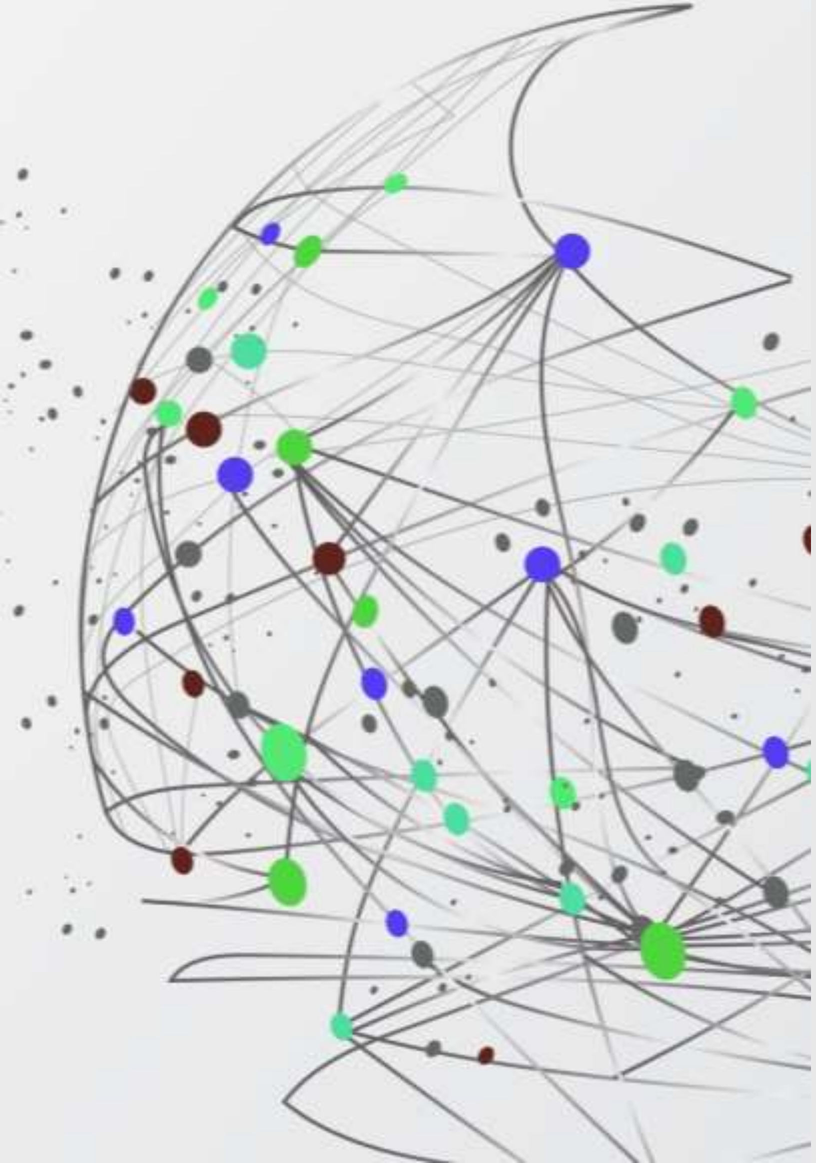


Beyond the Teacher–Student Paradigm: The Role of Network Structure in Knowledge Distillation

Nasif Fahmid Prangon and Jie Wu

Department of Computer and Information Sciences
Temple University, USA

International Conference on Computing, Networking and Communications (ICNC 2026)

Outline

- Motivation and Problem Setting
- From Static KD to Networked Distillation
- Topology-Aware Learning Framework
- Dynamic Reinforcement Strategy
- Theoretical Insights on Diffusion
- Experimental Evaluation
- Conclusions and Future Directions

