

Graph-Native Normalization

Reducing Redundancy in Labeled Property Graphs

Johannes Schrott

21st GDC TUC Meeting | September 4, 2026



A Graph-Native Approach to Normalization

Accepted and published at EDBT 2027

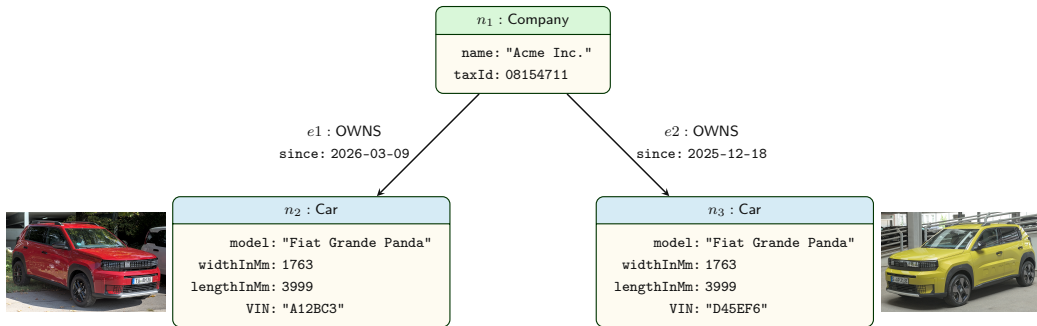
Authors:

- Johannes Schrott
- Maxime Jakubowski
- Katja Hose



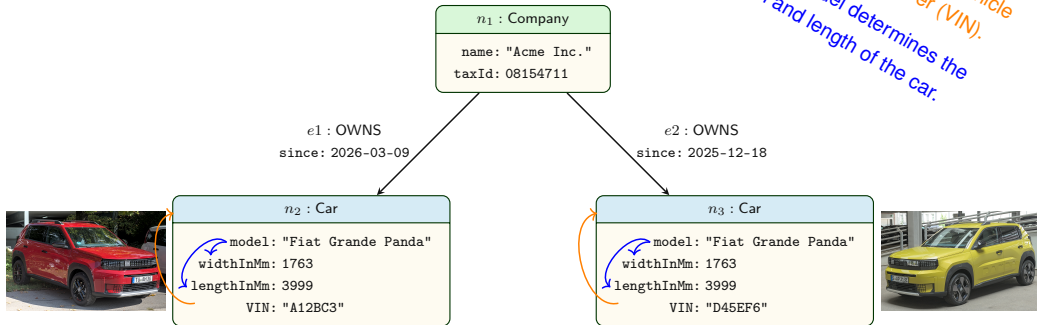
<https://dx.doi.org/10.48786/edbt.2027.06>

Redundancy and Anomalies in LPGs

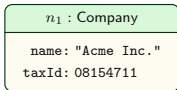
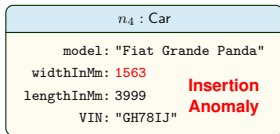


Redundancy and Anomalies in LPGs

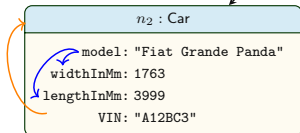
A car is identified by its Vehicle Identification Number (VIN).
The car model determines the width and length of the car.



