

Graph Neural Network-Based Traffic Prediction and Topology Optimization for Optical Data Center Networks

Mieszko Serwa

School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, OR, USA.

contactmieszko@oregonstate.edu

Feline Molenaar

Department of Computer Science and Engineering, University of Nevada, Reno, Reno, NV, USA.

felinemolenaar105@unr.edu

Marissa Poole

Department of Computer Science and Engineering, University at Buffalo, Buffalo, NY, USA.
marissap@buffalo.edu

Arnulfo Tartini

Department of Computer Science, University of Central Florida, Orlando, FL, USA.
arnulfo.tartini@ucf.edu

Abstract

Optical data center networks are increasingly expected to support workloads that fluctuate rapidly across racks, pods, and geographically distributed facilities. Static or quasi-static topologies often waste capacity during off-peak periods and become congestion bottlenecks during synchronized training, storage migration, or bursty inference workloads. This paper examines how graph neural networks can be used not only for traffic prediction but also for topology optimization in optical data center networks. The central argument is that prediction and reconfiguration should be treated as a coupled socio-technical control problem rather than as separate machine learning and networking tasks. Graph representations can encode racks, switches, transceivers, wavelength channels, and service dependencies as interrelated entities, while temporal graph learning can capture evolving demand patterns. When integrated with reconfigurable optical switching, such models may inform circuit provisioning, wavelength assignment, load balancing, and failure-aware rerouting. However, the paper also emphasizes structural trade-offs involving latency, energy, hardware complexity, telemetry overhead, fairness across tenants, and governance of automated infrastructure. Through conceptual analysis, cross-domain comparison, and deployment-oriented discussion, the paper identifies architectural requirements for robust and sustainable optical data center control. It concludes that graph-based intelligence is most valuable when embedded within transparent, auditable, and policy-aware orchestration frameworks that balance performance gains against operational risk, tenant equity, and long-term infrastructure resilience.

Keywords

graph neural networks; optical data center networks; traffic prediction; topology optimization; reconfigurable optical interconnects; socio-technical infrastructure; sustainability; network governance.

1. Introduction

Optical data center networks have emerged as a critical response to the bandwidth, latency, and energy pressures created by cloud computing, large-scale machine learning, and distributed storage. Electrical packet switching remains essential for fine-grained traffic, but optical circuit and wavelength switching can provide high-capacity paths with lower per-bit energy consumption when traffic patterns are sufficiently stable or predictable [13]. The challenge is that modern data center workloads are rarely stable for long periods. Training jobs, storage rebalancing, virtual machine migration, and tenant-driven microservice communication can shift demand across the topology within seconds or minutes. Consequently, the value of optical reconfiguration depends heavily on the ability to anticipate traffic shifts before congestion materializes. This creates a natural opening for graph neural networks, which are designed to learn from relational data and can model the data center as a dynamic graph of computing, storage, and networking elements [1], [4].

The title of this paper positions graph neural networks at the intersection of traffic prediction and topology optimization. That intersection is important because prediction alone does not improve a network unless it is translated into actionable configuration decisions. Conversely, topology optimization without reliable prediction can produce fragile circuits that perform well under expected demand but fail under uncertainty. The paper therefore treats prediction and reconfiguration as a closed-loop system in which telemetry informs forecasting, forecasting informs topology decisions, and topology decisions generate new traffic patterns that reshape future telemetry. This loop has technical, economic, and governance dimensions. It affects how much telemetry is collected, who can access it, how automated decisions are audited, and how infrastructure sustainability is measured. Recent work on low-latency wavelength-switched clock-synchronized intra-data center interconnects enabled by hollow core fiber further illustrates that physical-layer advances can reduce latency budgets, but they also increase the importance of precise control and synchronization [7].

2. Background and Motivation

Optical data center architectures have evolved through several design philosophies. Hybrid electrical and optical designs such as Helios demonstrated that optical circuit switching can complement electrical packet switching by handling large elephant flows while leaving mice flows to electrical fabrics [14]. Other architectures explored optical switching fabrics with centralized or distributed control, including designs that use arrayed waveguide gratings, tunable lasers, and microsecond or nanosecond switching to create flexible interconnects [15], [16], [17]. RotorNet and related systems pursue simplified optical fabrics with scheduled reconfiguration, while Sirius and similar platforms investigate flat optical switching with very low latency [16], [17]. More recent work continues to examine how optical interconnection networks can reduce energy per bit and improve scalability, although challenges remain in buffering, synchronization, and control-plane complexity [20].

