Graph Databases and Railway Operations Research Requirements

Alexander Kuckelberg
VIA Consulting & Development GmbH
Römerstr. 48-50
D-52064 Aachen
+49 241 463 662-16
a.kuckelberg@via-con.de

ABSTRACT

In this work-in-progress paper we describe requirements, scenarios and mandatory functionalities of graph databases within the application field of railway operations research (ROR).

The underlying railway infrastructure data of all ROR tasks can naturally be described by graph structures and can therefore be managed by graph databases; railway operations research functionalities might consequently be described as database functions on its graphs.

While the infrastructure data should remain persistent, a graph database might be a good choice to match the persistence needs quite close or even identical to the data structures to be managed. Moreover, the functionalities might be transformed into database functionality.

In current, productive systems, relational databases respectively models are still the most widely-used models, on which current infrastructure persistence is realized.

The work-in-progress focuses on the question, if graph databases with database supported functionalities might be a good alternative compared to current solutions on top of relational models.

This paper tries to outline a generic graph model as it can be used in ROR, to define requirements and framework conditions. It tries to summarize generic demands and to describe the query and functionality requirements that have to be satisfied by such databases. This paper presents basic ideas and the origin point of intended and starting database research projects and cooperation with universities in the next month and years.

Keywords

Railway infrastructure data, persistence of graph topology, railways operations research functionality, infrastructure database.

1. INTRODUCTION

The application field of railway operations research is a quite special field which is usually associated to other topics than graph databases or graph structured data and their management.

Usually railway operations research deals with topics like delay propagation, robustness of timetables, capacity of infrastructure, capacity allocation or evaluation of infrastructure modification effects. Such topics can be analyzed and answered on behalf of analytical, constructive or simulative approaches [5].

While the main focus of ROR activities usually lies on sufficient algorithms, formulas or modelling approaches the mostly unmentioned basic of all functionality is an infrastructure graph, which acts as the basis for running time computations, blocking or minimum headway time determination and the ability to select alternative routes, additional stops or to perform rescheduling operation.

Therefore, an essential but usually unconsidered component of all railway operations research activities is an infrastructure network graph, on which all functionalities are based.

The railway infrastructure network consists of rails, switches, crossings, buffer stops etc. and can mathematically be described as a (directed) graph, which is the most static part of a railway operations research project[1].

Based on this graph – the infrastructure graph – functionalities are defined and tasks are performed. The most elementary functionality is the determination of running times and the determination of infrastructure occupation but also the search for matching routes, alternative stop policies or the evaluation of infrastructure capacities based on timetables or queueing theory.

The following chapters try to introduce approaches to infrastructure graph modelling, existing exchange formats and perspectives onto such graphs.

A generic graph definition as a consensus of different views and approaches is derived and typical functionalities performed on such graphs are outlined.

The last chapter finally describes our work in progress and summarizes currently ongoing database research activities, primarily targeting a performant prototype and accompanying prove of concept of the approach described in this paper and their suitability with respect to manageability on behalf of graph databases.

---

[1] Usually several timetables and their robustness of delay behavior are evaluated for a given network infrastructure, therefore the infrastructure is considered as "most static" within a project.

2. DATA MODEL AND FUNCTIONALITY

To describe the requirements and functionalities a graph database that is tightly fitting the application field of railway operations research should satisfy, it is worth to take short looks into existing models, data exchange formats and systems, focusing on infrastructure models.

To introduce more aspects from railway operations research – not only the basic infrastructure network and graph-like topology – some typical tasks and questions are shortly described to allow a better understanding of required functionalities.

From that starting point it hopefully becomes clearer, what is expected from graph databases for this quite specific task, in functionality, data model and performance requirements.

2.1 Proprietary models

Proprietary data models for railway operations research were introduced decades ago. While old systems for timetabling support in the 80s used quite specific track and infrastructure models like sequences of elements for single or double track lines, specific configuration records to describe the characteristic (and track existences) of stations etc. the graph topology approach became popular in the 90s with the increased availability of personal computers and their performance.

It became clear, that graph models are the most flexible structures and best fitting representations of the real network, but computation power and the acceptance of computer based systems still had to grow.

