Novelty Assessment Report

Paper: On The Surprising Effectiveness of a Single Global Merging in Decentralized Learning
PDF URL: https://openreview.net/pdf?id=zrFnwRHuQo
Venue: ICLR 2026 Conference Submission
Year: 2026
Report Generated: 2025-12-27

Abstract

Decentralized learning provides a scalable alternative to parameter-server-based training, yet its performance is often hindered by limited peer-to-peer communication. In this paper, we study how communication should be scheduled over time to improve global generalization, including determining when and how frequently devices synchronize. Counterintuitive empirical results show that concentrating communication budgets in the later stages of decentralized training remarkably improves global generalization. Surprisingly, we uncover that fully connected communication at the final step, implemented by a single global merging, can significant improve the generalization performance of decentralized learning under serve high data heterogeneity. Our theoretical contributions, which explains these phenomena, are first to establish that the globally merged model of decentralized SGD can match the convergence rate of parallel SGD. Technically, we reinterpret part of the discrepancy among local models, which were previously considered as detrimental noise, as constructive components essential for matching this rate. This work provides promising results that decentralized learning is able to generalize under high data heterogeneity and limited communication, while offering broad new avenues for model merging research. The code will be made publicly available.

Core Task Landscape

This paper addresses: **Communication Scheduling in Decentralized Learning Under Limited Peer-to-Peer Communication**

A total of **49 papers** were analyzed and organized into a taxonomy with **19 categories**.

Taxonomy Overview

The research landscape has been organized into the following main categories:

- **Communication Scheduling and Synchronization Strategies**
- **Communication Efficiency via Model Compression and Sparsification**
- **Parameter Aggregation and Synchronization Mechanisms**
- **Resource-Aware Decentralized Learning**
- **Decentralized Learning Frameworks and Architectures**
- **Convergence, Optimization, and Theoretical Analysis**
- **Application-Specific Decentralized Learning**
- **Robustness and Resilience in Decentralized Learning**

Complete Taxonomy Tree

- Communication Scheduling in Decentralized Learning Under Limited Peer-to-Peer Communication Survey Taxonomy
- Communication Scheduling and Synchronization Strategies
    - Temporal Communication Scheduling ★ (3 papers)
    - [0] On The Surprising Effectiveness of a Single Global Merging in Decentralized Learning (Anon et al., 2026) View paper
    - [34] Scheduling and Communication Schemes for Decentralized Federated Learning (FernÃ¡ndez Vilas, 2023) View paper
    - [38] Loss Based Byzantine Resilience for Decentralized Learning (Shinnosuke Masuda, 2025) View paper
    - Asynchronous Communication Approaches (3 papers)
    - [13] Decentralized federated learning with asynchronous parameter sharing for large-scale IoT networks (Xie Haihui, 2024) View paper
    - [14] : Accelerating Asynchronous Communication in Decentralized Deep Learning (A Nabli, 2023) View paper
    - [31] Decentralized Machine Learning with Asynchronous Communication (Tavonput Luangphasy, 2025) View paper
    - Adaptive Peer Selection and Topology Manipulation (4 papers)
    - [8] Accelerating the decentralized federated learning via manipulating edges (Mingyang Zhou, 2024) View paper
    - [16] Travellingfl: Communication efficient peer-to-peer federated learning (Vansh Gupta, 2023) View paper
    - [29] Efficient Model Propagation for Peer-to-Peer Federated Learning using Minimum Spanning Tree and Gossip Networks (Alka Luqman, 2025) View paper
    - [41] Communication-Efficient Decentralized Learning with Sparsification and Adaptive Peer Selection (Zhenheng Tang, 2020) View paper
- Communication Efficiency via Model Compression and Sparsification
    - Gradient Compression and Sparsification (3 papers)
    - [11] SparSFA: Towards robust and communication-efficient peer-to-peer federated learning (Han Wang, 2023) View paper
    - [44] Deep Gradient Compression: Reducing the Communication Bandwidth for Distributed Training (Yujun Lin, 2017) View paper
    - [45] Distributed Adaptive Learning Under Communication Constraints (Carpentiero, 2021) View paper
    - Dynamic and Adaptive Model Compression (2 papers)
    - [7] Dynamic distributed model compression for efficient decentralized federated learning and incentive provisioning in edge computing networks (Alia Asheralieva, 2025) View paper

