Machine Learning for Graph Data Management and Query Processing

Hanchen Wang
University of Technology Sydney
Sydney, Australia
hanchen.wang@uts.edu.au

Ying Zhang*
Common Prosperity Visualization and Policy Simulation Lab of Zhejiang Gongshang University
Hangzhou, China
ying.zhang@zjgsu.edu.cn

Wenjie Zhang
University of New South Wales
Sydney, Australia
wenjie.zhang@unsw.edu.au

ABSTRACT

Machine learning techniques have been proposed to optimize the performance of graph databases in recent years. Due to the NP-hardness of graph database tasks and the complexity of graph data, traditional exact solutions usually encounter efficiency issues, while the performance of approximation solutions can be affected by issues like sampling failure and local optimality. Empowered by the inherent advantages of machine learning, the learning-based techniques show the generalization ability and better performance in many scenarios, including graph data management and graph query processing. Despite the efficiency and accuracy brought by machine learning techniques, machine learning for graph database models still face several critical challenges, including scalability and adaptability. In this tutorial, we first provide an in-depth survey of learning-based graph data management and query processing techniques published in recent database and data mining conferences to sketch the frontier of the research of Machine Learning for Graph Database. We also discuss the open challenges and provide future directions.

PVLDB Reference Format:
Hanchen Wang, Ying Zhang, and Wenjie Zhang. Machine Learning for Graph Data Management and Query Processing. PVLDB, 18(12): 5499 - 5503, 2025.
doi:10.14778/3750601.3750702

PVLDB Artifact Availability:
The source code, data, and/or other artifacts have been made available at https://github.com/DaIL-UTS/Awesome-ML4GDB-Papers.

1 INTRODUCTION

Graph data management and query processing are fundamental to modern data science and data engineering. Traditional graph data management and query processing techniques often rely on heuristics or manual tuning, which makes it challenging to adapt to the dynamic requirements of real-life graph data management and query processing, including evolving graph structures, diverse query workloads, low data quality, high computational complexity, and real-time response requirements. Machine learning (ML) has emerged as a powerful paradigm to address these challenges by leveraging data-driven insights. For instance, ML models can accelerate query planning through reinforcement learning and provide accurate cardinality estimation using deep learning. However, applying ML to graph databases introduces unique challenges, such as *scalability* and *adaptability*. Existing techniques often lack scalability for large graphs and fail to generalize across different tasks. This tutorial systematically reviews recent advancements in ML for graph databases, covering techniques for data management (e.g., data quality management and graph generation) and query processing (e.g., subgraph isomorphism, graph similarity computation, and community search and detection), while highlighting open challenges and future directions.

Tutorial Overview. This tutorial will be an 1.5-hour for machine learning for graph data management and query processing. The program of this tutorial is as follows:

Background and Foundations (15 Minutes). We first introduce the background and foundations for this tutorial. Specifically, we will provide the background information about graph database, introduce critical tasks in graph database, and highlight the motivation of utilizing machine learning techniques for graph data management and query processing.

Machine Learning for Graph Data Management (25 Minutes). We review the machine learning-based techniques for graph data management in this section. Specifically, we include machine learning techniques for graph data quality management [3, 15, 24, 34, 41, 43, 44] and graph generation [7, 8, 16–18, 23, 25, 35].

Machine Learning for Graph Query Processing (30 Minutes). Next, we summarize the machine learning-based techniques for popular graph query processing tasks, including subgraph isomorphism [6, 27–29, 40], graph similarity computation [1, 9, 10, 14, 19–21, 39, 46], and community detection and search [22, 30, 31].

Open Challenges and Future Directions (20 Minutes). This section covers the open challenges and future directions of learning-based graph data management and query processing. We will discuss the challenges of existing techniques, including the ability to handle large-scale graphs and the capability of being applied to various tasks. We also provide future directions with a discussion about the utilization of foundation models, integrating graph prompt learning methods, and combining database and machine learning techniques to improve scalability.

