Does Egalitarian Fairness Lead to Instability? The Fairness Bounds in Stable Federated Learning Under Altruistic Behaviors

Jiashi Gao[1], Ziwei Wang[1,2], Xiangyu Zhao[3], Xin Yao[4], Xuetao Wei[1]*

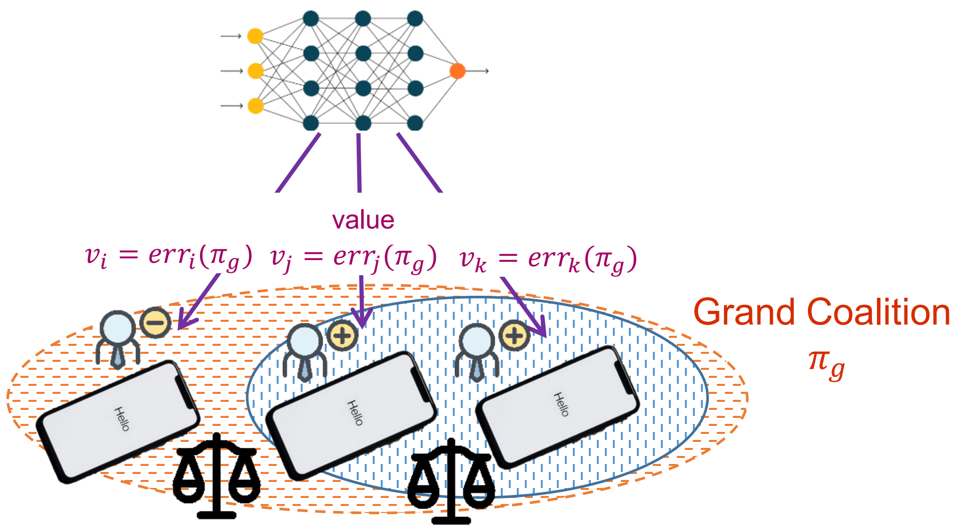
[1]Southern University of Science and Technology
[2]University of Birmingham
[3]City University of Hong Kong
[4]Lingnan University
{12131101,12250053}@mail.sustech.edu.cn
xy.zhao@cityu.edu.hk
xinyao@ln.edu.hk
weixt@sustech.edu.cn

Background & Motivation

➢ What is "egalitarian fairness" in federated learning?

- Ensuring that the performance of global model across the clients roughly comparable or even equal [1,2,3]

- **Welfare Scenario**: Enhance fairness in federated learning for clients with limited data due to unavoidable circumstances.



Definition 1 *(Egalitarian fairness) For the clients within a coalition π holding datasets of varying sizes $\{n_1, n_2, ..., n_N\}$ and experiencing errors $\{err_1(\pi), err_2(\pi), ..., err_N(\pi)\}$, the coalition structure π satisfy λ-egalitarian fairness if there exists a constant λ such that,*

$$\frac{err_i(\pi)}{err_j(\pi)} \geq \lambda, n_i \leq n_j. \tag{2}$$

Here, λ is the fairness bound. When $\lambda = 1$, the coalition π is said to satisfy strict egalitarian fairness.

Background & Motivation

- Why we care about "stability" and "egalitarian fairness"?

