2 PRELIMINARIES

Data model. A property graph is \( G = (V, E, s, t, L, T, L, T, P_e, P_v) \), where \( V \) is a set of vertices, \( E \) is a set of edges and \( s, t : E \to V \times V \) assigns the source and target vertices to edges. Vertices are labelled from \( L \) by function \( L \) and edges are typed from \( T \) by function \( T \).

Let \( D = \cup_i D_i \) be the union of atomic domains \( D_i \). \( P_v \) is a set of vertex properties. A vertex property \( p_i \in P_v \) is a partial function \( p_i : V \to D_i \). Edge properties \( P_e \) can be defined similarly.

Given a property graph \( G \), relation \( r \) is a graph relation if the following holds [13]: \( \forall A \in \text{sch}(r) : \text{dom}(A) \subseteq V \cup E \cup D \), where \( \text{sch}(r) \) is the schema of \( r \) (a list containing attribute names), \( \text{dom}(A) \) is the domain of attribute \( A \), and \( V/E \) are the vertices/edges of \( G \).

Running example. We use the following example graph:

```
MATCH t = (p:Post)-[r:REPLY*]->(c:Comm)
WHERE p.lang = c.lang
RETURN p, t
```

GRA. Graph queries can be formulated in graph relational algebra (GRA) [20], which introduces two graph-specific operators: (1) the get-vertices nullary operator \( \mathcal{G}(V) \), which returns vertices \( v \) with a label \( V \) to serve as a base relation for later operators, (2) the expand-out unary operator \( \uparrow^{(W)}_{(v)}[E](r) \) that navigates from \( v \) on an edge typed \( E \) to a vertex \( w \) with label \( W \). The expand-out operator can also define transitive closure patterns, denoted by the \( \ast \) symbol. GRA allows nested data structures, i.e. if \( x \) is an attribute of a graph relation, \( x.p \) accesses the value of property \( p \) in \( x \) [13].

NRA. To allow precise formalisation of nested data structures, we use nested relational algebra (NRA) [7, 14], which allows arbitrary nesting of relations. To access nested values, attribute \( A \) of a nested relation \( r \) can be unnest \( \text{using} \) the operator \( \mu_A(r) \). Nested relations can also represent properties of vertices/edges along with collections such as lists and maps. We present two nested relations \( \alpha \) and \( \beta \) that store the vertices and edges of the graph, respectively:

```
\begin{array}{c|c|c|c|c|c|c|c|c|c|c|c|c}
| id | label | properties |
\hline
\begin{array}{|c|c|c|}
1 & Post & key & value \\
2 & Comm & lang & en \\
\end{array}
\end{array}
```

We define operators formally as \( \mathcal{N}(V) \equiv \pi_{id \to v} \sigma_{\alpha.label=v(\alpha)} \) and \( \uparrow^{(W)}_{(v)}[E](r) \equiv \sigma_{r.v=\beta} \sigma_{\beta.type=\ast} \sigma_{r.t=\alpha} \cdot \pi_{id \to \alpha.label=W} \) \( r \equiv \beta \equiv \alpha \).
3 RELATED WORK

Cypher. Due to its novelty, there are only a few research works on the formalisation of (open)Cypher. An early attempt to provide a framework for the theoretical representation of openCypher queries was published in [13]. In [20], we published a formalisation of a subset of openCypher that mapped queries to GRA. The Cypher for Apache Spark project is an ongoing effort to adapt the Cypher language to Spark [22]. None of these works considers IVM. Grafholow [15] is an active graph database for incremental openCypher queries. However, it does not support nested data structures.

IVM of graph queries. The VIATRA framework [33] provides an incremental query engine over the object-oriented Eclipse Modeling Framework. However, it does not support FGN or ORD. Strider [26] is a system supporting continuous SPARQL queries. As the RDF data model does not handle collections as first class citizens (only head-tail style lists are supported), FGN is not supported.

Querying nested data structures. Paper [16] presents a method for incremental view maintenance in object-oriented databases, but ordering is not supported. Recently, the authors of [5, 6] formalised the language of the MongoDB document store using nested relational algebra, including ordering. However, IVM was not considered. An approach for incremental calculation of XQuery expressions is presented in [9] and its accompanying technical report [8].

4 APPROACH AND CONTRIBUTIONS

As discussed in Section 1, order-preserving lists are required to store paths. Henceforth, we propose a property graph query model that only allows (unordered) bags, except for paths that are still stored as a list but can only be updated as an atomic unit (i.e. the previous path has to be deleted and the new one has to be inserted). We argue that the distinction between collection properties and paths makes sense from a practical point of view: collection properties often receive updates, while paths only benefit from incremental updates in rare cases (e.g. when a single transaction deletes an edge in the path but adds another one that keeps the path from deleting).