Typical systems of this time use either tool specific, binary and size dense data and file formats or standard database models like the relational one to store and manage the network data persistently. While relational databases were considered to be performant, widely available and standardized such databases only support railway operations functionality in a quite limited manner. Databases are primarily used as persistence stores to guarantee ACID characteristics when working with network infrastructure data.

The functionality is usually implemented on top of the standard database system. The computation of e.g. possible routes cannot be implemented directly within the relational database domain[2]. So currently ROR functionality is performed on behalf of data loaded into main memory with corresponding performance and accessibility benefits but with the drawback, that database functionalities like transactional control is not available. Finally the evaluation of graph database based approaches for data persistency has to be compared against this "traditional" scenario: load from database and restore the network graph, perform functionalities within main memory, and probably store manipulated data back into the database in contrast to the directly performed functionalities within graph databases.

2.2 Graph models and UIC RailTopoModel

Even if the graph model is currently considered to be a sufficient and suitable approach to model network infrastructure, the content of these models differs, especially when considering different levels of granularity as Figure 1 illustrates.

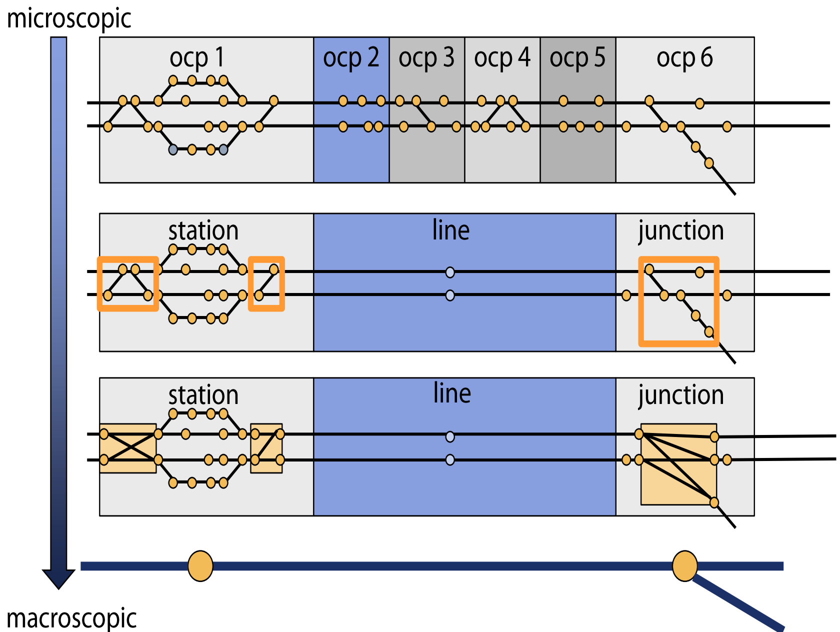

Figure 1: Microscopic, mesoscopic and macroscopic network infrastructure graphs and node aggregation into operational control points, stations, junctions and lines.

There are several more or less widely used approaches, philosophies and granularities used by different systems. Mostly, these models were implied by legacy systems, research prototypes or available data sources.

There are node weighted or edge weighted and attributed approaches, which both have advantages and disadvantages with respect to redundancies, performance or expressional strength.

There are microscopic and macroscopic models considering the network topology and network elements in varying detail depth. Microscopic models consider not only the track related topology but also signals, liberation equipment, curve radius, track gradients, tunnels, switches or stopping positions, balises, speed profiles etc., usually in precise of meters. Beside this, specific tools used for infrastructure planning might moreover contain much more elements and positioning precise.

Moreover, the application field a tool is designed for as well as the local technical requirements and circumstances determine the content of the infrastructure network graph[3].

[2] The relational model and SQL was extended by specific functionality, e.g. closure operators, but nevertheless these functionalities are not really used within current systems for different reasons like performance, portability or even availability within a specific RDBMS.

[3] In Germany there is e.g. a train protection system called LZB (lineare Zugbeeinflussung) which requires to model LZB areas, marker boards for area characteristics and which is not available in most other countries. The same is true for several other systems, which usually are country centered developments.