Target audience and assumed background. This tutorial is proposed for the audience, including VLDB attendees from both academic and industry communities who are interested in machine

*Corresponding Author

Proceedings of the VLDB Endowment, Vol. 18, No. 12 ISSN 2150-8097.
doi:10.14778/3750601.3750702

Table 1: Summary of Machine Learning for Graph Database Models

Objective	Task	Category	Models
Graph Data Management	Graph Data Quality Management	Data quality assessment	[3] [15] [43]
		Data quality enhancement	[24] [34] [41] [44]
	Graph Generation	Embedding similarity-based models	[7] [16] [18] [35]
		Statistics-based models	[17] [23] [25]
Graph Query Processing	Subgraph Isomorphism	Subgraph matching	[29] [40]
		Subgraph counting	[6] [27]
	Graph Similarity	Metric-agnostic models	[1] [14] [20]
		Metric-specific models	[19] [21] [46] [10] [9]
	Community Search	Community search	[30] [31]
		Overlapping community search	[22]

learning for graph database, graph data management, and graph query processing. This tutorial requires basic prior background knowledge in graph computation and graph database.

Recent related tutorials. There are recent tutorials about machine learning for database [2, 13]. However, these tutorials focus on relational database and database systems, rather than graph database. There are tutorials about machine learning for subgraph extraction [42] and graph generation [26]. This tutorial will cover broader contents about machine learning for graph data management and query processing. There are other tutorials that introduce the development of machine learning techniques for database [12, 38, 45]. These tutorials offer comprehensive overviews of machine learning applications in specific database tasks, whereas this one focuses on presenting learning-based advancements in graph database areas.

2 TUTORIAL OUTLINE

This is a 1.5-hour tutorial designed for VLDB attendees who are interested in machine learning for database, graph data management, and graph query processing. Below is the outline of the tutorial.

2.1 Background and Introduction

We first give the background knowledge required for this tutorial. First, the definitions of graph data management and query processing will be introduced, and the major tasks covered in this tutorial, including data quality management, graph data generation, and graph query optimization, will be introduced. Next, traditional solutions for graph data management and query processing will be provided, along with the motivation of developing machine learning techniques.

In this tutorial, we comprehensively introduce machine learning techniques for graph databases from two perspectives, i.e., graph data management and graph query processing. We provide a the categorization of the existing machine learning for graph database techniques, and summarize them in Table 1. Following the background knowledge and brief introduction, we give the detailed review of machine learning for graph database techniques.

2.2 Machine Learning for Graph Data Management

Graph Data Quality Management. Due to various quality issues, the graph data quality management is a crucial aspect of graph data management. In this tutorial, we introduce the machine learning-based graph data quality management techniques from two perspectives, data quality assessment and data quality improvement.

Data quality assessment. We focus on one major task, error detection, for data quality assessment in this tutorial. Specifically, the following representative works [3, 15, 43] for error detection on graph data will be introduced.

Data quality improvement. We focus on a specific data quality improvement task in this tutorial, i.e., missing data imputation. Because of the scarcity of real-life graph data, missing data imputation plays an important role in graph data management. In this tutorial, we introduce several advanced graph missing data imputation techniques along with a comprehensive overview [37] of the research in the area. Specifically, the learning-based models [24, 34, 41, 44] designed for missing data imputation will be introduced. We also introduce UnIMP, proposed in [32], which models the tabular data as a hypergraph and jointly utilizes both the graph neural network and LLM for imputation.

Graph Generation. Graph generation [4, 36], one of the most fundamental problems in graph data management and analytics, aims to generate new graphs from the approximate distribution of observed graph data. In this tutorial, we focus on similarity-based models which are optimized for structural and distributional fidelity between synthesized graphs and their real-world counterparts. The similarities are measured through metrics comparing embedding similarities and overall graph statistics.

Embedding similarity-based models. First, the graph generation model based on embedding similarities will be discussed. The similarity between the representations of the generated graph and the actual graph will be used as a training objective. Several representative works for graph generation based on similarities in graph embedding space will be introduced. The Variational Autoencoder (VAE) [11], recognized as a prominent generative learning framework, is extensively employed in graph generation [7, 16, 18, 35].

Statistics-based models. Numerous graph generation models also incorporate statistical similarity measurements between graphs. Recent advancements [17, 23, 25] have introduced novel metrics for assessing the similarity between generated and real graphs. These include maximum mean discrepancy (MMD), and Kullback-Leibler divergence, among others. These metrics provide a more nuanced evaluation of the generated graphs, considering both the global structure and local node-level features.

2.3 Machine Learning for Graph Query Processing

Subgraph Isomorphism. Subgraph isomorphism, a fundamental graph query processing task, aims to find subgraphs of a data graph that are isomorphic to the query graph. Due to the NP-completeness, it is time consuming to find exact solutions to subgraph isomorphism, therefore, machine learning techniques are developed to accelerate the computation and provide accurate approximations. In this tutorial, we focus on two tasks, subgraph isomorphism matching and subgraph isomorphism counting.

