Homomorphisms to model navigational queries of graph databases

Florent Foucaud
(LIFO, Université d'Orléans, France)

joint work with:
Laurent Beaudou, Florent Madelaine, Lhouari Nourine, Gaétan Richard

April 2019

Database (DB) : storage system + organisation of numerical data.

Relational database (RDB) : model defined by Codd, 1970.
Set of tables with keys identifying each data tuple.

PubID	Publisher	PubAddress
03-4472822	Random House	123 4th Street, New York
04-7733903	Wiley and Sons	45 Lincoln Blvd, Chicago
03-4859223	O'Reilly Press	77 Boston Ave, Cambridge
03-3920886	City Lights Books	99 Market, San Francisco

AuthorID	AuthorName	AuthorBDay
345-28-2938	Haile Selassie	14-Aug-92
392-48-9965	Joe Blow	14-Mar-15
454-22-4012	Sally Hemmings	12-Sept-70
663-59-1254	Hannah Arendt	12-Mar-06

ISBN	AuthorID	PubID	Date	Title
1-34532-482-1	345-28-2938	03-4472822	1990	Cold Fusion for Dummies
1-38482-995-1	392-48-9965	04-7733903	1985	Macrame and Straw Tying
2-35921-499-4	454-22-4012	03-4859223	1952	Fluid Dynamics of Aquaducts
1-38278-293-4	663-59-1254	03-3920886	1967	Beads, Baskets & Revolution

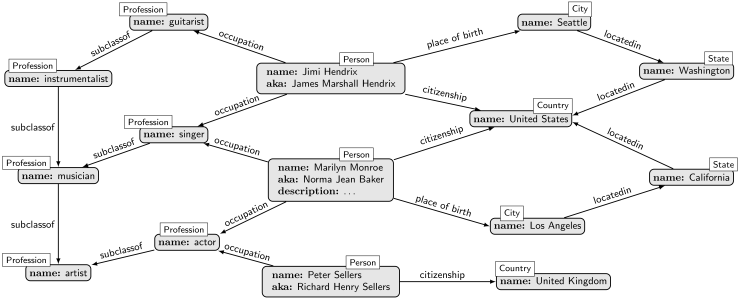
Query : interrogation of the DB to obtain some information

Typical query langage : SQL, Structured Query Language ("sequel")
Boyce-Chamberlain, IBM, 1979

ex : SELECT AuthorID FROM Books WHERE Date > 1970 ;

Graph database (GDB) : all relations are binary.

Data is stored as a graph ("data graph").



Fast-developing technology → Neo4J, OrientDB, GraphDB...

Query languages : Cypher, SPARQL, XPath...

Part of the recent trend for graph-based IT systems
→ Google (Pregel), Facebook (Graph API), Twitter (Cassovary)...

Standards are being written

Conference ACM SIGMOD'18 : industry and academics try to design future standards for "Graph Query Languages"

Industry 4: Graph databases & Query Processing on Modern Hardware SIGMOD'18, June 10-15, 2018, Houston, TX, USA

Cypher: An Evolving Query Language for Property Graphs

Nadime Francis*
Université Paris-Est

Alastair Green
Neo4j

Paolo Guagliardo
University of Edinburgh

Leonid Libkin
University of Edinburgh

Tobias Lindaaker
Neo4j

Victor Marsault
University of Edinburgh

Stefan Plantikow
Neo4j

Mats Rydberg
Neo4j

Petra Selmer
Neo4j

Andrés Taylor
Neo4j

Industry 4: Graph databases & Query Processing on Modern Hardware SIGMOD'18, June 10-15, 2018, Houston, TX, USA

G-CORE
A Core for Future Graph Query Languages

Designed by the LDBC Graph Query Language Task Force*

Renzo Angles
Universidad de Talca

Marcelo Arenas
Pontificia Universidad Católica de Chile

Pablo Barceló
DCC, Universidad de Chile

Peter Boncz
CWI, Amsterdam

George Fletcher
Technische Universiteit Eindhoven

Claudio Gutierrez
DCC, Universidad de Chile

Tobias Lindaaker
Neo4j

Marcus Paradies†
DLR

Stefan Plantikow
Neo4j

Juan Sequeda
Capsenta

Oskar van Rest
Oracle

Hannes Voigt
Technische Universität Dresden

neo4j

Cypher (2011) + Open Cypher (2015) on Property graphs

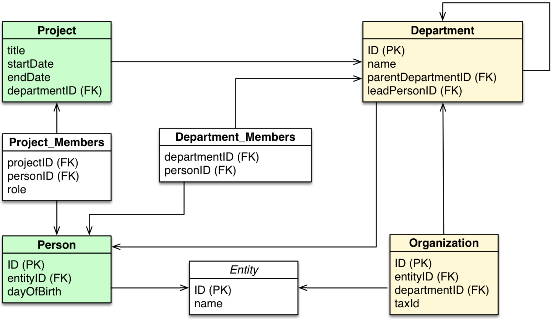
Examples of Cypher queries :

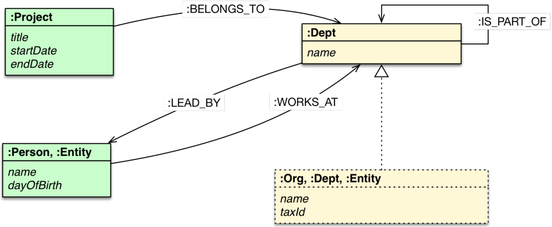
MATCH(:Person {name :'Jennifer'})-[:WORKS_FOR]->(c :Company)
RETURN c

MATCH p=(a :Person)-[*1..3]-(b :Person)
WHERE a.name='Alice' AND b.name='Bob'
RETURN p

