Entity/Relationship Modeling for Property Graphs

Philipp Skavantzos, Sebastian Link

School of Computer Science, University of Auckland, Auckland 1010, New Zealand

Abstract

This tutorial shows how traditional Entity/Relationship modeling and modern graph data modeling can be combined to bring forward well-designed graph data models that process workloads and maintain data integrity efficiently.

Keywords

Conceptual modeling, Entity/Relationship diagrams, Entity integrity, Property graphs, Referential integrity

1. Topic Motivation

Despite the popularity and growing maturity of graph database systems, they continue to rank well below trusted relational technology[1]. The reasons for this are not just historical. Indeed, an inhibitor to the further uptake of graph databases is the lack of principled methodologies for their design, rigorous schema and data integrity support. As a consequence, academics and practitioners have worked together to develop proposals for emerging standards of query languages [1, 2], schema [3], and data integrity [4]. For the emerging standards *PG-Schema* [3] for graph schemata and *PG-Key* [4] for graph integrity, it is not yet well understood which of their fragments supports which applications. PG-Schema, in particular, has been developed to support basic features of Entity/Relationship (E/R) models, but its expressiveness is well beyond those capabilities.

Chen's E/R model [5] constitutes the best breed of conceptual data models. The model captures entities and their relationships in an easy-to-understand framework powerful enough to derive a formal data model. E/R models visually represent complex requirements for the target database [6]. In fact, the graphical depiction of an E/R diagram is invaluable for effective communication between experts with different expertise. E/R modeling is a methodology for generating well-designed E/R diagrams as they guarantee directed acyclicity, data integrity and do not exhibit redundancy of data or anomalies of updates in any instance [7, 8, 9].

Given these developments, the tutorial addresses the following ambitious research question.

What is a methodology for designing property graph schemata that can process workloads and maintain data integrity efficaciously on each of their graph instances?

The main goals are to show what Entity/Relationship modeling can do for graph data, and what graph data can do for Entity/Relationship modeling. An important point is to bring together different communities for their mutual benefit.

In what follows, Sec. 2 will describe the target audience and knowledge outcomes. The latter are fundamental for conceptual, graph and logical data modeling. A detailed outline and timetable of the tutorial will be given in Sec. 3. The tutorial method is explained in Sec. 4, before a short biography for the presenters is given in Sec. 5.

ER2025: Companion Proceedings of the 44th International Conference on Conceptual Modeling: Industrial Track, ER Forum, 8th SCME, Doctoral Consortium, Tutorials, Project Exhibitions, Posters and Demos, October 20-23, 2025, Poitiers, France

philipp.skavantzos@auckland.ac.nz (P. Skavantzos); s.link@auckland.ac.nz (S. Link)

https://profiles.auckland.ac.nz/philipp-skavantzos (P. Skavantzos); https://profiles.auckland.ac.nz/s-link (S. Link)

0000-0001-6034-7485 (P. Skavantzos); 0000-0002-1816-2863 (S. Link)

[1] https://db-engines.com/en/ranking_trend

2. Target Audience and Knowledge Outcomes

The specific target audience is people with an interest in conceptual data modeling or graph data modeling. The tutorial showcases fundamentals for these disciplines, and focuses on how they can bring out the best in one another when combined. The topic is timely as graph databases have grown larger in popularity among researchers and practitioners, but also because the design of graph databases has emerged as a new and important direction. The tutorial is accessible to students with a general computer science education at undergraduate level because fundamental concepts are introduced. This is even of interest to experts in the area, due to the diversity of models and languages available, which makes it necessary to fix concepts and notation.

The overarching outcome targeted for this tutorial is to survey the fundamental impact E/R modeling and graph databases have on one another.

Property graphs that comply with E/R diagrams, called E/R graphs, constitute the first E/R databases that are graphs. Hence, instead of translating E/R diagrams to other data models, we can directly manage E/R graphs as graph databases. That is, graph databases bring E/R modeling concepts to life. The tutorial illustrates how E/R diagrams form a core fragment of PG-Schema that captures well-designed (graph) databases. Hence, PG-Schema can be viewed as a general data modeling tool.

It is demonstrated that reasoning about PG-Key is infeasible, but keys of E/R models form a fragment of PG-Key for efficiently managing entity integrity in well-designed property graphs.

The audience learns about principled approaches to managing referential integrity efficiently in E/R graphs. This contrasts relational and graph-based approaches, and pinpoints what opportunities for improving integrity management in graph database systems exist.

Overall, the tutorial conveys how E/R diagrams and property graphs can be combined to unify conceptual, logical and graph data modeling, how integrity management can be taken to the next level by eliminating property redundancy, and why relational benchmarks perform very well when translated into property graphs. Indeed, the tutorial shows how a major inhibitor to the uptake of graph databases can be turned into a strong driver.

