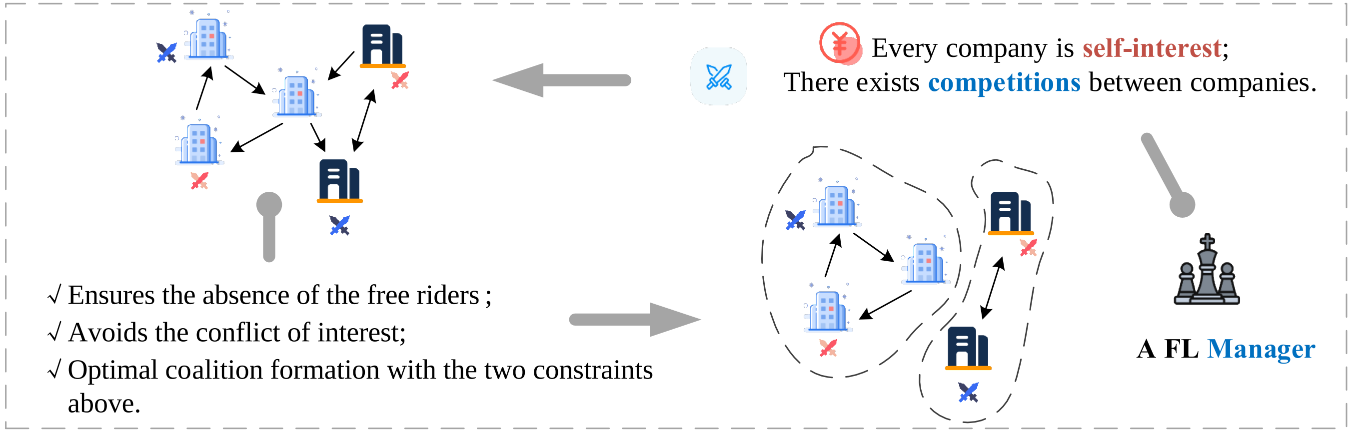
- Digital banking, ridesharing, recommender systems, health care, and Electric Vehicle(EV) charging services, among others

Scenario

NEURAL INFORMATION PROCESSING SYSTEMS

Two features: Self–interest, Competition

- The free-riding problem is in which some FL-PTs benefit from the contribution by others without making any contribution to the FL ecosystem.
- There is a potential conflict of interest between some two FL-PTs.



Motivating Example 1: Banks

- Regional banks have different user groups from their respective regions and are independent, while the banks in the same region can compete for users

Motivating Example 2: Drug Discovery

- An FL platform, **MELLODDY**, has been developed for drug discovery, currently comprised of 10 pharmaceutical companies, academic institutions, large industrial companies and startups, where competition exists when there are multiple organizations that are in the same market area.

Outline

- ❑ Background & Motivation
- ❑ **Problem Description**
- ❑ Solution
- ❑ Experiments

☐ Relationships among FL-PTs V={v_1,v_2,…,v_n}

NEURAL INFORMATION PROCESSING SYSTEMS

Known Parameters:

- Benefit Graph $G_b = (V, E_b)$. If v_i can benefit from v_j's data, then there is a directed edge from v_j to v_i(i.e.,(v_j , v_i)∈E_b) and the weight of this edge is $w_{j,i}$>0.
- Competing Graph $G_c = (V, E_c)$. For any two FL-PTs v_i and v_j, if they compete against each other, then there is an undirected edge between v_i and v_j (i.e.,(v_j , v_i)∈E_c) and if they are independent of each other, then (v_j , v_i)∉E_c.

Decision Variables $X = (x_{j,i})$:

- Data Usage Graph $G_u = (V, E_u)$. Let X=($x_{j,i}$) be an n×n matrix where $x_{j,i}$∈{0,1}: for two different FL-PTs v_i and v_j, $x_{j,i}$ is set to one if v_j will contribute to v_i in the FL training process and $x_{j,i}$ is set to zero otherwise.
- G_u will be a subgraph of the benefit graph G_b.

