Graph Databases as a Paradigm Shift in Behavior Science

Juan C. Correa[1,2*]

[1*]Departmento de Estudios Empresariales, Universidad Iberoamericana, Mexico City, Mexico, 1219.
[2*]Research and Development Unit, Critical Centrality Institute.

Corresponding author(s). E-mail(s): j.correa.n@gmail.com;

Abstract

Behavioral scientists have multiple options for managing research data. While relational databases offer robust tools for storage and analysis, they impose structural constraints that limit the representation of complex behavioral phenomena. This article argues that graph databases provide a rigorous and conceptually richer alternative, enabling the modeling of human behavior as an interconnected system rather than a set of isolated variables. Beyond a technological shift, adopting graph-based approaches invites a paradigm change in behavior science—one that embraces complexity, dynamic relationships, and multi-level contingencies. Practical implications are illustrated through examples from clinical, consumer, and industrial/organizational psychology.

Keywords: Graph database, Network modeling, complex behavior

1 Introduction

In a recent paper, Soto (2025) introduced relational databases for behavior science and used real-world examples to illustrate how relational databases have been used by behavioral scientists. Although relational databases remain the dominant paradigm in research and industry, alternative approaches are gaining traction. The so-called 'Not Only SQL' (NoSQL) category includes database systems that employ non-tabular data models, such as key-value pairs, documents, wide columns, matrices, and graphs. Whereas relational databases store structured data in tables (i.e., columns as variables, rows as cases, and cells containing specific values), NoSQL systems accommodate flexible formats that better support evolving and highly connected data.

In this article, I examine graphs as a paradigm that deviates from the traditional lenses of relational databases, where nodes and edges (instead of tables and joins) represent the basic elements of any behavior that can be represented as a network or complex system. Networks have a long history in mathematics as "*graph theory*" (Estrada, 2011). In sociology and social sciences, graph theory is known as "social network analysis" (Wasserman & Faust, 1994). In this context, the term "social network" should not be confused with online platforms such as Facebook or Instagram, as they are technological implementations that do not necessarily represent all aspects of social networks as a discipline. Psychologists have leveraged this framework to analyze the structure of psychopathology (Borsboom & Cramer, 2013), conduct bibliometric analysis of cyberbehavior (Serafin, Garcia-Vargas, García-Chivita, Caicedo, & Correra, 2019), estimate the correct number of dimensions in psychological and educational instruments (Golino & Epskamp, 2017), or understand the measurement of organizational climate (Menezes, Menezes, Moraes, & Pires, 2021).

2 Network as a collection nodes and edges

One of the easiest way to grasp the idea of a network is by looking its visual representation (see Figure 1).

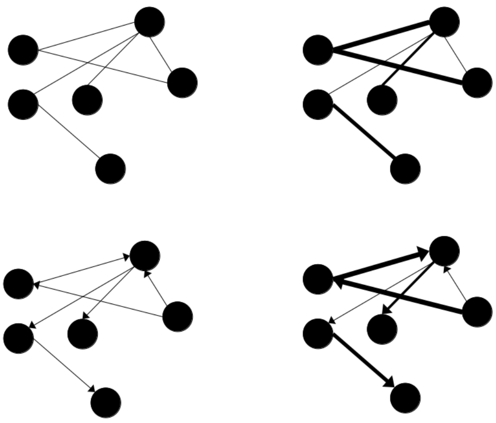

Fig. 1 A visual representation of four types of simple networks: non-directed unweighted network (upper left), non-directed weighted network (upper right), directed unweighted network (bottom left), and directed weighted network (bottom right).

According to Estrada (2011), a network is a collection of points (called nodes) joined together in pairs by lines (called arcs or edges). Despite this simplistic definition, networks provide a powerful framework to model any type of system from planets in a galaxy to neurons in the nervous system (Vazza & Feletti, 2020). In behavioral sciences, networks have been used to understand the mechanisms of team assembly and how these mechanisms determine collaboration structure and team performance (Guimerà, Uzzi, Spiro, & Amaral, 2005). Graphs offer fundamental concepts for understanding how entities (nodes) and their relationships (edges) form interconnected structures. From a data management viewpoint, the analysis of these networks requires tools that go beyond the rigid tabular constraints of relational databases. As the concepts of graphs are thoroughly covered in introductory texts (Newman, 2010), these will not be revisited here. Instead, this article aims to illustrate how graph-based databases can enrich the methodological toolbox of behavioral scientists, enabling analyses that embrace complexity, dynamic relationships, and multi-level contingencies (Robinson, Webber, & Eifrem, 2015).