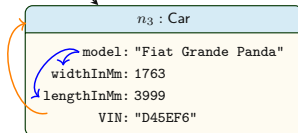
Redundancy and Anomalies in LPGs



e_1 : OWNS
since: 2026-03-09

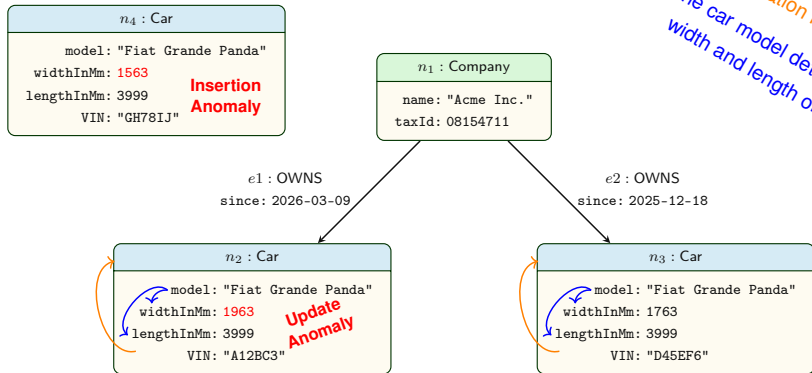


e_2 : OWNS
since: 2025-12-18



A car is identified by its Vehicle Identification Number (VIN).
The car model determines the width and length of the car.

Redundancy and Anomalies in LPGs



Redundancy and Anomalies in Relational Data

Original relational table:

CompanyName	CompanyTaxId	CarOwnedSince	CarVIN	CarModel	CarWidthInMm	CarLengthInMm
Acme Inc.	08154711	2025-12-18	D45EF6	Fiat Grande Panda	1763	3999
Acme Inc.	08154711	2026-03-09	A12BC3	Fiat Grande Panda	1763	3999

Redundancy and Anomalies in Relational Data

Original relational table:



CompanyName	CompanyTaxId	CarOwnedSince	CarVIN	CarModel	CarWidthInMm	CarLengthInMm
Acme Inc.	08154711	2025-12-18	D45EF6	Fiat Grande Panda	1763	3999
Acme Inc.	08154711	2026-03-09	A12BC3	Fiat Grande Panda	1763	3999

Redundancy and Anomalies in Relational Data

Original relational table:



CompanyName	CompanyTaxId	CarOwnedSince	CarVIN	CarModel	CarWidthInMm	CarLengthInMm
Acme Inc.	08154711	2025-12-18	D45EF6	Fiat Grande Panda	1763	3999
Acme Inc.	08154711	2026-03-09	A12BC3	Fiat Grande Panda	1963	3999
<i>null</i>	<i>null</i>	<i>null</i>	GH78IJ	Fiat Grande Panda	1563	3999

Redundancy and Anomalies in Relational Data

Original relational table:



CompanyName	CompanyTaxId	CarOwnedSince	CarVIN	CarModel	CarWidthInMm	CarLengthInMm
Acme Inc.	08154711	2025-12-18	D45EF6	Fiat Grande Panda	1763	3999
Acme Inc.	08154711	2026-03-09	A12BC3	Fiat Grande Panda	1763	3999

Redundancy and Anomalies in Relational Data

Original relational table:



CompanyName	CompanyTaxId	CarOwnedSince	CarVIN	CarModel	CarWidthInMm	CarLengthInMm
Acme Inc.	08154711	2025-12-18	D45EF6	Fiat Grande Panda	1763	3999
Acme Inc.	08154711	2026-03-09	A12BC3	Fiat Grande Panda	1763	3999

Normalized relational tables:

CompanyName	CompanyTaxId	CarOwnedSince	CarVIN
Acme Inc.	08154711	2025-12-18	D45EF6
Acme Inc.	08154711	2026-03-09	A12BC3



CarVIN	CarModel
D45EF6	Fiat Grande Panda
A12BC3	Fiat Grande Panda



CarModel	CarWidthInMm	CarLengthInMm
Fiat Grande Panda	1763	3999

Redundancy and Anomalies in Relational Data

Original relational table:



CompanyName	CompanyTaxId	CarOwnedSince	CarVIN	CarModel	CarWidthInMm	CarLengthInMm
Acme Inc.	08154711	2025-12-18	D45EF6	Fiat Grande Panda	1763	3999
Acme Inc.	08154711	2026-03-09	A12BC3	Fiat Grande Panda	1763	3999

Normalized relational tables:



CompanyName	CompanyTaxId	CarOwnedSince	CarVIN
Acme Inc.	08154711	2025-12-18	D45EF6
Acme Inc.	08154711	2026-03-09	A12BC3