Collaboration Principles

Principle 1. Absence of free riders

- For any FL-PT $v_i \in V$, there exists a FL-PT $v_j \in V$ that benefits v_i if and only if there exists at least one FL-PT v_k that can benefit from v_i. Each FL-PT $v_i \in S_k$ is only concerned with the contributions of other FL-PTs within the same S_k.

Coalitions: A partition $\pi = \{S_1, S_2, \ldots, S_K\}$ is said to be a set of coalitions if we have for any $S_k \in \pi$ with $|\ S_k\ | \geq 2$ and $v_i \in S_k$ that $\sum_{v_j \in S_k - \{v_i\}} w_{i,j} > 0\ and\ \sum_{v_j \in S_k - \{v_i\}} w_{j,i} > 0$

Principle 2. Avoiding conflict of interest

- For any two competing FL-PTs v_i and v_j, v_j is unreachable to v_i in the data usage graph G_u.

Problem to Be Solved

- The problem of this paper is to find a partition π of FL-PTs such that
- Principles 1 and 2 are satisfied.
- Subject to Principles 1 and 2, no coalitions of π (i.e., no subset π' of π) can collaborate together and be merged into a larger coalition $S' = \cup_{S_k \in \pi'} S_k$ with a higher utility $u(S')$. Formally, let

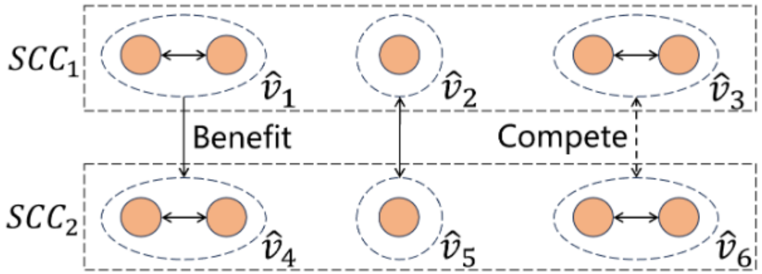
$\Pi = \left\{\pi' \subseteq \pi \middle| \sum_{S_k \in \pi'} u(S_k) < u(S'), Principles\ 1\ and\ 2\ are\ satisfied\ by\ S'\right\}$. Then $\Pi = \emptyset$.

Outline

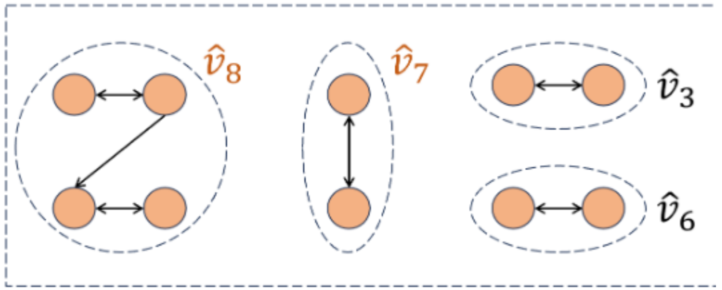
- ❑ Background & Motivation
- ❑ Problem Description
- ❑ **Solution**
- ❑ Experiments

Main Idea

- We find a partition $\hat{\pi} = \{\hat{S}_1, \hat{S}_2, \dots, \hat{S}_H\}$ of all FL-PTs V such that the FL-PTs of each subset $\hat{S}_h \in \hat{\pi}$ are independent of each other.
- $\hat{S}_h \in \hat{\pi}$ is further partitioned into several subsets/coalitions, denoted as $SCC_h = \{\hat{S}_{h,1}, \hat{S}_{h,2}, \dots, \hat{S}_{h,y_h}\}$ such that for all $l \in [1, y_h]$, $G_b(\hat{S}_{h,l})$ is a strongly connected component of $G_b(\hat{S}_h)$.
- For any coalitions of $\cup_{h=1}^{H} SCC_h$, we merge these coalitions into a larger one if doing so achieves a higher coalition utility without violating Principles 1 and 2.



(a) $\{SCC_h\}_{h=1}^{2}$ where $H = 2$.



