# Graph-Based Modeling of Diabetic Patient Data for Readmission Risk and Care Pattern Analysis

**Miloš Jovičić[1,*], Dragutin Ostojić[2], Nikola Prodanović[1], Nenad Djordjević[3], Nenad Janković[1]**

[1] University of Kragujevac, Institute for Information Technologies Kragujevac, Jovana Cvijića bb, 34000 Kragujevac, Serbia; e-mail: milos.jovicic@uni.kg.ac.rs , nikola.prodanovic@gmail.com , nenad.jankovic@uni.kg.ac.rs

[2] University of Kragujevac, Faculty of Science, Department of Mathematics and Informatics, Radoja Domanovića 12, 34000 Kragujevac, Serbia; e-mail: dragutin.ostojic@pmf.kg.ac.rs

[3] Brunel University London, Department of Mechanical and Aerospace Engineering, Centre for Assessment of Structures and Materials Under Extreme Conditions (CASMEC), London UB8 3PH, UK; e-mail: nenad.djordjevic@brunel.ac.uk

* *Corresponding author*

DOI: 10.46793/ICCBIKG25.071J

**Abstract**: Electronic health records (EHRs) contain rich relational data such as individual patient data, encounters, diagnoses, medications, etc. Healthcare systems often store EHR data in tabular form. However, traditional flat representations ("bag of features") can lose critical context. For example, treating a patient encounter as an unordered set of codes obscures the fact that a specific combination of drugs might have caused an adverse outcome. Knowledge graphs offer a robust alternative by organizing medical data into interconnected entities and relationships, capturing complex associations (e.g. between symptoms, treatments, diagnoses) for a more holistic understanding of patient history. In this work, we transform the Diabetes 130-US Hospitals dataset (a collection of ~100,000 inpatient encounters from 130 hospitals over 10 years) into a labeled property graph (LPG), and demonstrate the advantages both conceptual and quantitative of graph-based analysis, in a medical informatics context. Each encounter in this dataset includes patient demographics, diagnoses (ICD-9 codes), lab results (e.g. HbA1c), and 24 diabetes-related medications with change indicators ("up", "down", "steady" or "no change") among other features. Notably, the original study focused on 30-day readmissions, highlighting that poor glycemic control and suboptimal inpatient diabetes management lead to higher readmission rates and complications. Our graph model makes these clinical relationships explicit, enabling multi-hop reasoning (e.g. linking a patient's lab result to medication changes and subsequent readmission outcome) that is cumbersome with relational tables. We show that converting such EHR data into a graph can improve predictive modeling of readmissions and uncover insightful patterns of comorbidities and care processes that would be difficult to extract using SQL alone, aligning with recent trends in biomedical informatics to leverage networks for clinical data analysis.

**Keywords**: Knowledge graph, Electronic health records, Labeled property graph, Neo4j, Graph algorithms

## 1. Introduction

When modeling a labeled property graph, a well-designed domain specific graph schema plays a critical role [3], outside of the risk of having a low query performance, risk of context loss plays a critical role. In the construction of an EHR graph, key entities must be represented: patients, hospital encounters, diagnoses, and medications as first-class nodes. Additionally, key node labels can include: Patient, Encounter, Diagnosis, Drug, Hospital and Lab. An encounter node carries attributes like time_in_hospital, admission_type, discharge_disposition, and readmitted (30-day). As the case with all LPG, relationships (edges) between nodes are first class citizens [4]. Relationships encode the multi-relational structure of the data: each patient HAD_ENCOUNTER (edge) to their encounters; each encounter has one primary diagnosis (PRIMARY_DX) and up to two secondary diagnoses (modeled similarly). We capture diabetes medications by creating Drug nodes (e.g. insulin, metformin) and linking an encounter to each drug that was prescribed or adjusted during that stay (MED_CHANGE relationship). The MED_CHANGE relationship includes a property indicating dosage change ("up", "down" or "steady"). This was derived from 24 medication columns in the table by creating an edge for each non-"No" entry (for example, if a row indicates "insulin = Up", we create (Encounter)-[:MED_CHANGE {change:'up'}]->(Drug:Insulin)). In the same way, we optionally model lab test findings of interest (e.g. HbA1c > 8%) as nodes, connecting via LAB_RESULT relationships. This schema effectively turns the single-table hospital records into a network of entities: Patient → Encounter → Diagnosis/Drug, plus links between encounters and hospitals or lab measurements if available. By using a graph database (Neo4j), we can load the CSV data and create these nodes and relationships with Cypher queries in a batch ingestion [4]. For instance, using Cypher's UNWIND to iterate drug fields, we map each encounter's medication changes into corresponding (:Drug) nodes and [:MED_CHANGE] relationship. For instance, using Cypher's UNWIND to iterate drug fields, we map each encounter's medication changes into corresponding (:Drug) nodes and [:MED_CHANGE] relationship. The end result is a patient knowledge graph that explicitly links patients, their clinical encounters, conditions, treatments, and outcomes. This richer structure encodes the "geometry" of the clinical data that flat features would miss [1], allowing graph algorithms and queries to exploit the inherent relationships.