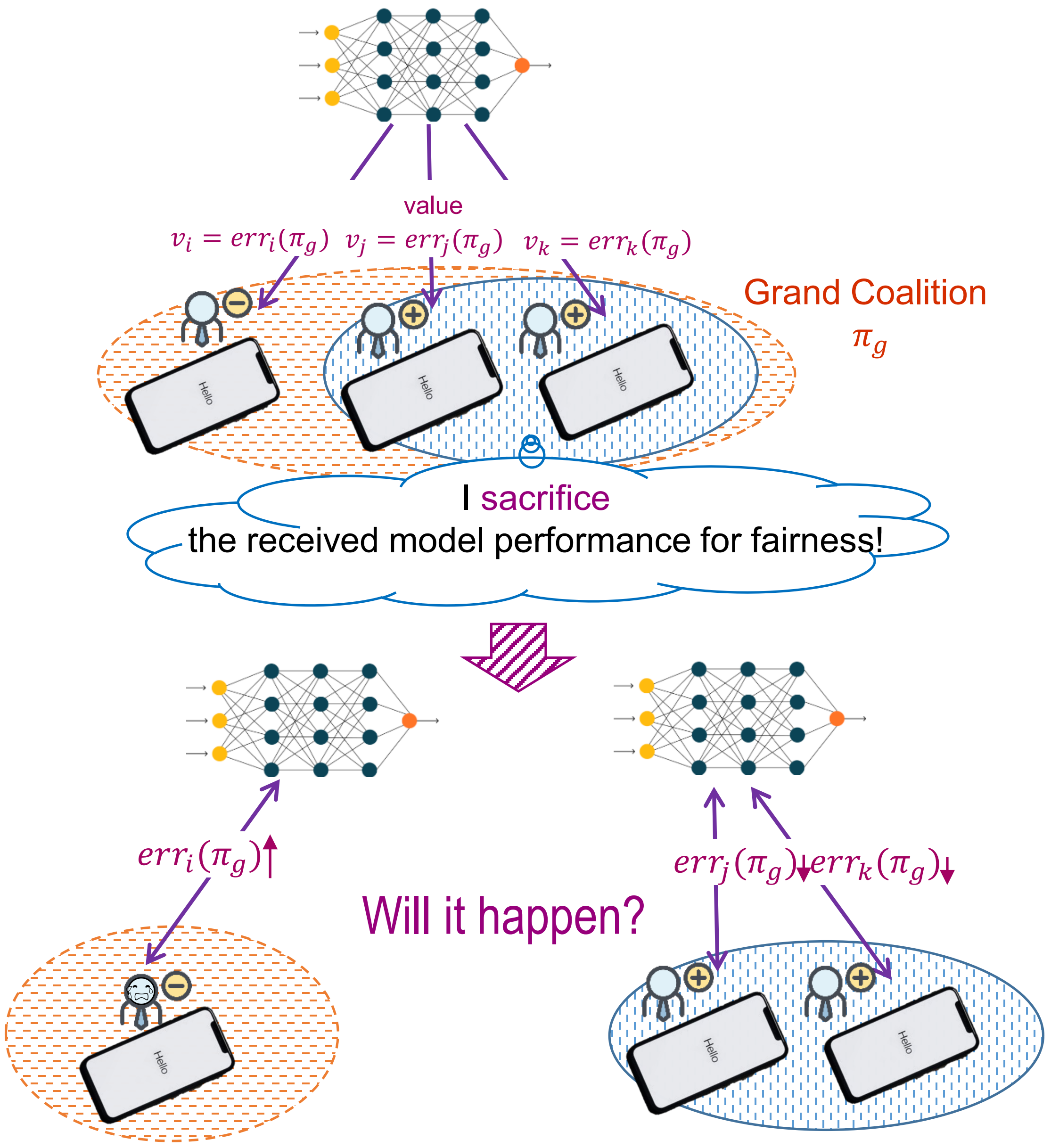
- **Observation:** Egalitarian fairness is misunderstood as unavoidably causing high-data-resource clients to leave the grand coalition and form sub-coalitions, thereby undermining the stability of federated learning.

- **Research Questions**

① How does egalitarian fairness affect the stability of FLs?

② How does this impact vary when clients exhibit altruistic behaviors?

③ What is the optimal egalitarian fairness that a stable FL can achieve?



Task model

- Mean estimation task with the **closed-form local errors** (Donahue et al. 2021.)

  (Necessary to determine a tight fairness bound)

Model-sharing Games:
Analyzing Federated Learning Under Voluntary Participation

Kate Donahue,[1] **Jon Kleinberg,** [1, 2]
[1] Department of Computer Science, Cornell University
[2] Department of Information Science, Cornell University
kdonahue@cs.cornell.edu, kleinber@cs.cornell.edu

In an FL setting with N clients, each client possesses a local dataset $\mathcal{D}_i$ of size n_i. The local dataset of each client $\mathcal{D}_i$ is with mean θ_i and standard deviation ϵ_i, where $(\theta_i, \epsilon_i^2) \sim \Theta$. When FL trains a global model for mean estimation and employs FedAvg for aggregation, the expected mean squared error (MSE) for a client with n_i samples within coalition π is as follows,

$$err_i(\pi) = \frac{\mu_e}{\sum_{j \in \pi} n_j} + \sigma^2 \cdot \frac{\sum_{j \in \pi, j\,\mathrm{neq}\,i} n_j^2 + \left(\sum_{j \in \pi, j\,\mathrm{neq}\,i} n_j\right)^2}{\left(\sum_{j \in \pi} n_j\right)^2},$$

where $\mu_e = \mathbb{E}_{(\theta_i, \epsilon_i^2) \sim \Theta}[\epsilon_i^2]$ denotes the expected value of the variance of the dataset distribution, and $\sigma^2 = var(\theta_i)$ denotes the variance between the means of the clients' local datasets.

Definitions

Definition 2 *(Value) In the context of collaborative gaming, the value quantifies the payoff accrued to the i-th player as a result of participating within the current coalition π. Within the framework of FL, the value is defined as the error of the global model evaluated on the i-th client's local dataset as* $v_i(\pi) = err_i(\pi)$.

