# Decentralized Federated Learning Over Edge Networks: a Coordination-Free Learning Substrate for Agentic AI

**Lorenzo Valerio, PhD**

**CNR-IIT**

**lorenzo.valerio@iit.cnr.it**

ISTITUTO
DI INFORMATICA
E TELEMATICA

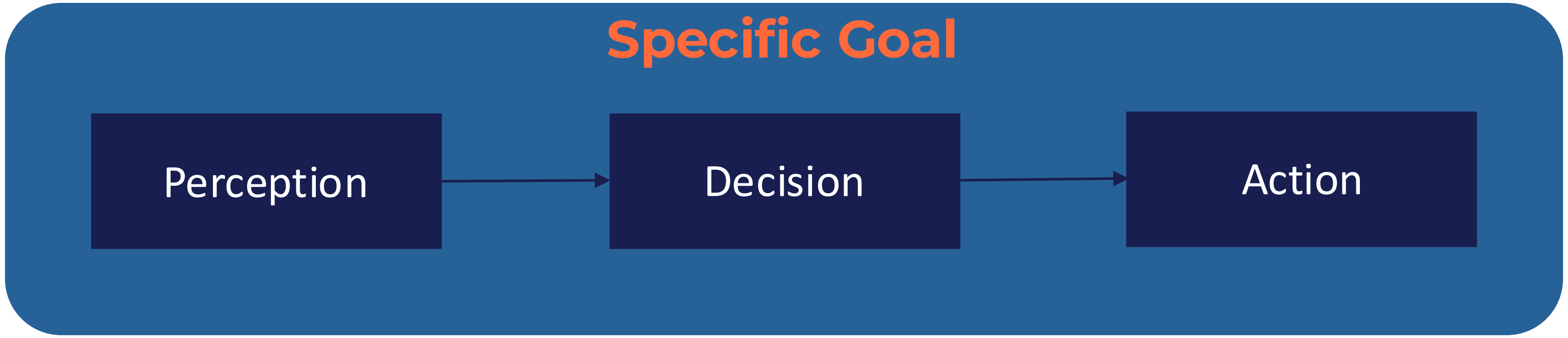
Consiglio Nazionale
delle Ricerche

# What is agentic AI: Definition

- Agentic AI is an **artificial intelligence system** that can accomplish a **specific goal** with **limited supervision**.
- It **consists** of **AI agents**
  - machine learning models that mimic human decision-making to solve problems in real time.
- In a **multiagent system**, each agent can perform a **specific subtask** required to reach the goal and their efforts are coordinated through AI orchestration

# Agentic AI: How do they do it?

Agentic AI systems can perceive, decide, and act to accomplish goals with minimal human intervention.



Key shift vs. traditional AI: closed-loop autonomy

# Autonomous agents: characteristics

- **Autonomy**: self-governing decision making
- **Reactivity**: timely response to environmental changes
- **Proactivity**: takes initiative to pursue goals
- **Goal-oriented behavior**: plan → execute → monitor → adapt
- **Social ability**: communicate, coordinate, negotiate

# Autonomous agents: architectural styles

**Reactive** Direct perception → action, minimal state (fast)

**Deliberative** Internal state + planning to reach goals

**Hybrid** Layered reactive + deliberative control

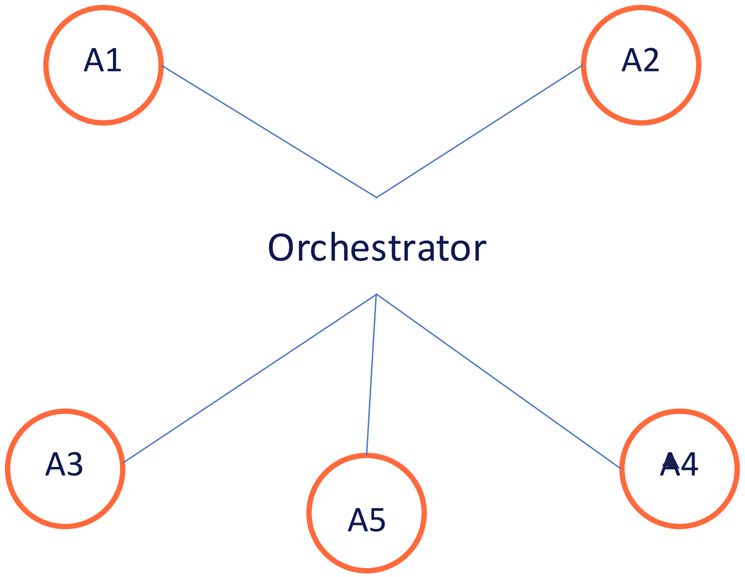
# Elements of an Agentic AI as a multi-agent system

**Coordination**: multiple agents work toward common or individual goals

**Communication** & **negotiation**: share information, allocate tasks, resolve conflicts

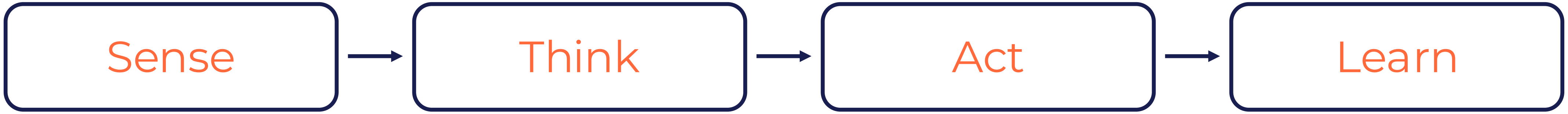
**Emergent behavior**: system-level patterns from local interactions

**Distributed AI**: decentralized control + local rules



# Moving Agentic AI @ the edge...

Considering:



Limited **compute** capability

Intermittent **connectivity**

**Lack of coordination**

**Locality**: fragmented local views

**RQ:** How do agents share and improve their internal models without a central coordinator?

# Decentralized federated learning

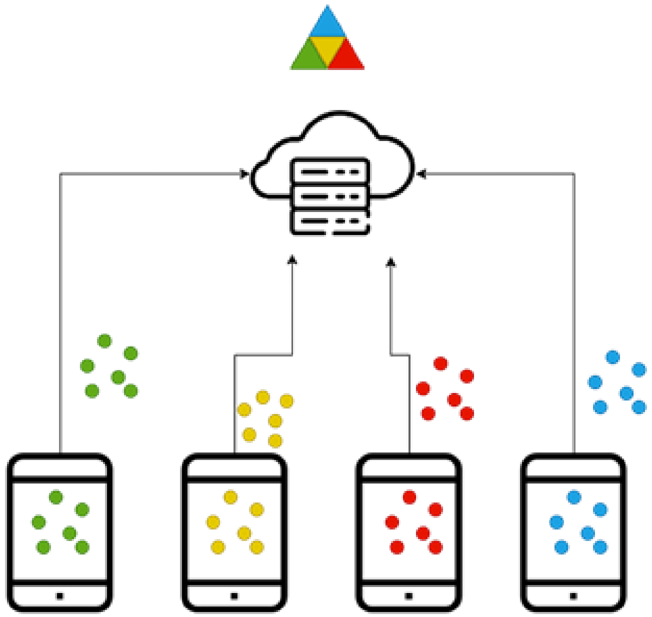
(as a coordination free learning layer)

The **goal** of **DFL @ the edge** is to go…

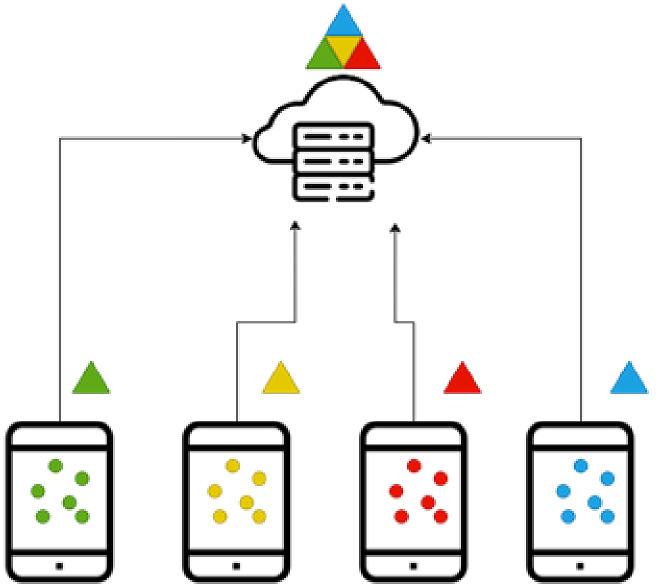
from this

to this

# AI network architectures



THE PAST



THE PRESENT

# Benefits of federated learning

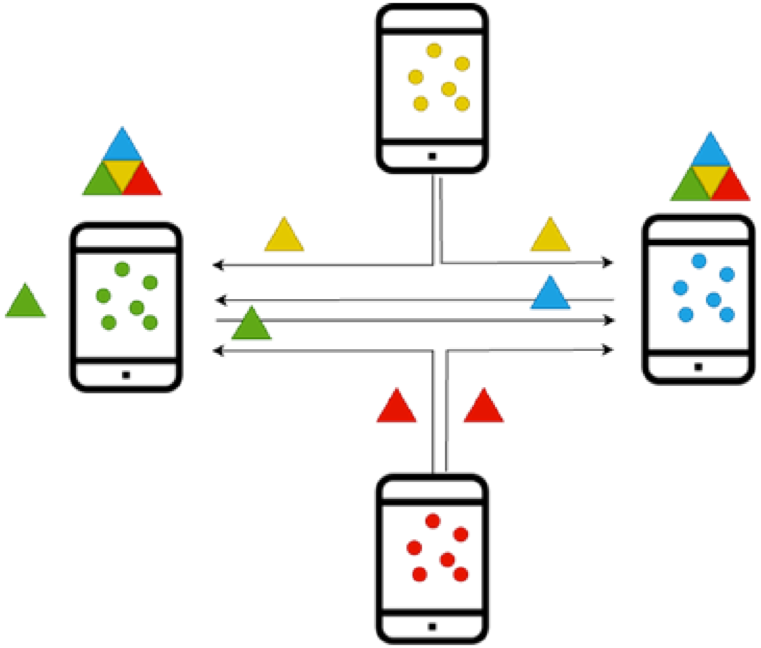
- **Privacy preservation**
- **Data security**
- **Collaboration without data sharing**
- Efficient data utilization
- Reduced communication costs
- Increased scalability

# Still, you need the central server to orchestrate the learning process

- Communication overhead
- Network dependency
- Centralized control and governance

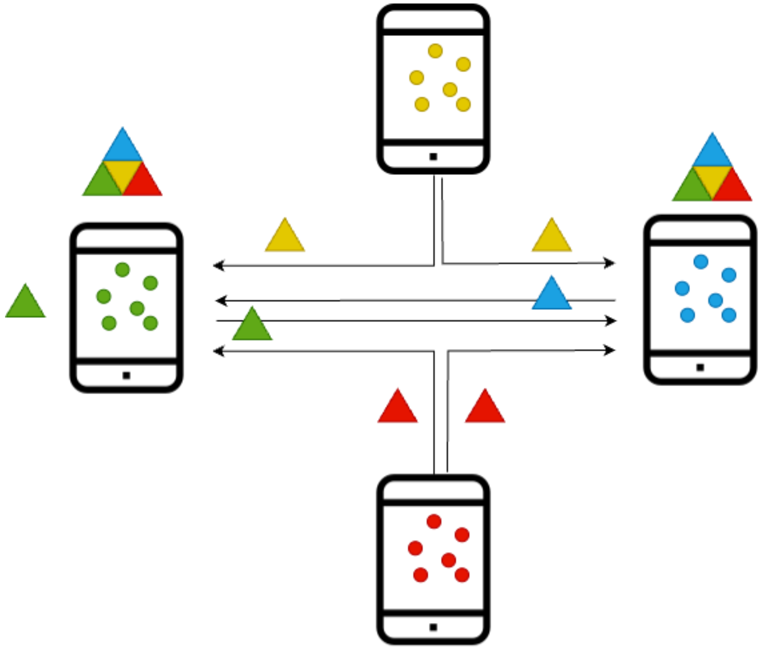
# Fully decentralized learning

Let’s get rid of the central controller, then!