## 2. Methodology

For the created graph Neo4j Graph Data Science (GDS) library was used for the graph analysis [5]. We project relevant subgraphs for algorithmic tasks. To analyze disease co-occurrence, we project a bipartite graph of encounters and diagnoses connected by PRIMARY_DX/SECONDARY_DX edges. To analyze drug co-prescription patterns, we project encounters and drugs connected by MED_CHANGE edges. The GDS library enables efficient computation of graph algorithms on these projections [6]. Graph projections are in-memory graphs that are created for the graph data science

operations. We applied community detection (Louvain modularity) and centrality measures [7] to the co-morbidity network, and embedding algorithms to patient/encounter graphs. All analyses were run on a standard machine using Neo4j 5.12 and GDS 2.2, with the memory configuration adjusted for the ~100k encounter graph size.

## 3. Results and Discussion

The created EHR graph can provide an explainable decision support system, allowing us to surface the chain of linked factors contributing to readmissions [8]. We also specifically examined interactions between medication changes and readmission. Using Cypher pattern queries on the created graph, without an advanced algorithmic search, one can easily retrieve cohorts such as "patients with HbA1c > 8 who had insulin dosage increased during the encounter" and compare their readmission frequency to others. Such a query-as-cohort approach is far more natural on the graph (one pattern expression) than in a relational framework with multiple joins and filters. Clinically, this enabled us to identify hypothesis-generating trends. For example, patients with intensification of insulin therapy and proper follow-up had slightly lower readmission rates, aligning with the notion that attention to glycemic control improves outcomes. In the drug co-prescription analysis, there are 20 different drug entities (nodes), additionally, there are co-administration nodes that connect multiple drugs. Using the medication change relationships, we merged encounters to build a drug–drug co-occurrence edge (counting how often two drugs were given together). An unsupervised algorithm grouped metformin with sulfonylureas and insulin with ACE inhibitors, reflecting common co-prescriptions for diabetes and its comorbid conditions.

## 4. Conclusions

Transforming the diabetes hospital dataset into a labeled property graph within the Neo4j platform and using the Neo4j GDS library unlocks analytical insight that is hard to produce with tabular data. The graph model enables intuitive exploration for domain experts, of highly complex multi-hop clinical patterns. LPG allows for the analysis of the important chain of co-occurring conditions and treatments related to readmission risk. Importantly, LPGs allow for discovering communities in electronic health records graphs, of co-morbid conditions to identifying central medications in treatment networks, the graph approach provides a richer lens to examine health data. As EHR datasets continue to grow in size and complexity, graph-based techniques such as knowledge graphs, graph algorithms, and advanced workflows relying on graph neural networks (GNN) or Graph Transformers, are positioned to play an increasingly important role in clinical decision support and healthcare analytics [2]. The results here motivate further development of graph-powered tools for patient risk prediction, care pathway optimization, and integrative analyses that ultimately support better data-driven healthcare.

**Acknowledgment**

This work has been supported by the Serbian Ministry of Science, Technological Development, and Innovations, Agreement No. 451-03-137/2025-03/ 200122. We gratefully acknowledge use of the "Diabetes 130-US Hospitals for Years 1999–2008" dataset, provided by the UCI Machine Learning Repository (DOI: 10.24432/C5230J). UCI Machine Learning Repository When citing this dataset, please also reference the introductory study describing its construction and variables: Strack B., DeShazo J.P., Gennings C., Olmo J.L., Ventura S., Cios K.J., Clore J.N. Impact of HbA1c Measurement on Hospital Readmission Rates: Analysis of 70,000 Clinical Database Patient Records, BioMed Research International, 2014, Article ID 781670. pmc.ncbi.nlm.nih.gov

**References**

[1] J. Schaff, C.C. Fink, B. Slepohenko, J.H. Carson, L.M. Loew., *A General Computational Framework for Modeling Cellular Structure and Function,* Biophysical Journal, 73 (1997) 1135-1146.

[2] B. Strack, J.P. DeShazo, C. Gennings, J.L. Olmo, S. Ventura, K.J. Cios, J.N. Clore., *Impact of HbA1c measurement on hospital readmission rates: analysis of 70,000 clinical database patient records*, BioMed Research International, 2014 (2014) 781670.

[3] Y. Li, H. Qiu., *Disease network analysis to reveal comorbidity patterns in hospitalized patients with COPD using large-scale administrative health data*, Scientific Reports, 15 (2025) 27082.

[4] R. Alotaibi, C. Lei, A. Quamar, V. Efthymiou, F. Özcan., *Property Graph Schema Optimization for Domain-Specific Knowledge Graphs*, 2021 IEEE 37th International Conference on Data Engineering (ICDE), (2021) 924–935.

[5] I. Robinson, J. Webber, E. Eifrem., *Graph Databases: New Opportunities for Connected Data*, 2nd ed., O'Reilly Media, (2015) 23–24.

[6] P. Marjai, B. Szabari, A. Kiss., *An Experimental Study on Centrality Measures Using Clustering*, Computers, 10 (2021) 115.

[7] T. Bratanič., *Graph Algorithms for Data Science: With examples in Neo4j*, Manning Publications, (2024).

[8] M.E.J. Newman, M. Girvan., *Finding and evaluating community structure in networks*, Physical Review E, 69 (2004) 026113.