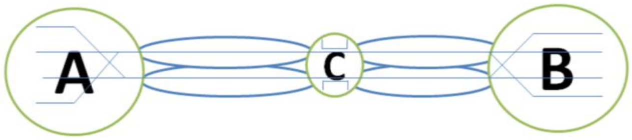

Figure 2: Micro-/macroscopic infrastructure modelling (UIC RailTopoModel).

Several country specific, national models and modelling approaches exists, e.g. in Germany, the DB Netz AG uses a node weighted graph model – the Spurplan – within their timetabling systems RUT-K [8] that defines a wide range of allowed elements and element instances to set up a network. The same is true for a Belgium specific approach to an infrastructure graph model, the INT graph [6].

To generalize and abstract the overall problem of infrastructure network graph modelling and its management within database systems, a scientific, generic model has to be used or to be derived.

One quite interesting project targeting such a generic infrastructure model is the RailTopoModel initiated by the UIC [1]. One of the ideas behind is to model network topologies for macroscopic as well as for microscopic approaches and to define mapping and transformation functionalities between different levels of granularity (Figure 2).

Consequently the UIC RailTopoModell is a (at least currently) promising approach to set up a generic network graph model that might cover a bright range of generic requirements. It therefore is one source of the overall graph model which should be implemented and supported by a graph database targeted by our ongoing work.

One crucial aspect which constantly causes problems with respect to generalization and universal usability and acceptance of such models are nationally affected (non-functional) requirements, e.g. a regulatory for clustering the network into operation control points (OCP), the aggregation of tracks within lines, separation of grids and intergrids and much more. Such classification criterion and requirements are often the background for a specific modelling and might be generally be described as graph clustering, coverage and overlapping problems (chapter 3.2).

2.3 Exchange Formats

Similar to graph models the exchange formats evolved. Besides "owning" a graph model it makes sense to define a sufficient exchange format for data fitting to this model.

While for the proprietary formats mentioned within section 2.1 usually binary file formats were used, nowadays data exchange formats are XML based, e.g. defined by XSD schemata.

The German Spurplan used by DB Netz AG implied the (company internal) XML-ISS standard for railway research operation tools. For more operation and planning centered systems other standards exists and are currently under development, e.g. within the PlanPro project [7].

The RailML project [2] is another example, where an international partnership tries to define an exchange format (not only for infrastructure) in a more or less generic and universally valid manner. Unfortunately, this approach again focusses on a quite specific model – a track oriented view – which contradicts the initially expressed universality. Moreover in practice, missing semantic specifications reduces the universal validity of exchange formats like RailML to a pairwise agreement and convention, which again strongly contradicts any standardization intention of this project.

3. GRAPH DATABASES

This paper wants to gain insight into ongoing work. This work focuses on graph databases and how such (new[4]) database approaches might be used in a beneficial manner to support, replace or extend the nowadays systems, their functionalities and performances.

The ongoing work focuses on research and evaluation activities and join-projects with universities and the determination of solutions which matches the application field requirements in a quite optimal manner.

In the following subsections requirements and demands are outlined, that have to be considered when designing and evaluating graph databases and their functionalities to be enabled to compare such rather new and alternative approaches to existing ones.