THE FUTURE

# Challenges of fully decentralized learning

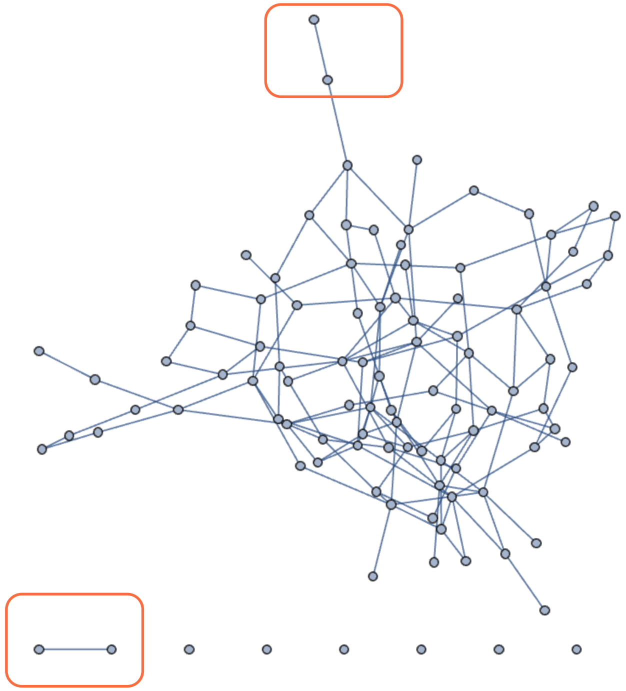


- **Peer-to-Peer Communication →** how information flows becomes critical for learning, which is a by-product of the **graph topology** connecting nodes
  **[Palmieri et al., 2024] [Palmieri et al., 2023]**
- **Data localization →** data partitioned across devices, usually in a **non-IID** way, issues of small data
  **[Ahmad et al., 2025]**
- **Resilience of collaborative learning →** issues of **trust**, low-quality data, malicious nodes
  **[Sabella et al., 2025]**
- **No centralized control →** lack of **coordination**
  **[Valerio et al., 2023] [Badie-Modiri et al., 2024]**
- **On-device learning →** address the **resource constraints**
  **[Valerio et al., 2022]**
- **Local decision making →** device and model **heterogeneity**, local resources limited

# Research direction #1:

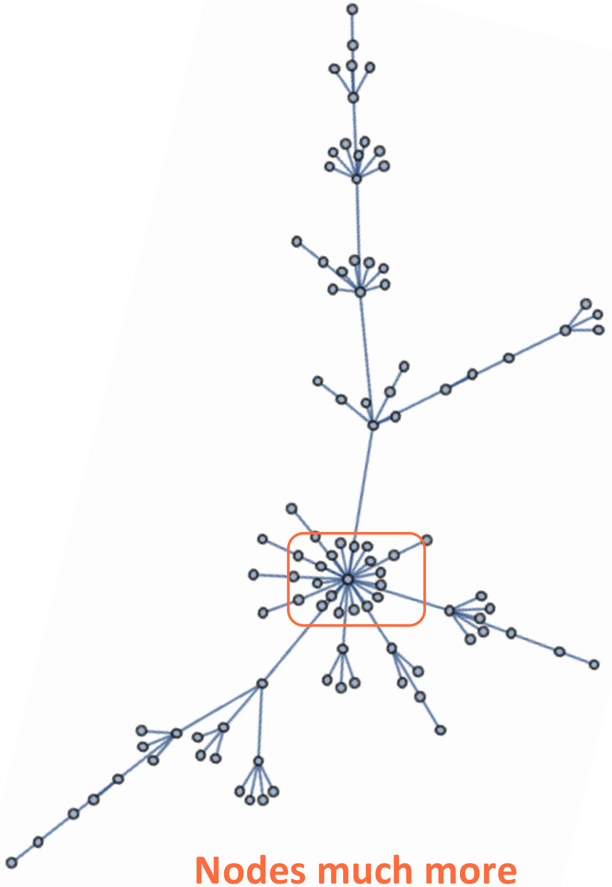
# What is the effect of different network topologies on the accuracy of decentralized learning?

# Why it is a crucial problem



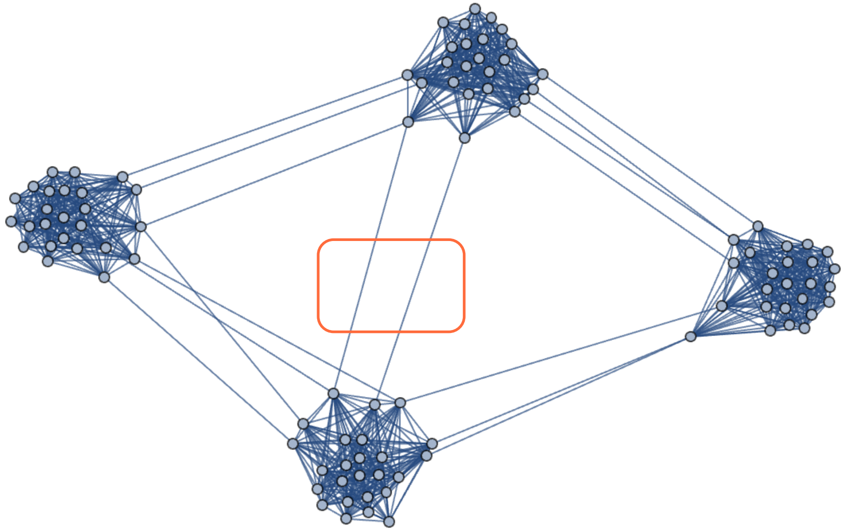
**Poorly connected nodes**

Erdős–Rényi graph



Barabási–Albert graph

**Nodes much more “important” than others**



Stochastic Block Model

**Communities that are well connected inside and poorly connected outside**

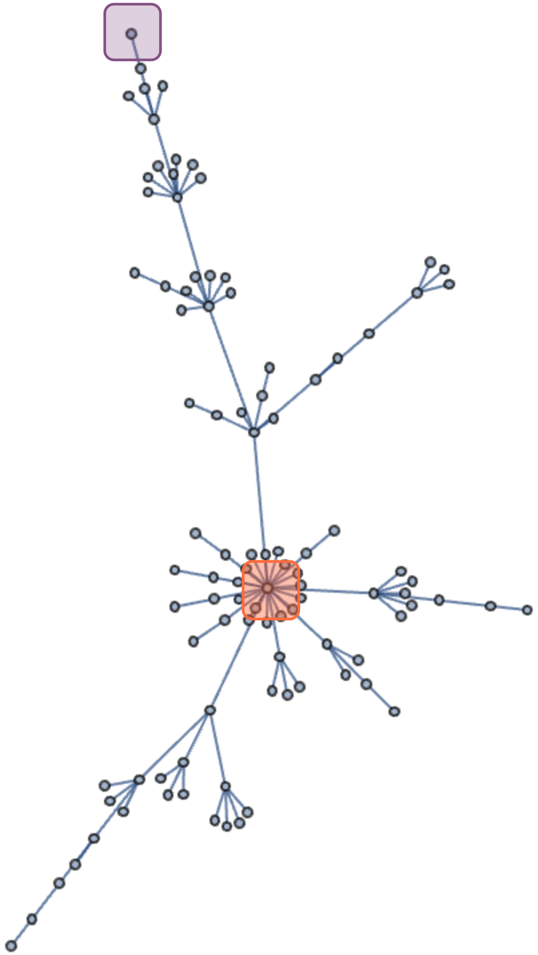
# Experiments

- Data allocation
  - 10 MNIST classes divided into two groups

G1

G2

  - **All nodes** receive an **equal share** (selected randomly) of data from **G1.**
  - Data from **G2 are allocated only to the 10%** highest-degree vs lowest-degree nodes

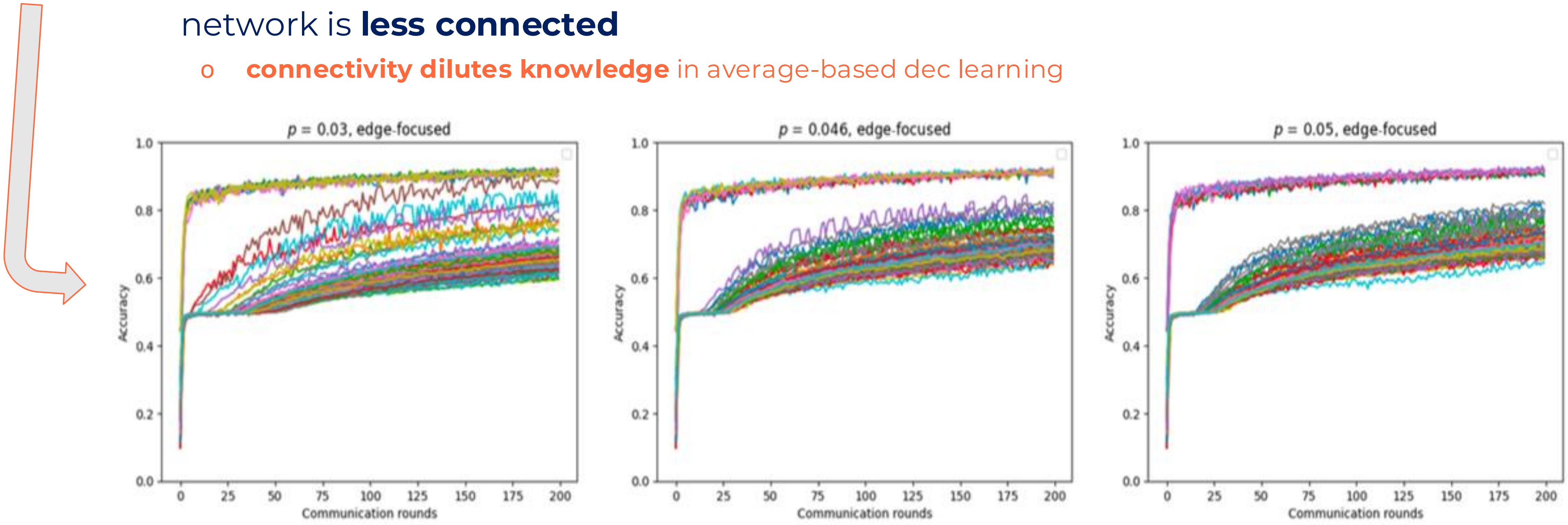


- Three topologies were considered (all with 100 nodes):
  - Erdős–Rényi, Barabási–Albert, Stochastic block model
- Model aggregation strategy: simple averaging

