# Towards Effective Clustered Federated Learning: A Peer-to-peer Framework with Adaptive Neighbor Matching

*Zexi Li, Chao Wu*
*Zhejiang University, Hangzhou 310007, China*

## Abstract

In federated learning (FL), clients may have diverse objectives, merging all clients' knowledge into one global model will cause negative transfers to local performance. Thus, **clustered FL** is proposed to group similar clients into clusters and maintain several global models. Nevertheless, current clustered FL algorithms **require the assumption of the number of clusters**, they are not effective enough to explore the latent relationships among clients. However, we take advantage of **peer-to-peer (P2P) FL**, where clients communicate with neighbors without a central server and propose an algorithm that enables clients to form an effective communication topology in a decentralized manner without assuming the number of clusters. Additionally, the P2P setting will release the concerns caused by the central server in centralized FL, such as reliability and communication bandwidth problems. Extensive experiments show that our method outperforms all P2P FL baselines and has comparable or even superior performance to centralized cluster FL. Moreover, results show that our method is much effective in mining latent cluster relationships under various heterogeneity without assuming the number of clusters and it is effective even under low communication budgets.

## Contributions

- We propose two efficient, effective, and privacy-preserving metrics to evaluate the pair-wise similarity of client objectives in P2P FL. They are based on losses and gradients, respectively.
- We present a novel P2P FL algorithm: **P**ersonalized **A**daptive **N**eighbour **M**atching (PANM), which enables clients to match neighbors with consistent objectives (same cluster identity), improving local performance. We devise two stages in PANM: confident neighbor initialization and heuristic neighbor matching based on **EM**.
- We conduct extensive experiments on a spectrum of Non-IID degrees and network settings, using different datasets. It is shown that PANM **outperforms all P2P baselines including Oracle** (with prior knowledge of cluster identities). Compared with centralized clustered FL algorithms, PANM is more effective in exploring latent cluster structure and has comparable, even better performance.
- Additionally, even under low communication budgets, PANM can still achieve superior performance to baselines.

## Personalized Adaptive Neighbour Matching

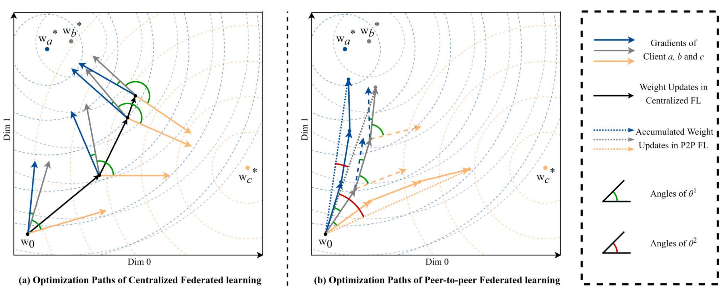


Fig 1. Schematic diagram of optimization paths in centralized FL (a) and P2P FL (b), respectively.

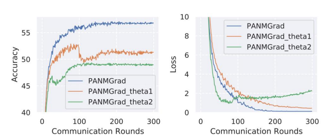


Fig 2. Ablation study of PANMGrad.

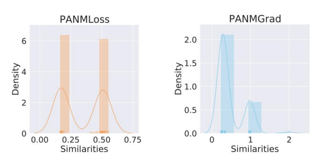


Fig 3. Distributions of similarities.

**Metrics for Measuring Client Similarity**: We develop an **efficient and effective** metric based on cosine similarity of **gradients** ($\theta^1$ in Fig 1) and cosine similarity of **accumulated weight updates** ($\theta^2$ in Fig 1). Two cosine functions are adopted in our metric; notably, the new metric is robust and effective in the P2P FL setting, we show the ablation study in Fig 2.

**Idea:** We solve the node clustering problem into a **binary classification problem**: from the perspective of the client-side, each client only needs to estimate an accessible client is whether in the same cluster as itself or not. Once the neighbor estimation is correct, a clustered communication topology will be inherently established without assuming the number of clusters. We devise a **two-stage** algorithm for P2P clustered FL. In the first stage, we enable clients to have few neighbors with **high precision** of being same-cluster, while in the second stage, we enable clients to match more neighbors with **high recall**.

***Stage 1:*** **Confident Neighbor Initialization (CNI)**
In the first stage of P2P FL training, clients have to initialize their collaborative neighbors from random sampled peers ($C^t_i$, $|C^t_i| = l$). In CNI, after the first round, we add the neighbors in the previous round ($N^{t-1}_i$) to the candidate list (N) in the current round, consequently, the confidence of same-cluster neighbors increases over round.