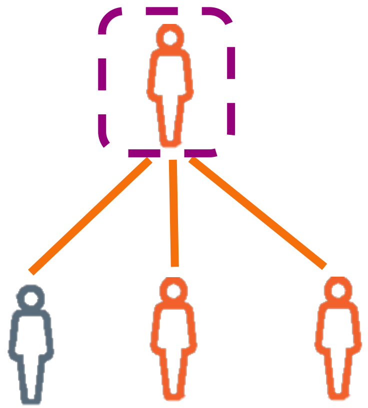
Definition 3 *(Friend) In a broader sociological context, the friend is considered the most intimate, trustful, and voluntarily chosen tie people maintain. Within the framework of FL, the friend set of the i-th client, denoted as F_i, is defined as the clients whose value is also expected to be better when i-th client makes a coalition participation decision.*

Definition 4 *(Core stability) The grand coalition π_g (the coalition consisting of all players) is considered to be core-stable if there does not exist nonempty sub-coalition $\pi_s \subset \pi_g$ such that $\pi_s \succ_i \pi_g$ for $\forall i \in \pi_s$, where $\succ$ is used to denote a preference relation. In other words, no nonempty sub-coalition $\pi_s \subset \pi_g$ blocks π_g.*

Game model

- Client behaviors

the i-th client

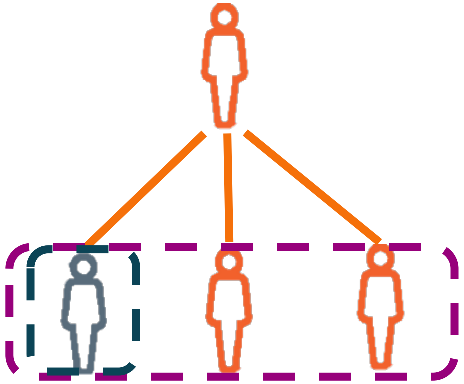


Friend set F_i

Purely selfish

$$u_i^{ps}(\pi) = v_i(\pi)$$

the i-th client



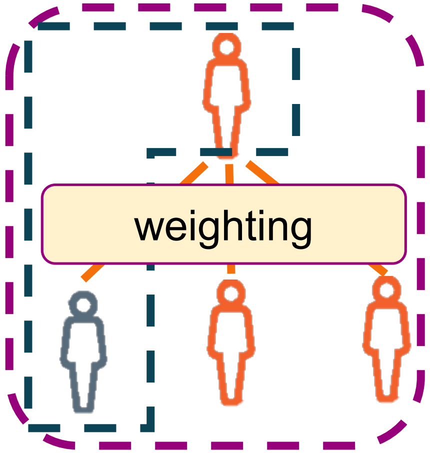
Friend set F_i

Purely welfare/
equal altruistic

$$u_i^{pa}(\pi) = \max_{f \in F_i}(\{v_f(\pi)\})$$

$$u_i^{pa}(\pi) = \frac{1}{|F_i|}\sum_{f \in F_i} v_f(\pi)$$

the i-th client



Friend set F_i

Friendly welfare
/equal altruistic

$$u_i^{fa}(\pi) = w \cdot v_i(\pi) + (1-w) \cdot \max_{f \in F_i \cup \{i\}}(\{v_f(\pi)\})$$

$$u_i^{fa}(\pi) = w \cdot v_i(\pi) + (1-w) \cdot \frac{1}{|F_i|+1}\sum_{f \in F_i \cup \{i\}} v_f(\pi)$$

Does egalitarian fairness lead to instability?

➢ Experimental findings

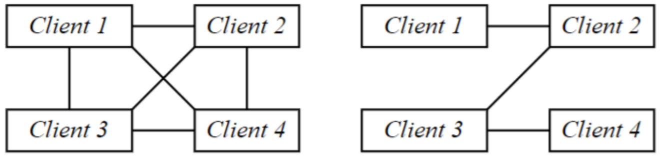


Figure 1: Friends-relationship networks: fully connected relation I (left) and partially connected relation II (right).