Motivation: The Limitation of Traditional KD

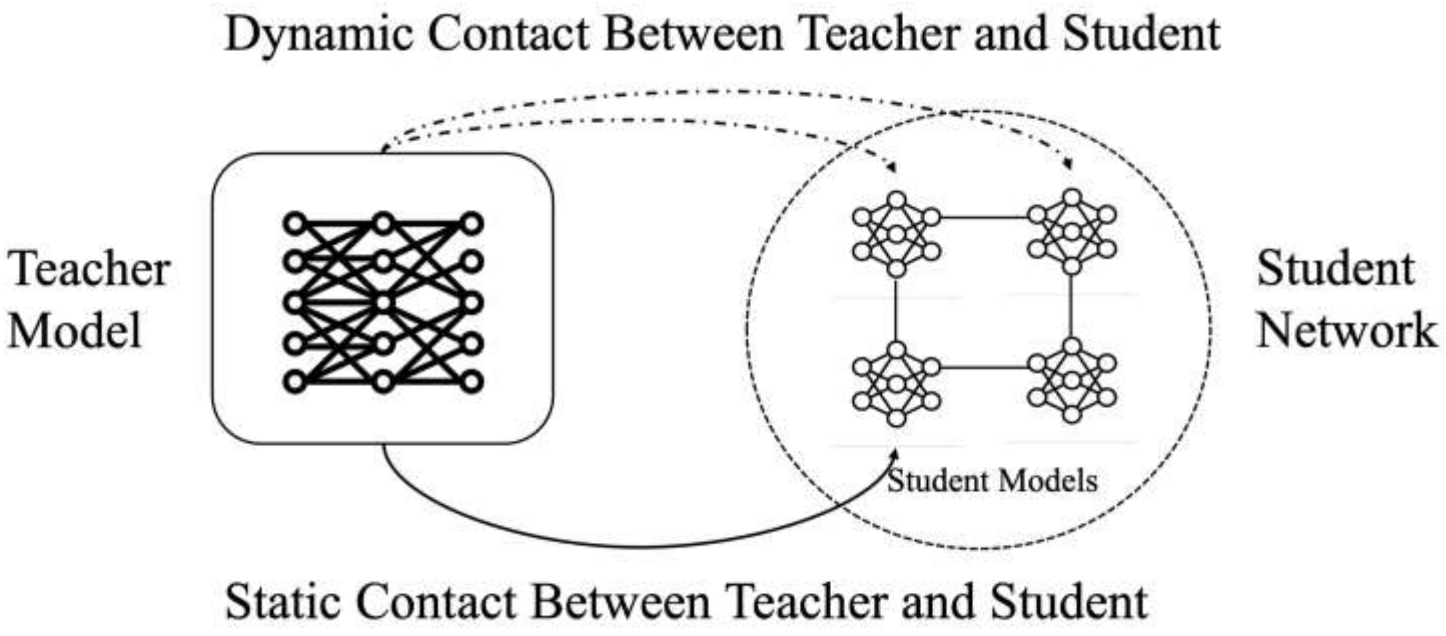
Traditional Knowledge Distillation

- Single teacher → single student
- One-shot interaction
- Static supervision
- Assumes continuous teacher availability

But Real Systems Are Different

- Teacher may be intermittently available
- Students operate in distributed environments
- Communication happens over structured networks
- Learning must continue even without the teacher

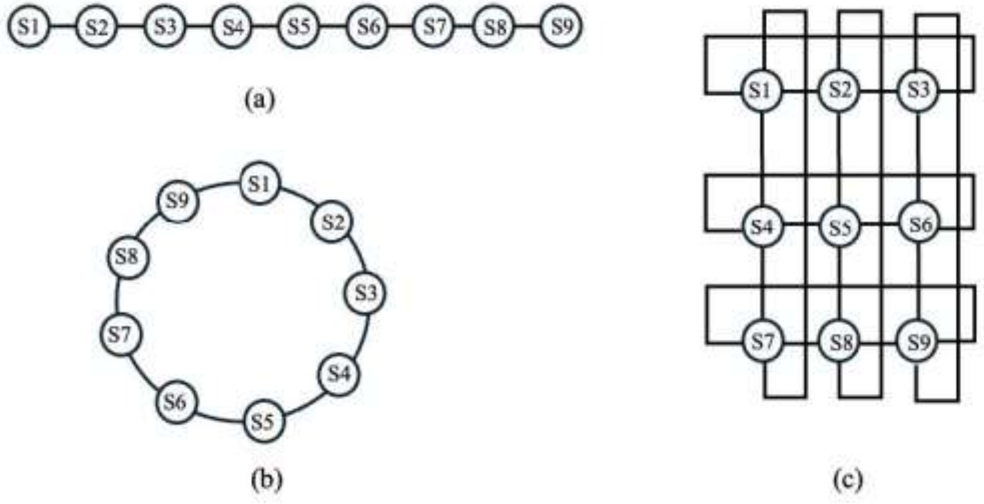
From Static Contact to Dynamic, Networked Distillation



- Static KD relies on one-shot supervision.
- Learning halts when teacher access is limited.
- We introduce multi-round dynamic reinforcement.
- Knowledge propagates through the student network in the form of peer-to-peer distillation.