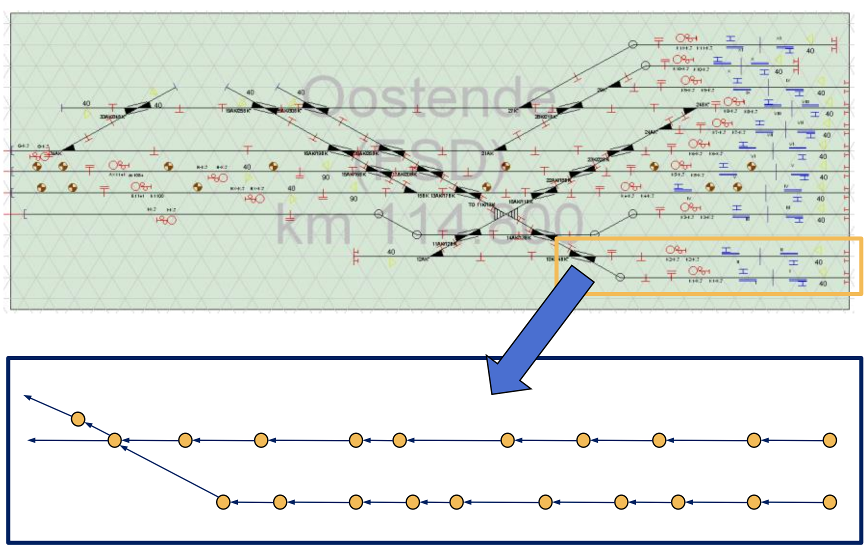

Figure 3: Rail network and network graphs.

3.1 Base Topology and requirements

As mentioned before, a generic graph model fits the network graph modelling requirements in a best way, similar to the mentioned UIC RailTopoModel (Figure 3). Additionally to this simple model, an infrastructure graph database must consider several more aspects outlined in the following subsections.

From our point of view, the core and elementary rail network topology should be modelled by a directed graph similar to the one proposed by the UIC RailTopoModel:

4 At least within the application field considered.

- A network graph is a graph G=(N, E) where N is a set of nodes and E is a set of (directed) edges with $E \subseteq N \times N$.
- The directed cardinality $|n|=(s_e, s_l)$ of a node $n \in N$ is defined by the number s_e of edges entering n and the number s_l of edges leaving n.
- A node $n \in N$ is called an inner node if $|n|=(1,1)$ and edge node otherwise.
- The graph is considered to be a node weighted graph, where characteristic values, e.g. speeds allowed, changing gradients or signaling functionality is assigned to nodes.
- Track sections are paths throughout the graph starting and terminating at edge nodes with only inner nodes within the path. For a path $P=(n_1, .., n_m)$ of nodes, n_i is an inner node for i=2…m-1 and $(n_i, n_{i+1}) \in E$ for i=1…m-1. A track section contains at least two nodes (m>1).
- All inner nodes are attributed by direction validity, e.g. a node is valid for train running within the direction of the associated edge of in opposite direction (or both). This validity has to be considered by all functionalities like running or occupation time computation as well as routing and route evaluation.
- For all nodes of a track section $P=(n_1,…,n_m)$ a layout position within a defined (from many possible) positioning system is given. This might be a GIS coordinate in case of GIS systems or the layout coordinates of a linearized or user friendly display of the network graph.
- For all nodes of a track section $P=(n_1,…,n_m)$ a (relative) positioning information pos(n) is assigned with $pos(n_1)=0$, $pos(n_m)=1$ and $pos(n_i) \leq pos(n_{i+1})$ for i=1…m-1.
- The mileage of nodes respectively section elements is derived from location information (GIS/meters/etc.) assigned to the section start and end due to the positioning information.

Nodes might additionally be distinguished due to their semantic, for which area or length they are valid. Most nodes respectively corresponding infrastructure element are usually point elements, whose semantic is related to a specific point, e.g. a speed change, a stopping position (the “H”-board) or the location of a signal.

But nevertheless semantic might be extended to area and length validity, e.g. speed restriction zones, level crossings etc. The validity semantic is expressed by node attributes.

3.2 Topology Coverage and Clustering

Some of the most problematic issues towards a unified topology model are national rules and regulations as mentioned before. E.g. in Germany infrastructure elements are logically organized with operation control points as the top-most classification criteria. In other countries grids or inter-grids are the primary structuring criteria; sometimes a track line is the major criteria.

The topology graph model we consider for the intended graph database implementation tries to generalize all these approaches on behalf of graph coverages and graph node clustering:

- The cluster C of a network graph G=(N, E) is a graph $G_C=(N_C, E_C) \subseteq G$ such that $N_C \subseteq N$ and for all $n_1,n_2 \in N_C$ with $(n_1,n_2) \in E$, also $(n_1,n_2) \in E_C$ holds.
- A coverage $CV=\{C_1, …C_m\}$ of a network graph G=(N, E) is a set of clusters of G.
- A total coverage $TC=\{C_1, …C_m\}$ of a network graph G=(N, E) is a set of clusters of G such that $C_i=(N_i, E_i)$ for i=1,…,m and N_i, N_j (i,j=1,…,m, $i \neq j$) are disjunctive node sets whose union is N.