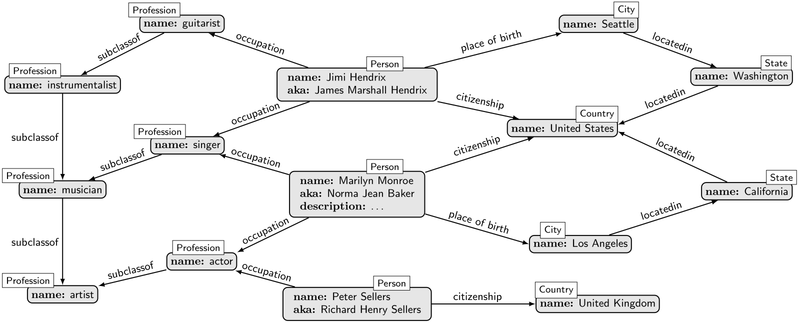
Also : shortest paths

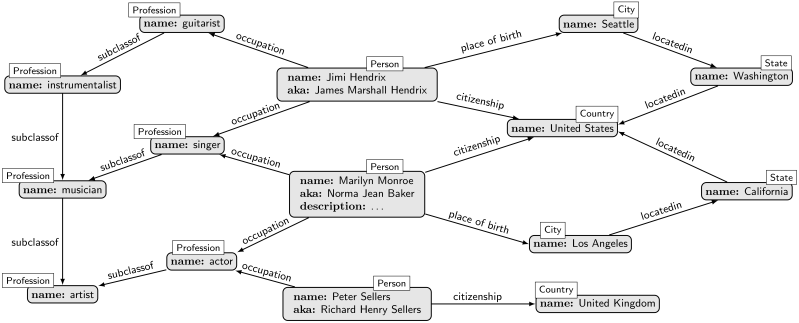
Graphs : both simple and expressive

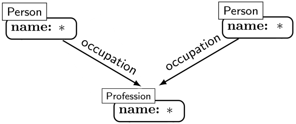


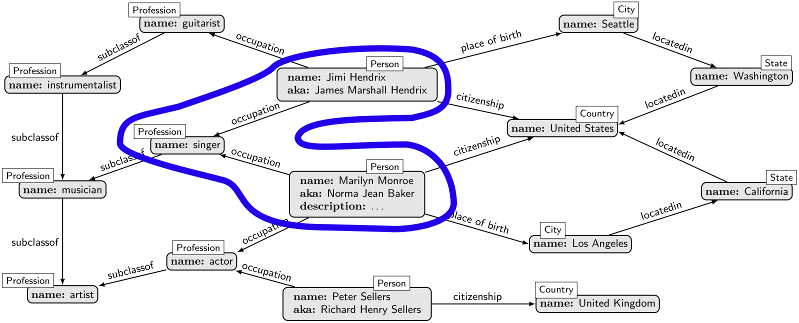


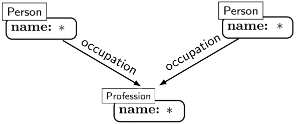
(example from Neo4J)


Profession
name: guitarist
subclassof
occupation
Profession
name: instrumentalist
Person
name: Jimi Hendrix
aka: James Marshall Hendrix
place of birth
City
name: Seattle
locatedin
State
name: Washington
citizenship
Country
name: United States
locatedin
subclassof
occupation
Profession
name: singer
subclassof
occupation
citizenship
Person
name: Marilyn Monroe
aka: Norma Jean Baker
description: . . .
locatedin
State
name: California
Profession
name: musician
place of birth
City
name: Los Angeles
locatedin
subclassof
occupation
Profession
name: actor
subclassof
occupation
Profession
name: artist
Person
name: Peter Sellers
aka: Richard Henry Sellers
citizenship
Country
name: United Kingdom

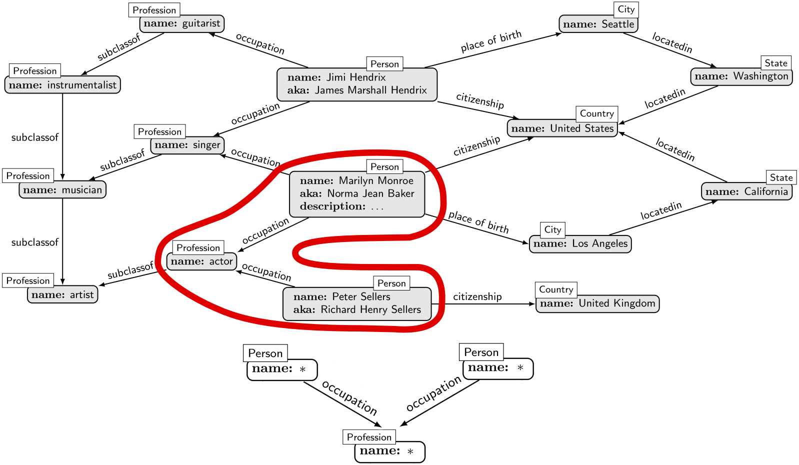

Profession
name: guitarist
subclassof
occupation
Profession
name: instrumentalist
Person
name: Jimi Hendrix
aka: James Marshall Hendrix
place of birth
City
name: Seattle
locatedin
State
name: Washington
citizenship
locatedin
Country
name: United States
subclassof
occupation
Profession
name: singer
subclassof
occupation
citizenship
locatedin
Profession
name: musician
Person
name: Marilyn Monroe
aka: Norma Jean Baker
description: . . .
State
name: California
place of birth
locatedin
City
name: Los Angeles
subclassof
occupation
Profession
name: actor
subclassof
occupation
Profession
name: artist
Person
name: Peter Sellers
aka: Richard Henry Sellers
citizenship
Country
name: United Kingdom