(b) π

Algorithm 1: Conflict-free Coalitions without Free Riders

Input: The benefit graph $\mathcal{G}_b$, the competing graph $\mathcal{G}_c$
Output: The set π of coalitions

```
π ← ∅;                    // Record the set of coalitions found by this algorithm.
Construct the inverse of G_c, denoted as G_c^-;
Find all maximal cliques of G_c^-, denoted as π̂ = {Ŝ_1, ··· , Ŝ_H}, by the Bron–Kerbosch
  algorithm;
for h ← 1 to H do
    Find all strongly connected components of G_b(Ŝ_h) by the Tarjan algorithm; // The node
      sets of the components of G_b(Ŝ_h) are denoted as SCC_h = {Ŝ_{h,1}, ··· , Ŝ_{h,y_h}}.
Let π = {v̂_1, v̂_2, ··· , v̂_Y} = ⋃_{h=1}^{H} SCC_h where Y = Σ_{h=1}^{H} y_h;
Construct by Definition 2 a directed graph Z_b and an undirected graph Z_c whose node sets are
  π;           // v̂_y is a node in Z_b and Z_c but also represents a subset of V.
/* Below, the node v̂_l of Z_b with |v̂_l| = 1 is processed.                              */
Let y ← Y + 1;             // y is the index of the new node v̂_y to be constructed.
(π, Z_b, Z_c, y) ← MergeCycle(π, Z_b, Z_c, y), presented as Algorithm 2;
(π, Z_b, Z_c, y) ← MergePath(π, Z_b, Z_c, y), presented as Algorithm 4;
/* Below, the edge (v̂_l, v̂_l') of Z_b with |v̂_l| ⩾ 2 and |v̂_l'| ⩾ 2 is processed.       */
(π, Z_b, Z_c, y) ← MergeNeighbors(π, Z_b, Z_c, y), presented as Algorithm 5;
```

New Graph: Z_b and Z_c

- In the graph Z_b, there is a directed edge from $\hat{v}_l$ to $\hat{v}_{l'}$ if and only if there exist two nodes $v_i \in \hat{v}_l$ and $v_j \in \hat{v}_{l'}$ such that (v_i, v_j) is a directed edge in the benefit graph G_b.
- In the graph Z_c, there is an undirected edge between $\hat{v}_l$ and $\hat{v}_{l'}$ if and only if there exist two nodes $v_i \in \hat{v}_l$ and $v_j \in \hat{v}_{l'}$ such that (v_i, v_j) is an undirected edge in the competing graph G_c.

Merge Operation

NEURAL INFORMATION PROCESSING SYSTEMS

Algorithm 3: Merge$(\mathcal{X}, \pi, \mathcal{Z}_b, \mathcal{Z}_c, y)$

1 $\hat{v}_y \leftarrow \bigcup_{\hat{v}_j \in \mathcal{X}} \hat{v}_j$, $y \leftarrow y+1$, $\pi \leftarrow \pi - \mathcal{X}$, and $\pi \leftarrow \pi \cup \{\hat{v}_y\}$;

2 Add $\hat{v}_y$ into $\mathcal{Z}_b$ as a new node, and all the edges in the graph $\mathcal{Z}_b$ that point to (resp. point from) the nodes of $\mathcal{X}$ change to point to (resp. point from) $\hat{v}_y$;

3 Add $\hat{v}_y$ into $\mathcal{Z}_c$ as a new node, and all the edges in the graph $\mathcal{Z}_c$ whose endpoints are the nodes of $\mathcal{X}$ change to become the edges whose endpoints are $\hat{v}_y$;

4 Remove the nodes of $\mathcal{X}$ from both $\mathcal{Z}_b$ and $\mathcal{Z}_c$;

5 **Return** $(\hat{v}_y, \pi, \mathcal{Z}_b, \mathcal{Z}_c, y)$;

