# The Future of Graph Analytics


Angela Bonifati
Lyon 1 University, CNRS & IUF
Lyon, France
angela.bonifati@univ-lyon1.fr

M. Tamer Özsu
University of Waterloo
Waterloo, Canada
tamer.ozsu@uwaterloo.ca

Yuanyuan Tian
Microsoft Gray Systems Lab
Seattle, USA
yuanyuantian@microsoft.com

Hannes Voigt
Neo4j
Leipzig, Germany
hannes.voigt@neo4j.com

Wenyuan Yu
Alibaba Group
Beijing, China
wenyuan.ywy@alibaba-inc.com

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



## ABSTRACT

In the last two decades, we have been witnessing high demand for graph-based technologies in industry. On the research side, several recent advances have been made about large-scale graph processing, graph analytical systems and graph databases. The landscape of graph query languages is currently evolving with the definition of new standards, and the need for domain-specific languages to express graph algorithmic and analytical primitives will continue and increase in the next future.

In this SIGMOD panel, we will discuss the impact of the above changes on the future of graph analytics. Is there a demand for more expressive languages and libraries for analyzing relationships in a graph? Are new hybrid OLTP/OLAP architectures required with improved performance and scalability? What are the graph analytical workloads and benchmarks that users expect on real-world graph applications? What will be the impact of graph ML on graph analytical systems? How to adapt these systems to the dynamic changes that are ubiquitous for graph-shaped data? These and other questions will be addressed in the panel.



## CCS CONCEPTS

• **Information systems** → **Data management systems**; **Database management system engines**; **Online analytical processing engines**; **Query languages for non-relational engines**.

## KEYWORDS

Graph databases; Graph processing systems; Graph query languages; OLAP; Graph analytics

**ACM Reference Format:**
Angela Bonifati, M. Tamer Özsu, Yuanyuan Tian, Hannes Voigt, Wenyuan Yu, and Wenjie Zhang. 2024. The Future of Graph Analytics. In *Companion of the 2024 International Conference on Management of Data (SIGMOD-Companion '24), June 9–15, 2024, Santiago, AA, Chile.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3626246.3658369




*SIGMOD-Companion '24, June 9–15, 2024, Santiago, AA, Chile*



ACM ISBN 979-8-4007-0422-2/24/06
https://doi.org/10.1145/3626246.3658369


## 1 INTRODUCTION

Graphs are ubiquitous data structures used in a large spectrum of applications [1, 4]. How might forthcoming graph processing and database systems deliver extensive scalability, efficiency, and versatile querying and analytical functionalities to meet the diverse demands of real-world scenarios?

A recent survey on the usage of graph applications from real users has highlighted the fact that analytics is the task on which users spent most of their time as opposed to testing, cleaning and ETL [6].

To understand the burden of graph analytics [9], one has to resort to the data processing pipeline behind it. Graph data is manipulated and queried with online transactional processing (OLTP) operations, such as selections, joins and transitive closures. The data is further analyzed, enriched, and condensed with online analytical processing (OLAP) operations, such as (1) grouping, aggregating, slicing, dicing, and rollup, (2) graph algoritms, such as shortest path counting [8], in-betweenness centrality and PageRank and (3) a handful of user-defined functions. Finally, graph data is disseminated and consumed by a variety of applications, including machine learning, such as ML libraries and processing frameworks. Current benchmarks [3] for graph analytics are read-only and on a limited set of operators, while it would be desirable to extend them for the above pipeline.

Whereas standardization is already taking place for graph query and update languages [2], it would also be worthwhile to reach a common understanding of the building blocks of graph algorithms and analytical APIs.

Novel graph processing and graph database architectures are then needed to ensure the unification of OLTP and OLAP operations [5]. The data injected in these future graph systems is expected to be of different format and might require multi-model, converged and/or polystore databases [7].

The panel will focus on the aforementioned challenges and the following open research questions, along with other research questions that might be suggested by the audience:

- What is needed in terms of DSL and APIs for enabling analytics on top of graph structures for subsequent data science and ML tasks?
- Do we need OLAP/OLTP architectures or their hybrid version thereof (HTAP for graphs) in order to execute graph analytical workloads?

- Are graph-only stores sufficient or are polystores needed for the future of graph analytics?
- What are the requirements in terms of scalability, performance and benchmarking?
- Is on cloud better than on premise for the infrastructure of graph analytical systems?
- Since graphs are continuously evolving data structures, what is desirable in terms of analytical operators for dynamic, incremental and streaming graphs?

## 2 PANELISTS AND MODERATORS

**Angela Bonifati (Moderator)** is a Distinguished Professor of Computer Science at Lyon 1 University and at the CNRS Liris research lab, where she leads the Database Group. She is also an Adjunct Professor at the University of Waterloo in Canada from 2020 and a Senior member of the French University Institute (IUF) from 2023. Her current research interests are on several aspects of data management, including graph databases, knowledge graphs, data integration and their applications to data science and machine learning. She has co-authored several publications in top venues of the data management field, including four Best Paper Awards, two books and an invited paper in ACM Sigmod Record 2018. She is the recipient of the prestigious IEEE TCDE Impact Award 2023 and a co-recipient of an ACM Research Highlights Award 2023. She was the Program Chair of ACM Sigmod 2022 and she is currently an Associate Editor for the Proceedings of VLDB and for several other journals, including the VLDB Journal, IEEE TKDE and ACM TODS. She is the President of the EDBT Executive Board (2020-2025) and a member of the PVLDB Board of Trustees (2024-2029).