The motivation for graph-based learning arises from the mismatch between static topology design and dynamic workload behavior. Traditional traffic engineering often relies on aggregated link utilization and coarse demand matrices. These methods can be effective for stable traffic, but they struggle when demand is spatially correlated across racks and

temporally bursty across services. Traffic forecasting models based on recurrent neural networks, convolutional neural networks, and graph convolution have shown that spatial dependencies among road segments or sensors can improve prediction accuracy [6], [11]. Data center traffic has different semantics, but it shares a fundamental property: communication patterns are not independent across locations. A training job may synchronize across many accelerators, a database may replicate across racks, and a microservice may fan out to dozens of downstream dependencies. These relationships form a graph that changes over time, making graph neural networks a natural modeling choice [5].

3. Graph Representation of Optical Data Center Networks

A graph representation for an optical data center network must capture multiple layers of structure. At the physical layer, nodes might represent switches, transceivers, wavelength channels, or ports, while edges represent fibers, waveguides, or available optical circuits. At the logical layer, nodes might represent racks, servers, virtual machines, containers, or services, while edges represent observed communication flows or declared dependencies. At the operational layer, nodes might represent tenants, jobs, or service-level agreements, while edges represent resource sharing or policy constraints. A single graph is rarely sufficient. Instead, a multi-layer or heterogeneous graph is often more appropriate because it can encode the fact that a physical link failure affects logical services differently depending on redundancy, priority, and tenant isolation [2], [4].

Temporal graph learning adds another dimension. Traffic in a data center is not merely a snapshot of who talks to whom. It is a sequence of interactions shaped by job scheduling, diurnal patterns, maintenance windows, and failure events. A graph neural network can incorporate temporal attention or recurrent memory to learn how edge weights evolve and how node states propagate across the topology [3], [8]. For optical networks, this temporal dimension is especially important because reconfiguration has a cost. Switching a wavelength path may require configuration time, synchronization, and guard bands. A prediction model that anticipates a traffic shift too late is less useful than one that provides a stable forecast horizon aligned with the reconfiguration latency. Thus, graph representation is not only a modeling choice but also a control-design decision. It determines what can be predicted, how quickly, and with what operational meaning.

4. Graph Neural Network Models for Traffic Prediction

Graph neural network models for traffic prediction have developed rapidly in transportation research and can inform data center forecasting. Spatio-temporal graph convolutional networks combine graph convolution with temporal convolution to capture both spatial and temporal dependencies [6]. Temporal graph convolutional networks extend this idea with recurrent units that model dynamic changes over time [8]. Attention-based spatial-temporal graph convolutional networks use attention to weight the influence of neighboring nodes, which is useful when some links or racks are more important than others [9]. Gated residual recurrent graph neural networks and diffusion convolutional recurrent neural networks further show how graph diffusion and recurrent memory can capture long-range dependencies in traffic networks [10], [11]. More recent spatial-temporal graph neural networks aim to unify heterogeneous dependencies and improve scalability [12].

For optical data center networks, these models must be adapted to a different physical and operational context. Road traffic is constrained by geography, whereas data center traffic is constrained by application placement, tenant policy, and switching hierarchy. A graph model

that treats all racks as equally adjacent may miss the fact that some racks share a top-of-rack switch, while others communicate through aggregation and core layers. Similarly, optical circuit availability depends on wavelength continuity, transceiver tuning range, and switching fabric state. The forecasting model must therefore be conditioned on the current topology and the feasible reconfiguration space. A recent study on low-latency wavelength-switched clock-synchronized intra-data center interconnects enabled by hollow core fiber highlights how physical-layer timing and synchronization can shape the practical envelope for such control [7]. Prediction accuracy alone is insufficient if the model does not respect the constraints of the optical layer.

5. Topology Optimization and Reconfiguration

Topology optimization in optical data center networks can be understood as a sequence of decisions about which logical and physical connections should be active, how much capacity they should carry, and how long they should remain in place. In hybrid electrical and optical fabrics, the optimization problem often involves deciding which flows should be routed over optical circuits and which should remain on the electrical packet network [14]. In wavelength-switched architectures, the problem includes wavelength assignment, path selection, and reconfiguration scheduling [15], [17]. In scheduled optical fabrics such as RotorNet, the topology may change according to a predetermined sequence, while more dynamic systems may attempt demand-aware reconfiguration [16]. Each approach embodies a different trade-off between predictability, hardware complexity, and control overhead.