Definition 2. *In the graph $\mathcal{Z}_b$, there is a directed edge from $\hat{v}_l$ to $\hat{v}_{l'}$ if and only if there exist two nodes $v_i \in \hat{v}_l$ and $v_j \in \hat{v}_{l'}$ such that (v_i, v_j) is a directed edge in the benefit graph $\mathcal{G}_b$. In the graph $\mathcal{Z}_c$, there is an undirected edge between $\hat{v}_l$ and $\hat{v}_{l'}$ if and only if there exist two nodes $v_i \in \hat{v}_l$ and $v_j \in \hat{v}_{l'}$ such that (v_i, v_j) is an undirected edge in the competing graph $\mathcal{G}_c$. For any two coalitions $\hat{v}_l$ and $\hat{v}_{l'}$ of π, $\hat{v}_l$ is said to benefit (resp. benefit from) $\hat{v}_{l'}$ if there is a directed edge $(\hat{v}_l, \hat{v}_{l'})$ (resp. $(\hat{v}_{l'}, \hat{v}_l)$) in the graph $\mathcal{Z}_b$; $\hat{v}_l$ and $\hat{v}_{l'}$ are said to be competitive if there is an undirected edge $(\hat{v}_l, \hat{v}_{l'})$ in the graph $\mathcal{Z}_c$ and independent of each other otherwise.*

- **MergeCycle**: While there is a node $\hat{v}_{y_i}$ of Z_b with $|\hat{v}_{y_i}| = 1$ such that (i) there is a cycle $(\hat{v}_{y_1}, \hat{v}_{y_2}, \ldots, \hat{v}_{y_\theta}, \hat{v}_{y_1})$ in the graph Z_b that contains $\hat{v}_{y_i}$ and (ii) the nodes $\hat{v}_{y_1}, \hat{v}_{y_2}, \ldots, \hat{v}_{y_\theta}$ of this cycle are independent of each other do Merge Operation.
- **MergePath**: While there is a node $\hat{v}_{y_i}$ of Z_b with $|\hat{v}_{y_i}| = 1$ such that (i) there is a simple path $(\hat{v}_{y_1}, \ldots, \hat{v}_{y_i}, \ldots, \hat{v}_{y_\theta})$ with $\hat{v}_{y_1} \geq 2$ and $\hat{v}_{y_\theta} \geq 2$ and (ii) the nodes $\hat{v}_{y_1}, \hat{v}_{y_2}, \ldots, \hat{v}_{y_\theta}$ of this cycle are independent of each other do Merge and MergeCycle Operation.
- **MergeNeighbors**: While there is an edge $(\hat{v}_l, \hat{v}_{l'})$ of Z_b with $|\hat{v}_l| \geq 2$ and $|\hat{v}_{l'}| \geq 2$ such that $\hat{v}_l$ and $\hat{v}_{l'}$ are independent of each other do Merge, MergeCycle and MergePath Operation.

Outline

- ❑ Background & Motivation
- ❑ Problem Description
- ❑ Solution
- ❑ **Experiments**

Experimental Settings

NEURAL INFORMATION PROCESSING SYSTEMS

A Naive Pre-Processing Procedure

- We use the operations in lines 1-4 of Algorithm 1 to generate a set of coalitions, denoted as $\cup_{h=1}^{H} SCC_h,\ where\ SCC_h = \{\hat{S}_{h,1}, \hat{S}_{h,2}, \dots, \hat{S}_{h,y_h}\}$. This makes the previous FL approaches applicable to the scenario of this paper.

Datasets

- **Synthetic data:** A randomly generated dataset for regression tasks, which is generated in a similar way that has been used in literature.
- **CIFAR-10 and CIFAR-100:** Both the CIFAR-10 and CIFAR-100 datasets contain 60,000 color images for image classification tasks but have different levels of complexity. CIFAR-10 images have 10 classes with 6,000 images per class, while CIFAR-100 is more complex and has 100 classes with only 600 images per class.
- **eICU:** A dataset collecting electronic health records (EHRs) from many hospitals across the United States admitted to the intensive care unit(ICU). The task is to predict mortality during hospitalization.

Baselines

- **FedAvg**: A vanilla FL algorithm.
- **FedProx and SCAFFOLD**: Represent two typical approaches that make the aggregated model at the CS.
- **pFedHN and pFedMe**: Two approaches based on hypernetworks and meta-learning respectively.
- **FedDisco and pFedGraph**: Based on data complementarity.
- **FedOra**: Assesses if a FL-PT generalization performance can benefit from knowledge transferred from others and maximizes it.
- **Local**: Each FL-PT simply takes local ML training without collaboration.