- [24] YOGA: Adaptive Layer-Wise Model Aggregation for Decentralized Federated Learning (Jun Liu, 2024) View paper
- Parameter Aggregation and Synchronization Mechanisms
  - Efficient Aggregation Protocols (3 papers)
  - [3] Efficient parameter synchronization for peer-to-peer distributed learning with selective multicast (Shouxi Luo, 2024) View paper
  - [23] Efficient parameter aggregation in federated learning with hybrid convergecast (Yangyang Tao, 2021) View paper
  - [37] Fast Parameter Synchronization for Distributed Learning with Selective Multicast (Shouxi Luo, 2022) View paper
  - Secure and Privacy-Preserving Aggregation (2 papers)
  - [15] Turbo-aggregate: Breaking the quadratic aggregation barrier in secure federated learning (Jinhyun So, 2021) View paper
  - [17] Sequoia: An Accessible and Extensible Framework for Privacy-Preserving Machine Learning over Distributed Data (Kaiqiang Xu, 2025) View paper
- Resource-Aware Decentralized Learning
  - Bandwidth-Constrained Learning (4 papers)
  - [9] Communication-efficient decentralized machine learning over heterogeneous networks (Pan Zhou, 2021) View paper
  - [12] Distributed learning for automatic modulation recognition in bandwidth-limited networks (Rashvand, 2025) View paper
  - [22] Learning and Optimization in Wireless Sensor Networks (Muhammad I. Qureshi, 2025) View paper
  - [32] Optimizing Data Distribution for Federated Learning Under Bandwidth Constraint (Kengo Tajiri, 2023) View paper
  - Heterogeneous Device and Network Management (4 papers)
  - [5] Communication-efficient training workload balancing for decentralized multi-agent learning (Seyed Mahmoud Sajjadi Mohammadabadi, 2024) View paper
  - [20] Accelerating decentralized federated learning in heterogeneous edge computing (Lun Wang, 2022) View paper
  - [25] Learning-Driven Decentralized Machine Learning in Resource-Constrained Wireless Edge Computing (Zeyu Meng, 2021) View paper
  - [35] MMDFL: Multi-Model-based Decentralized Federated Learning for Resource-Constrained AIoT Systems (Dengke Yan, 2025) View paper
  - Energy-Aware and Resource Allocation (2 papers)
  - [19] Clustered federated learning for energy-harvesting smart meters in P2P energy trading (Zi-ming Liu, 2025) View paper
  - [26] Power allocation and communication resource scheduling for federated learning in wireless IoT networks (Renan R. de Oliveira, 2025) View paper
- Decentralized Learning Frameworks and Architectures
  - Peer-to-Peer Federated Learning Frameworks (4 papers)
  - [18] BlockDFL: A blockchain-based fully decentralized peer-to-peer federated learning framework (Zhen Qin, 2024) View paper
  - [28] AROM: Control-Theoretic Learning for Resilient and Decentralized Edge Intelligence (E Jeong, 2025) View paper
  - [48] Secure Peer to Peer Learning Using Auto Encoders (Anirudh Kasturi, 2022) View paper
  - [49] Peer-to-Peer Deep Learning for Beyond-5G IoT (Srinivasa Pranav, 2023) View paper
  - Edge and IoT-Oriented Decentralized Learning (5 papers)
  - [1] Edge learning for B5G networks with distributed signal processing: Semantic communication, edge computing, and wireless sensing (Yang, 2023) View paper
  - [2] Communication-efficient and distributed learning over wireless networks: Principles and applications (J Park, 2021) View paper
  - [6] Client scheduling and resource management for efficient training in heterogeneous IoT-edge federated learning (Yangguang Cui, 2021) View paper
  - [10] Improving edge AI for industrial IoT applications with distributed learning using consensus (Samuel Fidelis, 2024) View paper
  - [42] Scale Wisely, Secure Wholly: P2P Swarm Learning Over Consortium Blockchain in Edge Networks (Feng Gao, 2024) View paper
  - Specialized Distributed Learning Architectures (2 papers)
  - [4] Snake Learning: A Communication-and Computation-Efficient Distributed Learning Framework for 6G (Xiao-Xue Yu, 2025) View paper
  - [21] AutoDDL: Automatic Distributed Deep Learning With Near-Optimal Bandwidth Cost (Jin-Fan Chen, 2023) View paper
- Convergence, Optimization, and Theoretical Analysis
  - Convergence Analysis and Guarantees (2 papers)
  - [30] Local training and scalability of federated learning systems (Syed Zawad, 2022) View paper
  - [33] Communication-efficient decentralized learning for intelligent networked systems (Jeong, 2024) View paper
  - Performance Evaluation and Comparison (1 papers)
  - [27] Performance analysis and comparison of distributed machine learning systems (Alqahtani, 2019) View paper
- Application-Specific Decentralized Learning
  - Wireless Communication and Networking Applications (2 papers)
  - [43] Joint Coding and Scheduling Optimization for Distributed Learning over Wireless Edge Networks (Van Huynh, 2021) View paper
  - [46] Intelligent Networking enabled Vehicular Distributed Learning (Wenchao Xu, 2021) View paper
  - Multi-Agent and Distributed Control (1 papers)
  - [40] Multi-Agent Bandit Learning through Heterogeneous Action Erasure Channels (Hanna, 2023) View paper
  - Peer-to-Peer Content Distribution (1 papers)
  - [47] Evolutionary Game for Cooperative Peer-to-Peer Streaming (Yan Chen, 2021) View paper
- Robustness and Resilience in Decentralized Learning (2 papers)
  - [36] Learnae: Distributed and Resilient Deep Neural Network Training for Heterogeneous Peer to Peer Topologies (Spyridon Nikolaidis, 2019) View paper
  - [39] Optimization strategies for federated learning (Zhang, 2025) View paper