Person
name: *
Person
name: *
occupation
occupation
Profession
name: *

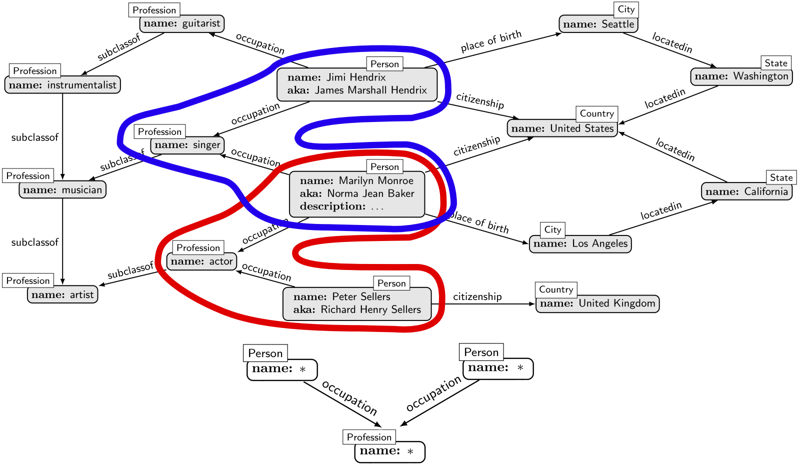

Profession
name: guitarist
subclassof
occupation
Profession
name: instrumentalist
Person
name: Jimi Hendrix
aka: James Marshall Hendrix
place of birth
City
name: Seattle
locatedin
State
name: Washington
citizenship
occupation
Profession
name: singer
subclassof
locatedin
Country
name: United States
subclassof
occupation
citizenship
locatedin
Profession
name: musician
Person
name: Marilyn Monroe
aka: Norma Jean Baker
description: . . .
State
name: California
place of birth
City
name: Los Angeles
locatedin
subclassof
Profession
name: actor
subclassof
occupation
Profession
name: artist
Person
name: Peter Sellers
aka: Richard Henry Sellers
citizenship
Country
name: United Kingdom

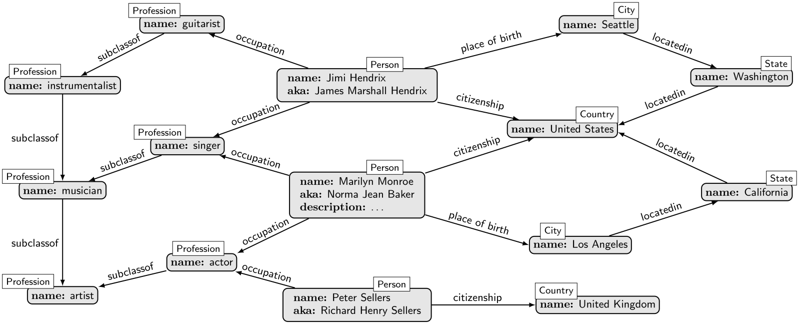

Person
name: *
occupation
Person
name: *
occupation
Profession
name: *

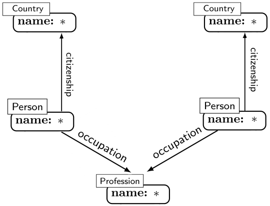
Graph databases - queries

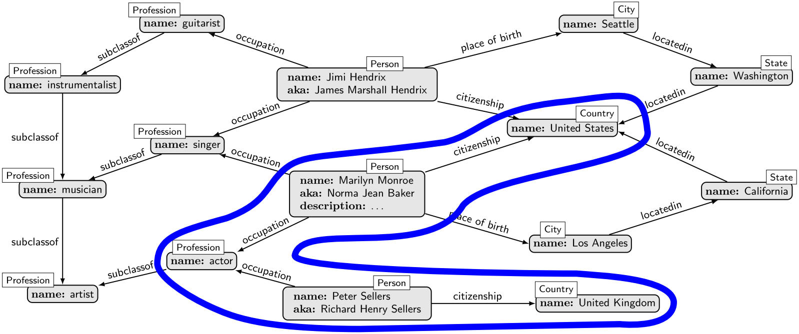


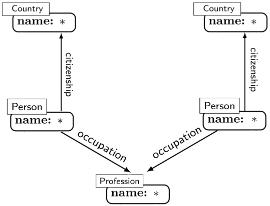
Graph databases - queries

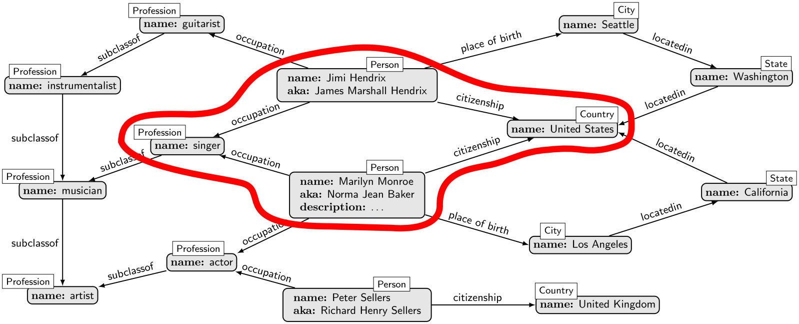


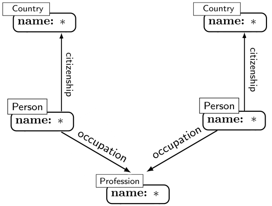

Profession
name: guitarist
subclassof
occupation
Profession
name: instrumentalist
Person
name: Jimi Hendrix
aka: James Marshall Hendrix
place of birth
City
name: Seattle
locatedin
State
name: Washington
citizenship
locatedin
Country
name: United States
subclassof
occupation
Profession
name: singer
subclassof
occupation
citizenship
locatedin
Profession
name: musician
Person
name: Marilyn Monroe
aka: Norma Jean Baker
description: . . .
State
name: California
place of birth
locatedin
City
name: Los Angeles
subclassof
occupation
Profession
name: actor
occupation
subclassof
Profession
name: artist
Person
name: Peter Sellers
aka: Richard Henry Sellers
citizenship
Country
name: United Kingdom