3. Detailed Outline and Timetable

We will give a detailed outline of what the tutorial covers, and indicate how much time is spent on each topic. The following subsections form the logical units of the tutorial, in the sequence they are presented.

After some motivation and outline of the goals **(5 minutes)** [10, 1, 2], we provide a background on Entity/Relationship modeling **(20 minutes)** [5, 9, 8, 7]. Utilizing a small toy example sufficiently large to illustrate all concepts, and also utilizing an E/R diagram for the TPC-H benchmark to illustrate some real-world like scenario, we fix the underlying syntax and semantics. There is a brief overview of recent approaches applying conceptual modeling to graph databases [11, 12, 13, 14, 15], with restrictions on the order and depth of object types utilized **(5 minutes)**. The remaining tutorial is based on the expressiveness of Entity/Relationship models to play out their key strength for the benefit of graph databases [16].

The concept of property graphs is introduced next, both in terms of formal definitions and visualizations **(10 minutes)**. We also illustrate the concepts on our running example.

We then discuss the recent proposal for the definition of key constraints in property graphs, called PG-Key [4, 3] **(10 minutes)**. We discuss the syntax and semantics of PG-Key expressions such that the audience is comfortable with understanding how expressive PG-Keys are and how they work. The discussion is led by examples.

Next we give a concise overview of PG-Schema [3], a recent proposal by academics and practitioners to help standardizing schema support for graph databases **(10 minutes)**. Similarly to PG-Key, we highlight the syntax and semantics of major features, predominantly by way of examples.

Our next aim is to show how traditional E/R modeling provides a methodology for designing property

graphs well **(20 minutes)**. In the first of two steps, we illustrate that E/R diagrams are property graphs themselves, but also form a fragment of PG-Schema. This situation is reminiscent of XML where XML Schema definitions constitute XML documents themselves [17]. As a result, E/R modeling is available as a mature and trusted methodology for the design of property graphs. In step two, we showcase that E/R graphs constitute the first graph semantics for E/R diagrams, as defined in [16]. As E/R modeling is a conceptual approach and property graphs constitute a logical data model, E/R graphs unify conceptual, logical, and graph data modeling. As consequences, modern graph database systems provide an operational platform for conceptual data models, offering a viable alternative to relational technology without having to translate conceptual models at all.

Next, we demonstrate that E/R graphs and diagrams offer principled concepts for maintaining entity and referential integrity efficiently within graph database systems **(20 minutes)**. In particular, translations of relational databases to E/R graphs offer various benefits for data integrity management. We illustrate the notion of an E/R key and demonstrate that every E/R key is also a PG-Key, but not vice versa. We demonstrate that E/R keys form an efficient fragment of PG-Key, namely for well-designed property graphs. In terms of referential integrity, we illustrate two principled approaches: one that always duplicates key properties as foreign key properties (relational semantics), and one that always uses references by directed edges in graphs (graph semantics). The first approach can directly benefit from graph technology, while the second one can only make use of current technology in case all key properties are defined locally on nodes without use of references. Hence, we propose a third approach (mixed semantics) which maximizes the opportunity for references while complying with current technology limits.

We then showcase TPC-H as an industry-like use case that illustrates the main concepts and quantifies how well the E/R approach works at operational level **(10 minutes)**. We present translations of the TPC-H schema, including keys and foreign keys, into an E/R diagram using the different semantics, translate its query and update operations, and compare the efficiency of their evaluation within Neo4j and MySQL. The results demonstrate to the audience how well our initial research question has been addressed. It also shows the audience how important data modeling is for core database operations.

Finally, we summarize main takeaways and outline future research directions **(10 minutes)**.

4. Tutorial Method

The presentation is based on slides to convey the knowledge outcomes in a visually appealing way that draws the audience's attention and promotes their understanding, but also facilitates interaction. Our timetable, proposed above, accommodates the presentation itself but also designates time for questions and discussions. In particular, we will prompt the audience to criticize our toy example and offer alternative modeling approaches, but also ask them to convert between some classical relational database instances and graph instances to facilitate understanding further. Lots of examples and visualizations illustrate ideas, techniques and findings. A 90 minute version was presented at VLDB 2025. The ER audience is quite different, the topic well aligned with ER, and a presentation of 2 hours conveys more details, discussion, and interesting direction of future research for the modeling community.

We only use a standard projector and a board for exploring ideas interactively with the audience.

5. Presenters

Philipp is a research fellow and teaching assistant with considerable experience teaching database topics to undergraduate and postgraduate students of different backgrounds. The topic of his PhD was on the design of graph databases.

