Internship: Topology Design for Decentralized Federated Learning

- **Research teams:** NEO (https://team.inria.fr/neo/) and PreMeDICaL (https://team.inria.fr/premedical/)
- **Supervisors:** Giovanni Neglia (giovanni.neglia@inria.fr)
  Aurélien Bellet (aurelien.bellet@inria.fr)
- **Location of the internship**: Inria Sophia-Antipolis
- **Keywords:** machine learning, federated learning, decentralized optimization

1 Context

The increasing size of data generated by smartphones and IoT devices motivated the development of Federated Learning (FL) [3], a framework for on-device collaborative training of machine learning models. FL algorithms 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 on their users' data, but the research community envisions also promising applications to learning across large data-silos, like hospitals that cannot share their patients' data.

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 datasets 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 P2P FL algorithms) work by having each client communicate directly with a subset of the clients (its neighbours): this process alternates between model updates and weighted averaging of the neighbours' models (consensus-based optimization). Decentralized algorithms can take advantage of good pairwise connectivity, avoid the potential communication bottleneck at the server [6] as well as provide better privacy guarantees [2].

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 [7]. 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 [5]. Most of the existing literature has focused on one aspect or the other, and few approaches exist to learn the communication topologies from data.

2 Objectives of the internship

The goal of this internship 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 affect both the number of rounds required and the duration of a single round.

The internship will start with a review of the recent literature on the impact of the communication topology on decentralized federated learning (see e.g., [6, 8, 9, 4]). Then, one or several of the following questions will be considered: (i) how to construct the topology in a pre-processing step (prior to learning)? (ii) how to construct the topology dynamically while learning, refining it as clients collect more knowledge or data? (iii) how to practically quantify the similarity of local data distributions during training in order to design data-driven topologies? and (iv) how to extend previous approaches to asymmetric communication links and other distributed optimization algorithms like push-sum ones [1]?

3 Expected 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/she should also be knowledgeable on machine learning and have good programming skills. Previous experiences with PyTorch or TensorFlow is a plus.

References

[1] M. Assran, N. Loizou, N. Ballas, and M. Rabbat. Stochastic gradient push for distributed deep learning. In *International Conference on Machine Learning*, pages 344–353. PMLR, 2019.

[2] E. Cyffers, M. Even, A. Bellet, and L. Massoulié. Muffliato: Peer-to-Peer Privacy Amplification for Decentralized Optimization and Averaging. In *NeurIPS*, 2022.

[3] P. Kairouz, H. B. McMahan, B. Avent, A. Bellet, M. Bennis, A. N. Bhagoji, K. Bonawitz, Z. Charles, G. Cormode, R. Cummings, et al. Advances and open problems in federated learning. *Foundations and Trends® in Machine Learning*, 14(1–2):1–210, 2021.

[4] B. Le Bars, A. Bellet, M. Tommasi, E. Lavoie, and A.-M. Kermarrec. Refined Convergence and Topology Learning for Decentralized SGD with Heterogeneous Data. In *AISTATS*, 2023.

[5] X. Lian, C. Zhang, H. Zhang, C.-J. Hsieh, W. Zhang, and J. Liu. Can decentralized algorithms outperform centralized algorithms? a case study for decentralized parallel stochastic gradient descent. *NeurIPS*, 2017.

[6] O. Marfoq, C. Xu, G. Neglia, and R. Vidal. Throughput-optimal topology design for cross-silo federated learning. In *NeurIPS*, 2020.

[7] A. Nedic, A. Olshevsky, and M. G. Rabbat. Network topology and communication-computation tradeoffs in decentralized optimization. *Proc. IEEE*, 106(5):953–976, 2018.

[8] G. Neglia, C. Xu, D. Towsley, and G. Calbi. Decentralized gradient methods: does topology matter? In *AISTATS*, 2020.

[9] T. Vogels, H. Hendrikx, and M. Jaggi. Beyond spectral gap: the role of the topology in decentralized learning. In *NeurIPS*, 2022.