Coalition Structure	Error (= u^{ps})				Utility u^{fa} in AHG (Relation I)				Utility u^{fa} in ACFG (Relation I)				Utility u^{fa} in ACFG (Relation II)			
	err_1	err_2	err_3	err_4	u_1	u_2	u_3	u_4	u_1	u_2	u_3	u_4	u_1	u_2	u_3	u_4
{1}	2.0	/	/	/	2.0	/	/	/	2.0	/	/	/	2.0	/	/	/
{2}	/	2.0	/	/	/	2.0	/	/	/	2.0	/	/	/	2.0	/	/
{3}	/	/	1.0	/	/	/	1.0	/	/	/	1.22	/	/	/	1.22	/
{4}	/	/	/	0.666	/	/	/	0.666	/	/	/	1.020	/	/	/	0.770
{1,2}	1.5	1.5	/	/	1.5	1.5	/	/	1.5	1.5	/	/	1.5	1.5	/	/
{2,3}	/	1.555	0.888	/	/	1.555	1.222	/	/	1.590	1.256	/	/	1.590	1.222	/
{3,4}	/	/	1.12	0.72	/	/	1.12	0.92	/	/	1.31	1.11	/	/	1.31	0.92
{1,3}	1.555	/	0.888	/	1.555	/	1.222	/	1.590	/	1.256	/	1.590	/	1.256	/
{1,4}	1.625	/	/	0.625	1.625	/	/	1.125	1.625	/	/	1.125	1.625	/	/	0.756
{2,4}	/	1.625	/	0.625	/	1.625	/	1.125	/	1.625	/	1.125	/	1.625	/	0.756
{1,2,3}	1.375	1.375	0.875	/	1.375	1.375	1.125	/	1.375	1.375	1.125	/	1.375	1.375	1.125	/
{1,2,4}	1.44	1.44	/	0.64	1.44	1.44	/	1.04	1.44	1.44	/	1.04	1.44	1.44	/	0.82
{1,3,4}	1.388	/	1.055	0.722	1.388	/	1.222	1.055	1.694	/	1.527	1.361	1.694	/	1.527	0.888
{2,3,4}	/	1.388	1.055	0.722	/	1.388	1.222	1.055	/	1.694	1.527	1.361	/	1.694	1.222	0.888
{1,2,3,4}	1.306	1.306	1.020	0.734	1.306	1.306	1.163	1.020	1.306	1.306	1.163	1.020	1.306	1.306	1.163	0.877

Takeaways from experiments

① Whether "egalitarian fairness leads to instability" is influenced by the clients' behavior;

② Whether "egalitarian fairness leads to instability" is influenced by the diverse friends-relationship networks.

How to establish appropriate egalitarian fairness in FL implementation?

- Preliminary

- Distance function

$$d(\pi, n_j) = \left(\sum_{i \in \pi} n_i^2 - n_j^2\right) + \left(\sum_{i \in \pi} n_i - n_j\right)^2.$$

measure the dataset size of a client relative to all other clients within the same coalition π.

- Notations

Table 2: Notation Definitions.

Notation	Description
π_c	The complement coalition of a coalition π_s: $\pi_c = \pi_g \setminus \pi_s$.
N_s	The sum of the dataset sizes in π_s: $N_s = \sum_{i \in \pi_s} n_i$.
N_c	The sum of the dataset sizes in π_c: $N_c = \sum_{i \in \pi_c} n_i$.
N_g	The sum of the dataset sizes in the grand coalition: $N_g = \sum_{i \in \pi_g} n_i$.
m	The index of the client with the smallest dataset size in π_g: $m = \arg\min_{i \in \pi_g} \{n_i\}$.
l	The index of the client with the largest dataset size in π_g: $l = \arg\max_{i \in \pi_g} \{n_i\}$.

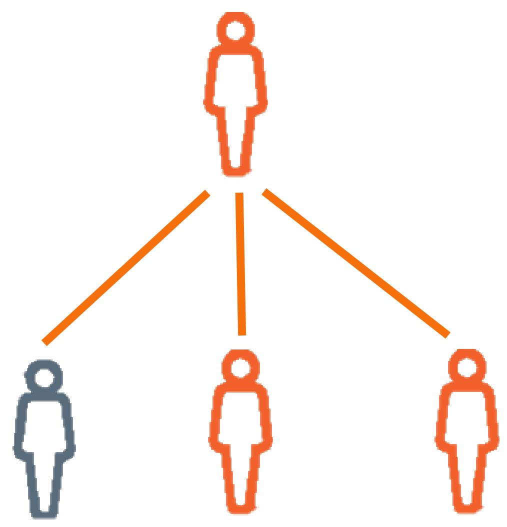
How to establish appropriate egalitarian fairness in FL implementation?

- Theoretical results showing **how the achievable bounds of egalitarian fairness vary under different client behaviors**

  - **Proposition 2** Considering all clients are purely selfish, the grand coalition π_g remains core-stable if the achieved egalitarian fairness is bounded by:

$$\lambda \geqslant \max_{\pi_s \subset \pi_g} \left\{ \frac{N_s^{\ 2}}{N_g^{\ 2}} \cdot \frac{N_g \cdot n_l + d(\pi_g, n_m)}{N_s \cdot n_l + d\left(\pi_s, n_{k_{\pi_s}}\right)} \right\}, \ where\ k_{\pi_s} = \mathrm{argmin}_{i \in \pi_s} \{n_i\}.$$

Insights: increase in the heterogeneity—the achievable egalitarian fairness of a core-stable grand coalition becomes poorer.



  - Sufficient condition for achieving strict egalitarian fairness (λ = 1)

Corollary 2 The core-stable grand coalition π_g comprising all selfish clients, can asymptotically achieve strict egalitarian fairness, provided that the local dataset sizes of all clients are equal.

How to establish appropriate egalitarian fairness in FL implementation?