Subgraph isomorphism matching. Two recent learning-based subgraph matching models will be introduced. RL-QVO [29] is a reinforcement learning-based model for optimizing subgraph matching by considering long-term rewards and generating optimal query vertex orders. GNN-PE [40] proposes a novel Path Dominance Embedding technique, where paths in the data graph are transformed into high-dimensional embeddings that follow a strict dominance relationship to ensure efficient candidate pruning.

Subgraph isomorphism counting. Subgraph isomorphism counting determines the frequency of subgraphs of the data graph that are isomorphic to the query graph. We discuss two works in this tutorial. LearnSC [6] decomposes both query and data graphs. The model employs a cross-graph learning model that embeds matching node features and introduces loss functions based on Direction Similarity and Projection Length. NeurSC [27] adaptively generates representative substructures from the data graph for each query graph, focusing only on relevant candidate structures. The inter- and intra-graph relationships are captured by the neural networks. FlowSC [5] simulates the candidate tree-based counting process using a bottom-up flow-learning approach, which explicitly captures the relationship between the structure and the counts. Combined with efficient candidate filtering, it achieves accurate subgraph counting.

Graph Similarity. Recently, machine learning techniques have been developed for the computation of graph similarity [39] to alleviate the computational costs of this task. Most learning-based models are designed for graph edit distance (GED) and maximum common subgraph (MCS). In this tutorial, we categorize these methods into two classes, i.e., *metric-agnostic* and *metric-specific*. We will present an overview of related research works.

Metric-agnostic models. SimGNN [1] pioneers a GNN-based approach for graph similarity computation, utilizing GCN as its foundational architecture. GMN [14] modifies the conventional GNN message-passing mechanism by incorporating cross-graph interactions, enabling simultaneous information aggregation from both input graphs. EGSC [20] introduces a layered processing framework, where outputs from a multi-layer GCN are progressively transformed through an Embedding Fusion Network (EFN) to integrate node embeddings across graph pairs.

Metric-specific models. Most existing metric-specific methods predominantly focus on learning-based GED computation. GREED [21] primarily considers several properties that GED as a metric is supposed to possess with GIN as the backbone. ERIC [46] conducts an in-depth analysis of the characteristics of GED and points out that when a pair of input graphs is aligned according to certain rules, GED can be derived based on the adjacency matrices and features of the aligned graphs. GEDGNN [19] takes a different perspective on GED computation by proposing the use of matching relationships and cost matrices to estimate the similarity between input graphs. DiffGED [10] leverages the diffusion model to generate a transition matrix for graph edit distance computation. This method provides an accurate estimation of GED with an associated GED path. GEDRanker [9] provides a novel unsupervised solution for GED approximation, which avoids the time-costly computation of ground-truth GED values.

Community Search. Community search is a representative graph processing task that identifies a specific community with a given query. In this tutorial, we discuss the following learning-based solutions.

Community search. CSGphormer [30] proposed a pre-trained graph transformer-based community search framework that uses zero label. Therefore, CSGphormer supports unsupervised community search. ALICE [31] extracts a candidate subgraph to reduce the search scope and subsequently predicts the community by the Consistency-aware Net, termed *ConNet*. ALICE is able to be applied on billion-scale graphs. EnMCS [33] is proposed for multilayer community search, which first searches communities in each layer without labels and then merges communities from different layers by the expectation-maximization algorithm

Overlapping community search. In real-world graphs, every node may belong to multiple communities, which motivates the development of overlapping community search methods. Research work [22] will be introduced, which proposed Sparse Subspace Filter (SSF) and Simplified Multi-hop Attention Networks (SMN) for overlapping community search.

2.4 Open Challenges and Future Directions

We will discuss the challenges of current research and several future research directions in learning-based graph data management and query processing.

Challenges. We will focus on the following two major challenges.

(1) Scalability. The scalability is a common challenge for machine learning-based techniques, especially with the million- or billion-scale real-life graphs.

(2) Adaptability. Adaptability poses a significant challenge when applying machine learning techniques to real-world graph data management systems. Most existing techniques are designed for a specific task or a specific dataset, which has the limitation of being adapted to dynamic graph data or various downstream tasks.

Future Directions. In conclusion to this tutorial, we introduce the future direction for machine learning for graph data management and query processing. We focus on three key directions promise to further advance the field.