# Main results

- the **initial data distribution** on high vs low-degree nodes plays a **key role**

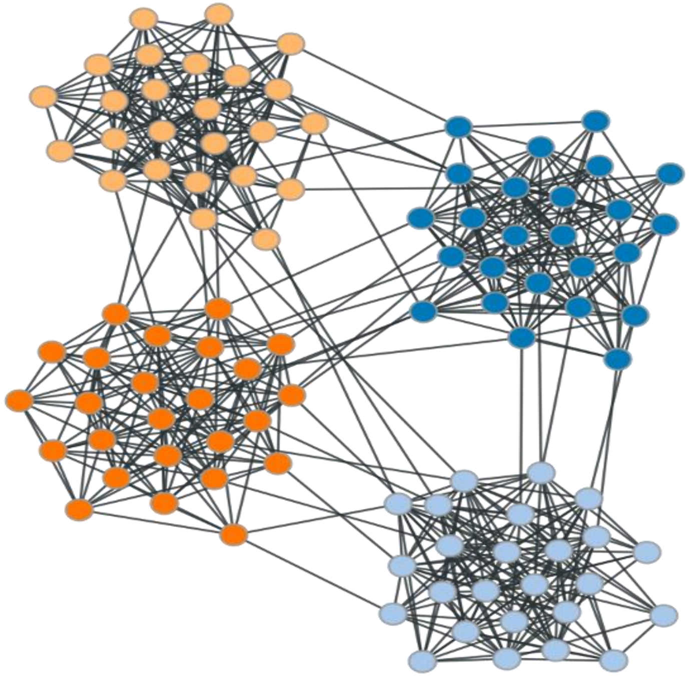
- when low-degree nodes have more knowledge, knowledge spreads **better** when the network is **less connected**
  - **connectivity dilutes knowledge** in average-based dec learning

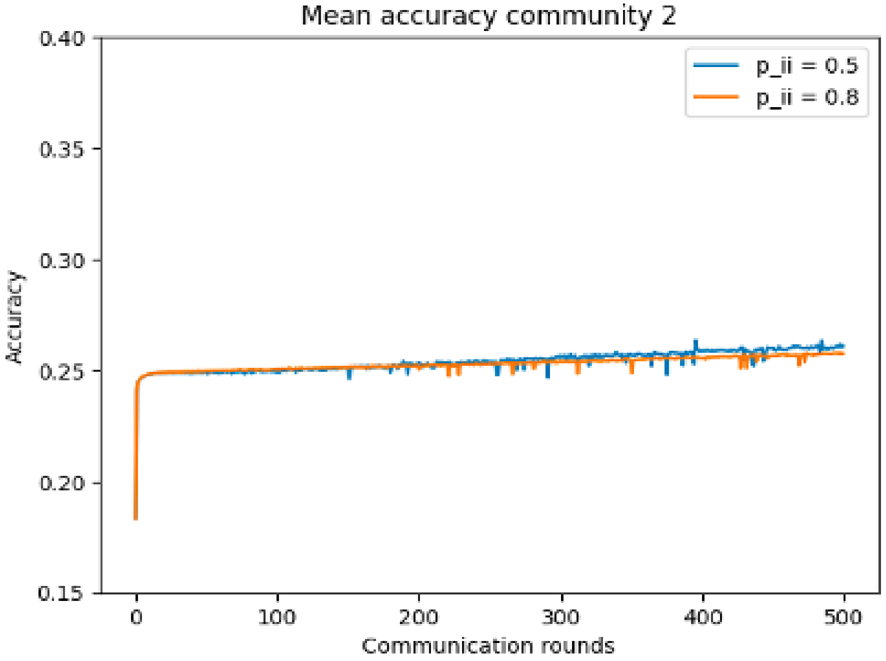


Palmieri, L., Boldrini, C., Valerio, L., Passarella, A., & Conti, M. (2024). Impact of network topology on the performance of Decentralized Federated Learning. Computer Networks 2024.

# Main results

- the **initial data distribution** on high vs low-degree nodes plays a **key role**
- when users are grouped in tightly knit communities, it is very **difficult for knowledge to circulate outside** of the community





Palmieri, L., Boldrini, C., Valerio, L., Passarella, A., & Conti, M. (2024). Impact of network topology on the performance of Decentralized Federated Learning. Computer Networks 2024.

# Selection strategy counts too

- **Betweenness centrality** → How much the node bridges together distant part of the network
- **Degree centrality** → How much connections does the node have

Global

- **Clustering coefficient** → How much the node is influential within its neighbors

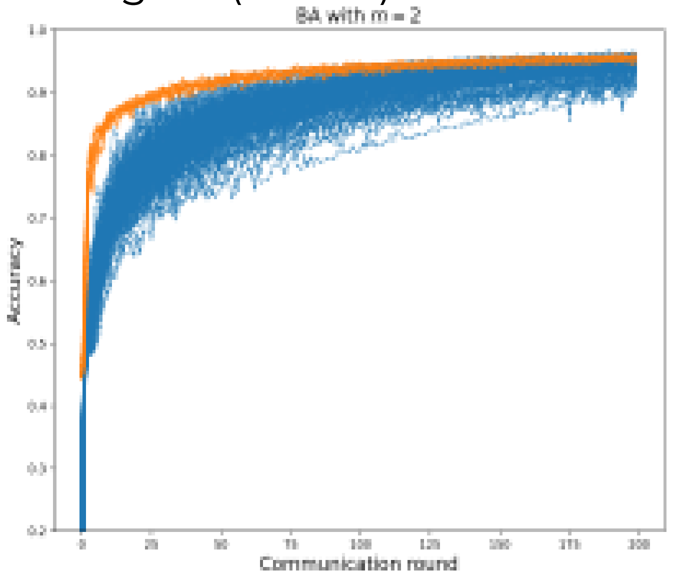
Local

# Selection metric counts too

Information flows better when more data is given to nodes that are **globally** more influencial.

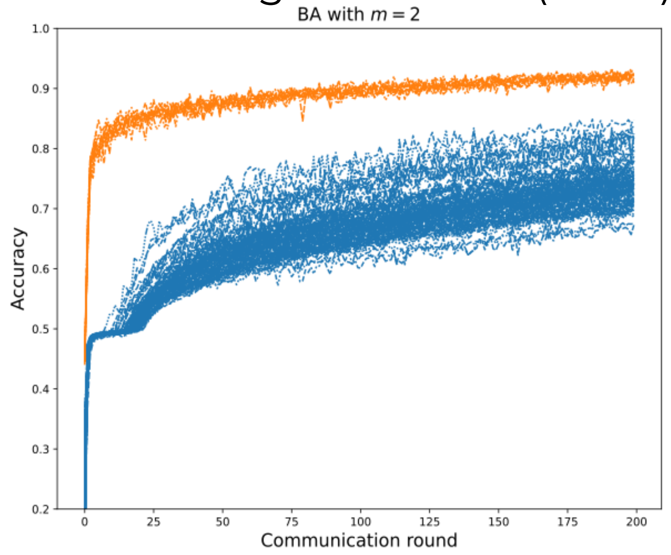
Highest-focus

Degree (Global)



VS

Clustering coefficient (Local)

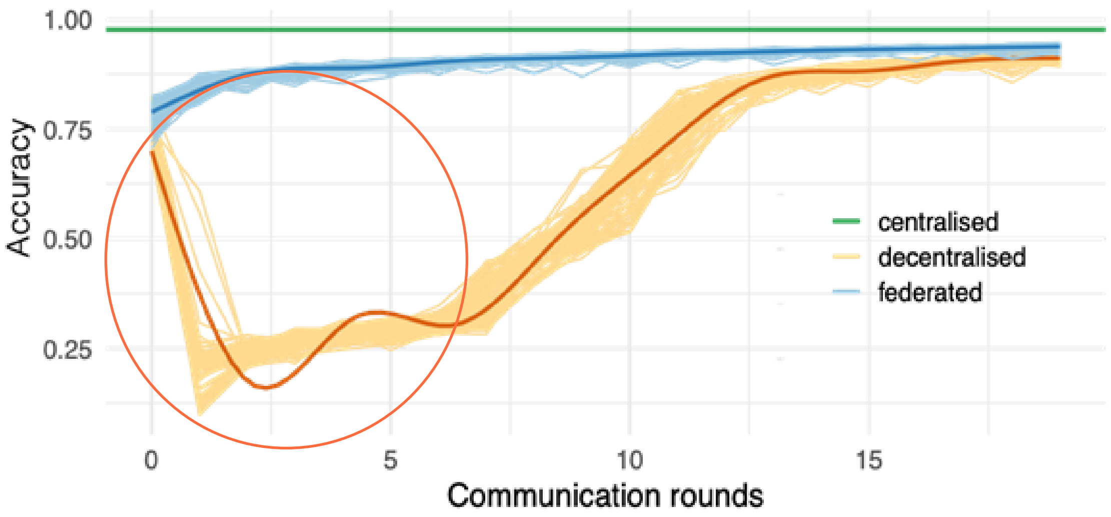


Orange curves: nodes with more data

Blue curves: all the other nodes in the network

# Research direction #2: Enhancing local AI models through decentralized collaboration

# Aggregation is not easy w/o coordination



**The accuracy drops instead of increasing!**

## Why the drop? It's the lack of coordination

- each node has a **different initialization** of the local (e.g., MLP) model
- due to the **permutation invariance of the hidden layers** of the neural network, coordinate-wise averaging can be detrimental without a common initialization
- Non-IID data worsen this effect

# Our solution #1: mitigation

Aggregate

Heterogeneity-aware aggregation function (**DecDiff**)
**Intuition**: give less importance to models that are very different from yours

Train

Boost the learning with a **virtual teacher**
**Intuition**: introduces a regularization element

# Aggregation strategy: beyond Decentralised Federated Average

Aggregate

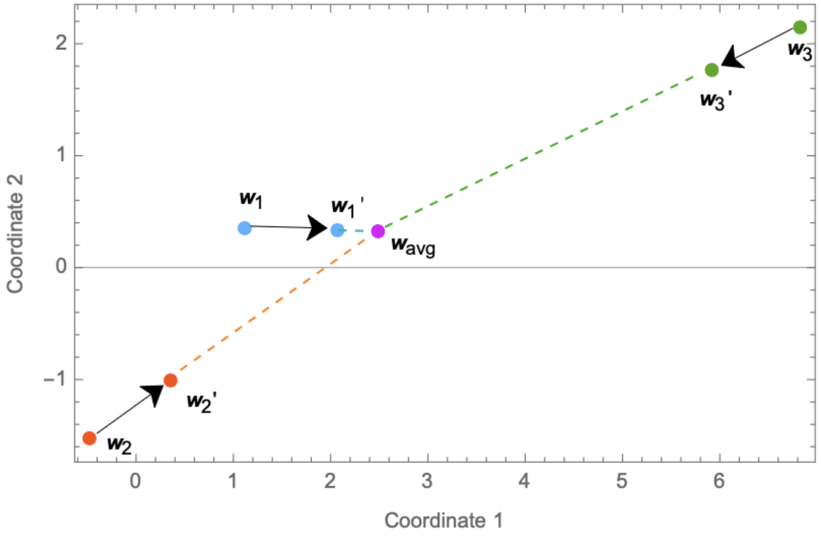
- DecDiff

$$\mathbf{w}_i^{(t)} = \mathbf{w}_i^{(t-1)} - \beta_{i,t}\frac{\mathbf{w}_i^{(t-1)} - \bar{\mathbf{w}}_i^{(t-1)}}{\|\mathbf{w}_i^{(t-1)} - \boxed{\bar{\mathbf{w}}_i^{(t-1)}}\|_2 + s}$$

where

$$\boxed{\bar{\mathbf{w}}_i^{(t-1)} = \frac{\sum_{\forall j \in \mathcal{N}_i} \omega_{ij} p_{ij} \mathbf{w}_j^{(t-1)}}{\sum_{\forall j \in \mathcal{N}_i} \omega_{ij} p_{ij}}}$$

