Walk Logic as a framework for path query languages on graph databases

Jelle Hellings, Bart Kuijpers
Jan Van den Bussche, and Xiaowang Zhang
Hasselt University and transnational University of Limburg

19 March 2013
Overview

Introduction

Walk Logic

Expressive Power

Regular Walk Logic

Open Problems and Conclusion
Motivation

- Expressing graph-queries
- Properties of paths, walks, ...

Route planning

We want to travel from our office to a cafetaria and from this cafetaria get back to the office using a different route
General logics

- First-order logic: limited to local reasoning
- Monadic second-order logic:
  - Focus on sets: *bipartite graph*
    
    \[
    \exists S \exists T (\forall x (x \in S \iff x \not\in T) \land \forall y \text{ edge}(x, y) \implies ((x \in S \land y \in T) \lor (y \in T \land x \in S)))
    \]
  - Paths non-straightforward: *y is reachable from x*
    
    \[
    \forall S [(x \in S) \land \forall u \forall v (u \in S \land \text{edge}(u, v) \implies v \in S) \implies y \in S]
    \]
  - *Nodes versus nodes and edges*
Specific logics

- Family of Conjunctive Regular Path Queries (CRPQs)
  - Focus on labelling of paths ('regular expression')
    \[ Q(a, b) := a\pi b, (\alpha\beta + \gamma\delta)^*(\pi) \]
  - Limited reasoning between paths ('equal length')
    \[ Q(\pi_1, \pi_2) := a\pi_1 b \land a\pi_2 b, [\frac{\alpha}{\beta}]^*(\frac{\pi_1}{\pi_2}) \]
- Family of verification logics (CTL* and hybrid extensions)
  - Focus on behaviour single/independent paths
    \[ AF(produce \lor break \lor no-resources) \]
Idea: extend first-order logic

- Add walks
- Add positions on walks
- Necessary operators to compare positions

Route planning

We want to travel from our office to a cafeteria \((W)\) and from this cafeteria get back to the office using a different route \((W')\)

\[
\exists W \exists W' \exists t_1^W \exists t_2^W \exists u_1^W \exists u_2^W \exists u_3^W' \\
(\text{office}(t_1) \land t_1 < t_2 \land \text{cafeteria}(t_2) \land u_1 < u_3 < u_2 \land u_1 \sim t_2 \land u_2 \sim t_1 \land \forall t_3^W (t_1 < t_3 < t_2 \implies t_3 \not\sim u_3))
\]
Overview

Introduction

Walk Logic

Expressive Power

Regular Walk Logic

Open Problems and Conclusion
Definitions

Definition (Directed node-labeled graph)

A directed node-labeled graph is a triple $G = (N, E, l)$:

- $N$ is a finite set of nodes
- $E \subseteq N \times N$ is the set of edges
- $l : N \rightarrow 2^{\mathcal{AP}}$ is a node-label function

Definition (Walk)

A walk in $G$ is a finite nonempty sequence $v_1 \ldots v_n$ of nodes such that $(v_i, v_{i+1}) \in E$ for each $1 \leq i < n$

Definition (Path)

A path in $G$ is a walk without node repetition
Walk Logic

- Quantification over walks and positions on walks
- Atomic formulae: properties on positions
  \[ a(t) \quad \text{Node referred to by position variable } t \text{ has labelling } a \]
  \[ t_1 \sim t_2 \quad \text{Position variables } t_1, t_2 \text{ refer to the same node} \]
  \[ t_1 < t_2 \quad \text{Position variable } t_1 \text{ comes before } t_2 \text{ in walk } W \]
  Position variables \( t_1 \) and \( t_2 \) must be of the same sort
- Logical connectives
- Optionally: syntactic sugar (quantification over nodes, \( = \), \ldots)
Path logic: Walk Logic with *path*-semantics

- Paths are useful themselves (*Hamiltonian path*):

\[ \exists P \forall Q \forall t^Q \exists u^P (t \sim u) \]

- Walk logic can express walk *P* is a path:

\[ \text{isPath}(P) \equiv \forall t^P \forall u^P (t^P \sim u^P) \implies (t^P = u^P) \]

- Set of edges can describe a path
  
  *MSO over nodes and edges subsumes Path Logic*

- Can we also express Walk Logic in Path Logic or MSO?
Overview

Introduction

Walk Logic

Expressive Power

Regular Walk Logic

Open Problems and Conclusion
## Walk-based Graph Properties - 1

<table>
<thead>
<tr>
<th>Property</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strongly Connected</strong></td>
<td>$$\forall P \forall Q \forall t^P \forall u^Q \exists R \exists v^R \exists w^R (v &lt; w \land t \sim v \land u \sim w)$$</td>
</tr>
<tr>
<td><strong>Hamiltonian Path (in Path Logic)</strong></td>
<td>$$\exists P \forall Q \forall t^Q \exists u^P (t \sim u)$$</td>
</tr>
<tr>
<td><strong>Eulerian Trail</strong></td>
<td>$$\exists W (W \text{ is a trail} \land \text{every edge is part of } W)$$</td>
</tr>
</tbody>
</table>
Walk-based Graph Properties - 2

Theorem

Weakly Connected is not expressible on directed graphs

Proof.

\[ n_1 \leftarrow n_2 \rightarrow n_3 \leftarrow n_4 \rightarrow n_5 \leftarrow n_6 \]

All walks contain at most 2 nodes: reduce to first-order logic

- Direction matters!
- On undirected graphs:
  - Weakly Connected same way as strongly connected
  - Planar Graph using Kuratowski’s Theorem
Set-based Graph Properties