Narrative

Core task: communication scheduling in decentralized learning under limited peer-to-peer communication. The field addresses how distributed agents can collaboratively train models when direct communication is constrained by bandwidth, latency, or energy budgets. The taxonomy reveals several complementary research directions: Communication Scheduling and Synchronization Strategies explore when and how nodes exchange updates, including temporal scheduling approaches like Global Merging Decentralized[0] and Scheduling Communication Schemes[34]; Communication Efficiency via Model Compression and Sparsification reduces payload sizes through techniques such as Deep Gradient Compression[44] and SparSFA[11]; Parameter Aggregation and Synchronization Mechanisms design protocols for merging distributed updates, exemplified by Turbo Aggregate[15] and Asynchronous Parameter Sharing[13]; Resource-Aware Decentralized Learning optimizes for heterogeneous compute and energy constraints, as seen in Clustered Energy Harvesting[19] and Resource Constrained Edge[25]; while Convergence, Optimization, and Theoretical Analysis provide formal guarantees, and Application-Specific branches tailor methods to domains like vehicular networks (Vehicular Distributed Learning[46]) or IoT (Client Scheduling IoT[6]).

A central tension emerges between synchronous schemes that ensure consistency but suffer from stragglers, and asynchronous or event-driven approaches that improve throughput at the cost of staleness. Works like Asynchronous Communication Acceleration[14] and Snake Learning[4] explore dynamic topologies and flexible timing, while Training Workload Balancing[5] and Selective Multicast Synchronization[3] address load imbalance and targeted communication. Global Merging Decentralized[0] sits within the Temporal Communication Scheduling branch, emphasizing periodic global synchronization phases to balance convergence speed and communication overhead. Compared to purely asynchronous methods like Asynchronous Communication Learning[31], it likely enforces stricter coordination intervals, and relative to fine-grained scheduling in Scheduling Communication Schemes[34], it may favor coarser global merge events. This positioning reflects a pragmatic middle ground: leveraging temporal structure to reduce redundant transmissions while maintaining sufficient alignment across peers for stable convergence in bandwidth-limited settings.

Related Works in Same Category

The following **2 sibling papers** share the same taxonomy leaf node with the original paper:

1. Scheduling and Communication Schemes for Decentralized Federated Learning

Authors: FernÃ¡ndez Vilas, Ana, FernÃ¡ndez Veiga, Manuel, El-Bendary Nashwa | **Year/Venue**: 2023 • arXiv (Cornell University) | **URL**: View paper

Abstract

Federated learning (FL) is a distributed machine learning paradigm in which a large number of clients coordinate with a central server to learn a model without sharing their own training data. One central server is not enough, due to problems of connectivity with clients. In this paper, a decentralized federated learning (DFL) model with the stochastic gradient descent (SGD) algorithm has been introduced, as a more scalable approach to improve the learning performance in a network of agents with...

Relationship Analysis

Both papers belong to the Temporal Communication Scheduling category, investigating when and how frequently to synchronize during decentralized training. The candidate paper proposes three scheduling policies for communications between clients and parallel servers in a decentralized federated learning setting with SGD, focusing on convergence speed and final model accuracy. In contrast, the original paper specifically investigates the surprising effectiveness of concentrating communication budgets in later training stages, particularly through a single global merging step, and provides theoretical analysis showing that the globally merged model can match parallel SGD convergence rates under high data heterogeneity.

---

2. Loss Based Byzantine Resilience for Decentralized Learning

Authors: Shinnosuke Masuda, Kazuyuki Shudo | **Year/Venue**: 2025 | **URL**: View paper

Abstract

â¦ Decentralized learning [1], [2] has gained attention as a peer-to-peer (P2P) network-based â¦ Shudo, âCommunication scheduling for gossip sgd in a wide area network,â IEEE Access, â¦

Relationship Analysis

Both papers belong to the Temporal Communication Scheduling category, investigating optimal timing and frequency of synchronization in decentralized learning. While the original paper focuses on concentrating communication budgets in later training stages and demonstrates the effectiveness of a single global merging under data heterogeneity, the candidate paper addresses Byzantine resilience in decentralized learning through loss-based mechanisms for handling malicious agents. The key difference is that the original paper optimizes communication timing for performance under benign conditions, whereas the candidate paper focuses on security and robustness against adversarial participants in the network.

Contributions Analysis

Overall novelty summary. The paper proposes concentrating communication budgets in later training stages, culminating in a single global merging step, to improve generalization in decentralized learning under severe data heterogeneity. It resides in the Temporal Communication Scheduling leaf, which contains only three papers total, indicating a relatively sparse research direction within the broader taxonomy of 49 papers across 19 leaf nodes. This leaf focuses specifically on optimal timing and frequency of synchronization events, distinguishing it from asynchronous methods or adaptive topology approaches that populate neighboring branches.