Experimental Results under Synthetic Data

Two settings are considered:

- **Weakly Non-IID setting**: There exists a quantity skew, i.e., a significant difference in the sample quantities of FL-PTs.
- **Strongly Non-IID setting**: Conflicting learning tasks are generated by flipping over the labels of some FL-PTs.

Table 1: Experiments with synthetic data(Weakly Non-IID,MSE) under fixed competing graphs

	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8
LOCAL	0.32 ± 0.05	0.28 ± 0.00	1.00 ± 0.07	0.69 ± 0.08	0.28 ± 0.02	0.28 ± 0.01	0.72 ± 0.06	0.90 ± 0.11
FEDAVE	0.25 ± 0.01	0.25 ± 0.01	0.79 ± 0.05	0.55 ± 0.05	0.23 ± 0.01	$\mathbf{0.23 \pm 0.00}$	0.61 ± 0.04	0.74 ± 0.07
FEDPROX	0.26 ± 0.01	0.27 ± 0.01	0.90 ± 0.10	0.67 ± 0.06	0.26 ± 0.01	0.26 ± 0.01	0.76 ± 0.11	1.02 ± 0.17
SCAFFOLD	0.27 ± 0.01	0.28 ± 0.00	0.90 ± 0.03	0.67 ± 0.06	0.25 ± 0.01	0.26 ± 0.01	0.72 ± 0.09	0.92 ± 0.10
PFEDME	0.28 ± 0.02	0.29 ± 0.03	1.13 ± 0.55	0.86 ± 0.58	0.33 ± 0.13	0.33 ± 0.12	0.74 ± 0.02	0.82 ± 0.04
PFEDHN	0.35 ± 0.07	0.31 ± 0.05	0.91 ± 0.07	0.61 ± 0.06	0.33 ± 0.04	0.31 ± 0.05	0.70 ± 0.09	0.90 ± 0.18
PFEDGRAPH	0.26 ± 0.01	0.27 ± 0.01	0.90 ± 0.04	0.67 ± 0.08	0.26 ± 0.01	0.26 ± 0.00	0.74 ± 0.08	0.99 ± 0.05
FEDEGOISTS	$\mathbf{0.23 \pm 0.01}$	$\mathbf{0.24 \pm 0.00}$	$\mathbf{0.24 \pm 0.01}$	$\mathbf{0.22 \pm 0.02}$	$\mathbf{0.22 \pm 0.00}$	0.23 ± 0.01	$\mathbf{0.25 \pm 0.01}$	$\mathbf{0.25 \pm 0.02}$

Table 2: Experiments with synthetic data(Strongly Non-IID, MSE) under fixed competing graphs

	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8
LOCAL	0.29 ± 0.03	0.29 ± 0.02	0.26 ± 0.00	0.29 ± 0.04	0.27 ± 0.01	0.27 ± 0.04	0.27 ± 0.02	0.27 ± 0.01
FEDAVE	0.25 ± 0.00	$\mathbf{0.25 \pm 0.01}$	$\mathbf{0.23 \pm 0.01}$	0.23 ± 0.01	0.23 ± 0.01	$\mathbf{0.22 \pm 0.00}$	0.23 ± 0.02	0.24 ± 0.02
FEDPROX	0.27 ± 0.02	0.26 ± 0.01	0.26 ± 0.01	0.26 ± 0.01	0.24 ± 0.01	0.24 ± 0.01	0.25 ± 0.01	0.25 ± 0.01
SCAFFOLD	0.26 ± 0.01	0.26 ± 0.01	0.26 ± 0.01	0.26 ± 0.01	0.24 ± 0.01	0.24 ± 0.01	0.25 ± 0.01	0.25 ± 0.01
PFEDME	0.36 ± 0.12	0.37 ± 0.12	0.25 ± 0.00	0.25 ± 0.01	0.28 ± 0.02	0.27 ± 0.01	0.27 ± 0.01	0.28 ± 0.01
PFEDHN	0.33 ± 0.05	0.34 ± 0.03	0.32 ± 0.05	0.28 ± 0.03	0.34 ± 0.03	0.29 ± 0.03	0.29 ± 0.05	0.29 ± 0.06
PFEDGRAPH	0.26 ± 0.01	0.27 ± 0.01	0.26 ± 0.02	0.26 ± 0.02	0.24 ± 0.01	0.24 ± 0.01	0.25 ± 0.01	0.25 ± 0.01
FEDEGOISTS	$\mathbf{0.24 \pm 0.00}$	0.27 ± 0.05	0.24 ± 0.03	$\mathbf{0.22 \pm 0.01}$	$\mathbf{0.22 \pm 0.00}$	$\mathbf{0.22 \pm 0.00}$	$\mathbf{0.22 \pm 0.01}$	$\mathbf{0.22 \pm 0.01}$