Graph neural networks can support topology optimization by predicting future edge demands, node congestion, or failure probabilities. These predictions can then be used to rank candidate reconfigurations. However, the optimization objective is rarely singular. Operators may seek to minimize latency, maximize throughput, reduce energy, preserve fairness, or maintain resilience. A reconfiguration that improves average latency may harm tail latency for a latency-sensitive tenant. A topology that reduces energy by consolidating traffic onto fewer optical paths may create a single point of failure. A control policy that reacts aggressively to predicted bursts may cause oscillation if predictions are noisy. Therefore, graph-based prediction must be coupled with multi-objective optimization and stability constraints. The system should prefer reconfigurations that are robust across plausible forecasts rather than optimal for a single predicted future [17], [18].

6. System Architecture and Deployment Considerations

A practical system architecture for graph-based prediction and topology optimization requires several layers. The telemetry layer collects measurements from switches, transceivers, servers, and applications. The graph construction layer transforms these measurements into nodes, edges, and features while respecting privacy and tenant isolation. The prediction layer applies graph neural networks to forecast demand and anomalies. The optimization layer evaluates candidate topologies and generates reconfiguration commands. The control layer executes these commands through software-defined networking or vendor-specific interfaces. The assurance layer monitors outcomes and detects drift, failures, and policy violations [4], [18]. This architecture is not merely a software stack. It is an operational regime that determines who can see what, who can change what, and how automated decisions are justified.

Deployment considerations vary by facility scale and ownership. Hyperscale operators may have custom hardware, detailed telemetry, and centralized control, while colocation providers may need to coordinate across tenants and equipment vendors. Edge data centers may have

fewer optical paths and more variable workloads. In all cases, the control plane must handle partial observability. Some traffic may be encrypted, some switches may expose limited counters, and some applications may not declare dependencies. Graph neural networks can tolerate missing data to some degree, but deployment requires explicit uncertainty estimates. A topology change should not be triggered solely by a high-confidence point forecast. It should also consider confidence intervals, worst-case scenarios, and fallback paths. The hollow core fiber interconnect work cited earlier reinforces the need for tight synchronization and low-latency control, because faster physical paths can make control-plane delays more visible [7].

7. Robustness, Fairness, and Governance

Robustness is a central concern when machine learning is used to reconfigure physical infrastructure. A graph neural network trained on historical traffic may encounter distribution shifts caused by new applications, hardware failures, cyberattacks, or policy changes. If the model silently degrades, the network may make poor reconfiguration decisions. Robustness therefore requires redundancy in prediction, conservative fallback modes, and continuous validation against operational baselines. It also requires adversarial thinking. An attacker might manipulate traffic patterns to induce harmful reconfigurations, or a tenant might intentionally create bursts to gain disproportionate optical capacity. Graph-based models that learn only from observed traffic may reproduce these strategic behaviors unless fairness and security constraints are embedded in the optimization objective [10], [19].

Fairness in optical data center networks is multifaceted. At the tenant level, fairness may mean preventing one tenant from monopolizing reconfigurable optical paths. At the service level, fairness may mean protecting latency-sensitive applications even when they generate less traffic. At the infrastructure level, fairness may mean distributing reconfiguration burden across switches and transceivers to avoid premature wear. Governance must address who is accountable when an automated topology change causes an outage or violates a service-level agreement. This requires audit logs, explainable predictions, human override mechanisms, and clear escalation paths. A graph neural network can provide useful signals, but it should not become an opaque authority over critical infrastructure. The socio-technical framing is essential because optical data centers are not just technical systems; they are sites of economic exchange, regulatory obligation, and organizational trust [4], [5].

8. Sustainability and Policy Implications

Sustainability is increasingly central to data center design. Optical interconnects can reduce energy per bit compared with electrical alternatives, but they also introduce new manufacturing, cooling, and disposal considerations. Reconfigurable optical components may have different lifetime and reliability profiles than fixed electrical links. A topology optimization system that maximizes throughput without considering embodied energy may produce short-term efficiency gains while increasing long-term material consumption. Graph neural networks can incorporate energy-related features such as switch utilization, transceiver power states, and cooling load, but doing so requires standardized telemetry and lifecycle accounting [13], [20]. The optimization objective should therefore include energy intensity, carbon intensity of the local grid, and hardware wear rather than only latency and throughput.