The taxonomy reveals that Temporal Communication Scheduling sits alongside Asynchronous Communication Approaches and Adaptive Peer Selection within the Communication Scheduling and Synchronization Strategies branch. Neighboring branches address orthogonal concerns: Gradient Compression reduces payload sizes, Efficient Aggregation Protocols optimize merging mechanics, and Bandwidth-Constrained Learning tackles resource limits. The paper's emphasis on when to communicate rather than how to compress or which peers to select places it squarely in the temporal scheduling domain, though its global merging strategy shares conceptual overlap with aggregation protocols that coordinate distributed updates.

Among 29 candidates examined, the theoretical convergence analysis matching parallel SGD rates encountered two refutable candidates from 10 examined, suggesting moderate prior work in this specific theoretical claim. The empirical demonstration of single global merging effectiveness and the theoretical explanation for temporal communication allocation each examined 10 and 9 candidates respectively, with zero refutable matches, indicating these contributions appear more novel within the limited search scope. The statistics reflect a focused literature search rather than exhaustive coverage, so unexamined work may exist beyond the top-K semantic matches and citation expansions performed.

Given the sparse Temporal Communication Scheduling leaf and the limited refutation evidence across most contributions, the work appears to occupy a relatively underexplored niche within decentralized learning. The theoretical convergence claim shows some overlap with prior analysis, but the empirical focus on late-stage global merging and the reinterpretation of local model discrepancy as constructive rather than detrimental noise seem less directly addressed in the examined candidates. These impressions are bounded by the 29-paper search scope and may shift with broader literature coverage.

This paper presents **3 main contributions**, each analyzed against relevant prior work:

Contribution 1: Empirical demonstration of single global merging effectiveness

Description: The authors empirically demonstrate that performing a single global merging (parameter averaging) at the final training step significantly improves global generalization in decentralized learning, even under extremely limited communication budgets and high data heterogeneity. This finding holds across diverse experimental settings including different datasets, architectures, and optimizers.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. K-DUMBs IoRT: Knowledge driven unified model block sharing in the Internet of Robotic Things

URL: View paper

Brief Assessment

K-DUMBs IoRT[52] focuses on federated learning in IoRT with one-shot weight sharing for knowledge transfer across robotic devices, not on decentralized learning with parameter averaging under communication constraints and data heterogeneity.

2. DIMAT: Decentralized Iterative Merging-And-Training for Deep Learning Models

URL: View paper

Brief Assessment

DIMAT[59] focuses on iterative merging-and-training with periodic model merging throughout training, not a single final global merging. The candidate's merging occurs every 2 training epochs as part of the training process, fundamentally different from the original's single final-step merging approach.

3. MOHFL: Multi-Level One-Shot Hierarchical Federated Learning With Enhanced Model Aggregation Over Non-IID Data

URL: View paper

Brief Assessment

MOHFL[57] focuses on hierarchical federated learning with multi-level aggregation over non-IID data, not on decentralized learning with single global merging at the final training step. The candidate's context fragments do not provide sufficient detail about single global merging strategies or their effectiveness under communication constraints.

4. One-shot federated learning for LEO constellations that reduces convergence time from days to 90 minutes

URL: View paper

Brief Assessment

LEO Constellations Oneshot[50] focuses on one-shot federated learning for satellite communications using synthetic data generation and knowledge distillation, not on decentralized learning with parameter averaging under communication constraints.

5. Communication-Efficient Distributed Deep Learning via Federated Dynamic Averaging

URL: View paper

Brief Assessment

Federated Dynamic Averaging[51] focuses on dynamically triggering synchronization based on model variance in federated settings, not on demonstrating the effectiveness of a single final global merging step in decentralized learning under extreme communication constraints and high data heterogeneity.

6. Ravnest: Decentralized Asynchronous Training on Heterogeneous Devices

URL: View paper

Brief Assessment

Ravnest[58] focuses on asynchronous decentralized training with model parallelism across heterogeneous devices, not on single global merging strategies for parameter averaging in decentralized learning under communication constraints.

7. Multi-Device Cooperative Fine-Tuning of Foundation Models at the Network Edge

URL: View paper

Brief Assessment

The candidate paper (Cooperative Foundation Finetuning[53]) focuses on multi-device cooperative fine-tuning with bandwidth allocation optimization at the network edge. The provided context does not contain sufficient technical detail about parameter merging strategies or decentralized training dynamics to challenge the original paper's novelty claim regarding single global merging in decentralized learning under communication constraints.

8. OSGAN: One-shot distributed learning using generative adversarial networks: A. Kasturi, C. Hota

URL: View paper

Brief Assessment

OSGAN[54] focuses on one-shot distributed learning using generative adversarial networks for data aggregation at a central server. The candidate's context is too limited to assess whether it addresses single global merging in decentralized training under communication constraints and data heterogeneity.