With this clustering, it is possible to define the varying logical orderings and classifications as mentioned before:

- The logical separation of a network graph G into operation control points is a coverage of G.
- A network graph G can be separated into grids and inter-grids. A grid-inter-grid-approach is a total coverage $TC=\{C_1, …C_m\}$ (m>0) of G such that a cluster C_i is a grid, whenever there is a node $n \in C_i$ with $|n|=(x_{in}, x_{out})$ and $x_{in}>1$ or $x_{out}>1$ (switch or crossing), and an inter-grid otherwise. All edges from G not contained in TC always connect nodes from grids to nodes from inter-girds or wise versa.
- Lines $L=\{C_1, …C_m\}$ (m>0) of a network graph G are a (not total) coverage of G where all nodes of each cluster C_i (i=1…m) are part of at least one path within C_i.
- Power supply areas of a network graph G are areas within the corresponding network, where (electrical) power is supplied by one or more transformer substations. Therefore the power supply areas of G can be modelled as a (not total) coverage of G.

All examples stated before are examples of different logical clustering of the overall network graph and should illustrate the functionality which has mandatorily to be supported by database, especially the support of clustering and additional cluster constraints.

3.3 Interlocking Routes

Several existing infrastructure data models for railway networks concentrate on a quite limited view on the rail (and graph) topology as a primary (and only) modeling aspect, as e.g. RailML does until nowadays.

For railway operations research tools this view is not sufficient. Track related systems like railways basically rely on interlocking techniques and therefore this aspect has to be supported by models and consequently by databases as well.

A route of an infrastructure graph G is a path within G corresponding to the technical circumstances given by the concrete settings of an interlocking station and its ability to control signals, switches and track accessibility.

So one additional requirement a graph database for infrastructure graph management has to fulfill is to support coverages representing routes and paths within the graph.

In contrast to "usual" routing and path finding functionality (which nevertheless is required but considered later on within the paper) specific route data has to be stored, because such routing information has to be enriched by application field specific attributes. Therefore it could be said that the graph database has to be able to manage attributed routing information. Such attributes might be information about the usability (electrification, axle weights, stopping positions offered etc.) that are available in addition to the pure infrastructure information, the classification of certain routes or the relevance for different train types respectively train classes[5]. With such route information several railway operations research functionalities are supported like computer based routing or rescheduling.

Routes typically start at one graph node and describe a path to another graph node. Such nodes can be signals, track ends/boundaries or even specific reference nodes[6].

3.4 Temporal Validity

One aspect typically not considered by infrastructure models is the spatio-temporal validity of the infrastructure network. Railway operations research functionality typically concentrates on a specific network graph, but this consideration is not necessarily true in any case.

Within timetabling periods there are more or less important changes somewhere in the network. Switches are added or removed, interlocking stations are extended or modified, tracks or even complete areas are closed for maintenance work etc.

Therefore it must be ensured, that a universally usable graph network considers temporal validities and retrieves network graphs and topologies depending on requested times respectively time periods.

So one essential question for this work in progress is how graph databases can be used to access different topologies changing over time and with which performance.

3.5 Routing and path finding

Last but not least another elementary functionality for the considered application field is the routing functionality as known from several similar application fields like route guidance and navigation systems.

The graph database has to offer this functionality on behalf of interlocking routes as described in section 3.3. In practice, queries for routes between two graph nodes have to consider the train characteristic and priority of the routes as well as the attributed routing information.

While this "plain" routing functionality is used e.g. for timetabling, the routing for railway operation simulation or analytical evaluation has to perform this search slightly different. Usually along the train run overtaking sections for the specific train have to be determined. Overtaking sections are areas of a network graph, where in practice no change of train order can be performed. The two ends of an overtaking section are characterized by the ability to change this train order, concretely to allow one train to overtake or to be overtaken by another train.

This is again determined by alternative routing selecting sufficient routes at the section ends which e.g. offer a sufficient stopping position and electrification.