CarVIN	CarModel
D45EF6	Fiat Grande Panda
A12BC3	Fiat Grande Panda

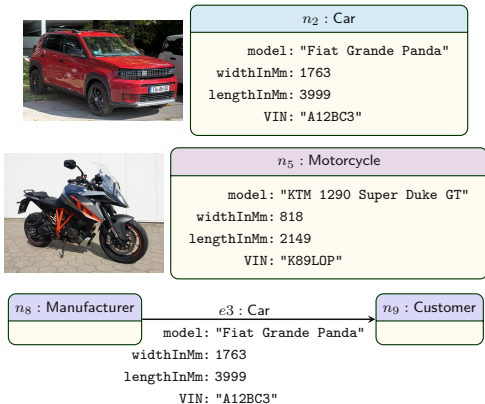
CarModel	CarWidthInMm	CarLengthInMm
Fiat Grande Panda	1763	3999

Anomalies are avoided!

Normalization for LPGs

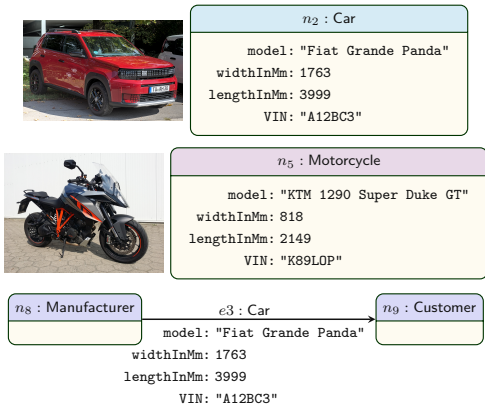
Normalization for LPGs

Scoping: What to consider?

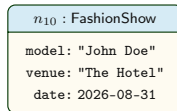
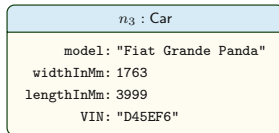


Normalization for LPGs

Scoping: What to consider?



Meaning of Keys: What means what?



Functional Dependencies for Graphs

$$\underbrace{Q}_{\text{Scope}} :: \underbrace{A \Rightarrow B}_{\text{Descriptor}}$$

Dependency Type	Scope	Descriptor
Graph Functional Dependencies (Fan et al., 2016)	Pattern	Predicates referencing property keys
Graph Entity Dependencies (Fan et al., 2017)	Pattern	Predicates referencing property keys or graph objects
Graph-tailored Functional Dependencies (Skavantzios et al., 2023)	Nodes	Property keys
Graph Functional Dependencies (Egger et al., 2026)	Pattern	Property keys
Graph Object Functional Dependencies (Schrott et al., 2026)	Pattern	Property keys and graph objects

Functional Dependencies for Graphs

$$\underbrace{Q}_{\text{Scope}} :: \underbrace{A \Rightarrow B}_{\text{Descriptor}}$$

Dependency Type	Scope	Descriptor
Graph Functional Dependencies (Fan et al., 2016)	Pattern	Predicates referencing property keys
Graph Entity Dependencies (Fan et al., 2017)	Pattern	Predicates referencing property keys or graph objects
Graph-tailored Functional Dependencies (Skavantzios et al., 2023)	Nodes	Property keys
Graph Functional Dependencies (Egger et al., 2026)	Pattern	Property keys
Graph Object Functional Dependencies (Schrott et al., 2026)	Pattern	Property keys and graph objects

Graph-Native Normalization

Core Idea – Redundancy reducing transformations

within-n

Redundancy
pattern



Legend: $\langle \rangle$ angle brackets denote new node IDs, underline denotes fresh labels, patterns also exist for edges of the other direction
n: node, np: node property, e: edge, ep: edge property

Graph-Native Normalization

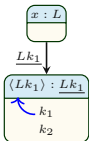
Core Idea – Redundancy reducing transformations

within-n

Redundancy
pattern



Resolved
redundancy
pattern



Legend: $\langle \rangle$ angle brackets denote new node IDs, underline denotes fresh labels, patterns also exist for edges of the other direction
n: node, np: node property, e: edge, ep: edge property