9. Optimizing quantum federated learning: addressing non-I I D data challenges with global data sharing in weighted model averaging and clustering-based parameter selection

URL: View paper

Brief Assessment

Quantum Federated Clustering[55] focuses on quantum federated learning with weighted averaging and clustering methods for non-IID quantum data. It does not address single global merging at the final training step in classical decentralized learning settings.

10. One-shot federated learning-based model-free reinforcement learning

URL: View paper

Brief Assessment

Oneshot Model Free[56] focuses on one-shot federated learning for model-free reinforcement learning on IoT devices with bandwidth allocation optimization. This is fundamentally different from the original paper's investigation of single global merging in decentralized learning for supervised tasks under communication constraints and data heterogeneity.

Contribution 2: Theoretical convergence analysis matching parallel SGD rate

Description: The authors establish the first theoretical result proving that the globally merged model from decentralized SGD can achieve the same convergence rate as parallel SGD. They reinterpret part of the model discrepancy among local models as constructive components rather than purely detrimental noise, enabling this rate matching.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. A(DP)SGD: Asynchronous Decentralized Parallel Stochastic Gradient Descent With Differential Privacy

URL: View paper

Brief Assessment

Asynchronous Differential Privacy[63] focuses on asynchronous decentralized SGD with differential privacy guarantees. The candidate's theoretical analysis addresses privacy-preserving convergence in asynchronous settings, which is technically distinct from the original paper's analysis of synchronous decentralized SGD with global merging strategies under non-IID data.

2. Does worst-performing agent lead the pack? analyzing agent dynamics in unified distributed sgd

URL: View paper

Brief Assessment

Worst Performing Agent[62] focuses on asymptotic analysis of unified distributed SGD with different sampling strategies (i.i.d., shuffling, Markovian) and their impact on convergence through agent dynamics. The original paper establishes convergence rate matching for decentralized SGD's globally merged model, while this candidate analyzes how individual agent sampling strategies affect overall convergence in distributed settings. These are complementary perspectives on distributed learning rather than overlapping novelty claims.

3. Decentralized asynchronous nonconvex stochastic optimization on directed graphs

URL: View paper

Brief Assessment

Directed Graphs Asynchronous[65] focuses on asynchronous optimization over directed graphs with stochastic gradients, achieving standard sublinear rates. The original paper addresses synchronous decentralized SGD with a novel reinterpretation of model discrepancy to match parallel SGD rates—a fundamentally different theoretical contribution and setting.

4. Tackling Data Heterogeneity: A New Unified Framework for Decentralized SGD with Sample-induced Topology

URL: View paper

Brief Assessment

Sample Induced Topology[66] focuses on decentralized SGD convergence rates but addresses a different problem setting (data heterogeneity with sample-induced topology) and does not claim to match parallel SGD rates through global merging or reinterpretation of model discrepancy as constructive components.

5. Can decentralized algorithms outperform centralized algorithms? a case study for decentralized parallel stochastic gradient descent

URL: View paper

Prior Art Analysis

Decentralized Outperform Centralized[60] demonstrates that prior work established theoretical results showing decentralized SGD can match the convergence rate of parallel SGD. The candidate paper explicitly proves that D-PSGD achieves the same convergence rate as C-PSGD (mini-batch SGD), with a rate of $O(1/k + 1/\sqrt{nk})$ when k is large enough. This directly refutes the novelty claim that the original paper is the first to establish such a theoretical result, as the candidate paper was published at NIPS 2017, predating the original submission to ICLR 2026.

Evidence

Evidence 1 - **Rationale**: The candidate paper explicitly states that D-PSGD achieves the same convergence rate as C-PSGD (centralized parallel SGD), which is equivalent to the claim in the original paper about matching parallel SGD rates. - **Original**: our theoretical contributions, which explains these phenomena, are first to establish that the globally merged model of decentralized sgd can match the convergence rate of parallel sgd. - **Candidate**: note that this rate is the same as c-psgd (or mini-batch sgd with mini-batch size n) [agarwal and duchi, 2011, dekel et al., 2012, lian et al., 2015]. the advantage of d-psgd over c-psgd is to avoid the communication traffic jam.

6. Improving the transient times for distributed stochastic gradient methods

URL: View paper

Brief Assessment

Transient Times Improvement[67] focuses on strongly convex optimization with transient time characterization in distributed settings, whereas the original paper addresses non-convex decentralized learning with data heterogeneity and proves convergence rate matching for the globally merged model.

7. Asymptotic network independence in distributed stochastic optimization for machine learning

URL: View paper

Prior Art Analysis

Network Independence Asymptotic[69] demonstrates that prior work has already established theoretical results showing decentralized SGD can match the convergence rate of parallel/centralized SGD. The candidate paper explicitly proves that distributed stochastic gradient descent achieves comparable convergence rates to centralized methods with the same computational power, stating this property as 'asymptotic network independence.' Multiple sections provide formal convergence analysis showing DSGD matches centralized SGD's rate of $\sigma^2/(n\mu^2 k)$, which directly challenges the novelty claim that the original paper is 'first to establish' this rate-matching result.