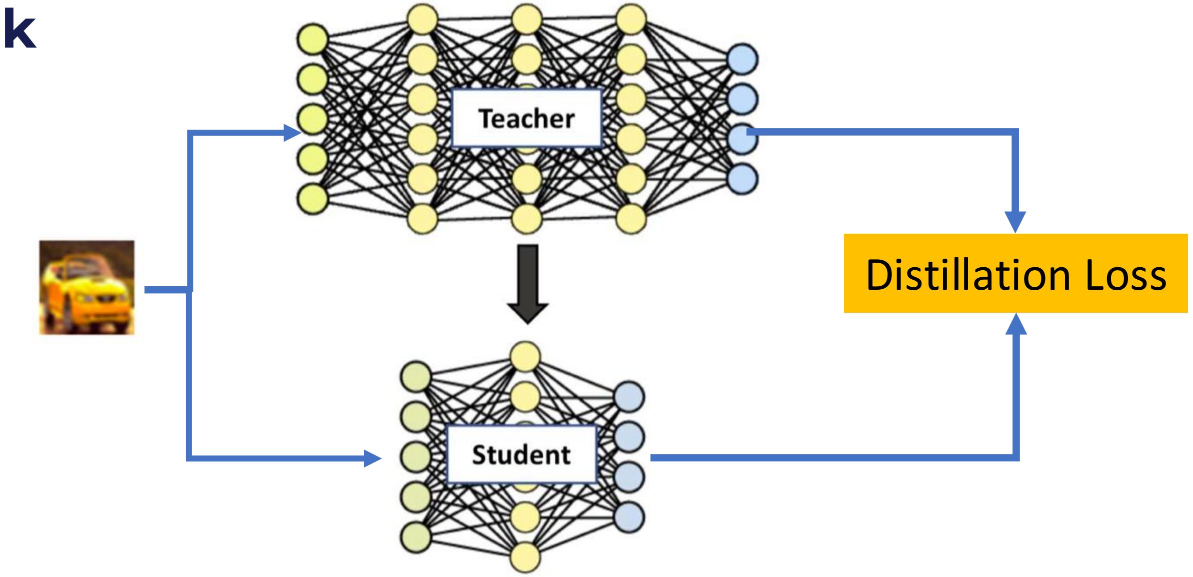
Average model
(from neighbors)



Valerio, L., Boldrini, C., Passarella, A., Kertész, J., Karsai, M., & Iñiguez, G. (2023). Coordination-free Decentralised Federated Learning on Complex Networks: Overcoming Heterogeneity. *arXiv preprint arXiv:2312.04504*.

# Local training: Based on distillation

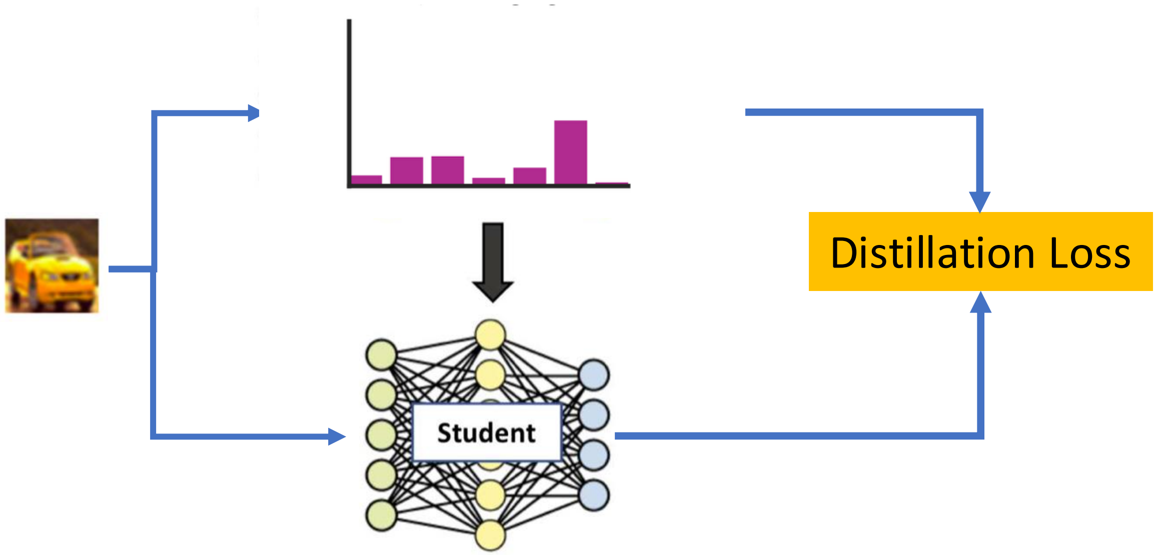
- Standard distillation
    - A **student** network tries to **mimic** a **Teacher** network
- Basic *assumptions* on the *Teacher* network
    - **Larger** and more capable **network**
    - Trained on **more data**



Valerio, L., Boldrini, C., Passarella, A., Kertész, J., Karsai, M., & Iñiguez, G. (2023). Coordination-free Decentralised Federated Learning on Complex Networks: Overcoming Heterogeneity. *arXiv preprint arXiv:2312.04504*.

# Local training: Based on distillation

- In <u>decentralised</u> settings:
  - **All devices** are both **teachers and students**
  - Trained on local (small data)
- <u>Potential Issues</u>: **computational bottleneck** for devices
- <u>Solution</u>: Self-distillation
  - Replace Teacher network with a virtual teacher

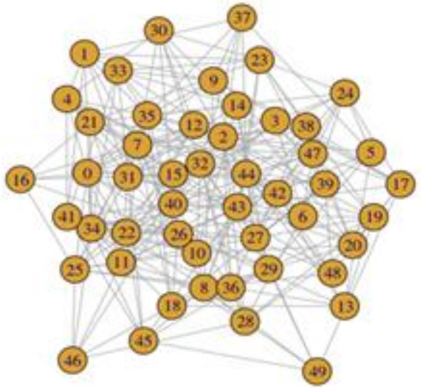


Valerio, L., Boldrini, C., Passarella, A., Kertész, J., Karsai, M., & Iñiguez, G. (2023). Coordination-free Decentralised Federated Learning on Complex Networks: Overcoming Heterogeneity. *arXiv preprint arXiv:2312.04504*.

# Results

- Our DFL vs FL: close performance

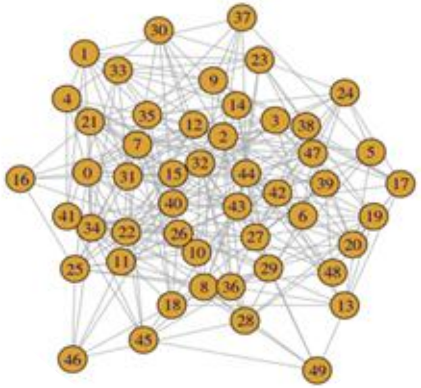
| Strategy | Accuracy (avg) | Conf. int. |
|---|---:|---:|
| Centralized | 0.918 | 0 |
| Federated | 0.896 | 0.00204 |
| DecDiff + Virtual Teacher | 0.894 | 0.00206 |
| DecDiff | 0.887 | 0.00463 |
| DecAvg | 0.886 | 0.00173 |
| SOTA benchmark #1 | 0.859 | 0.0033 |
| SOTA benchmark #2 | 0.859 | 0.0118 |
| No cooperation | 0.769 | 0.0396 |



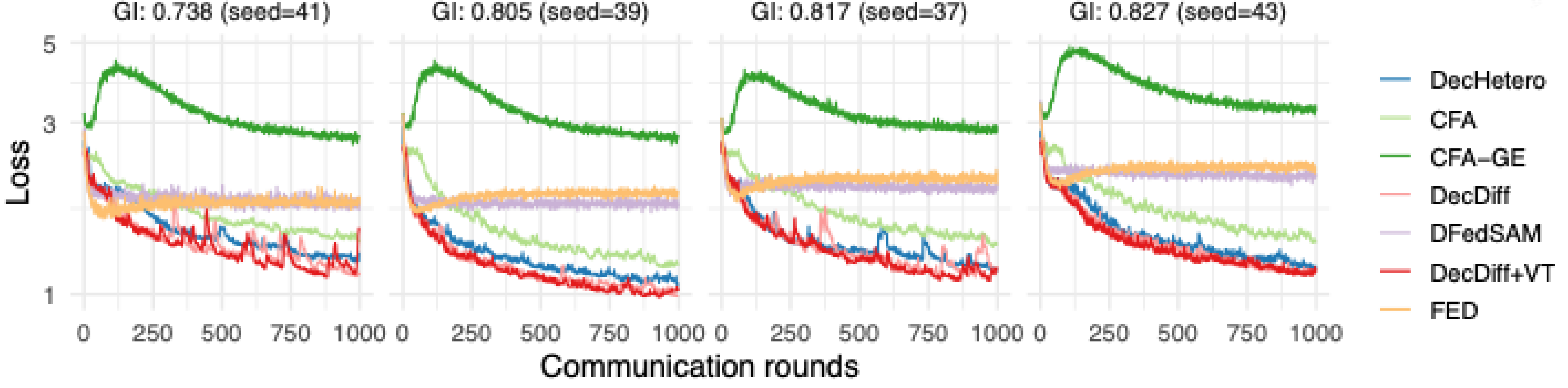
Valerio, L., Boldrini, C., Passarella, A., Kertész, J., Karsai, M., & Iñiguez, G. (2023). Coordination-free Decentralised Federated Learning on Complex Networks: Overcoming Heterogeneity. *arXiv preprint arXiv:2312.04504*.