➢ Theoretical results showing **how the achievable bounds of egalitarian fairness vary under different client behaviors**

- **Proposition 3** Considering all clients are purely welfare altruistic, the grand coalition π_g remains core-stable if the achieved egalitarian fairness is bounded by:

$$\lambda \geqslant \max_{\pi_s \in \pi_g} \left\{ \min \left(\frac{N_s^2}{N_g^2} \cdot \frac{N_g \cdot n_l + d(\pi_g, n_m)}{N_s \cdot n_l + d(\pi_s, f^{opt}_{\pi_s,1})}, \frac{N_c^2}{N_g^2} \cdot \frac{N_g \cdot n_l + d(\pi_g, n_m)}{N_c \cdot n_l + d(\pi_c, f^{opt}_{\pi_s,2})} \right) \right\},$$

where

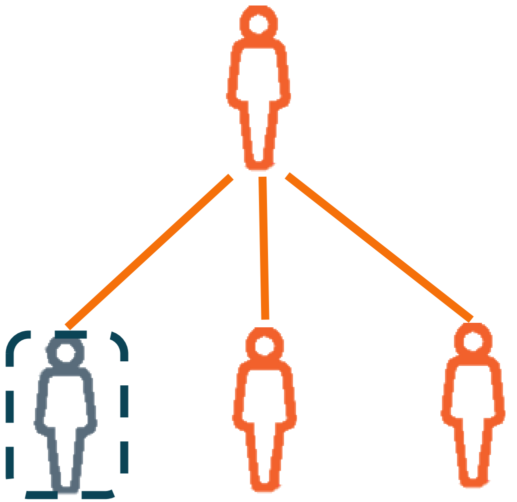
$$k_{\pi_s,1} = \mathrm{argmin}_{i \in \pi_s} \{\min_{f \in F_i \cap \pi_s} n_f\}, k_{\pi_s,2} = \mathrm{argmin}_{i \in \pi_s} \{\min_{f \in F_i \cap \pi_c} n_f\},$$
$$f^{opt}_{\pi_s,1} = \mathrm{argmin}_{f \in F_{k_{\pi_s,1}} \cap \pi_s} n_f, f^{opt}_{\pi_s,2} = \mathrm{argmin}_{f \in F_{k_{\pi_s,2}} \cap \pi_c} n_f.$$

Insights: the achieved egalitarian fairness declines as the gap between the smallest dataset size overall and the smallest dataset size within any given friends-relationship network increases.

- More relaxed condition for achieving strict egalitarian fairness (λ = 1)

Corollary 3 The core-stable grand coalition π_g consisting of purely welfare clients, can asymptotically achieve strict egalitarian fairness if all clients are friends with the client possessing the smallest dataset size

the i-th client



Friend set F_i

Purely welfare altruistic

$$u_i^{pa}(\pi) = \max_{f \in F_i} (\{v_f(\pi)\})$$

How to establish appropriate egalitarian fairness in FL implementation?

- Achievable bounds of egalitarian fairness under more complex client behaviors
  - **Proposition 4** Considering all clients are purely equal altruistic, the grand coalition π_g remains core-stable if the achieved egalitarian fairness is bounded by:

$$\lambda \geqslant \max_{\pi_s \in \pi_g} \left(\frac{|F_{k_{\pi_s}}| \cdot {N_s}^2 {N_c}^2}{{N_g}^2} \cdot \frac{N_g \cdot n_l + d(\pi_g, n_m)}{\mathbf{Q}} \right),$$

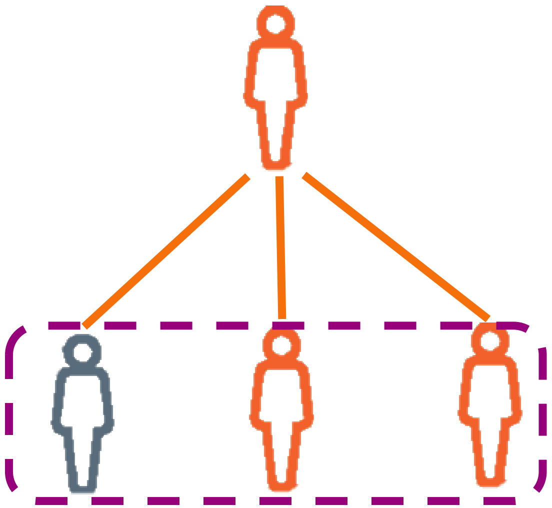
where

$$k_{\pi_s} = \operatorname{argmin}_{i \in \pi_s} \frac{1}{|F_i|} \left(\sum_{f \in F_i \cap \pi_s} n_f + \sum_{f \in F_i \cap \pi_c} n_f \right),$$

$$\mathbf{Q} = N_c^2 \cdot \sum_{f \in F_{k_{\pi_s}} \cap \pi_s} (N_s \cdot n_l + d(\pi_s, n_f)) + N_s^2 \cdot \sum_{f \in F_{k_{\pi_s}} \cap \pi_c} (N_c \cdot n_l + d(\pi_c, n_f)).$$

  - Insights: the egalitarian fairness bound for purely equal altruistic clients is influenced by the gap between the smallest dataset size overall and the weighted sum of dataset sizes within any given friends-relationship network.

the i-th client



Friend set F_i

Purely equal altruistic

$$u_i^{pa}(\pi) = \frac{1}{|F_i|} \sum_{f \in F_i} v_f(\pi)$$

How to establish appropriate egalitarian fairness in FL implementation?