Results: FedEgoists has the best performance compared with baselines.

Experimental Results under CIFAR-10 and CIFAR-100

NEURAL INFORMATION PROCESSING SYSTEMS

CIFAR10:

Table 3: Accuracy comparisons(MTA) under different α on CIFAR10.

α		LOCAL	FEDAVG	FEDPROX	SCAFFOLD	PFEDME	PFEDHN	FEDDISCO	PFEDGRAPH	FEDORA	FEDEGOISTS
0.05	PAT	80.47±2.06	36.86±3.00	36.62±6.17	36.61±6.18	48.66±6.38	66.53±2.00	36.61±6.18	52.04±8.66	69.73±1.62	**81.35±0.30**
0.05	Dir	61.59±0.53	44.98±1.91	46.94±2.12	46.76±2.92	44.64±2.61	55.61±0.45	46.74±2.99	46.56±2.55	55.28±0.75	**63.06±0.64**
0.1	PAT	80.47±2.06	49.40±5.50	48.19±5.17	48.18±5.16	56.56±1.66	66.61±1.62	48.19±5.17	55.35±4.51	68.65±2.02	**80.73±1.35**
0.1	Dir	61.59±0.53	46.77±1.96	48.71±1.97	48.61±2.02	46.65±2.74	54.21±0.83	48.56±1.99	49.10±3.19	55.97±0.22	**62.74±1.09**
0.2	PAT	80.47±2.06	63.67±2.10	57.26±1.48	57.24±2.34	79.27±1.35	76.08±2.20	57.25±2.15	60.27±2.33	72.74±1.91	**81.30±1.46**
0.2	Dir	61.59±0.53	55.69±1.90	53.79±1.07	54.16±0.79	53.64±0.79	61.31±0.56	54.08±1.43	53.85±1.07	55.67±0.96	**66.62±1.23**
0.3	PAT	80.47±2.06	57.95±2.37	59.82±4.88	59.83±4.87	63.09±3.26	65.11±2.4	59.82±4.88	62.12±4.51	71.51±2.40	**81.37±1.41**
0.3	Dir	61.59±0.53	50.48±0.87	49.99±1.15	50.09±1.29	49.33±1.94	53.21±0.49	50.17±1.29	50.66±1.59	55.9±1.01	**63.39±0.89**
0.4	PAT	80.47±2.06	58.47±5.87	63.28±4.54	63.27±4.54	66.36±3.88	67.51±3.04	63.28±4.55	63.30±4.61	72.89±1.67	**82.54±0.30**
0.4	Dir	61.59±0.53	50.14±2.2	51.20±2.16	51.23±2.09	51.00±0.94	53.04±0.80	51.14±2.09	51.14±2.16	57.26±0.32	**62.81±0.88**

Setting: We show the performance of the proposed approach when α takes different values in $\{0.05, 0.1, 0.2, 0.3, 0.4\}$, representing different levels of competing intensity between FL-PTs.

CIFAR100:

Table 4: Accuracy comparisons(MTA) under different α on CIFAR100.

