Dynamic Topology Optimization for Non-IID Data in Decentralized Learning

Antreas Ioannou

Delft University of Technology

TU Delft

1. Motivation

Decentralized Learning (DL) enables multiple nodes to collaboratively train a model without a central server [1,2]. In fully decentralized DL, nodes communicate only with their neighbors in a network topology, without any coordinator.

A key challenge arises when data is **non-IID**: static communication structures limit learning quality [3–5]. Existing methods often assume fixed topologies or rigid strategies, which struggle when node connectivity is suboptimal [5].

Our work addresses this gap by introducing a guided, fully decentralized algorithm **(DissDL)** that dynamically adapts the communication graph in response to non-IID data.

2. Research Question

Can fully decentralized, dissimilarity-guided topology adaptation improve convergence and model performance in decentralized learning under non-IID data?

Sub-Questions:

- **RQ1:** Is model dissimilarity an effective criterion for guiding dynamic topology adaptation under non-IID data?
- **RQ2:** Can nodes effectively discover and select diverse peers using only local and peer-propagated information, without any global knowledge?
- **RQ3:** How does our algorithm perform in terms of convergence speed, final accuracy, and communication efficiency compared to static and non-static decentralized baselines?

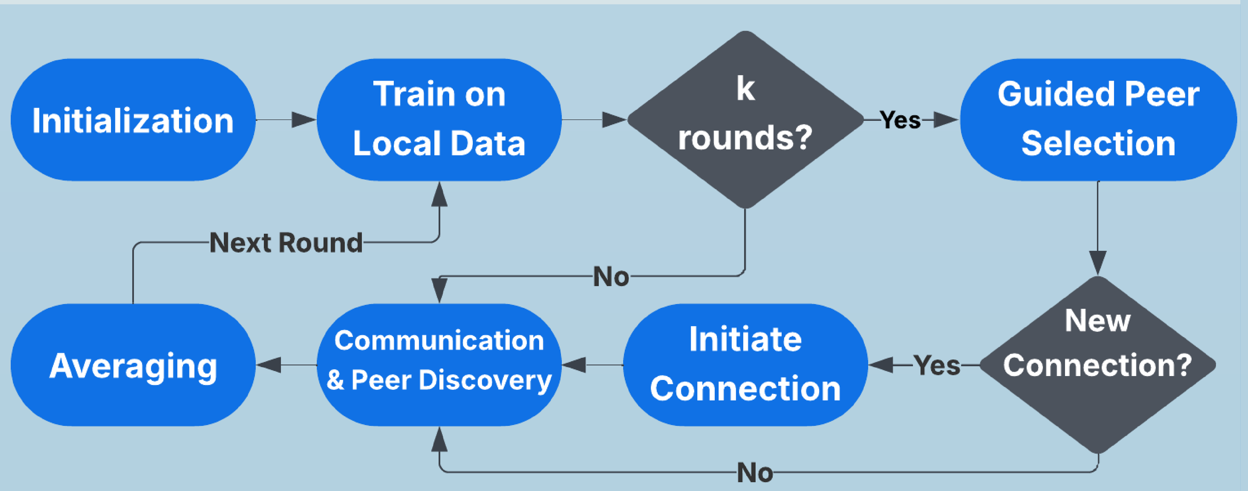

Fig. 1: Workflow of each node in the training process. The cycle runs for a predetermined number of rounds, with periodic topology adaptation.

3. DissDL

We propose **DissDL**, a **fully decentralized** algorithm where each node dynamically selects peers based on **model dissimilarity**, without any global knowledge required.

Communication Cycle:

- **Intent Phase:** Each node requests model updates from selected peers.
- **Exchange Phase:** Nodes only share models with those who requested them.
- **Aggregation:** Uniform averaging of received models and the local one.
- **Peer discovery via gossip**: Nodes exchange known peer lists during communication.

Topology Adaptation (every k rounds):

- Nodes estimate **cosine dissimilarity** between models.
  - If a peer's model is available, compute cosine similarity layer-wise and average.
  - If not, estimate it indirectly via **gossip-based similarity propagation** using other peers' reports.
- Each node uses **Softmax sampling** to probabilistically :
  - **Add** one new **dissimilar** peer.
  - **Remove** one existing **similar** peer.

Why It Works:

- Boosts exposure to **diverse data** under non-IID conditions, improving global generalization.
- Maintains a **fixed in-degree**, preventing node isolation.
- Enables **peer discovery** via gossip, expanding local network awareness.
- Fully **decentralized:** no central coordination or global topology required.
- Improves **accuracy, convergence**, and **training stability** adaptively.