- Achievable bounds of egalitarian fairness under more complex client behaviors
  - **Proposition 5** Considering all clients are friendly welfare altruistic, the grand coalition π_g remains core-stable if the achieved egalitarian fairness is bounded by:

$$\lambda \geqslant \max_{\pi_s \in \pi_g} \left\{ \min \left(\frac{{N_s}^2}{{N_g}^2} \cdot \frac{N_g \cdot n_l + d(\pi_g, n_m)}{\mathbf{Q}_1}, \frac{{N_s}^2 {N_c}^2}{{N_g}^2} \cdot \frac{N_g \cdot n_l + d(\pi_g, n_m)}{\mathbf{Q}_2} \right) \right\},$$

where

$$k_{\pi_s,1} = \operatorname{argmin}_{i \in \pi_s} \left\{ w \cdot n_i + (1-w) \cdot \min_{f \in F_i \cap \pi_s \cup \{i\}} n_f \right\},$$

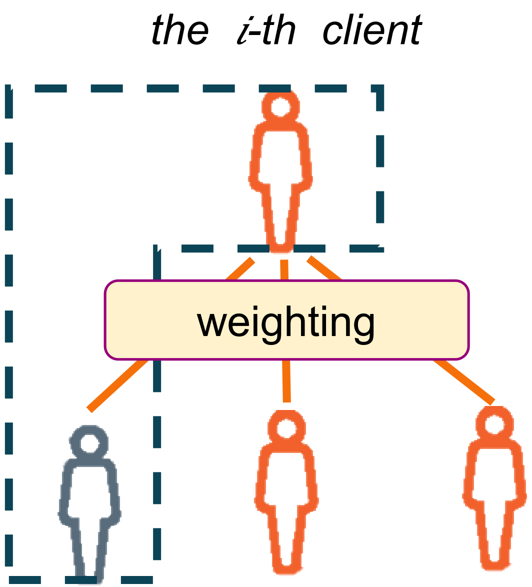
$$k_{\pi_s,2} = \operatorname{argmin}_{i \in \pi_s} \left\{ w \cdot n_i + (1-w) \cdot \min_{f \in F_i \cap \pi_c} n_f \right\},$$

$$f^{opt}_{\pi_s,1} = \operatorname{argmin}_{f \in F_{k_{\pi_s,1}} \cap \pi_s \cup \left\{ k_{\pi_s,1} \right\}} n_f, f^{opt}_{\pi_s,2} = \operatorname{argmin}_{f \in F_{k_{\pi_s,2}} \cap \pi_c} n_f,$$

$$\mathbf{Q}_1 = N_s \cdot n_l + w \cdot d\left(\pi_s, n_{k_{\pi_s,1}}\right) + (1-w) \cdot d\left(\pi_s, f^{opt}_{\pi_s,1}\right),$$

$$\mathbf{Q}_2 = N_c^2 \cdot w \cdot \left(N_s \cdot n_l + d\left(\pi_s, n_{k_{\pi_s,2}}\right)\right) + N_s^2 \cdot (1-w) \cdot \left(N_c \cdot n_l + d\left(\pi_c, f^{opt}_{\pi_s,2}\right)\right).$$

  - Insights: the egalitarian fairness bounds in the context of friendly altruism behavior are shaped by two factors:
    ① the heterogeneity of clients' local dataset sizes;
    ② the difference between the smallest dataset size in the grand coalition and the smallest dataset size within established friends-relationship networks.



Friend set F_i

Friendly welfare altruistic

$$u_i^{fa}(\pi) = w \cdot v_i(\pi) + (1-w) \cdot \max_{f \in F_i \cup \{i\}} (\{v_f(\pi)\})$$

Balanced by the selfishness degree parameter (w)

How to establish appropriate egalitarian fairness in FL implementation?