Policy implications follow from these technical choices. Governments and industry bodies may require data centers to report energy use, carbon emissions, and water consumption. If automated topology optimization shifts load across regions, it may also shift emissions and

resource burdens. Transparency is necessary to ensure that efficiency claims are verifiable. Operators may need to document how machine learning models influence infrastructure reconfiguration and whether those models consider tenant equity or only aggregate performance. Standardization could help by defining interfaces for telemetry, reconfiguration, and audit. Without such standards, graph-based control may remain proprietary and difficult to compare across vendors. A sustainable optical data center strategy should therefore treat prediction and topology optimization as part of a broader governance framework that includes energy policy, competition policy, and infrastructure resilience [17], [20].

9. Case Illustrations and Cross-Domain Comparisons

Cross-domain comparisons can clarify both the promise and the limits of graph neural networks for optical data centers. In transportation, graph-based forecasting has benefited from abundant sensor data, well-defined road networks, and recurring congestion patterns [6], [8]. In supply chains, graph learning can model dependencies among suppliers, factories, and distribution centers, but it must handle confidential data and strategic behavior. In power grids, graph learning can support load forecasting and topology control, but safety constraints are stringent and failures can be catastrophic. Optical data centers resemble all three domains in different ways. They have recurring traffic patterns like transportation, dependency structures like supply chains, and critical-infrastructure constraints like power grids. However, they also have unique properties such as rapid workload migration, tenant isolation, and programmable switching [19].

These comparisons suggest that graph neural networks should not be imported without adaptation. A model that works well for road traffic may not capture the bursty, synchronized nature of distributed training. A model that works well for power grids may be too conservative for commercial data center optimization. The most promising approach is hybrid. Graph learning can provide demand forecasts and anomaly signals, while rule-based or optimization-based controllers enforce hard constraints. Human operators can review high-impact reconfigurations. This hybrid design aligns with the broader lesson from graph neural network surveys: expressive models are valuable, but their deployment depends on data quality, interpretability, and integration with domain constraints [4], [5], [19].

10. Open Challenges and Research Agenda

Several open challenges remain. First, data scarcity and privacy limit the construction of rich graphs. Tenant traffic may be encrypted or aggregated, and cross-tenant dependencies may be hidden. Federated or privacy-preserving graph learning could help, but it introduces communication overhead and trust assumptions. Second, generalization is difficult. A model trained in one data center may not transfer to another with different topology, workload mix, or hardware. Transfer learning and meta-learning may help, but they require careful validation. Third, explainability is essential for operators. A prediction that cannot be explained is difficult to trust when it triggers a topology change. Attention mechanisms and counterfactual explanations may improve transparency, but they also add complexity [3], [4].

Fourth, standardization and interoperability remain underdeveloped. Optical switches, transceivers, and network operating systems expose different interfaces. A graph-based controller must integrate with heterogeneous equipment. Fifth, evaluation is challenging. Simulation may not capture real hardware timing, while production experiments risk service disruption. Digital twins and shadow modes can provide safer testing environments. Sixth, sustainability metrics need refinement. Energy per bit is useful but incomplete without

lifecycle accounting. Finally, governance must keep pace with automation. Operators, regulators, and tenants need shared expectations about fairness, accountability, and recourse. Addressing these challenges will require collaboration across machine learning, optical networking, systems engineering, and policy [19], [20].

11. Conclusion

Graph neural network-based traffic prediction and topology optimization offer a promising path toward more adaptive, efficient, and resilient optical data center networks. By representing racks, switches, wavelengths, services, and tenants as interrelated graph entities, these models can capture spatial and temporal dependencies that traditional traffic engineering often misses. When coupled with reconfigurable optical switching, they can inform circuit provisioning, wavelength assignment, load balancing, and failure-aware rerouting. However, the value of graph intelligence depends on how it is embedded in system architecture and governance. Prediction must respect physical-layer constraints, reconfiguration must be stable and auditable, and optimization must balance performance with fairness, robustness, and sustainability. The paper has argued that optical data centers should be understood as socio-technical infrastructures in which machine learning is one component of a larger control and accountability regime. Future research should therefore focus on hybrid control, privacy-preserving graph learning, standardized telemetry, lifecycle-aware optimization, and participatory governance. If these conditions are met, graph neural networks can help optical data centers respond to dynamic demand without sacrificing the resilience and equity on which critical digital infrastructure depends.

