Network Management with Graph Machine Learning: Challenges and Solutions

Yu Wang (Vanderbilt University) Ramakrishnan Durairajan (University of Oregon)

The exponential growth of network traffic and attacks, combined with operator scarcity, underscores the crucial role of machine learning (ML) for network management (e.g., detecting the onset of a DDoS attack, identifying network congestion, etc.). The application of ML for network performance and security tasks is henceforth referred to as ML4Nets. Today, ML4Nets is applied for a wide variety of network management tasks, including traffic classification, anomaly detection, and beyond [1, 2, 3, 4]. While great progress has been made in solving some of the key management tasks in networking, we posit that all these ML4Nets efforts focus *only* on the network's forwarding behavior (e.g., traffic patterns).

Ideally, networks can be conceptualized as graphs, with nodes representing devices (e.g., switches, routers, etc.) and edges denoting links (e.g., fiber-optic cables) or traffic paths [5, 6]. Such a graph-based representation not only captures the forwarding behavior (i.e., intricate traffic interactions) of a network but also the network connectivity (i.e., topological behavior), opening a new dimension for network management. For example, optical topology-aware traffic engineering has been shown to be effective against fiber cuts and flash crowds [7]. This makes graph machine learning (GML) an ideal tool for analyzing traffic patterns, attacks, and automating decision-making processes.

GML learns the node and edge embeddings by iteratively applying message-passing to capture topological patterns and neural transformations to extract task-specific insights from attributes [8, 9, 10]. These *learned embeddings* can then be applied for various tasks such as cyber-attack detection [11, 12] and traffic restoration [13, 14]. Despite these advancements, two key challenges persist which will be the focus of this talk.

First, a comprehensive exploration of graph construction methods and datasets for communication networks is lacking. Effective GML relies on appropriately constructed graphs tailored to specific applications. We address this gap by reviewing three prevalent graph construction techniques: the original network structure, the line graph capturing link homophily, and the byte-level correlation graph for identifying similar traffic patterns [5, 6, 11, 13, 15]. We analyze their advantages, disadvantages, and applications such as network disintegration, encrypted flow classification, and intrusion detection.

Second, the reliability and explainability of GML predictions require further investigation to achieve network operator buy-in. We review current research on explaining GML predictions [16] and quantifying the outcome uncertainty [17, 18]. Additionally, we apply a graph explanation framework, GNNExplainer [16], to understand traffic classification through flow subgraph and attribute mask analysis.

By addressing these challenges, our vision is to equip the network operators and researcher communities with improved graph construction techniques and a deeper understanding of graph-structured datasets for more effective GML4Nets. Furthermore, by exploring network explanation and uncertainty quantification methods, we strive to enhance the trustworthiness of GML in network management tasks.

References

[1] Calvin Ardi, Connor Aubry, Brian Kocoloski, Dave DeAngelis, Alefiya Hussain, Matt Troglia, and Stephen Schwab. The darpa searchlight dataset of application network traffic. In *Proceedings of the 15th Workshop on Cyber Security Experimentation and Test*, CSET '22, New York, NY, USA, August 2022. Association for Computing Machinery.

[2] Yukhe Lavinia, Ramakrishnan Durairajan, Reza Rejaie, and Walter Willinger. Challenges in using ml for networking research: How to label if you must. In *Proceedings of the Workshop on Network Meets AI & ML*, pages 21–27, 2020.

[3] Jared Knofczynski, Ramakrishnan Durairajan, and Walter Willinger. Arise: A multitask weak supervision framework for network measurements. *IEEE Journal on Selected Areas in Communications*, 40(8), 2022.

[4] Wenbo Zheng, Chao Gou, Lan Yan, and Shaocong Mo. Learning to classify: A flow-based relation network for encrypted traffic classification. In *Proceedings of The Web Conference 2020*, pages 13–22, 2020.

[5] Xingguo Ji and Qingmin Meng. Traffic classification based on graph convolutional network. In *2020 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA)*, pages 596–601. IEEE, 2020.

[6] Jingwei Zheng and Dagang Li. Gcn-tc: combining trace graph with statistical features for network traffic classification. In *ICC 2019-2019 IEEE International Conference on Communications (ICC)*. IEEE, 2019.

[7] Matthew Nance-Hall, Paul Barford, Klaus-Tycho Foerster, and Ramakrishnan Durairajan. Improving scalability in traffic engineering via optical topology programming. *IEEE Transactions on Network and Service Management*, 2023.

[8] Yu Wang and Tyler Derr. Tree decomposed graph neural network. In *Proceedings of the 30th ACM international conference on information & knowledge management*, pages 2040–2049, 2021.

[9] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

[10] Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, Yoshua Bengio, et al. Graph attention networks. *stat*, 1050(20):10–48550, 2017.

[11] Xiaoheng Deng, Jincai Zhu, Xinjun Pei, Lan Zhang, Zhen Ling, and Kaiping Xue. Flow topology-based graph convolutional network for intrusion detection in label-limited iot networks. *IEEE Transactions on Network and Service Management*, 20(1):684–696, 2022.

[12] Yuzhen Li, Renjie Li, Zhou Zhou, Jiang Guo, Wei Yang, Meijie Du, and Qingyun Liu. Graphddos: Effective ddos attack detection using graph neural networks. In *2022 IEEE 25th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, pages 1275–1280. IEEE, 2022.

[13] Marco Grassia, Manlio De Domenico, and Giuseppe Mangioni. Machine learning dismantling and early-warning signals of disintegration in complex systems. *Nature communications*, 12(1):5190, 2021.

[14] Changjun Fan, Li Zeng, Yizhou Sun, and Yang-Yu Liu. Finding key players in complex networks through deep reinforcement learning. *Nature machine intelligence*, 2(6):317–324, 2020.

[15] Haozhen Zhang, Le Yu, Xi Xiao, Qing Li, Francesco Mercaldo, Xiapu Luo, and Qixu Liu. Tfe-gnn: A temporal fusion encoder using graph neural networks for fine-grained encrypted traffic classification. In *Proceedings of the ACM Web Conference 2023*, pages 2066–2075, 2023.

[16] Zhitao Ying, Dylan Bourgeois, Jiaxuan You, Marinka Zitnik, and Jure Leskovec. Gnnexplainer: Generating explanations for graph neural networks. *Advances in neural information processing systems*, 32, 2019.

[17] Xujiang Zhao, Feng Chen, Shu Hu, and Jin-Hee Cho. Uncertainty aware semi-supervised learning on graph data. *Advances in Neural Information Processing Systems*, 33:12827–12836, 2020.

[18] Kexin Huang, Ying Jin, Emmanuel Candes, and Jure Leskovec. Uncertainty quantification over graph with conformalized graph neural networks. *Advances in Neural Information Processing Systems*, 36, 2024.