Sebastian is a Professor of Computer Science with vast experience in presenting database topics to audiences of various backgrounds. He has presented extensively at top database conferences, particularly on topics of the tutorial.

References

[1] R. Angles, M. Arenas, P. Barceló, P. A. Boncz, G. H. L. Fletcher, C. Gutierrez, T. Lindaaker, M. Paradies, S. Plantikow, J. F. Sequeda, O. van Rest, H. Voigt, G-CORE: A core for future graph query languages, in: Proceedings of the 2018 International Conference on Management of Data, SIGMOD Conference 2018, Houston, TX, USA, June 10-15, 2018, 2018, pp. 1421–1432.

[2] R. Angles, M. Arenas, P. Barceló, A. Hogan, J. L. Reutter, D. Vrgoc, Foundations of modern query languages for graph databases, ACM Comput. Surv. 50 (2017) 68:1–68:40.

[3] R. Angles, A. Bonifati, S. Dumbrava, G. Fletcher, A. Green, J. Hidders, B. Li, L. Libkin, V. Marsault, W. Martens, F. Murlak, S. Plantikow, O. Savkovic, M. Schmidt, J. Sequeda, S. Staworko, D. Tomaszuk, H. Voigt, D. Vrgoc, M. Wu, D. Zivkovic, Pg-schema: Schemas for property graphs, Proc. ACM Manag. Data 1 (2023) 198:1–198:25. URL: https://doi.org/10.1145/3589778. doi:10.1145/3589778.

[4] R. Angles, A. Bonifati, S. Dumbrava, G. Fletcher, K. W. Hare, J. Hidders, V. E. Lee, B. Li, L. Libkin, W. Martens, F. Murlak, J. Perryman, O. Savkovic, M. Schmidt, J. F. Sequeda, S. Staworko, D. Tomaszuk, Pg-keys: Keys for property graphs, in: SIGMOD '21: International Conference on Management of Data, Virtual Event, China, June 20-25, 2021, 2021, pp. 2423–2436. URL: https://doi.org/10.1145/3448016.3457561. doi:10.1145/3448016.3457561.

[5] P. P. Chen, The entity-relationship model - toward a unified view of data, ACM Trans. Database Syst. 1 (1976) 9–36.

[6] P. P. Chen, English, chinese and ER diagrams, Data Knowl. Eng. 23 (1997) 5–16.

[7] M. Levene, M. W. Vincent, Justification for inclusion dependency normal form, IEEE Trans. Knowl. Data Eng. 12 (2000) 281–291.

[8] H. Mannila, K. Räihä, Design of Relational Databases, Addison-Wesley, 1992.

[9] B. Thalheim, Entity-relationship modeling - foundations of database technology, Springer, 2000.

[10] A. Bonifati, G. H. L. Fletcher, H. Voigt, N. Yakovets, Querying Graphs, Synthesis Lectures on Data Management, Morgan & Claypool Publishers, 2018.

[11] E. Gelling, G. Fletcher, M. Schmidt, Bridging graph data models: Rdf, rdf-star, and property graphs as directed acyclic graphs, CoRR abs/2304.13097 (2023).

[12] G. Daniel, G. Sunyé, J. Cabot, Umltographdb: Mapping conceptual schemas to graph databases, in: Conceptual Modeling - 35th International Conference, ER 2016, Gifu, Japan, November 14-17, 2016, Proceedings, 2016, pp. 430–444.

[13] J. Pokorný, Conceptual and database modelling of graph databases, in: Proceedings of the 20th International Database Engineering & Applications Symposium, IDEAS 2016, Montreal, QC, Canada, July 11-13, 2016, 2016, pp. 370–377.

[14] V. M. de Sousa, L. M. del Val Cura, Logical design of graph databases from an entity-relationship conceptual model, in: Proceedings of the 20th International Conference on Information Integration and Web-based Applications & Services, iiWAS 2018, Yogyakarta, Indonesia, November 19-21, 2018, 2018, pp. 183–189.

[15] R. D. Virgilio, A. Maccioni, R. Torlone, Model-driven design of graph databases, in: Conceptual Modeling - 33rd International Conference, ER 2014, Atlanta, GA, USA, October 27-29, 2014. Proceedings, 2014, pp. 172–185.

[16] P. Skavantzos, S. Link, Entity/relationship graphs: Principled design, modeling and data integrity management of graph databases, Proc. ACM Manag. Data 3 (2025) 1–25.

[17] G. J. Bex, F. Neven, J. V. den Bussche, Dtds versus XML schema: A practical study, in: Proceedings of the Seventh International Workshop on the Web and Databases, WebDB 2004, June 17-18, 2004, Maison de la Chimie, Paris, France, Colocated with ACM SIGMOD/PODS 2004, 2004, pp. 79–84.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.