Evidence

Evidence 1 - **Rationale**: This pair shows that Network Independence Asymptotic[69] explicitly establishes the same core result—that decentralized methods match centralized performance—which the original paper claims to be 'first to establish.' - **Original**: our theoretical contributions, which explains these phenomena, are first to establish that the globally merged model of decentralized sgd can match the convergence rate of parallel sgd. - **Candidate**: we call this property asymptotic network independence: it is as if the network is not even there. asymptotic network independence provides an answer to the concerns raised in the previous section. we begin by illustrating these results with a simulation from [21], shown in figure 3. here the problem...

Evidence 2 - **Rationale**: Network Independence Asymptotic[69] cites multiple prior works ([4, 5, 6, 31]) that already established comparable convergence rates between distributed and centralized methods, directly refuting the claim of being 'first' to provide this convergence analysis. - **Original**: we investigate the underlying mechanism that enables themergeability of local models in decentralized learning. specifically, we provide the first convergence analysis showing that the globally merged model of decentralized sgd can match the rate of parallel sgd (theorem 1 and proposition 2). - **Candidate**: the papers [4, 5, 6, 31] gave the first crisp statement of the relationship between centralized and distributed methods in the setting of distributed optimization of smooth strongly convex functions in the presence of noise. under constant stepsizes, the papers [4, 5, 6] were the first to show that wh...

Evidence 3 - **Rationale**: Both papers present formal convergence theorems for decentralized SGD. Network Independence Asymptotic[69] provides Theorem 1 showing the rate $\sigma^2/(n\mu^2 k)$, which matches the parallel SGD rate, demonstrating that this theoretical result existed prior to the original paper's claim. - **Original**: theorem 1(non-convex convergence rate of dsgd).suppose assumption 2 and assumption 3 hold. consider decentralized sgd (dsgd) with initializations θ(0) k =θ (0) for all k∈ v, and a constant learning rate satisfying η≤ 1 l2 . let ¯θ(t) = 1 m pm k=1 θ(t) k denote the averaged model at the t-th step. to... - **Candidate**: theorem 1 under sgd (11), supposing assumptions 1-3 hold, we have r(k) ≤ σ2 nμ2k + ok (1 k2) . (13) to compare with the analysis for dsgd later, we briefly describe how to obtain (13).

---

8. Dâ(DP)2SGD: Decentralized Parallel SGD with Differential Privacy in Dynamic Networks

URL: View paper

Brief Assessment

Dynamic Networks Privacy[68] focuses on differential privacy in dynamic networks and achieves convergence rates, but does not address the specific theoretical insight about reinterpreting model discrepancy as constructive components for matching parallel SGD rates in decentralized learning without privacy constraints.

---

9. CEDAS: A Compressed Decentralized Stochastic Gradient Method With Improved Convergence

URL: View paper

Brief Assessment

CEDAS[61] focuses on compressed decentralized optimization with communication constraints, not on the specific phenomenon of global merging in decentralized learning that the original paper investigates.

---

10. An improved convergence analysis for decentralized online stochastic non-convex optimization

URL: View paper

Brief Assessment

Nonconvex Convergence Analysis[64] focuses on GT-DSGD (gradient tracking variant) achieving network-independent performance matching centralized minibatch SGD, while the original paper analyzes standard decentralized SGD with a novel reinterpretation of model discrepancy. The technical approaches and algorithmic variants differ fundamentally.

Contribution 3: Theoretical explanation for temporal communication allocation

Description: The authors provide theoretical justification showing why minimal but non-zero communication preserves model mergeability throughout training, and formally explain why allocating communication budgets toward later training stages improves performance. This is formalized through conditions on consensus violation and gradient norms.

This contribution was assessed against **9 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. A layer selection optimizer for communication-efficient decentralized federated deep learning

URL: View paper

Brief Assessment

Layer Selection Optimizer[76] focuses on spatial layer selection within DNNs for communication efficiency in decentralized FL, not temporal allocation of communication budgets across training stages or theoretical conditions on consensus violation and gradient norms for mergeability.

---

2. DIMAT: Decentralized Iterative Merging-And-Training for Deep Learning Models

URL: View paper

Brief Assessment

DIMAT[59] does not address temporal communication allocation or when to schedule communication during training. The candidate's theoretical contributions focus on convergence rates with periodic merging, not on optimizing communication timing or explaining why communication should be concentrated in later training stages.

---

3. Route-and-Aggregate Decentralized Federated Learning Under Communication Errors

URL: View paper

Brief Assessment

Route and Aggregate[77] focuses on routing strategies and adaptive aggregation coefficients in decentralized federated learning under communication errors, not on temporal communication allocation or mergeability conditions throughout training stages.