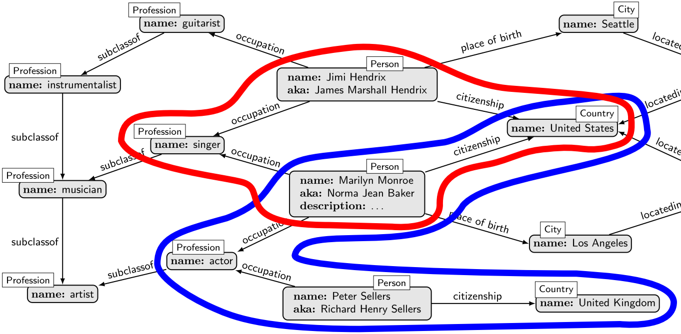

Country
name: *
citizenship
Person
name: *
occupation
Country
name: *
citizenship
Person
name: *
occupation
Profession
name: *

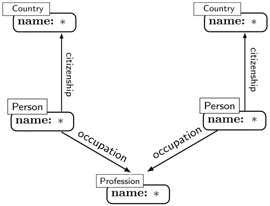

Profession
name: guitarist
subclassof
occupation
Profession
name: instrumentalist
Person
name: Jimi Hendrix
aka: James Marshall Hendrix
place of birth
City
name: Seattle
locatedin
State
name: Washington
citizenship
locatedin
Country
name: United States
subclassof
occupation
Profession
name: singer
subclassof
occupation
citizenship
locatedin
Profession
name: musician
Person
name: Marilyn Monroe
aka: Norma Jean Baker
description: ...
State
name: California
place of birth
locatedin
City
name: Los Angeles
subclassof
occupation
Profession
name: actor
subclassof
occupation
Profession
name: artist
Person
name: Peter Sellers
aka: Richard Henry Sellers
citizenship
Country
name: United Kingdom


Country
name: *
citizenship
Person
name: *
occupation
Country
name: *
citizenship
Person
name: *
occupation
Profession
name: *


Profession
name: guitarist
City
name: Seattle
subclassof
occupation
place of birth
locatedin
Profession
name: instrumentalist
Person
name: Jimi Hendrix
aka: James Marshall Hendrix
State
name: Washington
citizenship
locatedin
occupation
Country
name: United States
subclassof
Profession
name: singer
occupation
citizenship
locatedin
subclassof
Person
name: Marilyn Monroe
aka: Norma Jean Baker
description: ...
Profession
name: musician
State
name: California
place of birth
locatedin
subclassof
occupation
City
name: Los Angeles
Profession
name: actor
subclassof
occupation
Person
Profession
name: artist
name: Peter Sellers
aka: Richard Henry Sellers
citizenship
Country
name: United Kingdom


Country
name: *
Country
name: *
citizenship
citizenship
Person
name: *
Person
name: *
occupation
occupation
Profession
name: *


Profession
name: guitarist
subclassof
occupation
Profession
name: instrumentalist
Person
name: Jimi Hendrix
aka: James Marshall Hendrix
place of birth
City
name: Seattle
locatedin
State
name: Washington
citizenship
locatedin
Country
name: United States
occupation
subclassof
Profession
name: singer
occupation
citizenship
locatedin
Profession
name: musician
subclassof
Person
name: Marilyn Monroe
aka: Norma Jean Baker
description: ...
State
name: California
place of birth
locatedin
City
name: Los Angeles
subclassof
occupation
Profession
name: actor
occupation
Profession
name: artist
subclassof
Person
name: Peter Sellers
aka: Richard Henry Sellers
citizenship
Country
name: United Kingdom


Country
name: *
Country
name: *
citizenship
citizenship
Person
name: *
Person
name: *
occupation
occupation
Profession
name: *

Definition - Graph homomorphism of G to H

Mapping $h : V(G) \to V(H)$ which **preserves** adjacency, arc labels and orientation :

$$xy \in E(G) \Longrightarrow h(x)h(y) \in E(H)$$

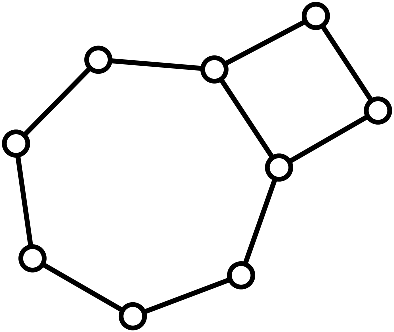
Notation : $G \to H$.

Definition - Graph homomorphism of G to H

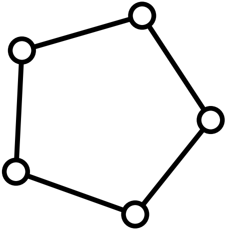
Mapping $h : V(G) \to V(H)$ which **preserves** adjacency, arc labels and orientation :

$$xy \in E(G) \Longrightarrow h(x)h(y) \in E(H)$$

Notation : $G \to H$.



Target graph : $H = C_5$

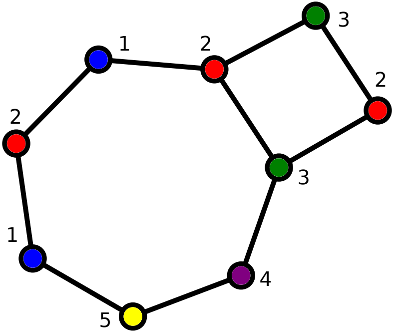


Definition - Graph homomorphism of G to H

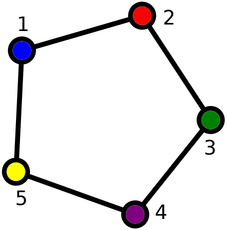
Mapping $h : V(G) \to V(H)$ which **preserves** adjacency, arc labels and orientation :

$$xy \in E(G) \Longrightarrow h(x)h(y) \in E(H)$$

Notation : $G \to H$.