**M. Tamer Özsu** is a University Professor at Cheriton School of Computer Science at University of Waterloo where he holds a Cheriton Faculty Fellowship. He has previously served as the Director of the Cheriton School and Associate Dean (Research) of the Faculty of Mathematics. His research is broadly in data science focusing on data engineering following two threads: large-scale data distribution, and management of non-traditional data (i.e., non-relational data). He is a Fellow of the Royal Society of Canada, American Association for the Advancement of Science, the Asia-Pacific Artificial Intelligence Association, and Balsillie School of International Affairs; he is also Life Fellow of the Association for Computing Machinery and the Institute of Electrical and Electronics Engineers. He is an elected member of the Science Academy of Türkiye and a member of Sigma Xi.

**Yuanyuan Tian** is a Principal Scientist Manager at Microsoft Gray Systems Lab (GSL), and an ACM Distinguished Member. Before Microsoft, she was a Principal Research Staff Member at IBM Almaden Research Center. She obtained her Ph.D. in computer science from the University of Michigan. She is currently leading the graph analytics platform and the ML-for-Systems research efforts in GSL. Her other research interests include HTAP, SQL-on-Hadoop, big data federation, and large-scale systems for machine learning. She has published two books and over 50 articles in top database venues with 5400+ citations. Dr. Tian has served in the editorial board for the new encyclopedia for Big Data, as an Associate Editor for VLDB Journal, PVLDB and SIGMOD, and chaired various tracks in top database conferences. She has also served in several NSF panels. She is the recipient of the DaMoN 2023 Best Short Paper Award, the SIGMOD 2019 Research Highlight Award, the EDBT 2018 Best Paper Award, the Outstanding Technical Achievement Award from IBM in 2020, 2019 and 2016, the Research Division Award from IBM in 2019, and the Distinguished Academic Achievement Award from the University of Michigan in 2008.

**Hannes Voigt** is a Staff Engineer at Neo4j, a Ph.D. graduate from the Technische Universität Dresden, the chair of INCITS/Data Management/Expert Group on GQL, and a member of ISO/IEC JTC 1/SC 32/WG3. At Neo4j, he works in standardization and design for the property graph query languages GQL and Cypher. His research interests are in property graph schema and graph analytics. He co-authored Industrial Track Best Paper at SIGMOD 2023, the CACM title page article "The Future Is Big Graphs", and one book.

**Wenyuan Yu** is a Senior Staff Engineer and Director at Alibaba Group and a Ph.D. graduate from the University of Edinburgh. At Alibaba, he leads the Fusion Computing team at the Institute for Intelligent Computing, focusing on machine learning systems and graph computing. Wenyuan is the founder and project lead of GraphScope, Alibaba's open-source large-scale graph computing system, and the CNCF's data sharing system, Vineyard. His research, published in top-tier international conferences and journals, has earned him recognition including Best Paper at SIGMOD 2017 and VLDB 2010, and the SIGMOD Research Highlight Award in 2018. Prior to Alibaba, Wenyuan was a founding member of 7Bridges Ltd and a Research Scientist at Facebook.

**Wenjie Zhang** is a full Professor in School of Computer Science and Engineering, the University of New South Wales, where she leads the Data and Knowledge Research Group. Her research focuses on developing scalable and efficient techniques for large-scale data processing and analytics, especially for graph and spatial data. Her papers were nominated as Best of SIGMOD and ICDE, and received ACM SIGMOD Research Highlight Award. She serves as an Associate Editor for TKDE and VLDB Journal, and area chair for ICDE/VLDB/ICDM.

## REFERENCES


[1] Angela Bonifati, George H. L. Fletcher, Hannes Voigt, and Nikolay Yakovets. 2018. *Querying Graphs.* Morgan & Claypool Publishers.

[2] Alin et al. Deutsch. 2022. Graph Pattern Matching in GQL and SQL/PGQ. In *Proceedings of the 2022 International Conference on Management of Data*, Vol. 46. ACM, New York, NY, USA, 2246–2258.

[3] Alexandru Iosup et al. 2016. LDBC Graphalytics: A Benchmark for Large-Scale Graph Analysis on Parallel and Distributed Platforms. *Proc. VLDB Endow.* 9, 13 (2016), 1317–1328.

[4] Sherif Sakr et al. 2021. The future is big graphs: a community view on graph processing systems. *Commun. ACM* 64, 9 (2021), 62–71.

[5] Sijie Shen et al. 2023. Bridging the Gap between Relational OLTP and Graph-based OLAP. In *2023 USENIX Annual Technical Conference, USENIX ATC 2023, Boston, MA, USA, July 10-12, 2023.* 181–196.

[6] Siddhartha Sahu, Amine Mhedhbi, Semih Salihoglu, Jimmy Lin, and M. Tamer Özsu. 2020. The ubiquity of large graphs and surprising challenges of graph processing: extended survey. *VLDB J.* 29, 2-3 (2020), 595–618.

[7] Yuanyuan Tian. 2022. The World of Graph Databases from An Industry Perspective. *SIGMOD Rec.* 51, 4 (2022), 60–67.

[8] Yiqi Wang, Long Yuan, Zi Chen, Wenjie Zhang, Xuemin Lin, and Qing Liu. 2023. Towards Efficient Shortest Path Counting on Billion-Scale Graphs. In *39th IEEE International Conference on Data Engineering, ICDE 2023, Anaheim, CA, USA, April 3-7, 2023.* 2579–2592.

[9] Da Yan, Yingyi Bu, Yuanyuan Tian, and Amol Deshpande. 2017. Big Graph Analytics Platforms. *Found. Trends Databases* 7, 1-2 (2017), 1–195.