References

1. Kipf, T. N., & Welling, M. (2017). Semi-supervised classification with graph convolutional networks. In International Conference on Learning Representations.
2. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., & Bengio, Y. (2018). Graph attention networks. In International Conference on Learning Representations.
3. Hamilton, W. L., Ying, R., & Leskovec, J. (2017). Inductive representation learning on large graphs. In Advances in Neural Information Processing Systems (pp. 1024–1034).
4. Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., & Yu, P. S. (2021). A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 32(1), 4–24.
5. Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., Wang, L., Li, C., & Sun, M. (2020). Graph neural networks: A review of methods and applications. *AI Open*, 1, 57–81.
6. Yu, B., Yin, H., & Zhu, Z. (2018). Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. In Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (pp. 3634–3640).
7. Zhou, Zichuan, et al. "Low-latency wavelength-switched clock-synchronized intra-data center interconnects enabled by hollow core fiber." *Optics Express* 31.15 (2023): 24739-24748.
8. Zhao, L., Song, Y., Zhang, C., Liu, Y., Wang, P., Lin, T., Deng, M., & Li, H. (2020). T-GCN: A temporal graph convolutional network for traffic prediction. *IEEE Transactions on Intelligent Transportation Systems*, 21(9), 3848–3858.

9. Guo, S., Lin, Y., Feng, N., Song, C., & Wan, H. (2019). Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 33, No. 1, pp. 922–929).
10. Chen, C., Li, K., Teo, S. G., Zou, X., Wang, K., Wang, J., & Zeng, Z. (2019). Gated residual recurrent graph neural networks for traffic prediction. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 33, No. 1, pp. 485–492).
11. Li, Y., Yu, R., Shahabi, C., & Liu, Y. (2018). Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. In International Conference on Learning Representations.
12. Wang, X., Ma, Y., Wang, Y., Jin, W., Wang, X., Tang, J., Jia, C., & Yu, J. (2020). Traffic flow prediction via spatial temporal graph neural network. In Proceedings of The Web Conference 2020 (pp. 1082–1092).
13. Kachris, C., & Tomkos, I. (2012). A survey on optical interconnects for data centers. *IEEE Communications Surveys & Tutorials*, 14(4), 1021–1036.
14. Zhou, D., & Yi, X. (2026). DISTRIBUTED AI-ASSISTED RISK ASSESSMENT FOR ZERO-TRUST MICROSERVICES: A TAXONOMY AND CONCEPTUAL FRAMEWORK. *AI and Data Science Journal*, 3(2), 1-13.
15. Chen, K., Singla, A., Singh, A., Ramachandran, K., Xu, L., Zhang, Y., Wen, X., & Chen, Y. (2014). OSA: An optical switching architecture for data center networks with unprecedented flexibility. *IEEE/ACM Transactions on Networking*, 22(2), 498–511.
16. Mellette, W. M., McGuinness, R., Roy, A., Forencich, A., Papen, G., Snoeren, A. C., & Porter, G. (2017). RotorNet: A scalable, low-complexity, optical datacenter network. In Proceedings of the Conference of the ACM Special Interest Group on Data Communication (pp. 267–280).
17. Ballani, H., Costa, P., Behrendt, R., Cletheroe, D., Haller, I., Jozwik, K., Karinou, F., Lange, S., Li, K., Mellette, W. M., Pappa, G., Popa, D., Sun, Y., Thomsen, B., & Williams, C. (2020). Sirius: A flat datacenter network with nanosecond optical switching. In Proceedings of the Annual Conference of the ACM Special Interest Group on Data Communication (pp. 782–797).
18. Porter, G., Strong, R., Farrington, N., Forencich, A., Chen-Sun, P., Rosing, T., & Vahdat, A. (2013). Integrating microsecond circuit switching into the data center. *ACM SIGCOMM Computer Communication Review*, 43(4), 447–458.
19. Jiang, W., & Luo, J. (2022). Graph neural network for traffic forecasting: A survey. *Expert Systems with Applications*, 207, 117921.
20. Kachris, C., Kanonakis, K., & Tomkos, I. (2013). Optical interconnection networks in data centers: Recent trends and future challenges. *IEEE Communications Magazine*, 51(9), 39–45.