Graph-Native Normalization

Core Idea – Redundancy reducing transformations



Legend: $\langle \rangle$ angle brackets denote new node IDs, underline denotes fresh labels, patterns also exist for edges of the other direction
 n: node, np: node property, e: edge, ep: edge property

Graph-Native Normalization

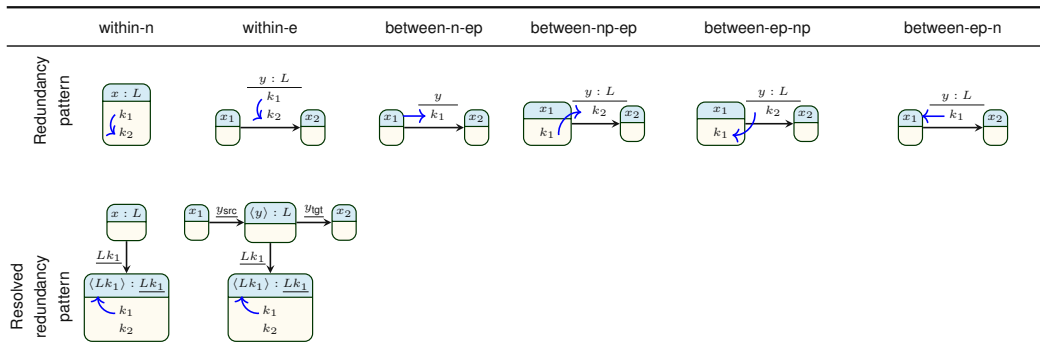
Core Idea – Redundancy reducing transformations



Legend: $\langle \rangle$ angle brackets denote new node IDs, underline denotes fresh labels, patterns also exist for edges of the other direction
 n: node, np: node property, e: edge, ep: edge property

Graph-Native Normalization

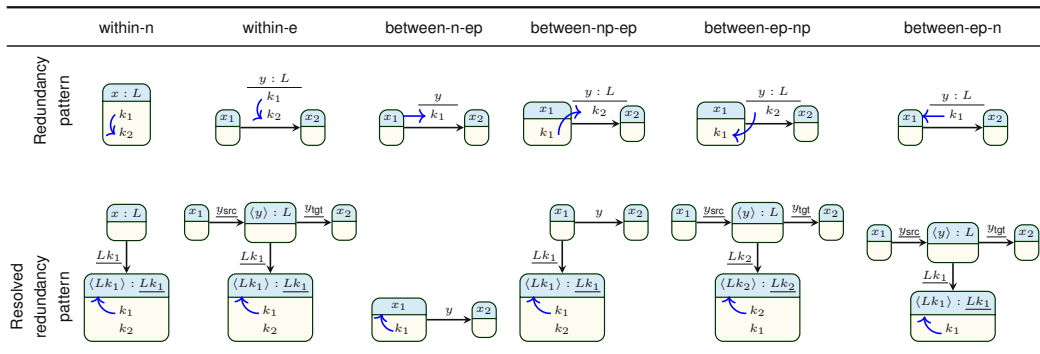
Core Idea – Redundancy reducing transformations



Legend: $\langle \rangle$ angle brackets denote new node IDs, underline denotes fresh labels, patterns also exist for edges of the other direction
 n: node, np: node property, e: edge, ep: edge property

Graph-Native Normalization

Core Idea – Redundancy reducing transformations



Legend: $\langle \rangle$ angle brackets denote new node IDs, underline denotes fresh labels, patterns also exist for edges of the other direction

n: node, np: node property, e: edge, ep: edge property

Graph-Native Normalization

Approach

Objective: Only key dependencies are present

Normalization Process:

1. **Compute minimal cover**
2. Sort dependencies by scope – more specific first!
3. Apply transformations

$$x.k_1 \Rightarrow x.k_2$$

$$x.k_2 \Rightarrow y.k_3$$

$$x.k_1 \Rightarrow y.k_3$$

Result: Normalized graph + transformed dependencies

Graph-Native Normalization

Approach

Objective: Only key dependencies are present

Normalization Process:

1. **Compute minimal cover**
2. Sort dependencies by scope – more specific first!
3. Apply transformations

$$x.k_1 \Rightarrow x.k_2$$

$$x.k_2 \Rightarrow y.k_3$$

$$x.k_1 \Rightarrow y.k_3$$

Result: Normalized graph + transformed dependencies

Graph-Native Normalization

Approach

Objective: Only key dependencies are present

Normalization Process:

1. Compute minimal cover
2. **Sort dependencies by scope – more specific first!**
3. Apply transformations

Result: Normalized graph + transformed dependencies

Convertible cars



Car



Vehicle

Graph-Native Normalization

Approach

Objective: Only key dependencies are present

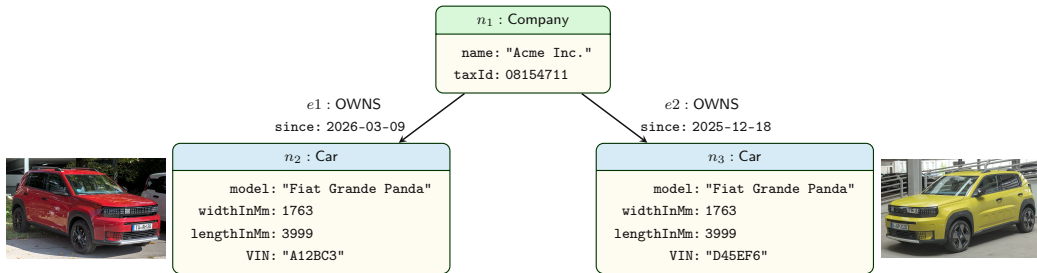
Normalization Process:

1. Compute minimal cover
2. Sort dependencies by scope – more specific first!
3. **Apply transformations**

Result: Normalized graph + transformed dependencies

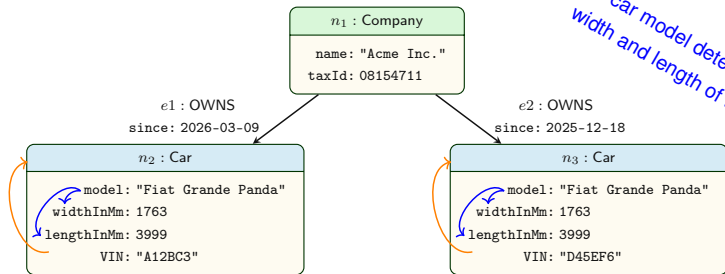
Graph-Native Normalization

In Practice



Graph-Native Normalization

In Practice

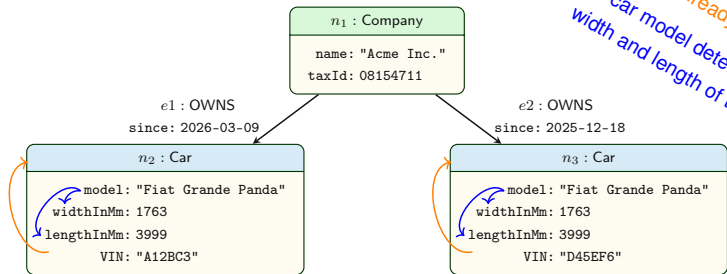


A car is identified by its Vehicle Identification Number (VIN).

The car model determines the width and length of the car.

Graph-Native Normalization

In Practice

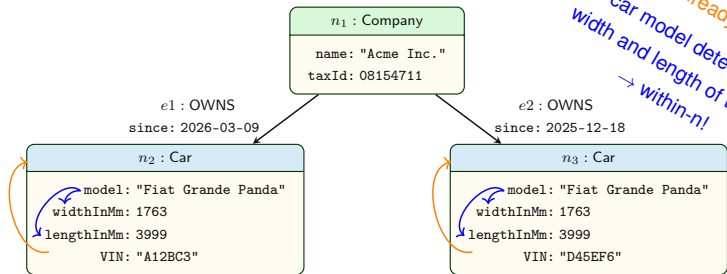


A car is identified by its Vehicle Identification Number (VIN).
→ Already a key!

The car model determines the width and length of the car.