(1) Foundation models for graphs. As discussed previously, the task-specific nature of current machine learning-based methods prevents them from application in real-life graph data management systems. Foundation models, which have great generalization ability, will enable unified, adaptable solutions across diverse tasks. Unlike task-specific models, these pretrained frameworks can generalize across different graph data, varying workloads, and multiple tasks, reducing the need for costly retraining.

(2) Integrating graph prompting frameworks. By unifying disparate learning-based models under a flexible prompting interface, broader applications can be supported by a single model. Meanwhile, adopting the graph prompting techniques can also minimize the customization overhead, e.g., retraining of the model, for various graph-related tasks.

(3) Combining database and machine learning techniques. Many existing techniques only rely on machine learning techniques to solve a database task, eventually leading to the scalability issue. Combining database techniques, such as candidate filtering, pruning, and sampling, is a promising future direction to improve the scalability of learning-based solutions to graph data management.

3 PRESENTERS

Hanchen Wang is a lecturer and ARC DECRA Fellow at AAII, FEIT, University of Technology Sydney, Australia. He received BSc degree from Zhejiang University in 2016, and PhD degree from the University of Technology Sydney in 2021. His research interests include graph analytics and machine learning for databases.

Ying Zhang is a Professor at Zhejiang Gongshang University. He received his BSc and MSc degrees in Computer Science from Peking University, and PhD in Computer Science from the University of New South Wales. He is the recipient of Best Paper Award ICDE 2016 and Best Paper Runner-up Award VLDB 2020. His research interests include query processing and analytics on large-scale data with focus on graphs and high dimensional data.

Wenjie Zhang is a full Professor and ARC Future Fellow in School of Computer Science and Engineering, the University of New South Wales, Australia. She is the recipient of ACM SIGMOD Research Highlight Award 2021, one of the Best of SIGMOD 2020 and ICDE 2013/2012/2010, and over ten Best Paper Awards from other conferences.

REFERENCES

[1] Yunsheng Bai, Hao Ding, Song Bian, Ting Chen, Yizhou Sun, and Wei Wang. 2019. SimGNN: A Neural Network Approach to Fast Graph Similarity Computation. In *WSDM*. 384–392.

[2] Gao Cong, Jingyi Yang, and Yue Zhao. 2024. Machine learning for databases: Foundations, paradigms, and open problems. In *Companion of the 2024 International Conference on Management of Data*. 622–629.

[3] Junnan Dong, Qinggang Zhang, Xiao Huang, Qiaoyu Tan, Daochen Zha, and Zhao Zihao. 2023. Active ensemble learning for knowledge graph error detection. In *Proceedings of the sixteenth ACM international conference on web search and data mining*. 877–885.

[4] Wenfei Fan, Xin Wang, Yinghui Wu, and Dong Deng. 2014. Distributed graph simulation: Impossibility and possibility. *(PVLDB)* 7, 12 (2014), 1083–1094.

[5] Qiuyu Guo, Jianye Yang, Wenjie Zhang, Hanchen Wang, Ying Zhang, and Xuemin Lin. 2025. Efficient and Accurate Subgraph Counting: A Bottom-up Flow-learning Based Approach. *Proceedings of the VLDB Endowment* 18 (2025), 2695–2708. Issue 8.

[6] Wenzhe Hou, Xiang Zhao, and Bo Tang. 2024. LearnSC: An Efficient and Unified Learning-Based Framework for Subgraph Counting Problem. In *2024 IEEE 40th International Conference on Data Engineering (ICDE)*. IEEE, 2625–2638.

[7] Han Huang, Leilei Sun, Bowen Du, and Weifeng Lv. 2023. Conditional Diffusion Based on Discrete Graph Structures for Molecular Graph Generation. *Proceedings of the AAAI Conference on Artificial Intelligence* 37, 4 (Jun. 2023), 4302–4311.

[8] Wei Huang, Hanchen Wang, Dong Wen, Xuefeng Chen, Wenjie Zhang, and Ying Zhang. 2024. GQ*: Towards Generalizable Deep Q-Learning for Steiner Tree in Graphs. In *2024 IEEE International Conference on Data Mining (ICDM)*. IEEE, 141–150.

[9] Wei Huang, Hanchen Wang, Dong Wen, Shaozhen Ma, Wenjie Zhang, and Xuemin Lin. 2025. Towards Unsupervised Training of Matching-based Graph Edit Distance Solver via Preference-aware GAN. *arXiv preprint arXiv:2506.01977* (2025).