3 Graph databases: A gentle introduction

Robinson et al. (2015) define a graph database as a system that implements **C**reate, **R**ead, **U**pdate, and **D**elete (CRUD) operations on a graph data model, where entities are represented as nodes and relationships as edges. Unlike relational databases, which organize data in tables, graph databases treat relationships as first-class elements rather than secondary links between tables. This design enables efficient traversal and pattern matching across highly connected data, making it ideal for modeling complex networks such as behavioral contingencies or social interactions.

Figure 2 shows a graph with eight nodes and eight edges. Nodes represent real-world entities such as persons (i.e., Ana, Pam, Carlos, and John) companies (i.e., Chevron and Rice University), and routes (i.e., I-10 and I-45). Edges represent the relationship between pairs of nodes. Thus, John and Pam work in Chevron but they commute distinct distances through different routes. Carlos and Ana work in Rice University, they both commute by the same route but they have to drive different distances. Interestingly, the information of nodes and edges can be enriched with attributes. These attributes also represent real-world characteristics like the distance each person has to commute, their sex, or the sector of the company they work for. A graph database like the one depicted in Figure 2 leverages the so-called "labeled property graph" which has the following elements: 1) it contains nodes and relationships, 2) nodes contain properties (key-value pairs), 3) nodes can be labeled with one or more labels, 4) relationships are named and directed, and always have a start and end node, 5) relationships can also contain properties. Although these elements offer significant benefits for behavioral scientists, they have been largely ignored. This gap presents an opportunity to enrich the methodological toolbox for behavior analysts working in both basic and applied settings, particularly those interested in integrating methods from other disciplines into behavioral science.

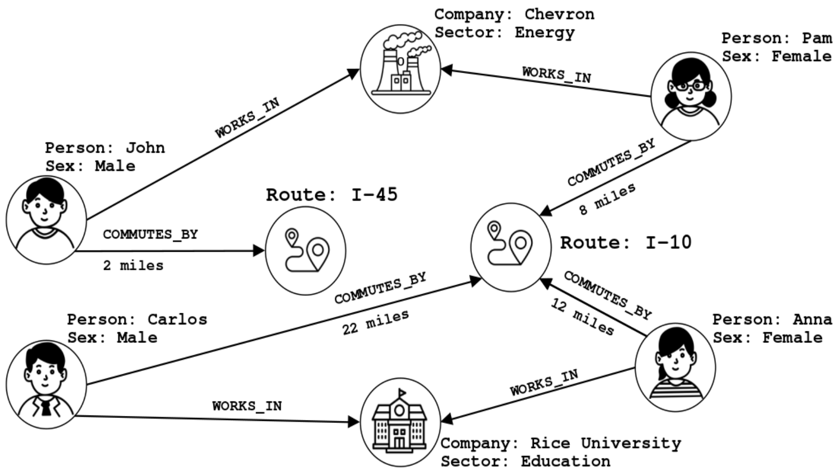

Fig. 2 A visual representation of a graph database that combines persons, companies, and routes

4 Benefits of graph databases

According to Robinson et al. (2015), an advantage of graph databases is the performance increase when dealing with connected data versus relational databases working with tabular data. Compared with relational databases, where join-intensive query performance deteriorates as the dataset gets bigger, a graph database performance remains relatively constant, even as the dataset grows with millions of records or variables. Big data issues might not be present for experimental behavior analysts who use to work with data coming from laboratory experiments. Yet, applied behavior analysts can enjoy the the benefits of graph databases. One example where graph databases can be particularly helpful is when urban traffic in a city is under scrutiny at an individual scale by tracking the movements of hundreds of thousands of drivers in a city every two hours (Gonzalez, Hidalgo, & Barabasi, 2008). In a case like the one just described, the outstanding performance of graph databases is in its queries as they are localized to a portion of the graph (e.g., if edges represent roads, the queries match the specific edges to be consulted). As a result, the execution time for each query is proportional only to the size of the part of the graph traversed to satisfy that query, rather than the size of the overall graph.