Target graph : $H = C_5$

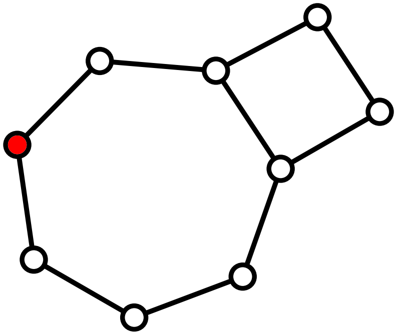


Definition - Graph homomorphism of G to H

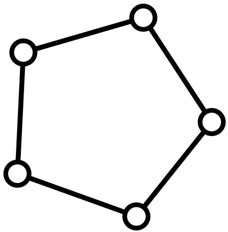
Mapping $h : V(G) \to V(H)$ which **preserves** adjacency, arc labels and orientation :

$$xy \in E(G) \Longrightarrow h(x)h(y) \in E(H)$$

Notation : $G \to H$.



Target graph : $H = C_5$

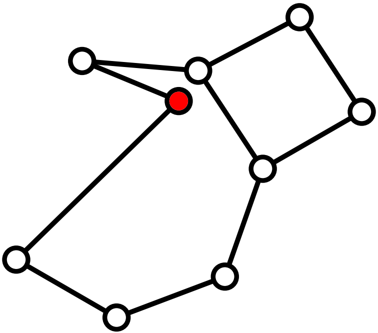


Definition - Graph homomorphism of G to H

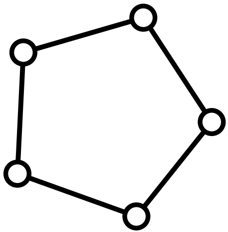
Mapping $h : V(G) \to V(H)$ which **preserves** adjacency, arc labels and orientation :

$$xy \in E(G) \Longrightarrow h(x)h(y) \in E(H)$$

Notation : $G \to H$.



Target graph : $H = C_5$

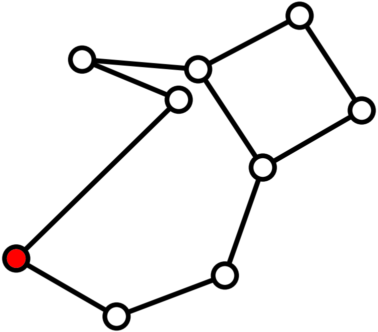


Definition - Graph homomorphism of G to H

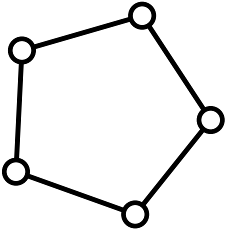
Mapping $h : V(G) \to V(H)$ which **preserves** adjacency, arc labels and orientation :

$$xy \in E(G) \Longrightarrow h(x)h(y) \in E(H)$$

Notation : $G \to H$.



Target graph : $H = C_5$

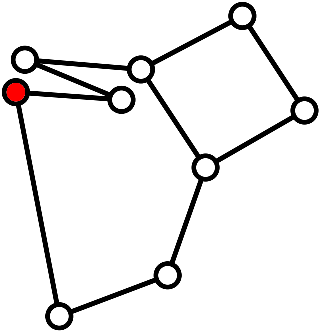


Definition - Graph homomorphism of G to H

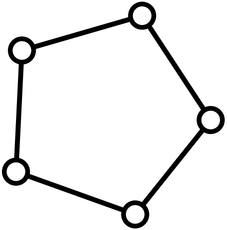
Mapping $h : V(G) \to V(H)$ which **preserves** adjacency, arc labels and orientation :

$$xy \in E(G) \Longrightarrow h(x)h(y) \in E(H)$$

Notation : $G \to H$.



Target graph : $H = C_5$

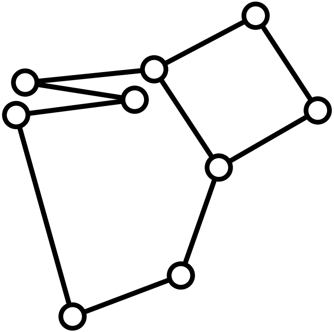


Definition - Graph homomorphism of G to H

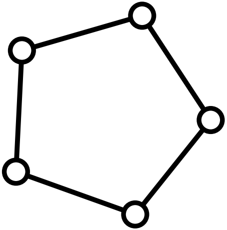
Mapping $h : V(G) \to V(H)$ which **preserves** adjacency, arc labels and orientation :

$$xy \in E(G) \Longrightarrow h(x)h(y) \in E(H)$$

Notation : $G \to H$.



Target graph : $H = C_5$

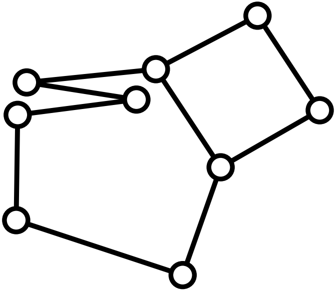


Definition - Graph homomorphism of G to H

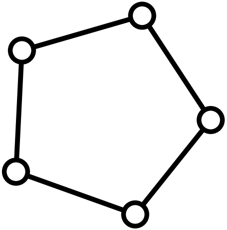
Mapping $h : V(G) \to V(H)$ which **preserves** adjacency, arc labels and orientation :

$$xy \in E(G) \Longrightarrow h(x)h(y) \in E(H)$$

Notation : $G \to H$.



Target graph : $H = C_5$

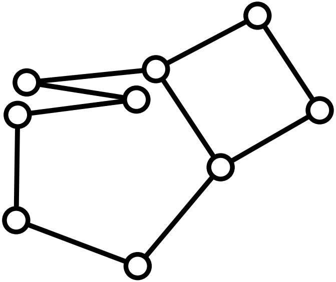


Definition - Graph homomorphism of G to H

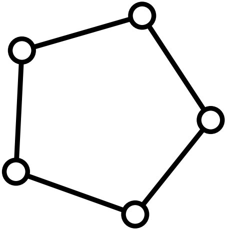
Mapping $h : V(G) \to V(H)$ which **preserves** adjacency, arc labels and orientation :

$$xy \in E(G) \Longrightarrow h(x)h(y) \in E(H)$$

Notation : $G \to H$.