- Achievable bounds of egalitarian fairness under more complex client behaviors
  - **Proposition 6** Considering all clients are friendly equal altruistic, the grand coalition π_g remains core-stable if the achieved egalitarian fairness is bounded by:

$$\lambda \geqslant \max_{\pi_s \in \pi_g}\left(\frac{\left(|F_{k_{\pi_s}}|+1\right)\cdot {N_s}^2 \cdot {N_c}^2}{{N_g}^2}\cdot\frac{N_g\cdot n_l + d(\pi_g, n_m)}{\mathbf{Q}}\right),$$

where

$$k_{\pi_s} = \text{argmin}_{i\in\pi_s}\left(w\cdot n_i + (1-w)\cdot\frac{1}{|F_i|+1}\cdot\left(\sum_{f\in F_i\cap\pi_s\cup\{i\}} n_f + \sum_{f\in F_i\cap\pi_c} n_f\right)\right),$$

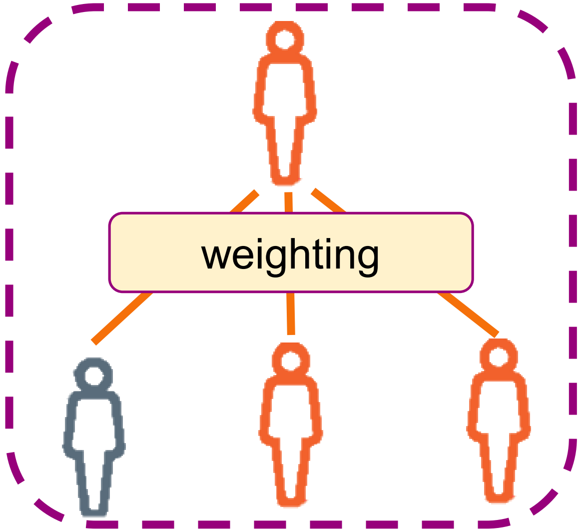
$$\hat{F}_s = F_{k_{\pi_s}}\cap\pi_s\cup\{k_{\pi_s}\}, \hat{F}_c = F_{k_{\pi_s}}\cap\pi_c,$$

$$\mathbf{Q} = w\cdot\left(|F_{k_{\pi_s}}|+1\right)\cdot N_c^2\cdot\left(N_s\cdot n_l + d\left(\pi_s, n_{k_{\pi_s}}\right)\right)+$$

$$(1-w)\cdot\left(N_c^2\cdot\sum_{f\in\hat{F}_s}(N_s\cdot n_l + d(\pi_s,n_f)) + N_s^2\cdot\sum_{f\in\hat{F}_c}(N_c\cdot n_l + d(\pi_c,n_f))\right).$$

  - Insights: the egalitarian fairness bounds in the context of friendly altruism behavior are shaped by two factors:
    ① the heterogeneity of clients‘ local dataset sizes;
    ② the difference between the smallest dataset size in the grand coalition and the weighted sum of dataset sizes within established friends-relationship networks.

the i-th client



Friend set F_i

Friendly equal altruistic

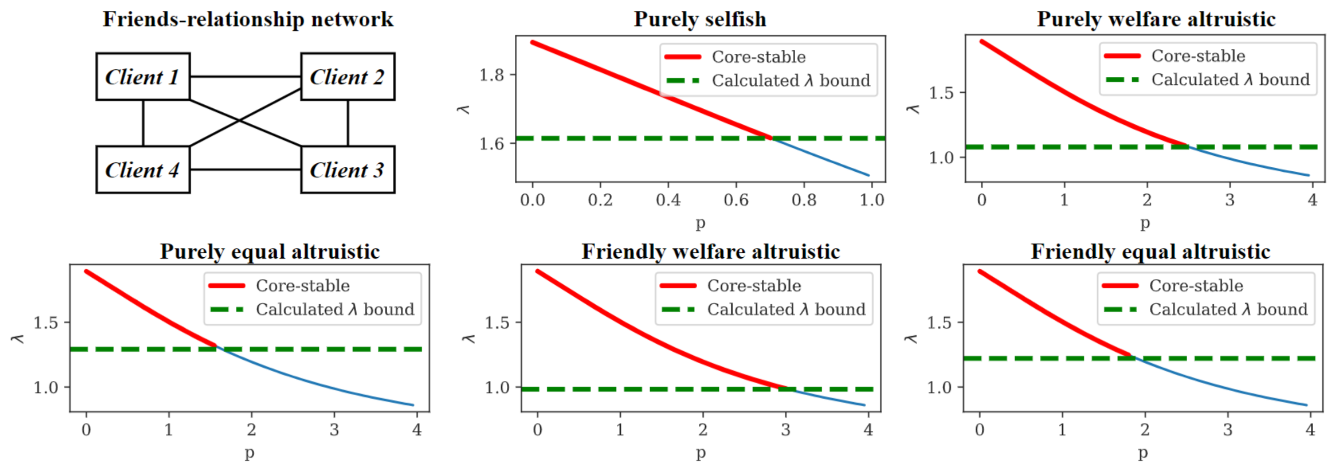
$$u_i^{fa}(\pi) = w\cdot v_i(\pi) + (1-w)\cdot\frac{1}{|F_i|+1}\sum_{f\in F_i\cup\{i\}} v_f(\pi)$$

Balanced by the selfishness degree parameter (w)