# Results

- Faster and best convergence



Dataset: CIFAR10



Valerio, L., Boldrini, C., Passarella, A., Kertész, J., Karsai, M., & Iñiguez, G. (2023). Coordination-free Decentralised Federated Learning on Complex Networks: Overcoming Heterogeneity. *arXiv preprint arXiv:2312.04504.*

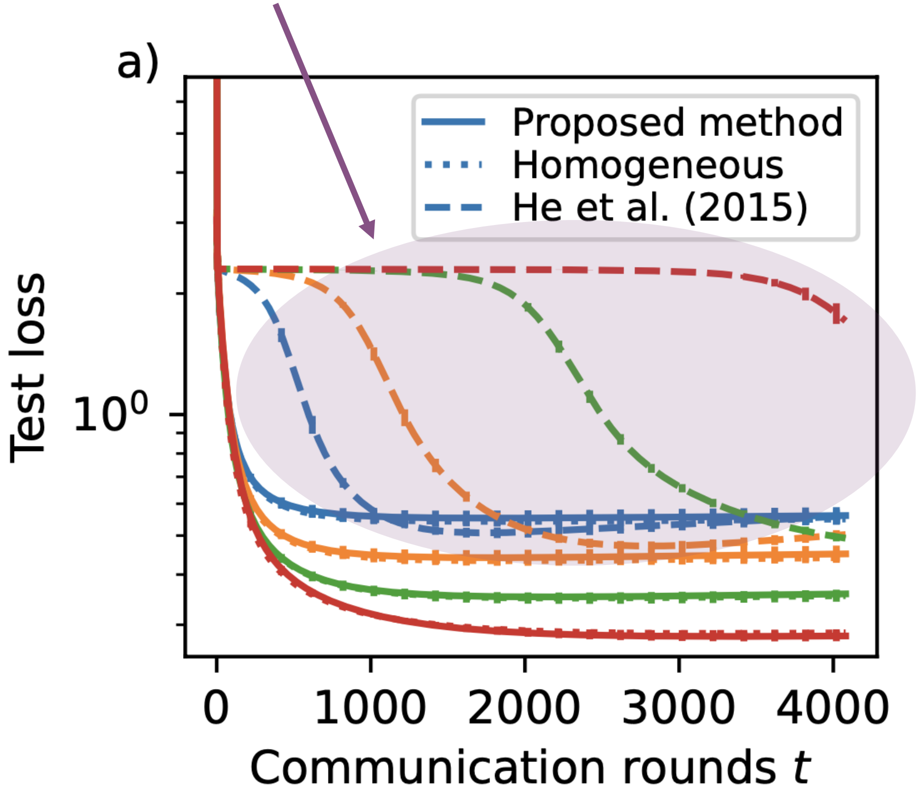
# Our solution #2: acceleration

- Standard model initialization (He et al. 2015)
  - weights of layer *l* ~ Gaussian(0, $\sigma_l^2$)

- In decentralized, uncoordinated settings, it results in **progressively poorer performance as the number *n* of nodes grows**

- We propose a novel initialization with gain correction

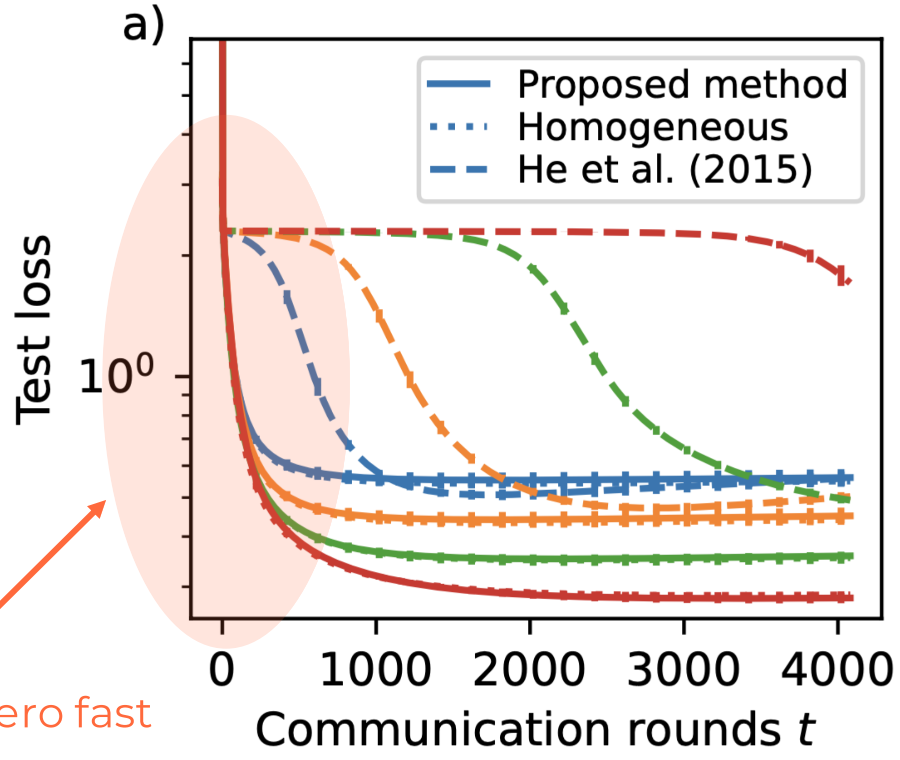
The loss decreases **very** slowly



Badie-Modiri, A., Boldrini, C., Valerio, L., Kertész, J., & Karsai, M. (2025). Initialisation and Topology Effects in Decentralised Federated Learning. Springer Applied Network Science. arXiv preprint arXiv:2403.15855.

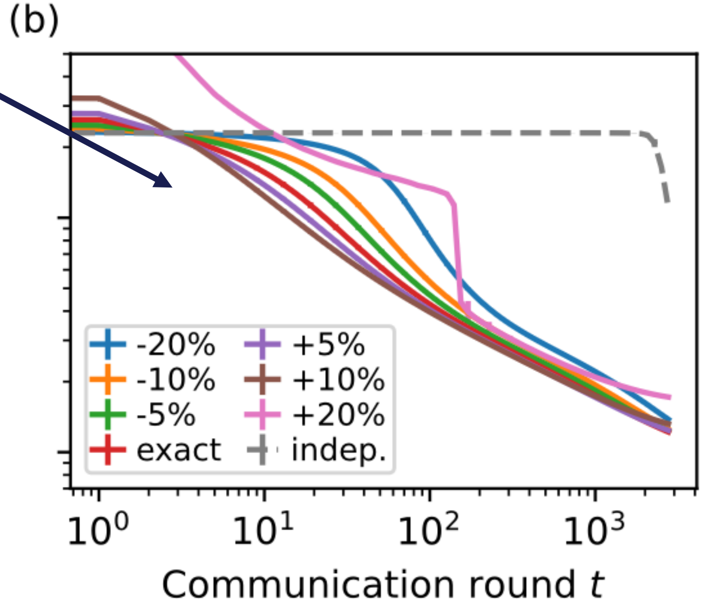
# Our solution #2: acceleration

- **HOW**: Use $\sigma_{init} \cdot \|v_{steady}\|^{-1}$, where $\|v_{steady}\|$ is the $\ell_2$-norm of the steady-state eigenvector (corresponding to eigenvalue 1) of the Markov matrix A associated with the communication graph G, normalized to have unit sum



Now, it goes to zero fast

It works better than std init even with large estimation errors!

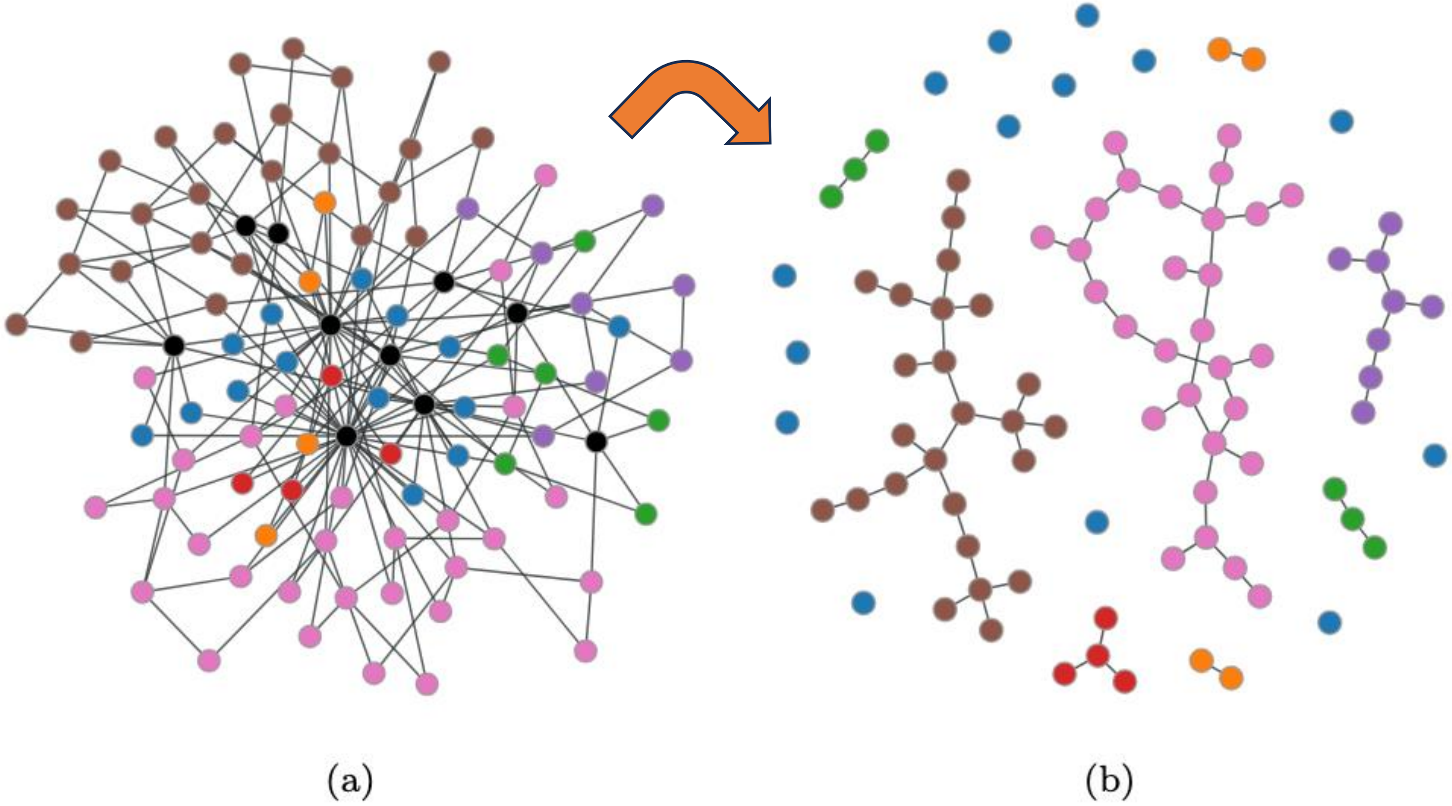


Badie-Modiri, A., Boldrini, C., Valerio, L., Kertész, J., & Karsai, M. (2024). Initialisation and Topology Effects in Decentralised Federated Learning. R1 under review at Applied Network Science. arXiv preprint arXiv:2403.15855.

# Research direction #4:
# Resilience of decentralized learning

# Resilience to data and node loss

- The **most central nodes** disappear from the network
  - They have data vs they don't have data (IID vs non-IID)

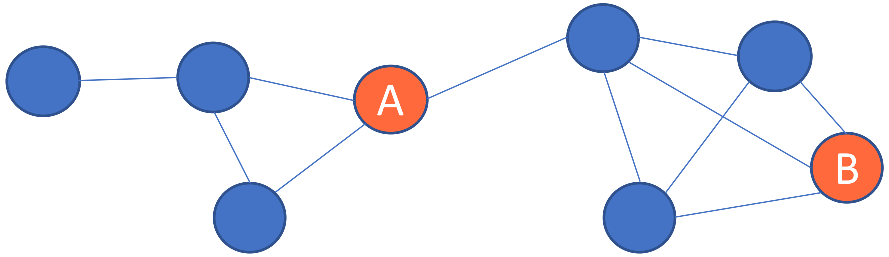


Palmieri, L., Boldrini, C., Valerio, L., Passarella, A., Conti, M., & Kertész, J. (2025). Robustness of decentralised learning to nodes and data disruption. Computer Communications, 2025.