[10] Wei Huang, Hanchen Wang, Dong Wen, Wenjie Zhang, Ying Zhang, and Xuemin Lin. 2025. Diffged: Computing graph edit distance via diffusion-based graph matching. *arXiv preprint arXiv:2503.18245* (2025).

[11] Diederik P Kingma and Max Welling. 2013. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114* (2013).

[12] Cheng-Te Li, Yu-Che Tsai, and Jay Chiehen Liao. 2023. Graph neural networks for tabular data learning. In *2023 IEEE 39th International Conference on Data Engineering (ICDE)*. IEEE, 3589–3592.

[13] Guoliang Li and Xuanhe Zhou. 2022. Machine learning for data management: A system view. In *2022 IEEE 38th International Conference on Data Engineering (ICDE)*. IEEE, 3198–3201.

[14] Yujia Li, Chenjie Gu, Thomas Dullien, Oriol Vinyals, and Pushmeet Kohli. 2019. Graph Matching Networks for Learning the Similarity of Graph Structured Objects. In *ICML*, Vol. 97. 3835–3845.

[15] Xiangyu Liu, Yang Liu, and Wei Hu. 2024. Knowledge graph error detection with contrastive confidence adaption. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 38. 8824–8831.

[16] Youzhi Luo, Chengkai Liu, and Shuiwang Ji. 2024. Towards Symmetry-Aware Generation of Periodic Materials. *NeurIPS* 36 (2024).

[17] Youzhi Luo, Keqiang Yan, and Shuiwang Ji. 2021. Graphdf: A discrete flow model for molecular graph generation. In *ICML*. PMLR, 7192–7203.

[18] Karolis Martinkus, Andreas Loukas, Nathanaël Perraudin, and Roger Wattenhofer. 2022. Spectre: Spectral conditioning helps to overcome the expressivity limits of one-shot graph generators. In *ICML*. PMLR, 15159–15179.

[19] Chengzhi Piao, Tingyang Xu, Xiangguo Sun, Yu Rong, Kangfei Zhao, and Hong Cheng. 2023. Computing Graph Edit Distance via Neural Graph Matching. *Proc. VLDB Endow.* 16, 8 (2023), 1817–1829.

[20] Can Qin, Handong Zhao, Lichen Wang, Huan Wang, Yulun Zhang, and Yun Fu. 2021. Slow Learning and Fast Inference: Efficient Graph Similarity Computation via Knowledge Distillation. In *NeurIPS*. 14110–14121.

[21] Rishabh Ranjan, Siddharth Grover, Sourav Medya, Venkatesan T. Chakaravarthy, Yogish Sabharwal, and Sayan Ranu. 2022. GREED: A Neural Framework for Learning Graph Distance Functions. In *NeurIPS*.

[22] Qing Sima, Jianke Yu, Xiaoyang Wang, Wenjie Zhang, Ying Zhang, and Xuemin Lin. 2025. Deep Overlapping Community Search via Subspace Embedding. *Proceedings of the ACM on Management of Data* 3, 1 (2025), 1–26.

[23] Gregor NC Simm and José Miguel Hernández-Lobato. 2019. A generative model for molecular distance geometry. *arXiv preprint arXiv:1909.11459* (2019).

[24] Wenxuan Tu, Bin Xiao, Xinwang Liu, Sihang Zhou, Zhiping Cai, and Jieren Cheng. 2024. Revisiting initializing then refining: an incomplete and missing graph imputation network. *IEEE Transactions on Neural Networks and Learning Systems* (2024).

[25] Clement Vignac, Igor Krawczuk, Antoine Siraudin, Bohan Wang, Volkan Cevher, and Pascal Frossard. 2023. Digress: Discrete denoising diffusion for graph generation. *ICLR* (2023).

[26] Hanchen Wang, Dawei Cheng, Ying Zhang, and Wenjie Zhang. 2025. AIGC for Graphs: Current Techniques and Future Trends. In *2025 IEEE 41st International Conference on Data Engineering (ICDE)*. IEEE Computer Society, 4504–4508.

[27] Hanchen Wang, Rong Hu, Ying Zhang, Lu Qin, Wei Wang, and Wenjie Zhang. 2022. Neural subgraph counting with Wasserstein estimator. In *Proceedings of the 2022 International Conference on Management of Data*. 160–175.

[28] Hanchen Wang, Jianke Yu, Xiaoyang Wang, Chen Chen, Wenjie Zhang, and Xuemin Lin. 2023. Neural similarity search on supergraph containment. *IEEE Transactions on Knowledge and Data Engineering* 36, 1 (2023), 281–295.