Target graph : $H = C_5$

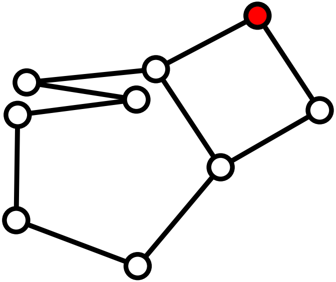


Definition - Graph homomorphism of G to H

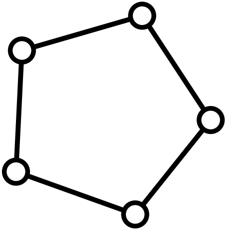
Mapping $h : V(G) \to V(H)$ which **preserves** adjacency, arc labels and orientation :

$$xy \in E(G) \Longrightarrow h(x)h(y) \in E(H)$$

Notation : $G \to H$.



Target graph : $H = C_5$

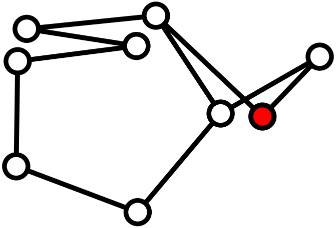


Definition - Graph homomorphism of G to H

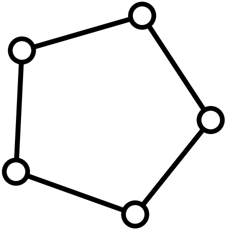
Mapping $h : V(G) \to V(H)$ which **preserves** adjacency, arc labels and orientation :

$$xy \in E(G) \Longrightarrow h(x)h(y) \in E(H)$$

Notation : $G \to H$.



Target graph : $H = C_5$

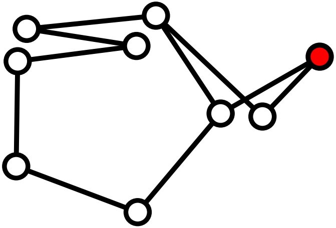


Definition - Graph homomorphism of G to H

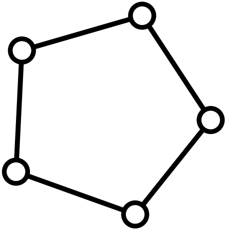
Mapping $h : V(G) \to V(H)$ which **preserves** adjacency, arc labels and orientation :

$$xy \in E(G) \Longrightarrow h(x)h(y) \in E(H)$$

Notation : $G \to H$.



Target graph : $H = C_5$

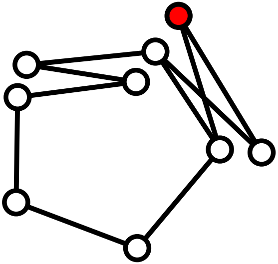


Definition - Graph homomorphism of G to H

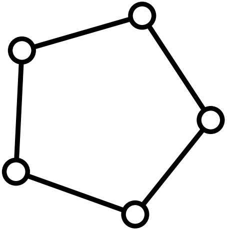
Mapping $h : V(G) \to V(H)$ which **preserves** adjacency, arc labels and orientation :

$$xy \in E(G) \Longrightarrow h(x)h(y) \in E(H)$$

Notation : $G \to H$.



Target graph : $H = C_5$

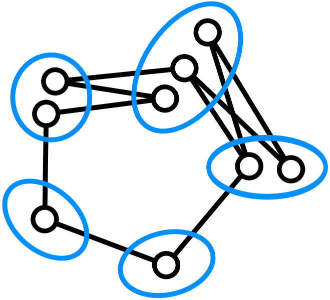


Definition - Graph homomorphism of G to H

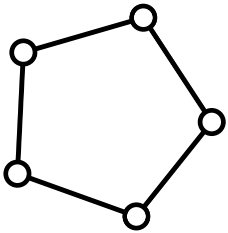
Mapping $h : V(G) \to V(H)$ which **preserves** adjacency, arc labels and orientation :

$$xy \in E(G) \Longrightarrow h(x)h(y) \in E(H)$$

Notation : $G \to H$.



Target graph : $H = C_5$



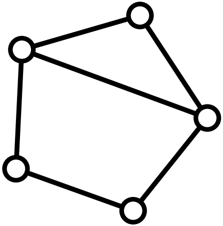
Definition - Graph homomorphism of G to H

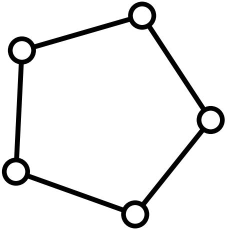
Mapping $h : V(G) \to V(H)$ which **preserves** adjacency, arc labels and orientation :

$$xy \in E(G) \Longrightarrow h(x)h(y) \in E(H)$$

Notation : $G \to H$.

Target graph : $H = C_5$



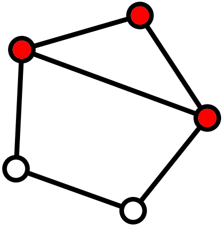


Definition - Graph homomorphism of G to H

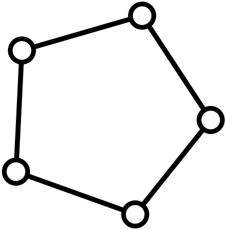
Mapping $h : V(G) \to V(H)$ which **preserves** adjacency, arc labels and orientation :

$$xy \in E(G) \Longrightarrow h(x)h(y) \in E(H)$$

Notation : $G \to H$.