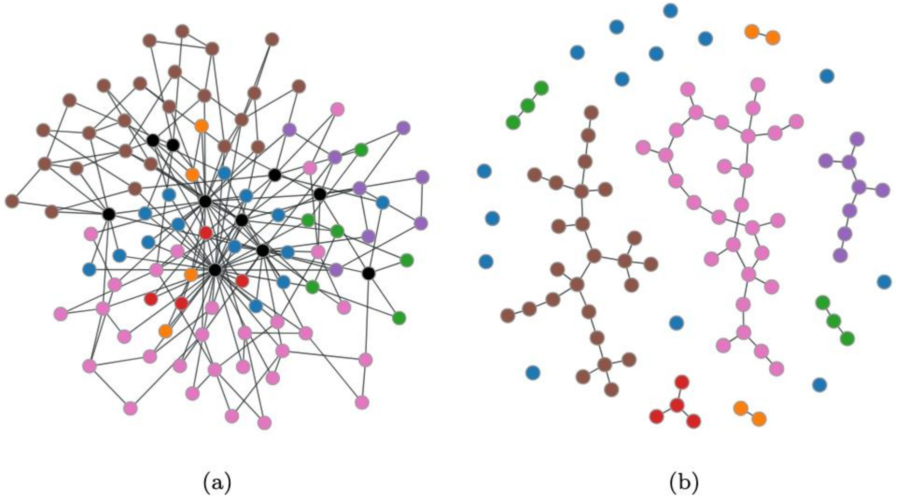
# Experimental settings: selection of cut off nodes

How do we disrupt: **switch off nodes** according to their **centrality score**

Centrality score: *Structural holes score (SH)*



We remove top 10% of nodes with highest SH



Initial

After disruption

Palmieri, L., Boldrini, C., Valerio, L., Passarella, A., Conti, M., & Kertész, J. (2025). Robustness of decentralised learning to nodes and data disruption. Computer Communications, 2025.

# Disruption analysis

## Case 1

**How**: Highly *central nodes have no data assigned*

**Central nodes role**: connectivity only

## Case 2

**How**: *central nodes have data*

**Central nodes role**: connectivity + training

Disruptions happens through time: the t=0, t=2 and t=10

10% accuracy curve

50% accuracy curve

Palmieri, L., Boldrini, C., Valerio, L., Passarella, A., Conti, M., & Kertész, J. (2025). Robustness of decentralised learning to nodes and data disruption. Computer Communications, 2025.

# Case 1: connectivity drops

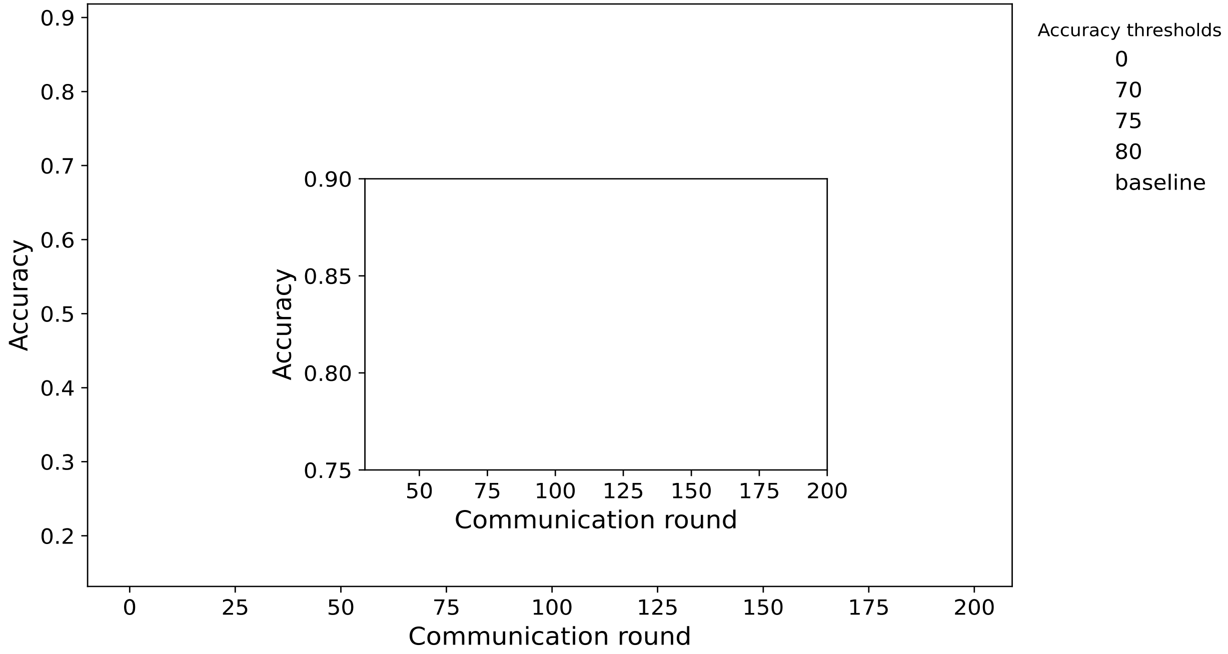
The mean overall accuracy does not change much with respect to the baseline

**DFL is robust against failures**

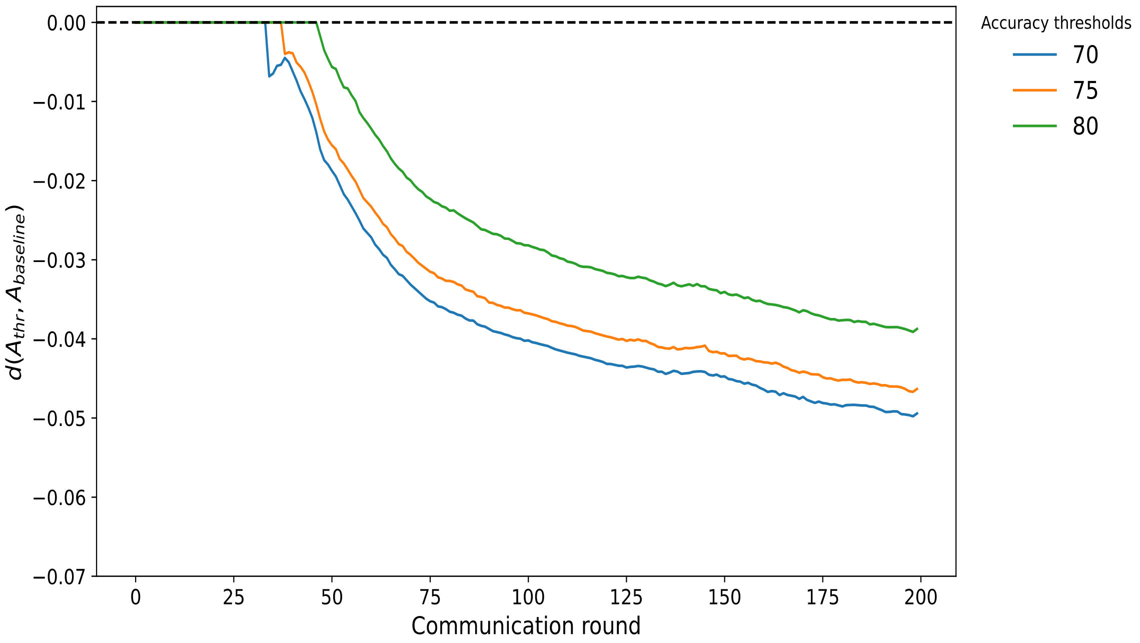
Mean accuracy

$$d(A_{thr}, A_{baseline}) = \frac{A_{thr}}{A_{baseline}} - 1$$

Mean accuracy over all surviving nodes
Case 1



Mean accuracy difference from baseline of all surviving nodes
Case 1

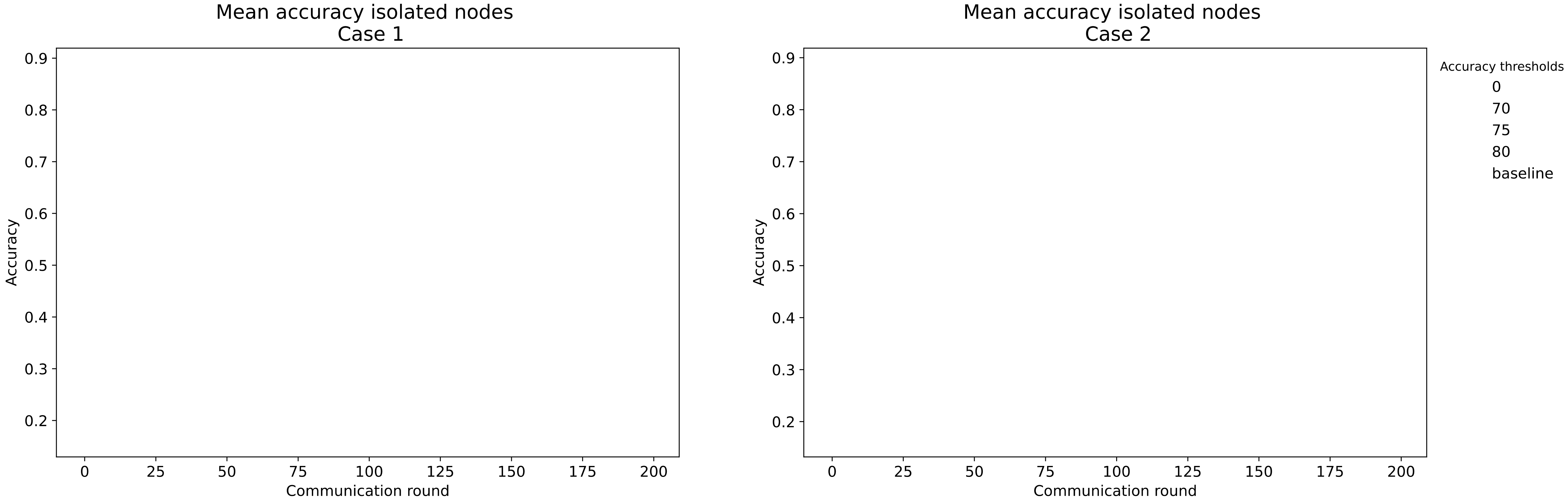


Palmieri, L., Boldrini, C., Valerio, L., Passarella, A., Conti, M., & Kertész, J. (2025). Robustness of decentralised learning to nodes and data disruption. Computer Communications, 2025.

