2023-06130 – PhD Position F/M Topology Design for Decentralized Federated Learning

Contract type: Fixed-term contract
Level of qualifications required: Graduate degree or equivalent
Function: PhD Position

About the research centre or Inria department

The Inria centre at Université Côte d'Azur includes 37 research teams and 8 support services. The centre's staff (about 500 people) is made up of scientists of different nationalities, engineers, technicians and administrative staff. The teams are mainly located on the university campuses of Sophia Antipolis and Nice as well as Montpellier, in close collaboration with research and higher education laboratories and establishments (Université Côte d'Azur, CNRS, INRIA, INSERM -...), but also with the region's economic players.

With a presence in the fields of computational neuroscience and biology, data science and modeling, software engineering and certification, as well as collaboration and coordination, the Inria Centre at Université Côte d'Azur is a major player in terms of scientific excellence through its results and collaborations at both European and international levels.

Context

This PhD thesis is in the framework of Inria-research initiative on Federated Learning, FedMalin https://project.inria.fr/fedmalin/

The PhD candidate will join NEO project-team https://team.inria.fr/neo/

NEO is positioned at the intersection of Operations Research and Network Science. By using the tools of Stochastic Operations Research, the team models member model situations arising in several application domains, involving networking in one way or the other.

The research activity will be supervised by
- Aurelien Bellet, http://researchers.lille.inria.fr/abellet/

Assignment

# Topology Design for Decentralized Federated Learning

## Context

The increasing size of data generated by smartphones and IoT devices motivated the development of Federated Learning (FL) [li20, kairouz21], a framework for on-device collaborative training of machine learning models. FL algorithms like FedAvg (mcmahan17) allow clients to train a common global model without sharing their personal data: FL reduces data collection costs and can help to mitigate data privacy issues, making it possible to train models on large datasets that would otherwise be inaccessible. FL is currently used by many big tech companies (e.g., Google, Apple, Facebook) for learning from their users' data, but the research community is also exploring applications to learning across large data silos, like hospitals that cannot share their patients' data [rieke20].

In the classic FL setting, a server coordinates the training phase. At each training round, the server sends the current model to the clients, which individually train on their local data and send model updates to the server, which in turn aggregates them (often through a simple averaging operation). In contrast to this client-server approach, decentralized FL algorithms (also called POP FL algorithms) work by having each client communicate directly with a subset of the clients (its neighbors): this process alternates between model updates and weighted averaging of the neighbors' models (consensus-based optimization). Decentralized algorithms can take advantage of good pairwise connectivity, avoid the potential communication bottleneck at the server [marfoq20] as well as provide better privacy guarantees [cyffers22].

The communication graph (i.e., the graph induced by clients' pairwise communications) and the local clients' aggregation strategies play a fundamental role in determining FL algorithms' convergence speed. In particular, the communication topology has two contrasting effects on training time. First, a more connected topology leads to faster convergence in terms of number of communication rounds (nedic18). Second, a more connected topology increases the duration of a communication round (e.g., because it may cause network congestion), motivating the use of degree-bounded topologies where every client sends and receives a small number of messages at each round (ianu17). Most of the existing literature has focused on one aspect or the other.

The classic literature on consensus-based optimization has quantified the effect of the communication topology on the number of rounds through worst-case convergence bounds in terms of the spectral gap of the consensus matrix (i.e., the matrix with the averaging weight), see [nedic18] and references there. Later papers have highlighted the convergence rate's insensitivity to the spectral gap for a large number of communication rounds and small learning rates [ianu17, koloskova20, pu20].

Another line of work has shown that the effect of the topology is less important if local data distributions [neglia20] or average data distributions in each neighborhood [lebris22, dardik22] are close to the average data distribution over the whole population. In the extreme case of homogeneous local distributions, one may even prefer consensus matrices with poor spectral properties because they enable the use of larger learning rates [vogel22].
A separate line of works has studied how to design the communication topology in order to minimize the duration of one round, taking into account the variability of the computation times [neglia19] or the characteristics of Internet connections [marfoq20].

## Research objectives

The goal of this PhD is to propose algorithms to design the communication topology for decentralized federated learning with the goal of minimizing the total training duration, taking into account how connectivity affects both the number of rounds required and the duration of a single round. Several settings will be considered: in particular, one may construct the topology in a pre-processing step (prior to learning), or dynamically while learning. Dynamic topology design can be a way to tackle online decentralized learning [asadi22,marfoq20], where the topology is adjusted and refined as clients collect more data. The candidate will also investigate how to practically quantify the similarity of local data distributions during training in order to exploit the advantage of having a neighborhood representative of the average population distribution [lebars23,dandi22]. Finally, he/she will also study to what extent the existing results can be extended to asymmetric communication links and other distributed optimization algorithms like push-sum ones [kempe03,benezit10].

## References


Main activities
Research.

Skills
The candidate should have a solid mathematical background (in particular on optimization) and in general be keen on using mathematics to model real problems and get insights. He should also be knowledgeable on machine learning and have good programming skills. Previous experiences with PyTorch or TensorFlow is a plus.
We expect the candidate to be fluent in English.

Benefits package
- Subsidized meals
- Partial reimbursement of public transport costs
- Leave: 7 weeks of annual leave + 10 extra days off due to RTT (statutory reduction in working hours) + possibility of exceptional leave (sick children, moving home, etc.)
- Possibility of teleworking and flexible organization of working hours
- Professional equipment available (videoconferencing, loan of computer equipment, etc.)
- Social, cultural and sports events and activities
- Access to vocational training
- Contribution to mutual insurance (subject to conditions)

Remuneration
Duration: 36 months
Location: Sophia Antipolis, France
Gross Salary per month: 2051€ brut per month (year 1 & 2) and 2158€ brut per month (year 3)