How Does Network Structure Affect Knowledge Diffusion?

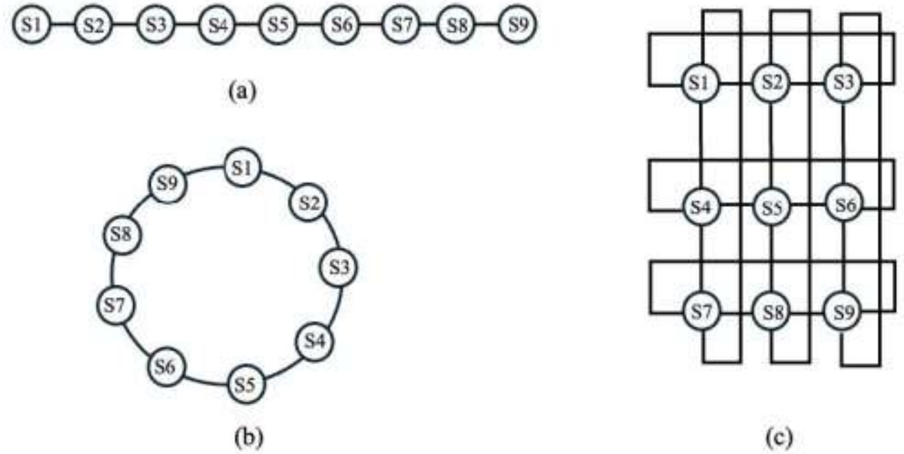
- Knowledge spreads hop-by-hop across the graph
- Diffusion speed depends on shortest-path distances
- Diameter upper-bounds full network propagation
- Lower diameter → Faster and more uniform convergence



- (a) Line topology — large diameter, long propagation paths
- (b) Ring topology — reduced diameter, bidirectional flow
- (c) Torus topology — minimal diameter, high connectivity

Topology and Graph Metrics

- **Diameter bounds worst-case diffusion time**
- **Eccentricity determines node-level influence radius**
- Line topology is structurally imbalanced and reinforcement-sensitive
- Ring topology is symmetric and more robust
- Torus topology minimizes both diameter and eccentricity
- Compact topologies enable faster and more uniform convergence



Topology	Diameter	Center	Min Ecc	Max Ecc
Line	8	5	4	8
Ring	4	5	4	4
Torus	2	5	2	2

Problem Formulation

- **Student Objective**

$$L_i^{(t)} = \alpha\, CE(f(w_i^{(t)}), y) + (1-\alpha)\, KL(f(w_i^{(t)}), g_i^{(t)})$$

- Where:
- First term: supervised learning
- Second term: distillation from peers + teacher
- α: trade-off parameter

- **Distillation Target**

$$g_i^{(t)} = \sum_{j \in \mathcal{N}_i} \gamma_{ij}^{(t)} f(w_j^{(t)}) + \gamma_{iT}^{(t)} f(w_T)$$

- Where:
- $\mathcal{N}_i$: neighbors of student i
- γ: topology-aware weighting
- Teacher influence is selective

Reinforcement Strategy and Optimization Objective

Global Objective

$$\max \sum_{i=1}^{N} A_i^{(t_{last})}$$

- Maximize final student accuracy

Selection Rule

$$S_{\text{reinforced}} = \arg\max_i R_i^{(t)}$$

• Reinforce highest-scoring student each round

Reinforcement Score

$$R_i^{(t)} = W_1(1 - A_i^{(t)}) + W_2 \frac{Deg_i}{Deg_{\max}} - W_3 Dist_i$$

• First term: improvement potential
• Second term: structural influence (degree / centrality)
• Third term: communication cost based on eccentricity