4. Results

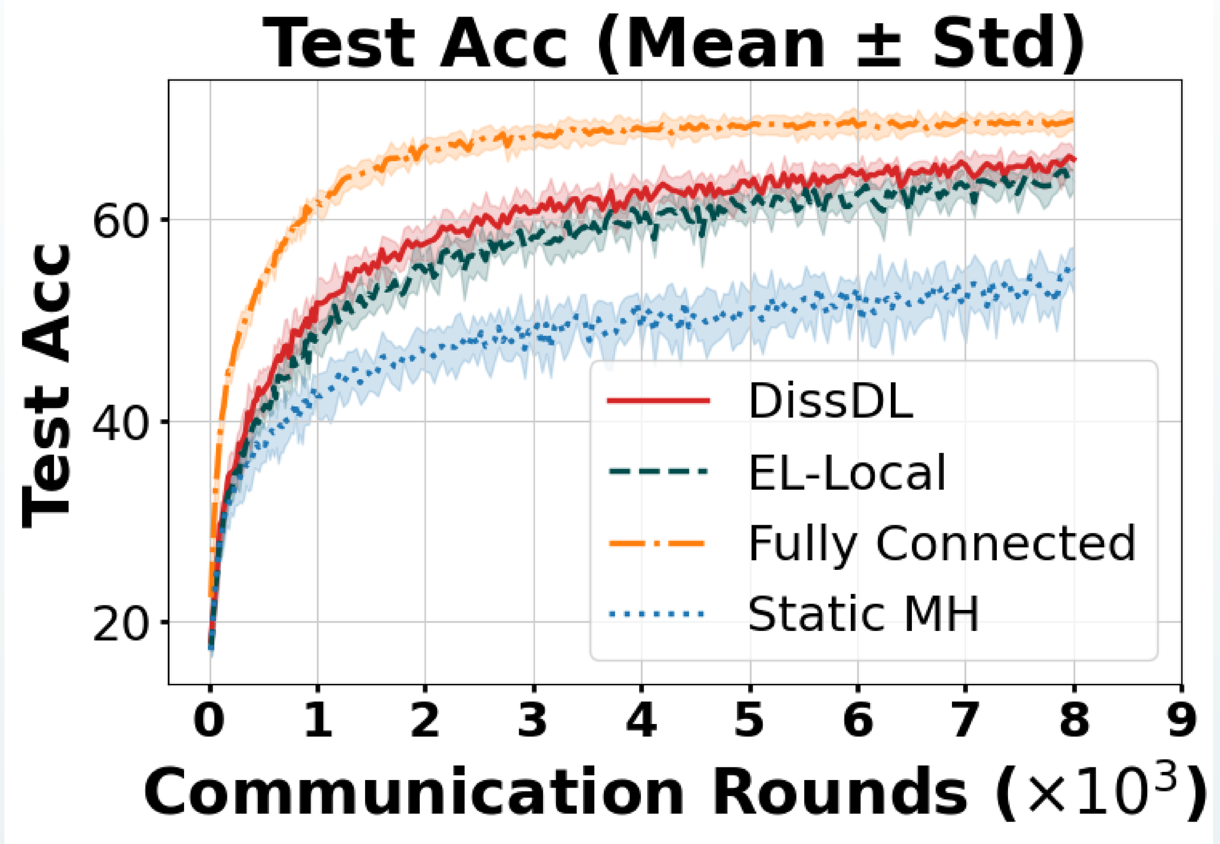

Fig. 2: Average accuracy progression across communication rounds

1. **16 node Decentralized Learning:**
   - Fully Connected: the upper bound (high communication)
   - EL-Local [6]: randomized connections (improves information mixing)
   - Static MH: 3-regular static topology
2. **CIFAR-10 non-IID Partitioning:**
   - Dirichlet Distribution (α = 0.1)
3. **GN-LeNet with SGD.**