Target graph : $H = C_5$

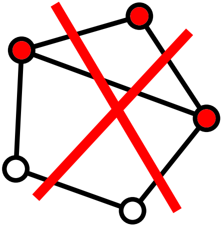


Definition - Graph homomorphism of G to H

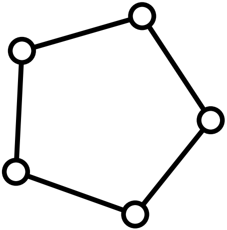
Mapping $h : V(G) \to V(H)$ which **preserves** adjacency, arc labels and orientation :

$$xy \in E(G) \Longrightarrow h(x)h(y) \in E(H)$$

Notation : $G \to H$.



Target graph : $H = C_5$

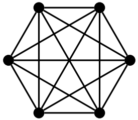


Definition - Graph homomorphism of G to H

Mapping $h : V(G) \to V(H)$ which **preserves** adjacency, arc labels and orientation :

$$xy \in E(G) \Longrightarrow h(x)h(y) \in E(H)$$

Notation : $G \to H$.



Complete graph K_6

Remark: Homomorphisms generalize proper colourings

$$G \to K_k \text{ if and only if } \chi(G) \leq k$$

Definition - Core

- **Core** of G : minimal subgraph H with $G \to H$
- G is a **core** if $core(G) = G$

Definition - Core

- **Core** of G : minimal subgraph H with $G \to H$
- G is a **core** if $core(G) = G$

Proposition

The core of a graph is unique (up to isomorphism)

Examples :
- the core of any nontrivial bipartite graph is K_2
- complete graphs and odd cycles are cores

Definition - Core

- **Core** of G : minimal subgraph H with $G \to H$
- G is a **core** if $core(G) = G$

Proposition

The core of a graph is unique (up to isomorphism)

Examples :
- the core of any nontrivial bipartite graph is K_2
- complete graphs and odd cycles are cores

Proposition

$G \to H$ if and only if $core(G) \to core(H)$

Definition - Core

- **Core** of G : minimal subgraph H with $G \to H$
- G is a **core** if $core(G) = G$

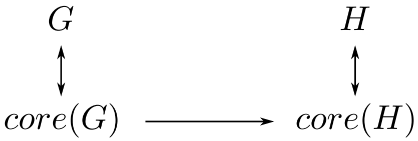
Proposition

The core of a graph is unique (up to isomorphism)

Examples : • the core of any nontrivial bipartite graph is K_2
• complete graphs and odd cycles are cores

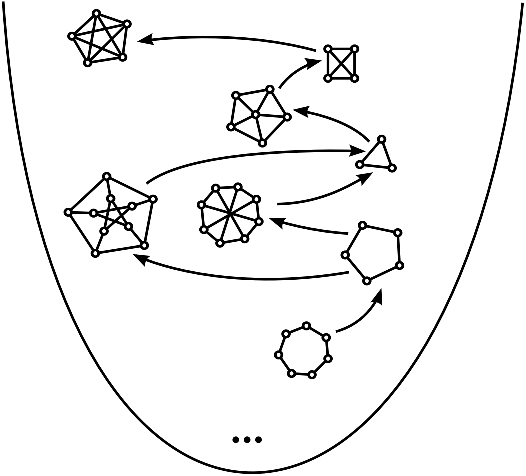
Proposition

$G \to H$ if and only if $core(G) \to core(H)$



Definition - Homomorphism quasi-order

Defined by $G \preceq H$ iff $G \to H$ (if restricted to cores : partial order).



$\preceq$ is :

- reflexive
- transitive
- antisymmetric (on cores)

The order is :

- dense
- universal
- fractal

H-COLOURING

INPUT : a graph G.
QUESTION : Does G have a homomorphism to H ?

Theorem (Hell-Nešetřil, 1990)

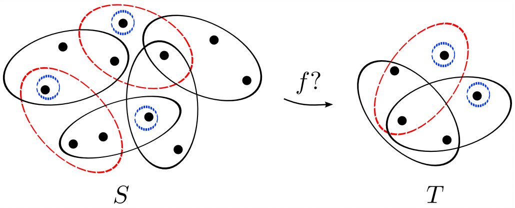
H-COLOURING is polynomial-time if H is bipartite or has a loop.
Otherwise, it is NP-complete.

Note : for directed graphs, such nice dichotomy is unknown (and probably does not exist) !

Relational structure $S = (X, V)$: domain X + relations $R_1, \ldots, R_k$ of arity $a_1, \ldots, a_k$ ($R_i \subseteq X^{a_i}$).
Ex : graphs, digraphs, k-SAT Boolean formulae...

CSP decision problem

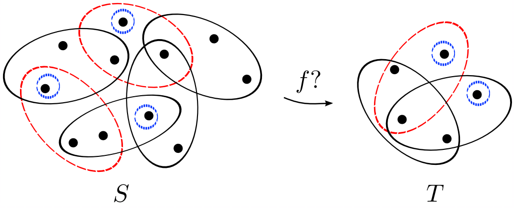
INPUT : two relational structures S et T.
QUESTION : Does S have a homomorphism to T ?



Relational structure $S=(X,V)$: domain X + relations $R_1,\dots,R_k$ of arity $a_1,\dots,a_k$ ($R_i \subseteq X^{a_i}$).
Ex : graphs, digraphs, k-SAT Boolean formulae...

(Non-uniform) CSP decision problem for a fixed template T

INPUT : a relational structure S.
QUESTION : Does S have a homomorphism to T ?



Relational structure $S = (X, V)$: domain X + relations $R_1, \ldots, R_k$ of arity $a_1, \ldots, a_k$ ($R_i \subseteq X^{a_i}$).
Ex : graphs, digraphs, k-SAT Boolean formulae...

(Non-uniform) CSP decision problem for a fixed template T

INPUT : a relational structure S.
QUESTION : Does S have a homomorphism to T ?