# What happens to isolated nodes? Case 1 vs Case 2

**Knowledge persists**

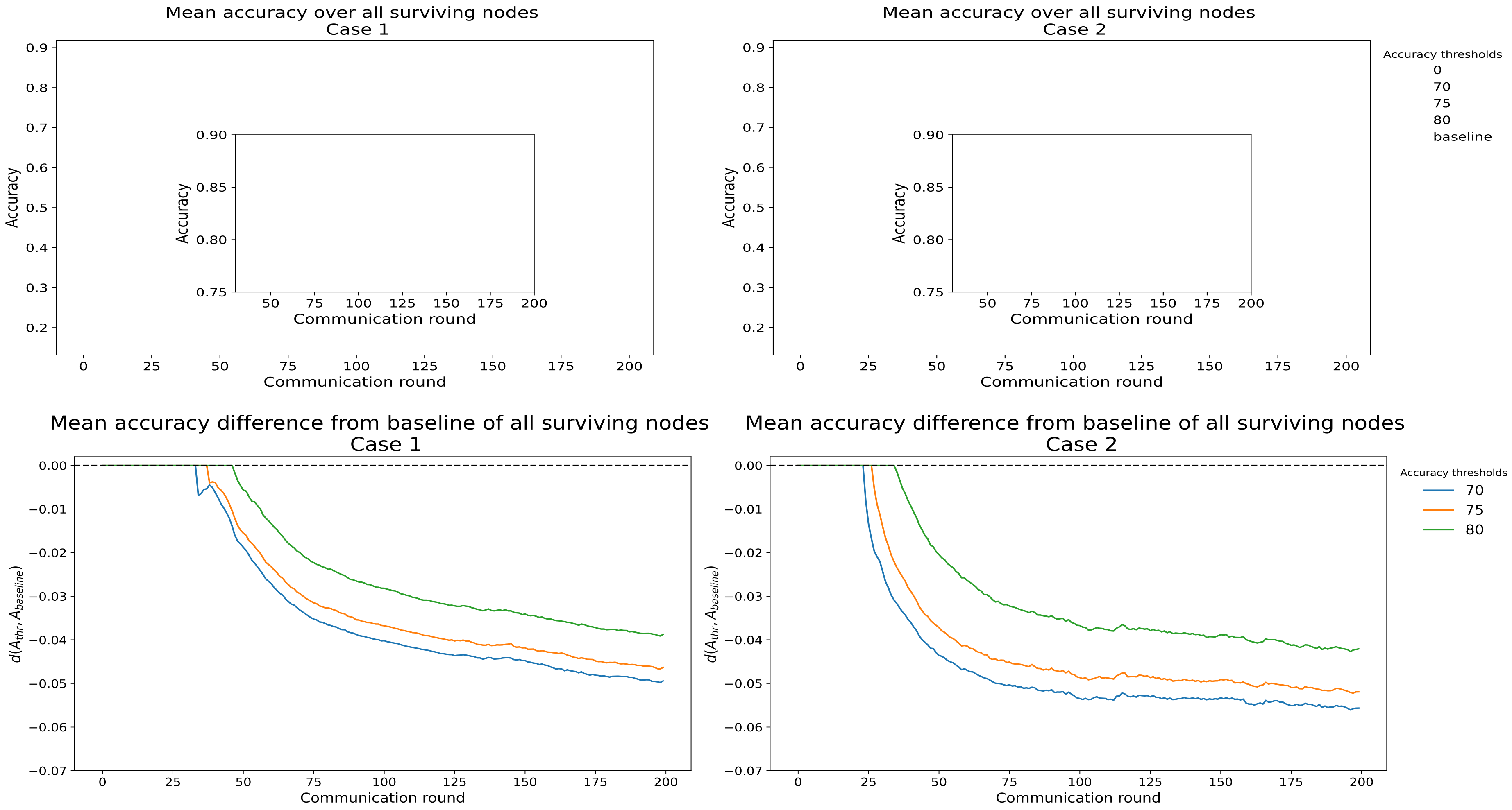
Mean accuracy of the isolated nodes is directly proportional to the accuracy threshold



Palmieri, L., Boldrini, C., Valerio, L., Passarella, A., Conti, M., & Kertész, J. (2025). Robustness of decentralised learning to nodes and data disruption. Computer Communications, 2025.

# The mean accuracy difference from the baseline is similar between Case 1 and Case 2

**DFL can tolerate large loss of data**



Palmieri, L., Boldrini, C., Valerio, L., Passarella, A., Conti, M., & Kertész, J. (2025). Robustness of decentralised learning to nodes and data disruption. Computer Communications, 2025.

# Concluding

**Key findings**

- **Knowledge** acquired before disruption **persists**, and is not lost even by isolated nodes
- **Accuracy** can be recovered **if data is present “somewhere”** in the network
- Even **modest connectivity supports efficient recovery** from failures

**Decentralized learning is robust to all types of disruption**

Palmieri, L., Boldrini, C., Valerio, L., Passarella, A., Conti, M., & Kertész, J. (2025). Robustness of decentralised learning to nodes and data disruption. Computer Communications, 2025.

# Resilience to low-quality data

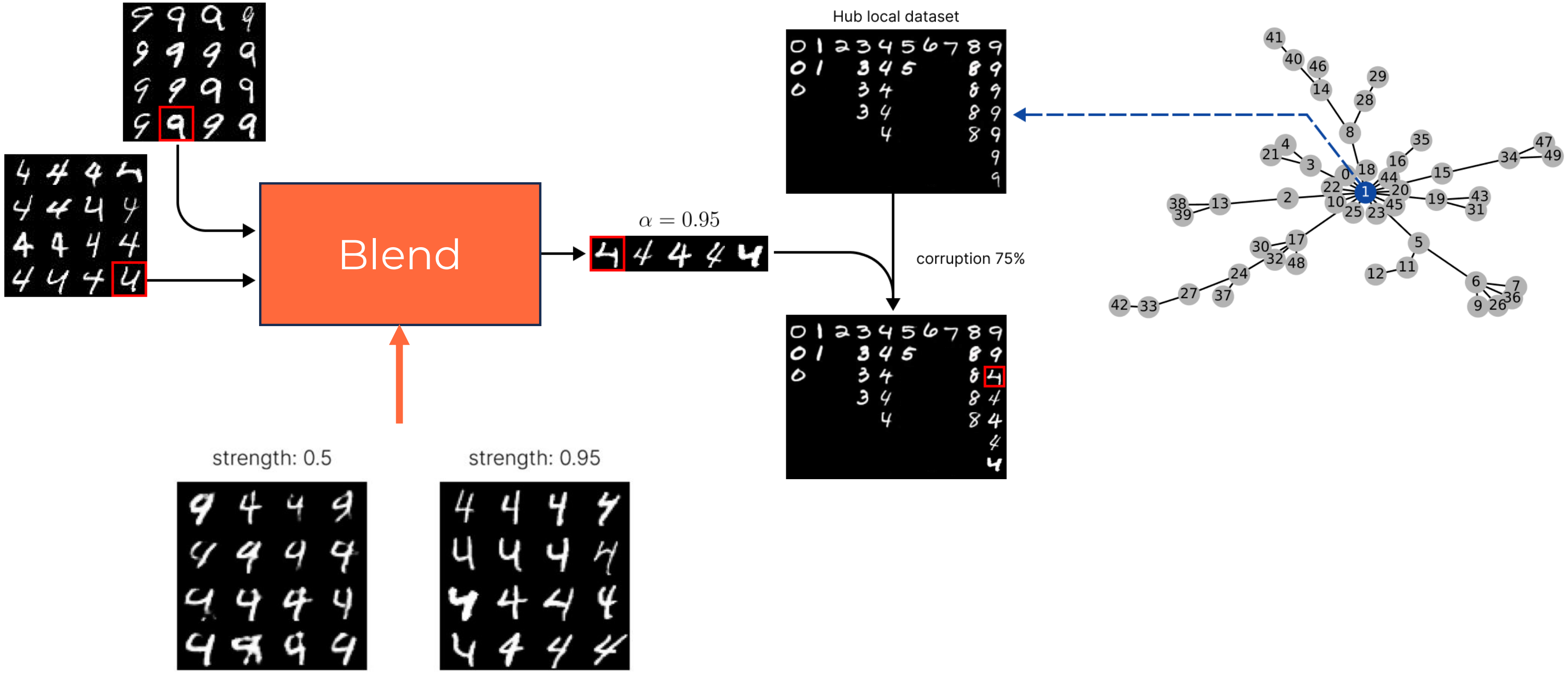
- **RQ1:** How **sensitive** is average-based decentralized federated learning **to low-quality or corrupted data**?

- **RQ2:** To what extent is this **sensitivity influenced by the underlying network topology**?

Sabella, S., Boldrini, C., Valerio, L., Passarella, A., & Conti, M. (2025). The Built-In Robustness of Decentralized Federated Averaging to Bad Data. IJCNN 2025 arXiv:2502.18097.

# Resilience to low-quality data

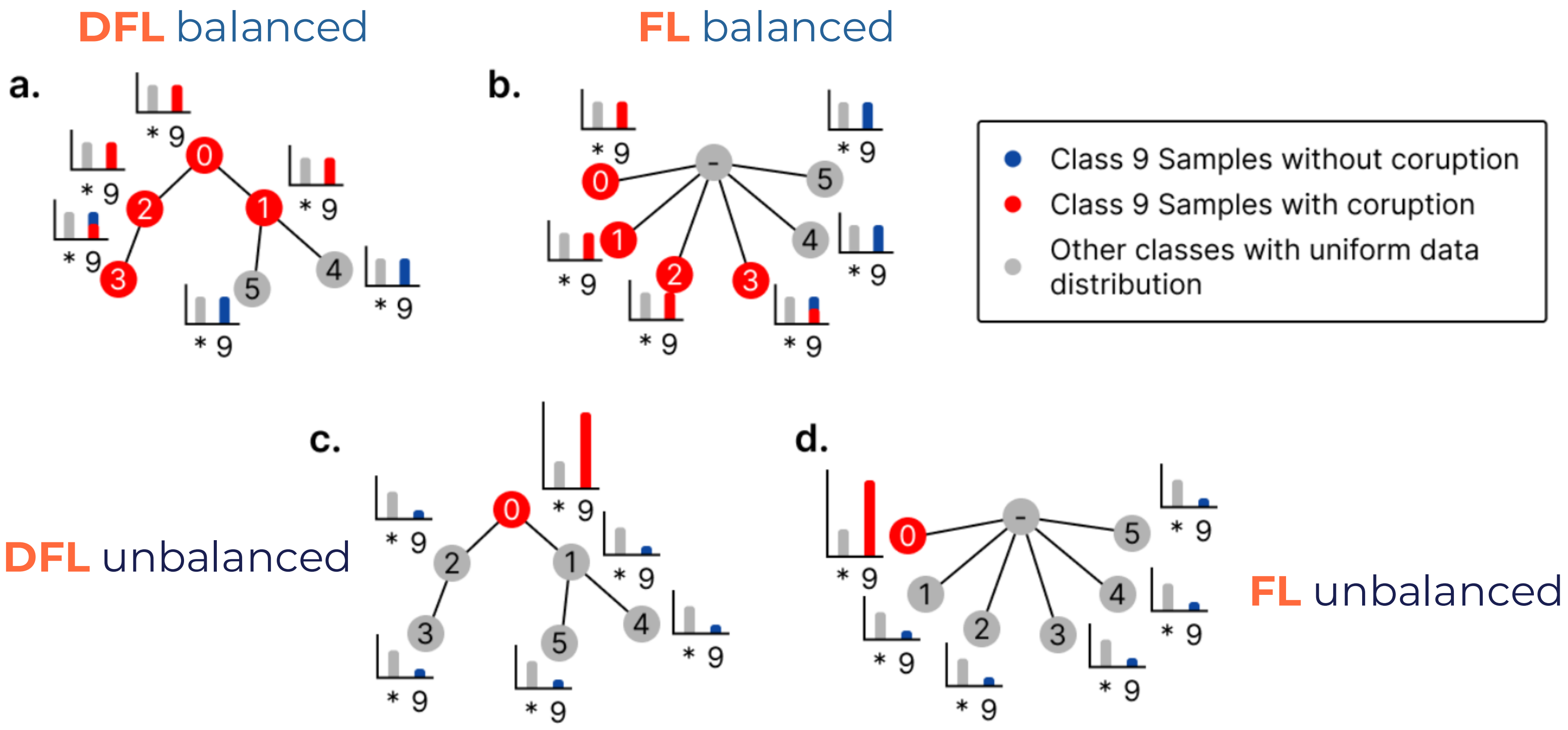
- **Low-quality data** **(i.e., 9s look like 4s, but labelled as 9)**