α		LOCAL	FEDAVG	FEDPROX	SCAFFOLD	PFEDME	PFEDHN	FEDDISCO	PFEDGRAPH	FEDORA	FEDEGOISTS
0.05	PAT	46.24±1.38	34.52±8.65	35.42±1.36	35.47±1.36	35.78±1.72	29.98±1.07	35.42±3.58	36.60±1.15	41.91±0.49	**47.00±1.81**
0.05	Dir	**30.31±0.48**	15.33±5.35	19.81±6.54	19.73±6.50	18.71±1.41	18.12±0.92	19.76±6.56	19.76±6.50	27.06±0.26	27.59±1.52
0.1	PAT	46.24±1.38	40.01±0.89	42.57±0.44	42.73±0.44	34.40±4.67	30.17±0.47	42.56±0.45	42.78±0.46	42.63±1.04	**46.28±1.05**
0.1	Dir	30.31±0.48	20.25±4.93	18.86±5.07	18.80±5.03	20.51±0.98	17.45±0.55	18.87±5.05	18.88±4.95	27.50±0.21	**32.01±1.66**
0.2	PAT	46.24±1.38	29.68±4.12	28.60±4.56	28.55±4.34	29.90±1.85	28.38±0.71	29.05±4.11	30.51±4.03	41.63±1.65	**50.21±2.24**
0.2	Dir	30.31±0.48	19.24±1.13	20.10±0.35	20.00±0.48	19.89±0.36	23.11±0.79	19.93±0.38	20.17±0.35	27.24±0.36	**32.86±1.53**
0.3	PAT	46.24±1.38	40.24±0.55	42.42±0.42	42.57±0.30	44.34±2.16	29.63±0.23	42.42±0.41	42.48±0.48	41.72±1.98	**46.38±1.83**
0.3	Dir	30.31±0.48	25.56±0.32	27.37±0.17	27.27±0.24	25.28±2.55	17.21±0.17	27.37±0.17	26.18±1.69	27.43±0.20	**34.30±0.44**
0.4	PAT	46.24±1.38	40.52±0.27	41.63±1.03	41.71±1.05	44.38±1.94	30.18±0.28	41.73±1.03	41.66±1.10	42.94±0.25	**48.16±1.61**
0.4	Dir	30.31±0.48	24.73±0.97	27.37±0.40	27.31±0.26	26.72±1.89	17.08±0.35	27.37±0.40	27.17±0.42	27.24±0.23	**34.15±0.96**

Results: FedEgoists has the best performance compared with the nine baselines.

New Metric

- $r_{\alpha,l,p}$: The performance of the proposed approach.
- $r_{\alpha,l,i}$: The performance of the i-th baseline approach. $i \in \{1,2,\dots,9\}$
- l^*: $l^* = \mathrm{argmax}_{l\in[1,5]}(\max_{i\in[1,9]} r_{\alpha,l,i} - r_{\alpha,l,p})$ where $\max_{i\in[1,9]} r_{\alpha,l,i}$ is the best performance of all the baseline approaches in the l^*-th trial and $\max_{i\in[1,9]} r_{\alpha,l,i} - r_{\alpha,l,p}$ is their performance improvement (or the difference) to the proposed approach, which may be negative if the proposed approach achieves a better performance.

Table 5: The worst-case performance of the proposed approach compared with the baseline approaches.

	0.05		0.1		0.2		0.3		0.4	
	PAT	Dir	PAT	Dir	PAT	Dir	PAT	Dir	PAT	Dir
CIFAR10	0.011000	-0.002903	0.022900	-0.000624	0.025800	-0.0006030	0.028800	-0.005725	-0.002399	-0.000100
CIFAR100	-0.000999	0.076002	0.011400	-0.000008	-0.000636	-0.0009356	-0.000020	-0.032153	-0.000699	-0.027078