***Stage 2:*** **Heuristic Neighbor Matching (HNM)**
It is obvious that for a client, the same-cluster clients may have high similarities while the ones of the different-cluster are low, so we assume the similarities obey two distinct Gaussian distributions (shown in Fig 3), thus we can formulate it into a **Gaussian Mixture Model** (GMM) problem. We devise our HNM algorithm based on Expectation Maximization (EM). Knowing that EM algorithm is sensitive to initialization, with the prior knowledge that most of neighbors in stage 1 are same-cluster (it is also possible that it includes outliers), so we can initialize a better parameter.

## Experiments

We evaluate our methods and compare them with baselines. P2P FL baselines include **PENS** (state-of-the-art personalized P2P FL algorithm), Random (gossip with random neighbors), Local (without communication), FixTopology (neighbors are randomly sampled at the beginning and fixed during training), **Oracle** (with prior knowledge of cluster identities, gossip with same-cluster clients). Centralized FL baselines include **IFCA** (state-of-the-art centralized clustered FL) and centralized Federated Averaging. Our methods include **PANMLoss** (PANM with metric based on loss), and **PANMGrad** (PANM with metric based on weight updates and gradients). The clustered setting is followed by the centralized clustered FL literature, clusters of clients are generated by image rotation or swapping labels. Results show our methods can effectively enable clients to inherently form clusters (Fig 4); our methods are robust while varying trainset size and number of clients (Fig 5); our methods have superior performance compared with the centralized clustered FL with incorrect cluster number estimation and are comparable to the centralized with correct estimation (Table 1); even under low communication budgets, PANM can still achieve superior performance to baselines (Table 2).

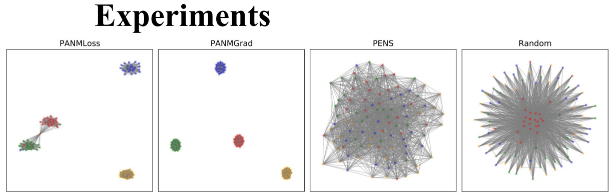


Fig 4. Neighbor topologies in stage two, compared with baselines. Each color denotes a cluster.

| Methods | CIFAR10(4) | FMNIST(4) | FMNIST(2) |
|---|---|---|---|
| Local | 25.27 ±1.21 | 76.24 ±0.22 | 76.24 ±0.22 |
| FedAvg | 37.03 ±0.74 | 83.54 ±0.08 | 86.86 ±0.16 |
| IFCA(c=2) | 40.64 ±2.18 | 86.19 ±0.04 | **88.06 ±0.20** |
| IFCA(c=3) | 41.05 ±1.09 | **86.78 ±0.36** | / |
| IFCA(c=4) | **43.65 ±0.77** | 86.50 ±0.07 | / |
| Oracle | 43.32 ±0.85 | 85.45 ±0.38 | 87.01 ±0.26 |
| PENS | 36.64 ±0.58 | 84.68 ±0.27 | 86.82 ±0.11 |
| PANMLoss | 41.43 ±1.83 | **86.09 ±0.31** | **87.33 ±0.17** |
| PANMGrad | **43.99 ±1.26** | 85.64 ±0.25 | 86.88 ±0.34 |

Table 1. Comparison with centralized FL.

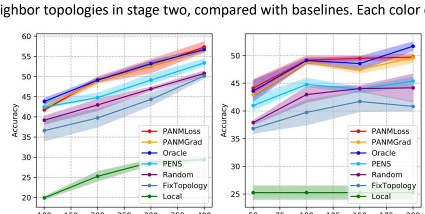


Fig 5. Left: Accuracies when changing trainset size.
Right: Accuracies when changing number of clients.

| Methods | Comm. costs | Max. req. band. | Test acc. |
|---|---|---|---|
| FedAvg | 600Δ | 1Δ | 37.03 ±0.74 |
| IFCA ($c$=2) | 900Δ | 2Δ | 40.64 ±2.18 |
| IFCA ($c$=3) | 1200Δ | 3Δ | 41.05 ±1.09 |
| IFCA ($c$=4) | 1500Δ | 4Δ | **43.65 ±0.77** |
| PANMLoss ($k$=2) | 1118Δ | 0.06Δ | 41.36±0.64 |
| PANMGrad ($k$=2) | 1118Δ | 0.06Δ | 42.78±1.68 |
| PANMLoss ($k$=3) | 1397Δ | 0.07Δ | 43.30±1.32 |
| PANMGrad ($k$=3) | 1397Δ | 0.07Δ | **43.34±0.85** |

Table 2. Performance under low communication budgets.

*Please contact us if you are interested in our work.*
*Email: zexi.li@zju.edu.cn, chao.wu@zju.edu.cn*