Global Optimization Objective and Constraints

Global Objective

$$\max \sum_{i=1}^{N} A_i^{(t_{last})}$$

- Maximize final student accuracy

Reinforcement Budget Constraint

$$|R^{(t)}| \leq R_{\max}$$

Limited teacher supervision per round

Communication Constraint

$$\sum_{S_i \in R^{(t)}} \kappa_i \leq C_{\max}$$

Total reinforcement cost must respect communication budget

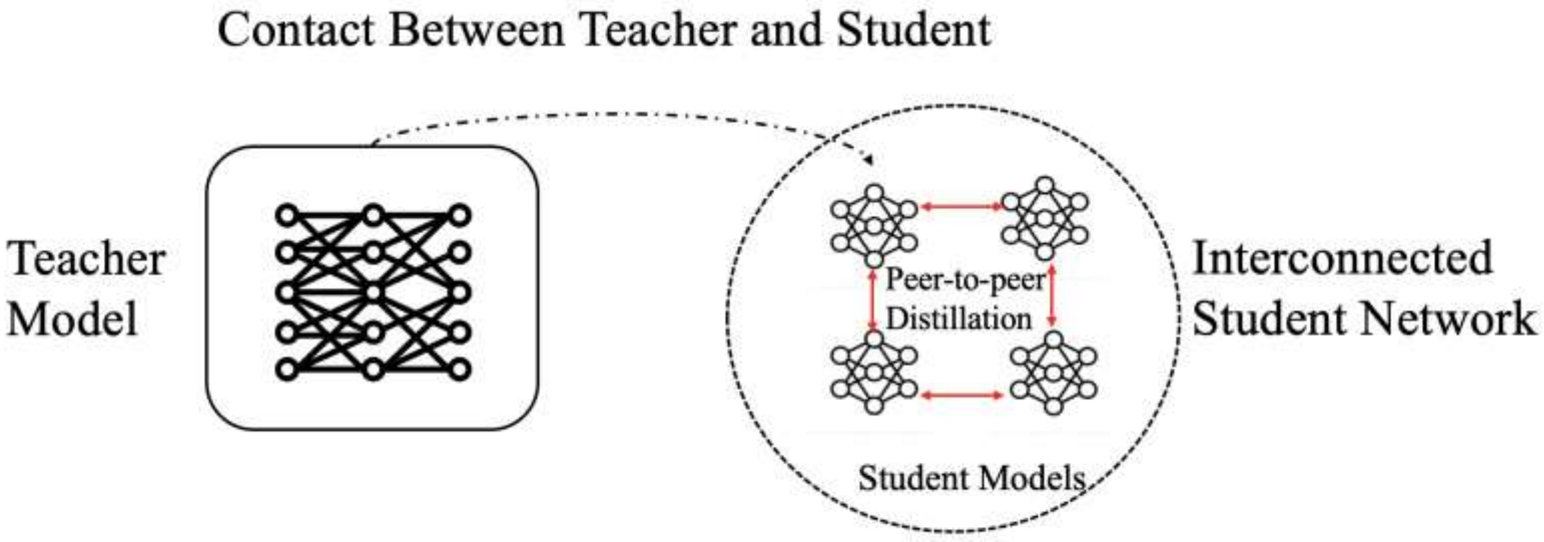
Monotonic Learning Assumption

$$A_i^{(t+1)} \geq A_i^{(t)}$$

Learning should be non-decreasing over rounds

Peer-to-Peer Propagation Mechanism

- Teacher reinforces one student
- Reinforced node updates with teacher supervision
- Knowledge spreads hop-by-hop
- Each round propagates one structural layer outward