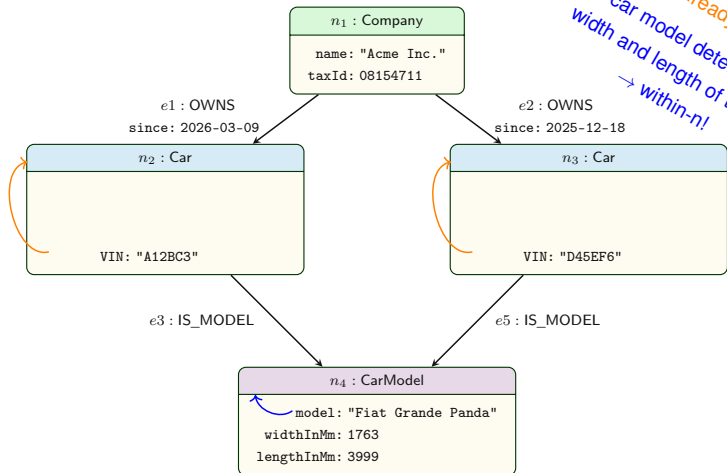
Graph-Native Normalization

In Practice



Graph-Native Normalization

In Practice

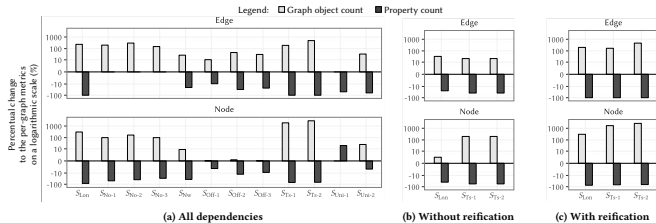


A car is identified by its Vehicle Identification Number (VIN).
→ Already a key!

The car model determines the width and length of the car.
→ within-n!

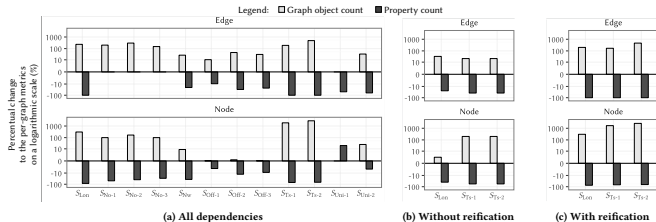
Evaluation

- Per-graph metrics:
 - Number of graph objects increases
 - Average number of property records decreases



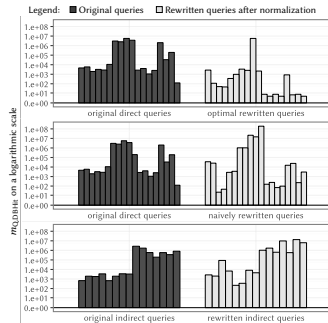
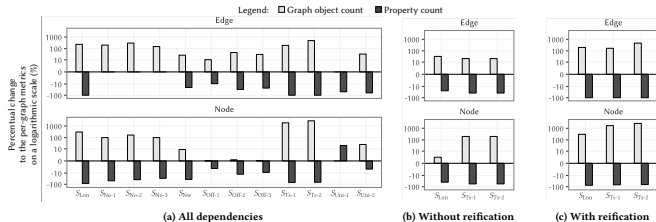
Evaluation

- Per-graph metrics:
 - Number of graph objects increases
 - Average number of property records decreases
- Per-dependency metrics:
 - Redundancies are eliminated, anomalies avoided



Evaluation

- Per-graph metrics:
 - Number of graph objects increases
 - Average number of property records decreases
- Per-dependency metrics:
 - Redundancies are eliminated, anomalies avoided
- Query performance:
 - Increase and decrease possible, depending on dependencies and queries



Outlook

- Consider more complex patterns
- Practical tooling
- Large language models for labeling



<https://dx.doi.org/10.48786/edbt.2027.06>



[https://dmki-tuwien.github.io/
lpg-normalization/evaluation_results/](https://dmki-tuwien.github.io/lpg-normalization/evaluation_results/)