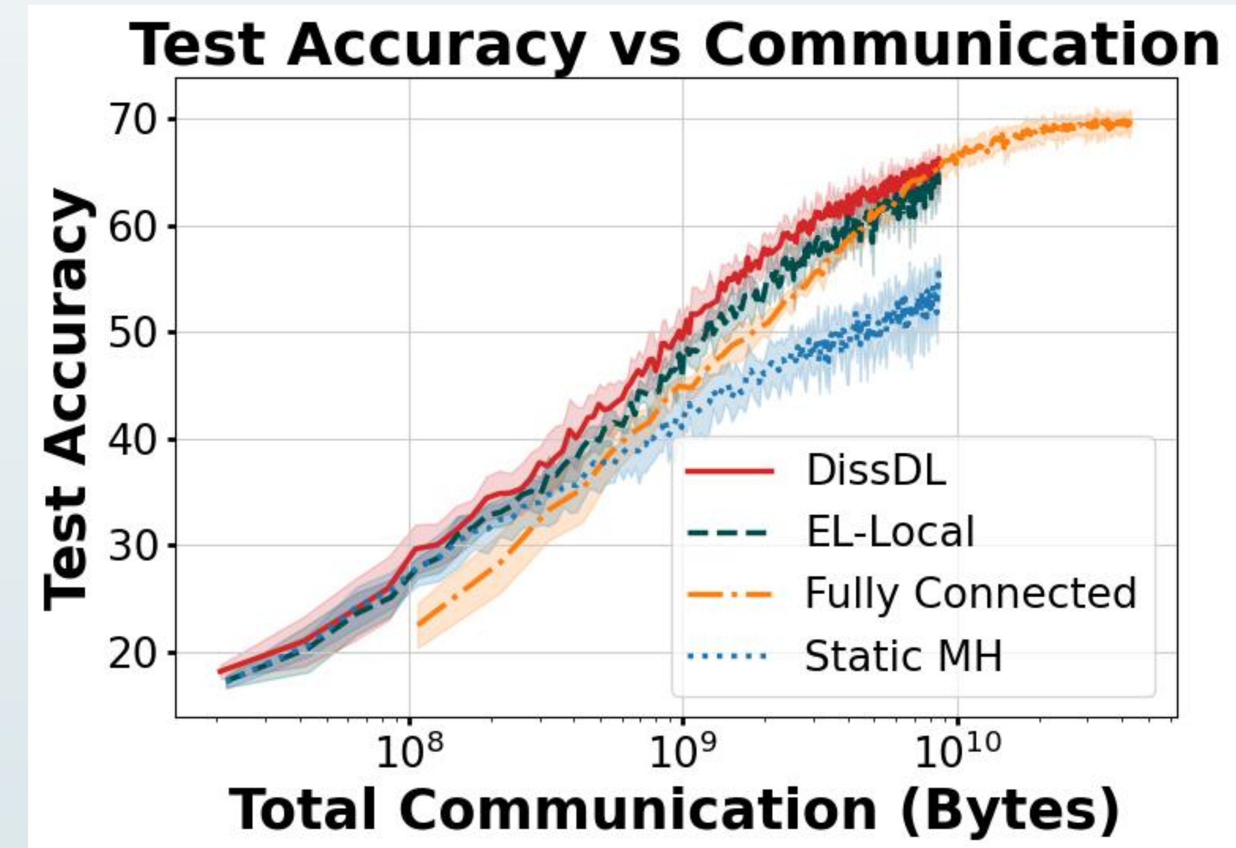

Fig. 3: Average accuracy progression across total communication overhead (in bytes)

DissDL:

- Converges the fastest among sparse methods
- Achieves higher accuracy across most communication budgets
- Matches EL-Local's highest accuracy with **1.34×** less communication
- Outperforms all sparse methods in final accuracy (**+1.44%** over EL-Local)

5. Conclusion & Limitations

- DissDL **improves accuracy and convergence** in non-IID decentralized learning.
- **Informed peer selection** can significantly boost learning quality in decentralized settings.
- **Model dissimilarity** is a promising criterion for **adaptive topology optimization** in decentralized learning, especially under non-IID data conditions.
- **Limitation:** Some nodes may receive excessive requests due to out-degree not being controlled, leading to potential bottleneck.

References

[1] Y. Lu and C. D. Sa. *Optimal Complexity in Decentralized Training*. ICML, 2021.
[2] X. Lian et al. *Can Decentralized Algorithms Outperform Centralized Algorithms?* NeurIPS, 2017.
[3] X. Gao et al. *Semantic-Aware Node Synthesis for Imbalanced HINs*. CIKM, 2023.
[4] B. Bars et al. *Refined Convergence and Topology Learning for Decentralized SGD*. AISTATS, 2023.
[5] K. Hsieh et al. *The Non-IID Data Quagmire of Decentralized Machine Learning*. ICML, 2020.
[6] M. De Vos et al. *Epidemic Learning: Boosting Decentralized Learning with Randomized Communication*. NeurIPS, 2023.

aioannou@tudelft.nl