Theoretical Diffusion Properties

- **Property 1**

Knowledge propagates one hop per round

$$\text{Rounds to reach student } j = d(S_c, j)$$

- **Property 2**

Full diffusion time equals eccentricity of reinforced node

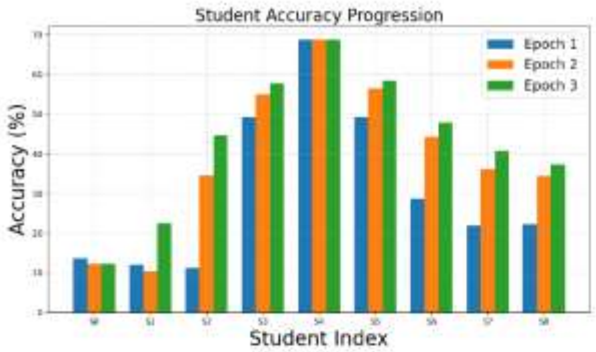
$$T = \epsilon(S_c)$$

- **Theorem 1**

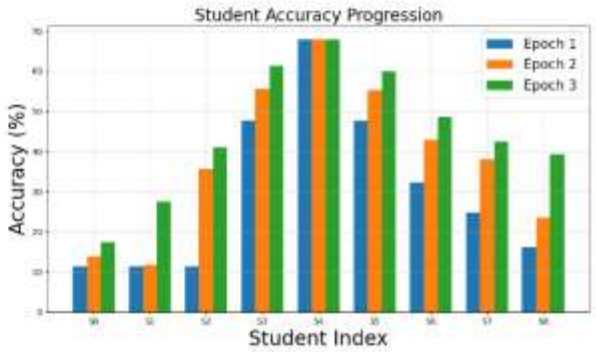
Eccentricity upper-bounded by diameter

$$T \geq \epsilon(S_c) \leq D$$

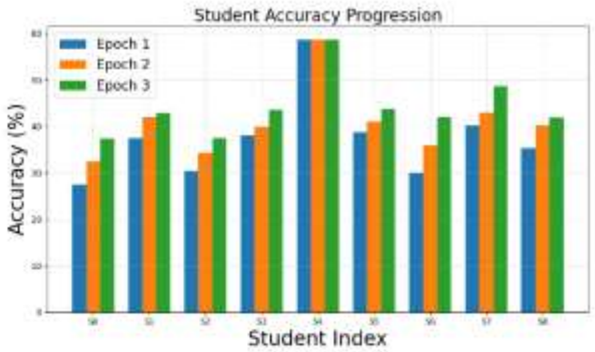
Simulation Results



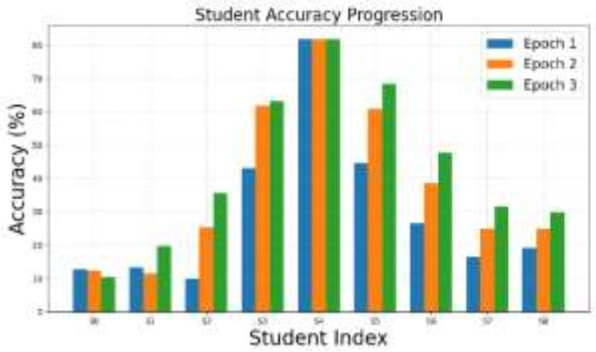
(a) Line Topology (CIFAR-10)



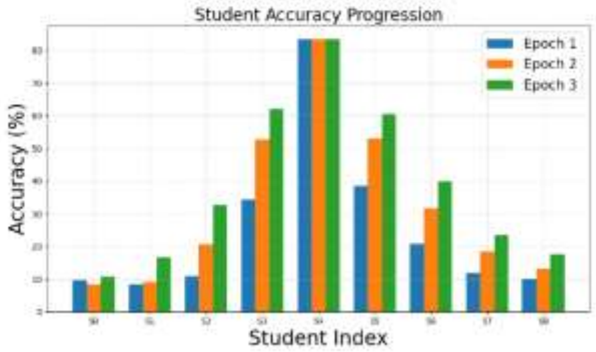
(b) Ring Topology (CIFAR-10)



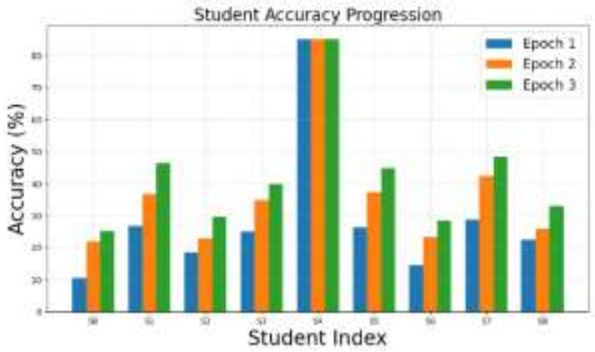
(c) Torus Topology (CIFAR-10)



(d) Line Topology (SVHN)