---

4. Adaptive transmission scheduling in wireless networks for asynchronous federated learning

URL: View paper

Brief Assessment

Adaptive Transmission Scheduling[75] focuses on wireless federated learning with asynchronous updates and transmission scheduling under channel uncertainties. The original paper addresses decentralized learning with peer-to-peer communication and provides theoretical analysis on consensus violation and gradient norms for temporal communication allocation, which is a fundamentally different setting and theoretical framework.

---

5. DIST: Distributed Learning-based Energy-Efficient and Reliable Task Scheduling and Resource Allocation in Fog Computing

URL: View paper

Brief Assessment

DIST[71] focuses on task scheduling and resource allocation in fog computing environments, not on temporal communication allocation strategies in decentralized learning or the theoretical conditions for model mergeability throughout training.

---

6. New scheduling approach using reinforcement learning for heterogeneous distributed systems

URL: View paper

Brief Assessment

Reinforcement Learning Scheduling[74] focuses on task scheduling in heterogeneous distributed systems using RL, not on communication timing or mergeability in decentralized training. The domains are fundamentally different.

7. Joint Coding and Scheduling Optimization for Distributed Learning over Wireless Edge Networks

URL: View paper

Brief Assessment

Joint Coding Scheduling[43] focuses on wireless edge networks with coded computing for distributed learning tasks (e.g., matrix multiplication), optimizing communication timing through MDS codes and scheduling. It does not address decentralized peer-to-peer training with model merging or provide theoretical analysis of consensus violation and gradient norms in the context of temporal communication allocation for preserving mergeability.

8. Communication optimization algorithms for distributed deep learning systems: A survey

URL: View paper

Brief Assessment

Communication Optimization Survey[73] is a survey paper that categorizes communication optimization strategies broadly (masking, compression, frequency reduction, hybrid). It does not provide theoretical analysis of temporal allocation effects on mergeability or gradient norms during training phases.

9. Optimizing Distributed Computing Resources with Federated Learning: Task Scheduling and Communication Efficiency

URL: View paper

Brief Assessment

Task Scheduling Resources[70] focuses on federated learning for distributed computing resource optimization, task scheduling, and communication efficiency in cloud/edge computing contexts. It does not address temporal communication allocation strategies or mergeability conditions in decentralized training, which are the core theoretical contributions of the original paper.

Appendix: Text Similarity Detection

No high-similarity text segments were detected across any compared papers.

References

- [0] On The Surprising Effectiveness of a Single Global Merging in Decentralized Learning View paper
- [1] Edge learning for B5G networks with distributed signal processing: Semantic communication, edge computing, and wireless sensing View paper
- [2] Communication-efficient and distributed learning over wireless networks: Principles and applications View paper
- [3] Efficient parameter synchronization for peer-to-peer distributed learning with selective multicast View paper
- [4] Snake Learning: A Communication-and Computation-Efficient Distributed Learning Framework for 6G View paper
- [5] Communication-efficient training workload balancing for decentralized multi-agent learning View paper
- [6] Client scheduling and resource management for efficient training in heterogeneous IoT-edge federated learning View paper
- [7] Dynamic distributed model compression for efficient decentralized federated learning and incentive provisioning in edge computing networks View paper
- [8] Accelerating the decentralized federated learning via manipulating edges View paper
- [9] Communication-efficient decentralized machine learning over heterogeneous networks View paper
- [10] Improving edge AI for industrial IoT applications with distributed learning using consensus View paper
- [11] SparSFA: Towards robust and communication-efficient peer-to-peer federated learning View paper
- [12] Distributed learning for automatic modulation recognition in bandwidth-limited networks View paper
- [13] Decentralized federated learning with asynchronous parameter sharing for large-scale IoT networks View paper
- [14] : Accelerating Asynchronous Communication in Decentralized Deep Learning View paper
- [15] Turbo-aggregate: Breaking the quadratic aggregation barrier in secure federated learning View paper
- [16] Travellingfl: Communication efficient peer-to-peer federated learning View paper
- [17] Sequoia: An Accessible and Extensible Framework for Privacy-Preserving Machine Learning over Distributed Data View paper
- [18] BlockDFL: A blockchain-based fully decentralized peer-to-peer federated learning framework View paper
- [19] Clustered federated learning for energy-harvesting smart meters in P2P energy trading View paper
- [20] Accelerating decentralized federated learning in heterogeneous edge computing View paper
- [21] AutoDDL: Automatic Distributed Deep Learning With Near-Optimal Bandwidth Cost View paper
- [22] Learning and Optimization in Wireless Sensor Networks View paper
- [23] Efficient parameter aggregation in federated learning with hybrid convergecast View paper
- [24] YOGA: Adaptive Layer-Wise Model Aggregation for Decentralized Federated Learning View paper
- [25] Learning-Driven Decentralized Machine Learning in Resource-Constrained Wireless Edge Computing View paper
- [26] Power allocation and communication resource scheduling for federated learning in wireless IoT networks View paper
- [27] Performance analysis and comparison of distributed machine learning systems View paper
- [28] AROM: Control-Theoretic Learning for Resilient and Decentralized Edge Intelligence View paper
- [29] Efficient Model Propagation for Peer-to-Peer Federated Learning using Minimum Spanning Tree and Gossip Networks View paper
- [30] Local training and scalability of federated learning systems View paper
- [31] Decentralized Machine Learning with Asynchronous Communication View paper
- [32] Optimizing Data Distribution for Federated Learning Under Bandwidth Constraint View paper
- [33] Communication-efficient decentralized learning for intelligent networked systems View paper
- [34] Scheduling and Communication Schemes for Decentralized Federated Learning View paper
- [35] MMDFL: Multi-Model-based Decentralized Federated Learning for Resource-Constrained AIoT Systems View paper
- [36] Learnae: Distributed and Resilient Deep Neural Network Training for Heterogeneous Peer to Peer Topologies View paper
- [37] Fast Parameter Synchronization for Distributed Learning with Selective Multicast View paper
- [38] Loss Based Byzantine Resilience for Decentralized Learning View paper
- [39] Optimization strategies for federated learning View paper
- [40] Multi-Agent Bandit Learning through Heterogeneous Action Erasure Channels View paper
- [41] Communication-Efficient Decentralized Learning with Sparsification and Adaptive Peer Selection View paper