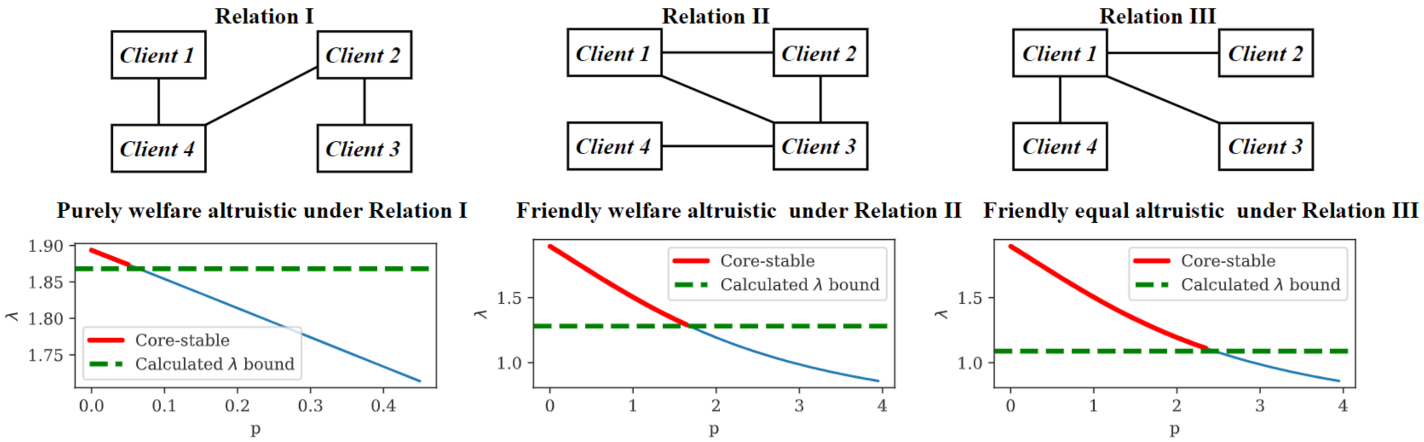
Evaluation

- Tightness validation

• Fully connected



• Partially connected



Theoretically derived egalitarian fairness bounds (green dashed line) align with empirically achieved egalitarian fairness within the core-stable grand coalition (red solid line) under different client behaviors.

Discussion and Limitations

Scalable Scenario 1： Heterogeneous behaviors

Example 1 *An example to calculate the achievable egalitarian fairness bound under heterogeneous behaviors is as follows: for a set of N clients, where clients $i \in \mathbf{C} = \{1, 2, ..., S\}$ act selfishly and the remaining act purely welfare altruistic, the achieved egalitarian fairness of π_g is bounded by,*

$$\lambda \geq \max_{\pi_s \subset \pi_g} \left\{ \max \left\{ \frac{{N_s}^2}{{N_g}^2} \cdot \frac{N_g \cdot n_l + d(\pi_g, n_m)}{N_s \cdot n_l + d(\pi_s, n_{k_{selfish}})}, \min \left(\frac{{N_s}^2}{{N_g}^2} \cdot \frac{N_g \cdot n_l + d(\pi_g, n_m)}{N_s \cdot n_l + d(\pi_s, f^{opt}_{altruistic,1})}, \frac{{N_c}^2}{{N_g}^2} \cdot \frac{N_g \cdot n_l + d(\pi_g, n_m)}{N_c \cdot n_l + d(\pi_c, f^{opt}_{altruistic,2})} \right) \right\} \right\},$$

where

$$k_{selfish} = \arg\min_{i \in \pi_s \cap \mathbf{C}} \{n_i\},$$
$$k_{altruistic,1} = \arg\min_{i \in \pi_s \setminus \mathbf{C}} \{\min_{f \in F_i \cap \pi_s} n_f\}, k_{altruistic,2} = \arg\min_{i \in \pi_s \setminus \mathbf{C}} \{\min_{f \in F_i \cap \pi_c} n_f\},$$
$$f^{opt}_{altruistic,1} = \arg\min_{f \in F_{k_{altruistic,1}} \cap \pi_s} n_f, f^{opt}_{altruistic,2} = \arg\min_{f \in F_{k_{altruistic,2}} \cap \pi_c} n_f. \tag{9}$$

Scalable Scenario 2： A broader class of utility functions in the form of generalized mean

$$u_i(\pi_g) = \left(\sum_{i=1}^{|F_i|} w_i err_i^q(\pi_g) \right)^{\frac{1}{q}}.$$

Scalability and Limitations

Limitation 1： More complex scenarios

- *more complex collaborative training tasks*
- *other notions of fairness*
- more complex client behavior

Limitation 2： Incentive Mechanisms

Limitation 3: Overfitting

Thank You!

Jiashi Gao

gaojs2021@mail.sustech.edu.cn

Southern University of Science and Technology
Shenzhen, China