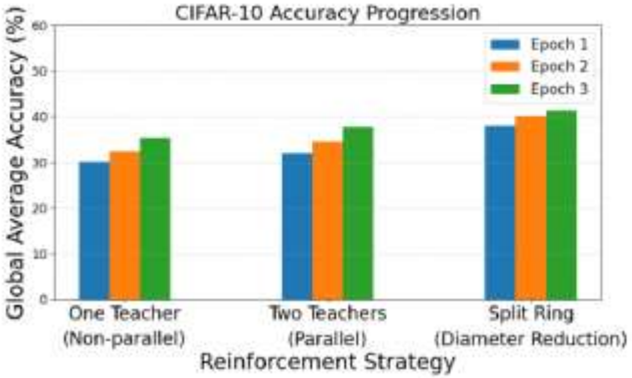
(e) Ring Topology (SVHN)



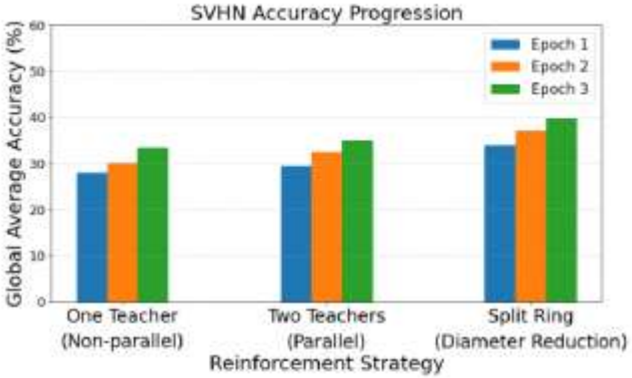
(f) Torus Topology (SVHN)