- [42] Scale Wisely, Secure Wholly: P2P Swarm Learning Over Consortium Blockchain in Edge Networks View paper
- [43] Joint Coding and Scheduling Optimization for Distributed Learning over Wireless Edge Networks View paper
- [44] Deep Gradient Compression: Reducing the Communication Bandwidth for Distributed Training View paper
- [45] Distributed Adaptive Learning Under Communication Constraints View paper
- [46] Intelligent Networking enabled Vehicular Distributed Learning View paper
- [47] Evolutionary Game for Cooperative Peer-to-Peer Streaming View paper
- [48] Secure Peer to Peer Learning Using Auto Encoders View paper
- [49] Peer-to-Peer Deep Learning for Beyond-5G IoT View paper
- [50] One-shot federated learning for LEO constellations that reduces convergence time from days to 90 minutes View paper
- [51] Communication-Efficient Distributed Deep Learning via Federated Dynamic Averaging View paper
- [52] K-DUMBs IoRT: Knowledge driven unified model block sharing in the Internet of Robotic Things View paper
- [53] Multi-Device Cooperative Fine-Tuning of Foundation Models at the Network Edge View paper
- [54] OSGAN: One-shot distributed learning using generative adversarial networks: A. Kasturi, C. Hota View paper
- [55] Optimizing quantum federated learning: addressing non-I I D data challenges with global data sharing in weighted model averaging and clustering-based parameter selection View paper
- [56] One-shot federated learning-based model-free reinforcement learning View paper
- [57] MOHFL: Multi-Level One-Shot Hierarchical Federated Learning With Enhanced Model Aggregation Over Non-IID Data View paper
- [58] Ravnest: Decentralized Asynchronous Training on Heterogeneous Devices View paper
- [59] DIMAT: Decentralized Iterative Merging-And-Training for Deep Learning Models View paper
- [60] Can decentralized algorithms outperform centralized algorithms? a case study for decentralized parallel stochastic gradient descent View paper
- [61] CEDAS: A Compressed Decentralized Stochastic Gradient Method With Improved Convergence View paper
- [62] Does worst-performing agent lead the pack? analyzing agent dynamics in unified distributed sgd View paper
- [63] A(DP)SGD: Asynchronous Decentralized Parallel Stochastic Gradient Descent With Differential Privacy View paper
- [64] An improved convergence analysis for decentralized online stochastic non-convex optimization View paper
- [65] Decentralized asynchronous nonconvex stochastic optimization on directed graphs View paper
- [66] Tackling Data Heterogeneity: A New Unified Framework for Decentralized SGD with Sample-induced Topology View paper
- [67] Improving the transient times for distributed stochastic gradient methods View paper
- [68] Dâ□□(DP)2SGD: Decentralized Parallel SGD with Differential Privacy in Dynamic Networks View paper
- [69] Asymptotic network independence in distributed stochastic optimization for machine learning View paper
- [70] Optimizing Distributed Computing Resources with Federated Learning: Task Scheduling and Communication Efficiency View paper
- [71] DIST: Distributed Learning-based Energy-Efficient and Reliable Task Scheduling and Resource Allocation in Fog Computing View paper
- [72] Communication-efficient large-scale distributed deep learning: A comprehensive survey View paper
- [73] Communication optimization algorithms for distributed deep learning systems: A survey View paper
- [74] New scheduling approach using reinforcement learning for heterogeneous distributed systems View paper
- [75] Adaptive transmission scheduling in wireless networks for asynchronous federated learning View paper
- [76] A layer selection optimizer for communication-efficient decentralized federated deep learning View paper
- [77] Route-and-Aggregate Decentralized Federated Learning Under Communication Errors View paper