Another advantage of graph databases has to do with their flexibility. Behavior scientists aim to connect data in ways that reflect the knowledge domain itself, allowing structure and schema to evolve alongside their understanding of the problem rather than being rigidly defined upfront, when knowledge is most limited. This is particularly important when a behavioral phenomenon lacks theoretical background or lacks replication (Burgos, 2025). Graph databases fulfill this need by providing a flexible model that adapts to changing requirements and evidence. Because graphs are inherently additive, new nodes, relationships, labels, and subgraphs can be introduced. The modifications introduced to the original graph do not imply a threat to existing queries or application functionality. This flexibility minimizes the need for exhaustive upfront modeling and lowers the frequency of costly migrations, thereby reducing maintenance overhead.

A third benefit of graph databases is the agility that they offer. Modern graph databases enable smooth development and easy system maintenance. Their schema-free design, combined with testable APIs and query languages, allows controlled evolution of applications. While the absence of rigid schemas means traditional governance mechanisms are missing, this is not a drawback. On the contrary, it encourages more transparent and actionable data governance. Typically, governance is enforced programmatically through tests that validate data models, queries, and business rules. This approach aligns well with agile and test-driven development practices, making graph database–based applications adaptable to changing business needs.

It is worth mentioning that relational databases acknowledge relationships, but only during modeling, where they serve as join mechanisms between tables. In graph databases, we often need to clarify the meaning of relationships and even qualify their strength. These are aspects that relational models cannot do (Robinson et al., 2015). From this viewpoint, graph databases demand ontological considerations such as those recently described for psychology and behavioral sciences (Burgos, 2025). As datasets grow more complex and less uniform, relational data management systems like Microsoft Access, PostgreSQL, or SQLite become burdened with large join tables, sparsely populated rows, and extensive null-checking logic. Increased connectedness in relational databases translates into more joins, which degrade performance and make adapting the database to evolving requirements difficult. These and other limitations of relational databases are highlighted as "challenges to adoption" (Soto, 2025).

5 Potential usages of graph databases in behavioral research

Even though network modeling has been used for several research purposes like the one mentioned above, to my knowledge graph databases remain underutilized by behavior scientists.

References

Borsboom, D., & Cramer, A. (2013, 03). Network analysis: An integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology*, *9*(Volume 9, 2013), 91–121, https://doi.org/10.1146/annurev-clinpsy-050212-185608

Burgos, J. (2025). Getting ontologically serious about the replication crisis in psychology. *Journal of Theoretical and Philosophical Psychology*, *45*(2), 79–100, https://doi.org/10.1037/teo0000281 (Place: US Publisher: Educational Publishing Foundation)

Estrada, E. (2011). *The structure of complex networks: Theory and applications.* Oxford University Press.

Golino, H.F., & Epskamp, S. (2017). Exploratory graph analysis: A new approach for estimating the number of dimensions in psychological research. *PLOS ONE*, *12*(6), e0174035, https://doi.org/10.1371/journal.pone.0174035

Gonzalez, M.C., Hidalgo, C.A., Barabasi, A.-L. (2008). Understanding individual human mobility patterns. *Nature*, *453*(7196), 779–782, https://doi.org/10.1038/nature06958

Guimerà, R., Uzzi, B., Spiro, J., Amaral, L. (2005, 04). Team assembly mechanisms determine collaboration network structure and team performance. *Science*, *308*(5722), 697–702, https://doi.org/10.1126/science.1106340 Retrieved from https://www.science.org/doi/10.1126/science.1106340 (Publisher: American Association for the Advancement of Science)

Menezes, I., Menezes, A.C., Moraes, E., Pires, P.P. (2021). Measuring organizational climate via psychological networks analysis. *International Journal of Organization Theory & Behavior*, *24*(3), 229–250, https://doi.org/10.1108/IJOTB-08-2020-0142

Newman, M.E.J. (2010). *Networks: An introduction.* Oxford University Press.

Robinson, I., Webber, J., Eifrem, E. (2015). *Graph databases: New opportunities for connected data.* Sebastopol, California: O'Reilly.

Serafin, M.J., Garcia-Vargas, G.R., García-Chivita, M., Caicedo, M.I., Correra, J.C. (2019). Cyberbehavior: A bibliometric analysis. *Annual Review of CyberTherapy and Telemedicine*, *17*, 17–24,

Soto, P.L. (2025, 12). Relational databases for behavior science. *Perspectives on Behavior Science*, , https://doi.org/10.1007/s40614-025-00486-w

Vazza, F., & Feletti, A. (2020). The quantitative comparison between the neuronal network and the cosmic web. *Frontiers in Physics*, *8*, 525731, https://doi.org/10.3389/fphy.2020.525731

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications.* Cambridge University Press.