Overview. We propose the following workflow for compiling property graph queries to an incrementally maintainable expression and use the example of Section 2 for illustration.

<table>
<thead>
<tr>
<th>Query spec</th>
<th>Graph RA plan</th>
<th>Nested RA plan</th>
<th>Flat RA plan</th>
<th>Incremental engine</th>
</tr>
</thead>
</table>

(1) Compile the queries to GRA. A mapping from openCypher was given in our earlier work [20]. The example query results in:

\[
\pi_{p.t} \sigma_{c.lang=p.lang} \left( \upsilon (\cdot . cComm) . REPLY+ . \bigcap_{p.Post} \right)
\]

(2) Transform GRA to NRA, which is the key step to allow incremental maintenance. As expand operators cannot be maintained incrementally, they are replaced with joins. For this, we introduce the nullary get-edges operator \( [E \rightarrow v(w)] \) that returns triples \((v, e, w)\) for each edge \(e\) of type \(E\) between \(v\) of label \(V\) and \(w\) of label \(W\). Using this, each expand-out is replaced with natural joins:

\[
\upsilon (w) . \left[ E \rightarrow v \right] (r) \equiv r \Rightarrow_{[E \rightarrow v]} \left[ w \rightarrow W \right]
\]

Similarly, transitive expand-outs are replaced with transitive joins:

\[
\upsilon (w) . \left[ E^* \right] (r) \equiv r \Rightarrow_{[E \rightarrow v]} \left[ w \rightarrow W \right]
\]

Unlike relational databases, property graphs do not have a predefined schema. Hence, we slightly modify the unnest operator (Section 2) so that defines specific attribute(s) to be unnested from the nested relation. For example, \( \mu_{c.lang\rightarrow c}\) extracts the lang property of \(c\). Using these rules, the example is transformed to:

\[
\pi_{p.t} \sigma_{c.lang=p.lang} \left( \sigma_{cComm} \left[ \cdot . REPLY \right] \left( \bigcap_{p.Post} \right) \right)
\]

(3) Transform NRA to FRA following the approaches presented in [7, 25]. However, a key difference is that due to their schema-free nature, the schema of the nested relations is not known for property graphs in advance and has to be inferred based on the query. Therefore, this step includes pushing down nested attributes to the \(\bigcup\) and \(\cap\) operators. On the example, this results in:

\[
\pi_{p.t} \sigma_{c.lang=p.lang} \left( \sigma_{cComm} \left[ \cdot . REPLY \right] \left( \bigcap_{p.Post} \right) \right)
\]

(4) Create an incremental view for the FRA expression. Incremental view maintenance algorithms for FRA are well studied both from a theoretical perspective [2, 4, 10, 11] and implementation-wise, with many practical tools [12, 33] and research prototypes [15, 26, 31]. While they are not expressible in first-order logic, it is possible to evaluate transitive operations incrementally [3, 23].

Based on this approach, we propose that a fragment of the open-Cypher language, with unordered bags (instead of lists) and atomic paths (which can only be inserted or deleted, and lose their ordering when unnested), can be evaluated using relational IVM techniques.

Evaluation. The presented approach allows IVM for property graph queries while allowing FGN and some degree of ORD (for paths). In particular, the proposed fragment still allows returning paths and path unwinding [1], a feature that permits the query to iterate over the nodes of a path variable. The main tradeoff of the approach is that it does not allow users to use lists in the data model and queries. It is also not possible to specify top-k style queries, e.g. get the top 3 messages, based on the number of replies received.

Summary of contributions. Up to our best knowledge, our research is the first to investigate challenges of incremental view maintenance for property graph queries. We put a particular emphasis on handling nested data structures and ordering; and propose to limit the usage of ordering for (atomic) paths. Formulating the queries in NRA and flattening it to an FRA expression allows us to infer the minimal schema required by each operator, based on the query specification. Our approach does not require a priori knowledge of the data schema, unlike the schema cleanup algorithm of [34] (defined in the context of evaluating XQuery expressions on XML documents) and the schema merging algorithm of [18] (defined for consolidating multiple schemas into a mediated one).

Limitations and future work. Property graph queries present numerous additional challenges not presented in this paper. In particular, aggregations, the optional MATCH, WITH, SKIP constructs were omitted, and are discussed (for non-incremental queries) in our earlier work [20]. Expressions were also left for future work.

ACKNOWLEDGMENTS

This work was partially supported by NSERC RGPIN-04573-16 and the MTA-BME Lendület Cyber-Physical Systems Research Group.
REFERENCES


[13] Jürgen Hohlsch and Michael Grossniklaus. 2016. An Algebra and Equivalences to Transform Graph Patterns in Neo4j. In GraphQ at EDBT/ICDT.


[27] Ian Robinson, Jim Webber, and Emil Eifrém. 2015. Graph Databases (2nd ed.). O’Reilly Media.