Different types of interpolation

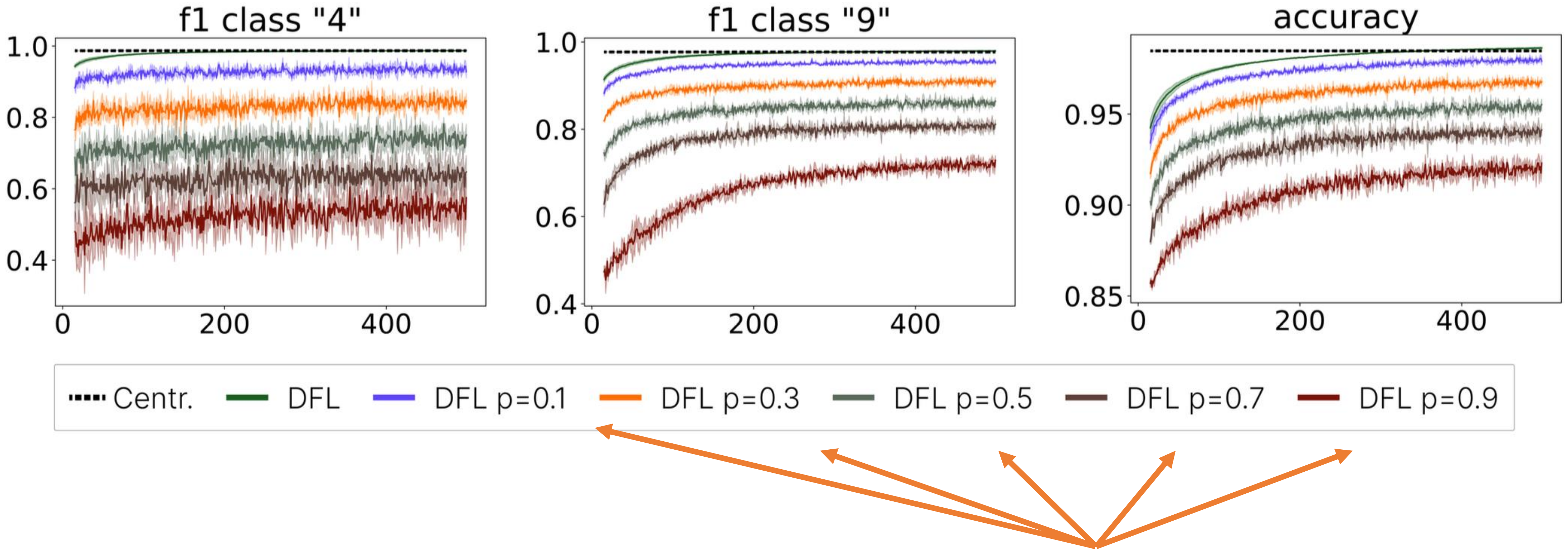
Sabella, S., Boldrini, C., Valerio, L., Passarella, A., & Conti, M. (2025). The Built-In Robustness of Decentralized Federated Averaging to Bad Data. IJCNN 2025 arXiv:2502.18097.

# Settings: how bad data is distributed


DFL balanced
FL balanced
DFL unbalanced
FL unbalanced
a.
b.
c.
d.
0
1
2
3
4
5
-
* 9
● Class 9 Samples without coruption
● Class 9 Samples with coruption
● Other classes with uniform data distribution

Sabella, S., Boldrini, C., Valerio, L., Passarella, A., & Conti, M. (2025). The Built-In Robustness of Decentralized Federated Averaging to Bad Data. IJCNN 2025 arXiv:2502.18097.

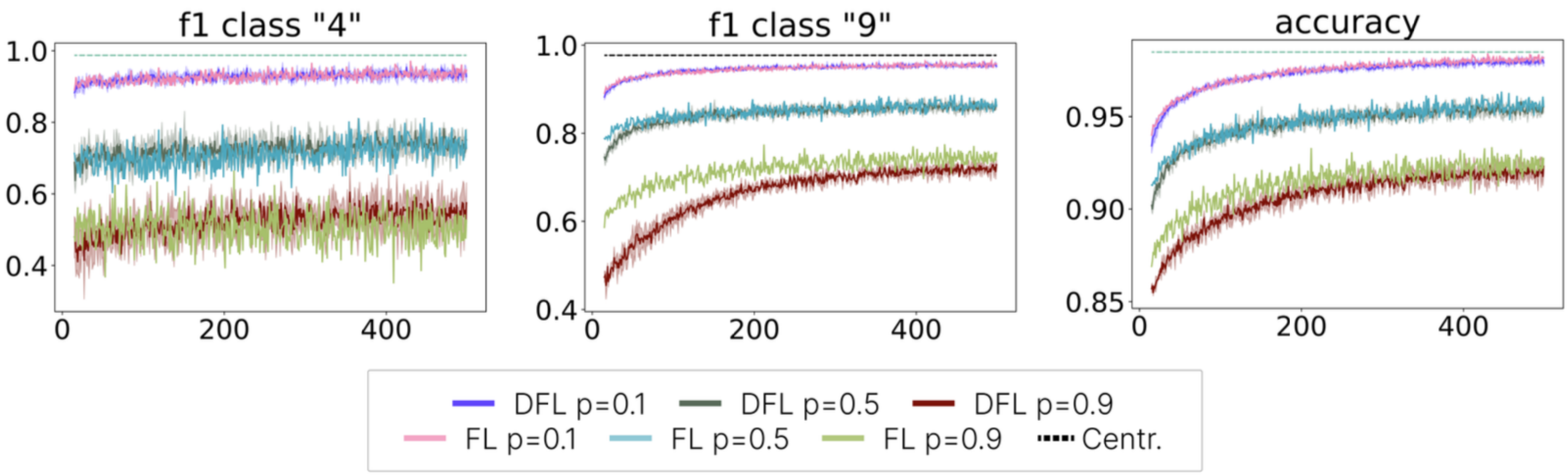
# Impact of corruption: Centralized VS DFL



Fraction of bad samples

Sabella, S., Boldrini, C., Valerio, L., Passarella, A., & Conti, M. (2025). The Built-In Robustness of Decentralized Federated Averaging to Bad Data. IJCNN 2025 arXiv:2502.18097.

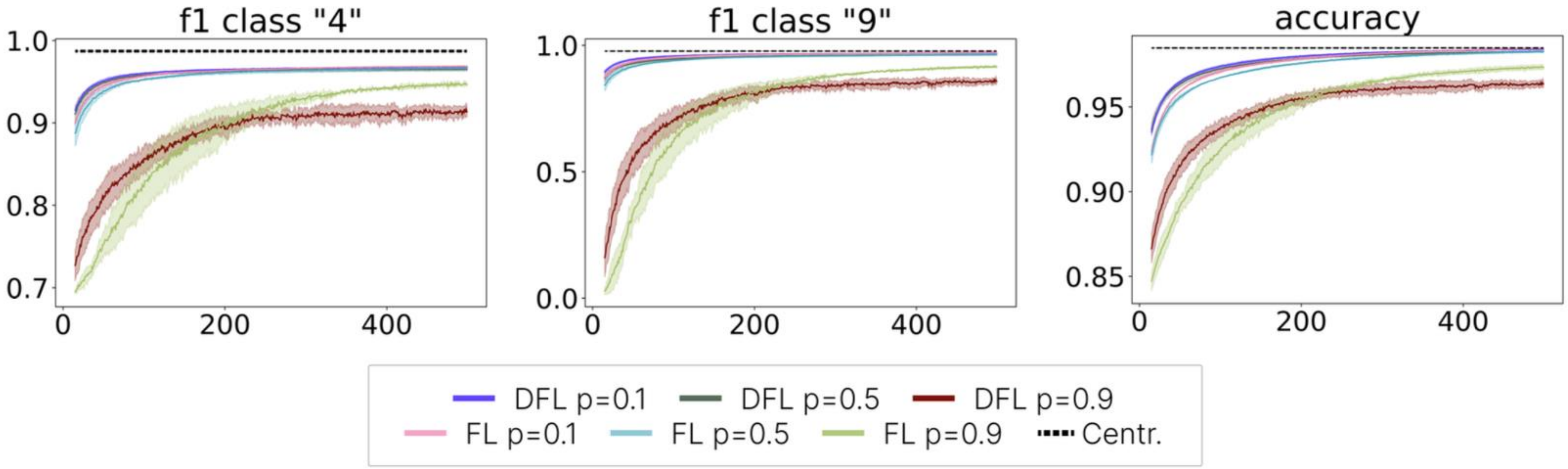
# Impact of corruption: **Balanced** corruption – DFL vs FL



No difference!

Sabella, S., Boldrini, C., Valerio, L., Passarella, A., & Conti, M. (2025). The Built-In Robustness of Decentralized Federated Averaging to Bad Data. IJCNN 2025 arXiv:2502.18097.

# Impact of corruption: **Unbalanced** corruption – DFL vs FL



FL less susceptible

Both show robustness

Sabella, S., Boldrini, C., Valerio, L., Passarella, A., & Conti, M. (2025). The Built-In Robustness of Decentralized Federated Averaging to Bad Data. IJCNN 2025 arXiv:2502.18097.

## To summarize

- **Corruption can hide in accuracy**
  - overall accuracy stays fairly stable
- **Spread beats spike**
  - the same bad-data budget is **far more damaging when dispersed across many nodes**
- **Coordination helps resilience:**
  - **federated (server-based) learning** shows **better long-run robustness** to corruption than fully decentralized learning.

**Again, we found that decentralized training is extremely resilient**

# What’s **left** & What’s **next**?

# To recap…

- We covered:
  - Impact of **network topology**
  - Cope with data and models’ **heterogeneity**
  - **Resilience** to bad data/low quality data

# What’s next

- **Resource limitations**, i.e. quite crucial at the edge
  - Model reduction (pruning/quantization/…)
  - Alternative models, e.g., neuromorphic models and hardware

- Go beyond correlations and introduce **collaborative causality**

# What’s next

- **Adaptation** to continuously **changing environments**
  - Continual learning
  - Unlearning, i.e., selectively forget concepts.

- Security and privacy:
  - Go **beyond resilience** and explore over **attacks to agents** in decentralized and collaborative environment

# Thank you for the attention!