3.6 Summary

In this chapter, requirements against a graph database to handle network graphs for railway operations research purpose were mentioned and introduced. Roughly speaking, the most important are:

- Support of (node weighted) graphs with different positioning and layout systems.
- Temporal validities and the ability to retrieve time specific graph topologies.
- Clustering and coverage of network graphs to ensure generality.
- Management of interlocking routes and routing functionality on top of these routes.

With this functional "specification" the evaluation of graph databases as a sufficient persistent storage system can be started.

4. WORK IN PROGRESS/NEXT STEPS

The handling of infrastructure network data and ensuring its persistency is an elementary component of nowadays railway operations research tools.

The current legacy system landscape usually uses "traditional, relational" approaches to store and manage such data. There is an obvious mismatch between the relation and set oriented paradigms of these databases and the topology, semantic and structure of graphs which are a "natural" model for railway network infrastructure.

If functionalities like simulation, timetabling, capacity evaluation or other tasks from the application field of railway operations research should be performed, they are currently performed on in-memory data structures which had been created from the relational data sets while loading them.

The existence of graph databases obviously closes the gap between the database model and the one of the specific application domain. A central question for commercial tools is if it is worth to shift to new, less evaluated approaches like graph databases.

For this reason we work on an evaluation of the performance and functional capabilities of graph databases in comparison to "traditional, relational" approaches.

At the Workshop on Querying Graph Structured Data 2015 (GraphQ 2015) we expect to be able to present and show first results, provide an insight into current settings of this ongoing

[5] Often there are tracks and lines dedicated e.g. to freight or passenger trains, even if both types are physically comparable (same gauge, same locomotive etc.) but routes are more relevant for on type than for the other. Therefore a route might have a high priority for freight trains and a very low one for passenger trains.

[6] This corresponds to the interlocking paradigms, where exactly one origin and one target have to be defined before the interlocking process – e.g. setting up switches and signals – is accepted and started.

evaluation or at least discuss aspects of this problem field at the workshop itself.

The evaluation is intended to start soon as a joint-project between VIA Consulting & Development GmbH as initiator of this work, different students and universities specialized on graph database techniques and the railway infrastructure manager DB Netz AG in Germany.

It is expected to provide graph and route data from existing systems with expected graph sizes up to several hundred-thousand nodes and thousands of interlocking routes for the whole German railway network. In this way it will be ensured, that the research and evaluation work is related to practical conditions and requirements.

The next steps from the current stage of the ongoing work are the definition and selection of different evaluation and comparison scenarios and modelling approaches with respect to specific databases. As a basis of comparison, a relational database as it is currently used in practice is considered.

5. REFERENCES/LITERATURE

[1] UIC – International Union of Railways, UIC RailTopoModel: Railway Network Description – A conceptual model to describe a railway network, http://documents.railml.org/science/280714_uic_railtopomodel_rc2.pdf

[2] railML.org Initiative, http://www.railml.org

[3] Hansen, I. A.; Pachl, J. (eds.): Railway Timetabling & Operations. Analysis - Modelling - Optimisation - Simulation - Performance Evaluation. Eurailpress 2014, ISBN 978-3777104621

[4] Kuckelberg, A.; Seybold, B.: “Adaptive Rule-Based Infrastructure Modelling” – In: Proc. of the 5th International Seminar on Railway Operations Modelling and Analysis, Copenhagen, 13.-15.05.2013.

[5] Janecek, D.; Kuckelberg, A.; Nießen, N.: “Kapazitätsermittlung von Eisenbahnknoten und Strecken” – In: Eisenbahntechnische Rundschau (ETR) 61 (2012) 10, pp. 30-36.

[6] INT – Graph Model Design, Infrabel internal working paper, INT RFT.04

[7] PlanPro, Durchgängige Datenhaltung der Leit- und Sicherungstechnik (LST) von der Planung bis zum Bestand (in German), DB Netze, http://fahrweg.dbnetze.com/fahrweg-de/start/technik/innovationen/planpro

[8] K. Wölfle, RUT-K – Computer-Aided Train-Path Management, Paris, 13.10.05, Talk at UIC