[29] Hanchen Wang, Ying Zhang, Lu Qin, Wei Wang, Wenjie Zhang, and Xuemin Lin. 2022. Reinforcement learning based query vertex ordering model for subgraph matching. In *2022 IEEE 38th International Conference on Data Engineering (ICDE)*. IEEE, 245–258.

[30] Jianwei Wang, Kai Wang, Xuemin Lin, Wenjie Zhang, and Ying Zhang. 2024. Efficient Unsupervised Community Search with Pre-Trained Graph Transformer. *Proceedings of the VLDB Endowment* 17, 9 (2024), 2227–2240.

[31] Jianwei Wang, Kai Wang, Xuemin Lin, Wenjie Zhang, and Ying Zhang. 2024. Neural attributed community search at billion scale. *Proceedings of the ACM on Management of Data* 1, 4 (2024), 1–25.

[32] Jianwei Wang, Kai Wang, Ying Zhang, Wenjie Zhang, Xu Xiwei, and Xuemin Lin. 2025. On LLM-Enhanced Mixed-Type Data Imputation with High-Order Message Passing. *Proceedings of the VLDB Endowment* 18, 10 (2025), 3421 – 3434.

[33] Jianwei Wang, Yuehai Wang, Kai Wang, Xuemin Lin, Wenjie Zhang, and Ying Zhang. 2025. Ensemble-based Deep Multilayer Community Search. *arXiv preprint arXiv:2501.02194* (2025).

[34] Jianwei Wang, Ying Zhang, Kai Wang, Xuemin Lin, and Wenjie Zhang. 2024. Missing data imputation with uncertainty-driven network. *Proceedings of the ACM on Management of Data* 2, 3 (2024), 1–25.

[35] Shiyu Wang, Xiaojie Guo, and Liang Zhao. 2022. Deep generative model for periodic graphs. *Advances in Neural Information Processing Systems* 35 (2022).

[36] Sheng Xiang, Dong Wen, Dawei Cheng, Ying Zhang, Lu Qin, Zhengping Qian, and Xuemin Lin. 2022. General graph generators: experiments, analyses, and improvements. *The VLDB Journal* (2022), 1–29.

[37] Bingcong Xue and Lei Zou. 2022. Knowledge graph quality management: a comprehensive survey. *IEEE Transactions on Knowledge and Data Engineering* 35, 5 (2022), 4969–4988.

[38] Zhengtong Yan, Valter Uotila, and Jiaheng Lu. 2023. Join order selection with deep reinforcement learning: fundamentals, techniques, and challenges. *Proceedings of the VLDB Endowment* 16, 12 (2023), 3882–3885.

[39] Peilun Yang, Hanchen Wang, Jianye Yang, Zhengping Qian, Ying Zhang, and Xuemin Lin. 2024. Deep learning approaches for similarity computation: A survey. *IEEE Transactions on Knowledge and Data Engineering* (2024).

[40] Yutong Ye, Xiang Lian, and Mingsong Chen. 2024. Efficient Exact Subgraph Matching via GNN-Based Path Dominance Embedding. *Proceedings of the VLDB Endowment* 17, 7 (2024), 1628–1641.

[41] Jaemin Yoo, Hyunsik Jeon, Jinhong Jung, and U Kang. 2022. Accurate node feature estimation with structured variational graph autoencoder. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 2336–2346.

[42] Kai Siong Yow, Ningyi Liao, Siqiang Luo, and Reynold Cheng. 2023. Machine learning for subgraph extraction: Methods, applications and challenges. *Proceedings of the VLDB Endowment* 16, 12 (2023), 3864–3867.

[43] Qinggang Zhang, Junnan Dong, Keyu Duan, Xiao Huang, Yezi Liu, and Linchuan Xu. 2022. Contrastive knowledge graph error detection. In *Proceedings of the 31st ACM international conference on information & knowledge management*. 2590–2599.

[44] Jiajun Zhong, Ning Gui, and Weiwei Ye. 2023. Data imputation with iterative graph reconstruction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 37. 11399–11407.

[45] Rong Zhu, Lianggui Weng, Bolin Ding, and Jingren Zhou. 2024. Learned query optimizer: what is new and what is next. In *Companion of the 2024 International Conference on Management of Data*. 561–569.

[46] Wei Zhuo and Guang Tan. 2022. Efficient Graph Similarity Computation with Alignment Regularization. In *NeurIPS*.