Student accuracy progression across epochs for different topologies

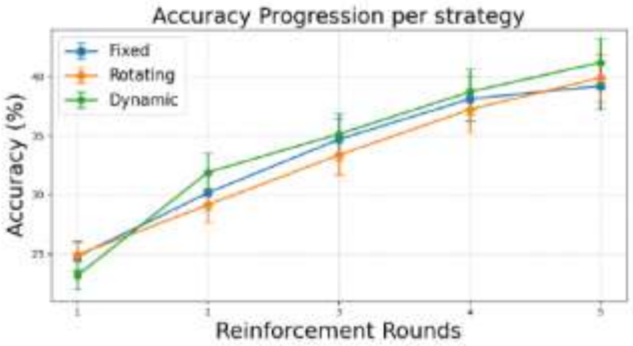
Simulation Results



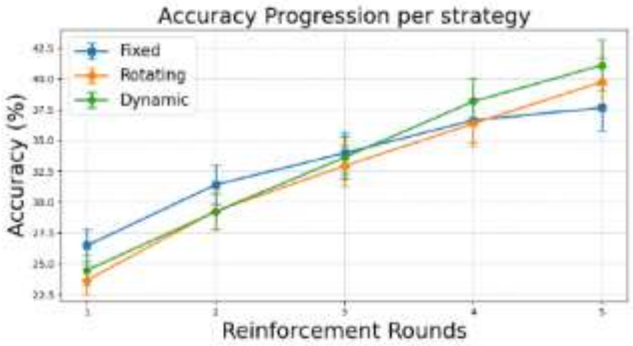
(a) CIFAR-10



(b) SVHN



(c) CIFAR-10



(d) SVHN

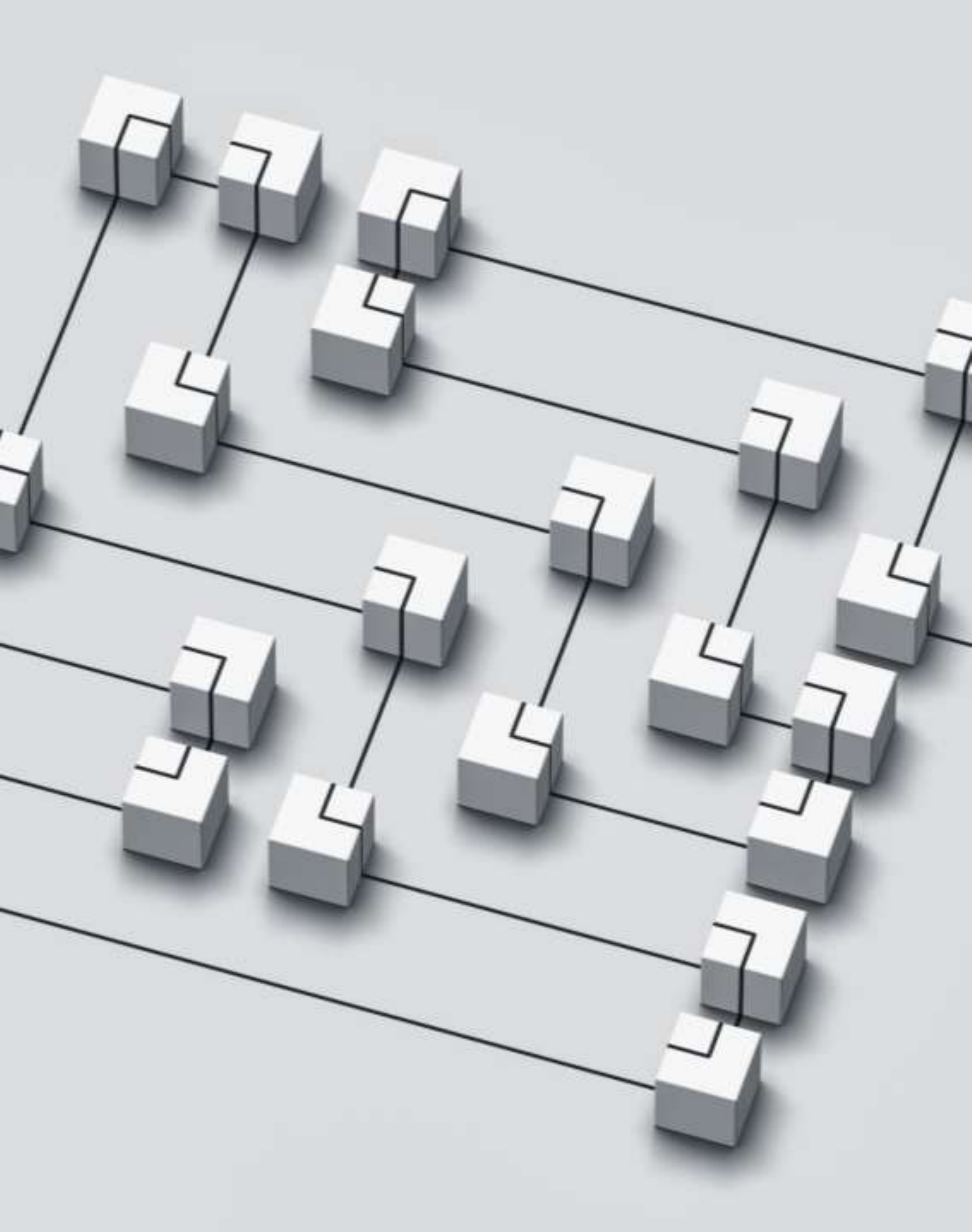
Data	Strategy	R1	R2	R3	R4	R5
CIFAR	Fixed	24.76	30.15	34.68	38.10	39.20
	Rotating	24.89	29.14	33.36	37.20	39.90
	Dynamic	23.13	31.89	35.13	38.70	41.20
SVHN	Fixed	26.48	31.41	33.97	36.63	37.64
	Rotating	23.67	29.29	32.93	36.36	39.74
	Dynamic	24.48	29.22	33.62	38.15	41.09

Global Accuracy per Round Across Strategies

Accuracy trends: (a–b) show the effect of decreasing diameter on ring topology; (c–d) compare fixed, rotating, and dynamic reinforcement on a line topology.

Key Experimental Insights

• Convergence speed strongly correlates with graph diameter
• Torus topology (D = 2) achieves fastest and most uniform diffusion
• Parallel reinforcement reduces effective propagation distance
• Dynamic reinforcement consistently yields highest global accuracy
• Results are consistent across CIFAR-10 and SVHN



Conclusion & Future Work

Conclusion

• Knowledge distillation is topology-dependent
• Diffusion time is bounded by diameter
• Reinforcement effectiveness depends on node position
• Topology-aware weighting improves convergence efficiency

Future Directions

• Larger and irregular network structures
• Stochastic or unreliable communication links
• Application to federated / edge systems
• Adaptive topology reconfiguration

Thank you!