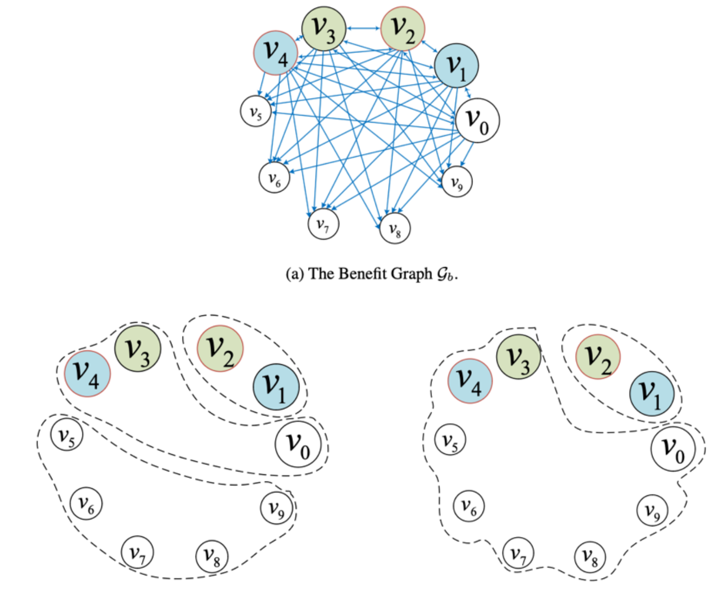
Experimental Results under eICU

Settings:

- There are ten hospitals in total, with $\{v_i\}_{i=0}^{4}$ as large hospitals and $\{v_i\}_{i=5}^{9}$ as small hospitals.
- Due to the extreme imbalance of data labels, where over 90% are negative labels, we use the AUC scores to evaluate the performance of the trained model

Table 6: eICU

AUC	LOCAL	FEDAVG	FEDPROX	SCAFFOLD	pFEDME	pFEDHN	FEDDISCO	pFEDGRAPH	FEDORA	FedEgoists
v_0	53.64±22.12	63.52±22.40	80.42±9.85	80.24±9.92	52.30±19.79	41.94±19.14	60.48±13.07	80.42±9.85	**90.36±2.26**	66.36±19.28
v_1	67.94±6.88	62.55±16.49	57.03±16.62	57.21±16.68	46.00±34.96	76.61±14.77	63.76±14.97	59.62±7.49	81.52±16.91	**81.58±6.65**
v_2	37.33±17.74	76.48±12.70	60.13±6.77	60.38±6.64	36.48±27.59	79.62±16.18	**92.70±4.60**	57.32±8.17	47.56±9.62	66.04±33.21
v_3	79.88±21.16	67.04±26.74	78.74±15.66	78.87±15.44	45.79±32.04	55.35±26.55	80.38±18.24	78.69±7.48	75.12±7.85	**84.40±5.76**
v_4	52.48±11.61	73.46±15.58	73.63±9.74	75.75±11.07	57.07±23.12	48.75±22.68	70.15±9.96	49.61±5.31	48.95±6.80	**75.84±11.26**
v_5	39.45±9.06	57.09±7.46	61.94±9.13	61.70±9.12	55.15±24.92	52.55±25.12	53.03±9.73	**89.37±7.71**	77.72±8.24	68.41±5.60
v_6	68.00±32.62	77.61±5.87	79.62±7.62	78.74±7.81	57.23±32.51	42.01±16.65	82.26±6.41	**98.80±0.76**	98.55±1.18	56.86±7.52
v_7	73.36±7.08	71.80±9.52	73.55±10.48	73.59±10.17	56.60±7.56	51.21±5.01	68.45±10.98	76.82±11.07	75.53±5.94	**77.97±14.94**
v_8	36.24±22.56	73.55±2.70	77.47±3.80	77.43±3.66	61.22±10.49	46.71±16.08	65.05±3.41	69.16±3.12	72.26±12.01	**90.60±10.57**
v_9	71.70±10.64	63.14±9.42	63.82±9.32	63.79±9.36	42.97±12.63	45.42±17.42	63.24±10.63	60.76±10.12	58.55±7.62	**79.88±8.29**
Avg	58.01	68.62	70.66	70.77	51.08	54.02	69.95	72.06	72.61	**74.79**



(a) The Benefit Graph $\mathcal{G}_b$.

(b) The set π of coalitions in the baseline algorithms.

(c) The set π of coalitions by Algorithm 1.

Figure 6: **Real-world Collaboration Example**

Results: Extensive experiments over real-world datasets have demonstrated the effectiveness of the proposed solution compared to nine baseline methods, and its ability to establish efficient collaborative networks in cross-silos FL with FL-PTs that engage in business activities.

Thank you for your listening!

Any questions?