Theorem

*Bipartite graph is not expressible on directed graphs*

Lemma (Dénes Kőnig)

*A graph is bipartite iff it does not contain an odd cycle*

Proof.

\[ n_2 \rightarrow n_3 \quad \downarrow \quad n_1 \]
\[ m_2 \rightarrow m_3 \rightarrow m_4 \quad \downarrow \quad m_6 \leftarrow m_5 \]

All walks contain at most 3 nodes: *reduce to first-order logic*

- MSO *can* express bipartite graph
- Is Walk Logic strictly subsumed by MSO?
Open questions

- Can we express Walk Logic in Path Logic?
- Can we express Walk Logic in MSO?
- Is Walk Logic strictly subsumed by MSO?
Eulerian Trail

Theorem

\[ \text{MSO}(\text{nodes, edges}) \text{ and Path Logic cannot express Eulerian Trail} \]

Lemma (well known result)

\[ \text{MSO cannot distinguish sets with } i \text{ from sets with } j \text{ elements} \]

Proof.

For MSO: existence of Eulerian Trail in the graph

\[
\begin{align*}
a_n & \leftrightarrow v_2 & b_m \\
\vdots & \leftrightarrow v_1 & \vdots \\
a_1 & \leftrightarrow b_1 & a_n & b_m \\
\end{align*}
\]

\[
\begin{align*}
\vdots & \leftrightarrow \vdots \\
a_1 & \leftrightarrow b_1 & a_1 & b_1 \\
\end{align*}
\]

Reduces to sets \( A \) and \( B \) having the equal number of elements
Relations with FO and MSO

Lemma (Courcelle and Engelfriet)

**MSO(nodes) cannot express Hamiltonian Path**

- FO and Path Logic are strictly subsumed by Walk Logic
- MSO(nodes) incomparable with Path Logic and Walk Logic
- MSO(nodes, edges) strictly subsumes Path Logic
- MSO(nodes, edges) incomparable with Walk Logic
Overview

Introduction

Walk Logic

Expressive Power

Regular Walk Logic

Open Problems and Conclusion
Regular walk logic

- Conjunctive Regular Path Queries (CRPQs)
  *Regular expressions over single walk*

- Extended Conjunctive Regular Path Queries (ECRPQs)
  *Regular expressions over n-tuples of walks*

- (Extended) Regular Walk Logic ((E)RWL)\(^1\):
  *Generalize (E)CRPQs by adding Boolean connectives*

\[
\exists \pi_1 \exists \pi_2 \exists v_1 \exists v_2 (v_1 \pi_1 v_2 \land v_1 \pi_2 v_2 \land [\alpha/\beta]^* (\pi_1/\pi_2))
\]

- Purpose: study open problems for (E)CRPQs

\(^1\)In the literature this variant is also called ECRPQ\(^\sim\)
ECRPQs with *path*-semantics

- Standard (E)CRPQs work with *walk* semantics
- Efficient query evaluations
- Under *path* semantics:
  \[\textit{No efficient query evaluation algorithm is known}\]
- SPARQL 1.1: property paths had path-based semantic
- Regular Path Logic (RPL) is RWL with path-based semantic
**Theorem**

*ERWL cannot express Hamiltonian Path*

**Definition (\(\overline{K}_n \times C_m\)-graphs)**

- \(a_1, \ldots, a_n, b_1, \ldots, b_m\)
- \(n\) point-nodes, \(m\) nodes on an undirected cycle
- Undirected edges between every point-node and cycle-node

**Lemma**

\[\forall \text{ length } l > 2 \text{ and nodes } v_1, v_2: \text{ there is a walk } v_1 \pi v_2 \text{ of length } l\]
Theorem (repeated)

ERWL cannot express Hamiltonian Path

Lemma (repeated)

\( \forall \ length \ l > 2 \ and \ nodes \ v_1, v_2: \ there \ is \ a \ walk \ v_1 \pi v_2 \ of \ length \ l \)

Corollary

Using a unary alphabet for the labelling:

- Regular expressions reduce to reachability in \( \overline{K}_n \times C_m \)-graphs
- ERWL on \( \overline{K}_n \times C_m \)-graphs reduces to FO-logic

Proof (de Rougemont).

FO logic on \( \overline{K}_n \times C_m \) graphs cannot express Hamiltonian Path.
Theorem

ERPL is not subsumed by ERWL

Proof.

- ERWL cannot distinguish $\overline{K_n} \times C_m$- from $\overline{K_{n'}} \times C_{m'}$-graphs
- ERPL can express ‘Longest path has even length’

$$\exists \pi_1((\alpha \alpha)^* \pi_1 \land \neg \exists \pi_2 \left[ \frac{\alpha}{\alpha} \right]^* \left[ \frac{\bot}{\alpha} \right]^+ (\pi_1, \pi_2))$$
Additional results

- Eulerian Path not expressible in RWL or RPL
- CRPQ and star-free ECRPQ are incomparable with WL
- Path-based CRPQ is not subsumed by ECRPQ
Overview

Introduction

Walk Logic

Expressive Power

Regular Walk Logic

Open Problems and Conclusion
Open Problems

» Relations with verification logic:
  » Infinite walks are the standard in verification logics
  » Can we express the verification logics in Walk Logic?
  » Walk Logic with infinite walks?

» Complexity bounds on model checking for WL:
  » WL model checking is decidable
  » Current approach has horrible complexity
Conclusion

- General walk-based reasoning on graphs
- Relates to practical graph languages
- Framework for studying expressivity