Conjecture (Feder-Vardi, 1998 - Dichotomy Conjecture)

For every relational structure T, T-CSP is either NP-complete or polynomial-time (i.e. not NP-intermediate).

Relational structure $S=(X,V)$: domain X + relations $R_1,\ldots,R_k$ of arity $a_1,\ldots,a_k$ ($R_i \subseteq X^{a_i}$).
Ex : graphs, digraphs, k-SAT Boolean formulae...

(Non-uniform) CSP decision problem for a fixed template T

INPUT : a relational structure S.
QUESTION : Does S have a homomorphism to T ?

Conjecture (Feder-Vardi, 1998 - Dichotomy Conjecture)

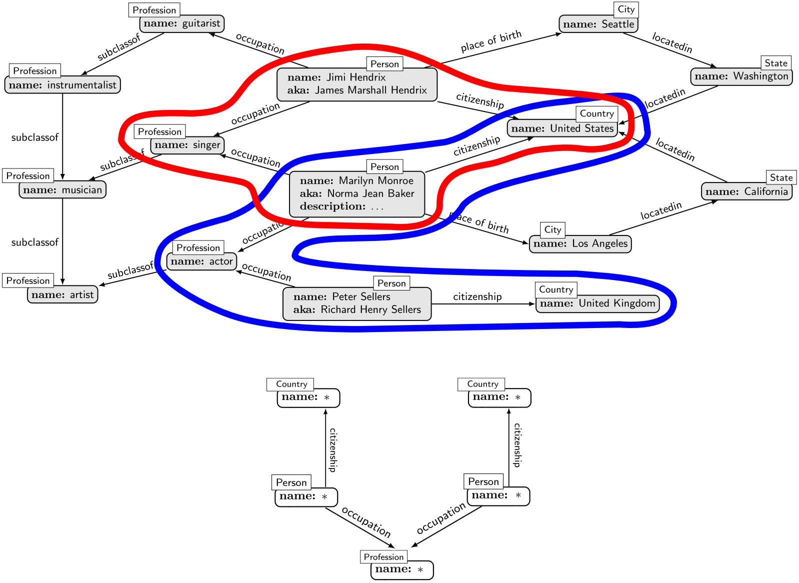
For every relational structure T, T-CSP is either NP-complete or polynomial-time (i.e. not NP-intermediate).

Siggers polymorphism : homomorphism $f : T^4 \to T$ with
$f(a,r,e,a) = f(r,a,r,e)$ for all $a,e,r \in T$.

Theorem (Bulatov 2017 + Zhuk 2017)

If T has a Siggers polymorphism, T-CSP is polynomial. Otherwise, NP-hard.

Proved after 20 years of intensive works using algebraic techniques


Profession
name: guitarist
Profession
name: instrumentalist
Profession
name: singer
Profession
name: musician
Profession
name: actor
Profession
name: artist
Person
name: Jimi Hendrix
aka: James Marshall Hendrix
Person
name: Marilyn Monroe
aka: Norma Jean Baker
description: ...
Person
name: Peter Sellers
aka: Richard Henry Sellers
City
name: Seattle
State
name: Washington
Country
name: United States
State
name: California
City
name: Los Angeles
Country
name: United Kingdom
subclassof
occupation
place of birth
locatedin
citizenship
Country
name: *
Person
name: *
Profession
name: *
citizenship
occupation

Query evaluation problem

INPUT : a database B and a query Q.
QUESTION : Does there exist a match of Q in B ?

Query evaluation problem

INPUT : a database B and a query Q.
QUESTION : Does there exist a match of Q in B ?

Enumerative query evaluation problem

INPUT : a database B and a query Q.
TASK : List all matches of Q in B.

Query minimization

INPUT : a query Q.
QUESTION : Is Q minimal (is a core) ?

Algorithmic tasks - queries

Query minimization

INPUT : a query Q.
QUESTION : Is Q minimal (is a core) ?

Q_1 and Q_2 are equivalent if for each database B, Q_1 matches B if and only if Q_2 matches B.

Query equivalence

INPUT : two queries Q_1, Q_2.
QUESTION : Are Q_1 and Q_2 equivalent ?

Query minimization

INPUT : a query Q.
QUESTION : Is Q minimal (is a core) ?

Q_1 and Q_2 are equivalent if for each database B, Q_1 matches B if and only if Q_2 matches B.

Query equivalence

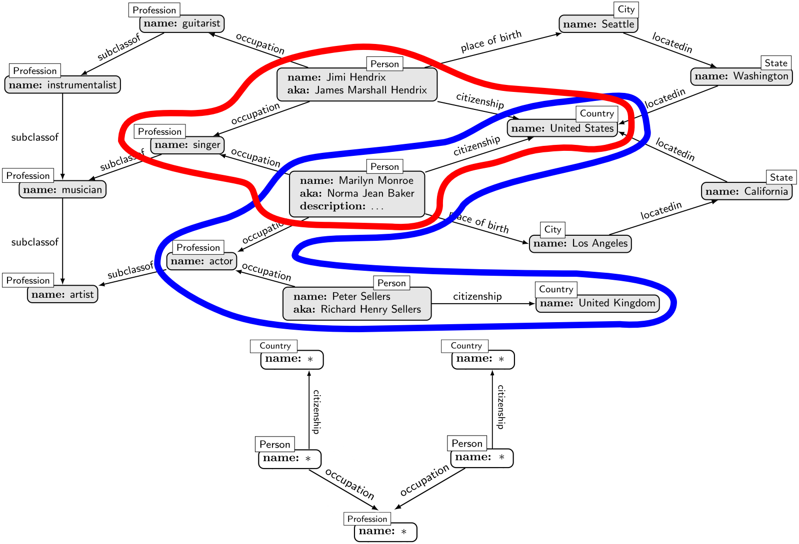
INPUT : two queries Q_1, Q_2.
QUESTION : Are Q_1 and Q_2 equivalent ?

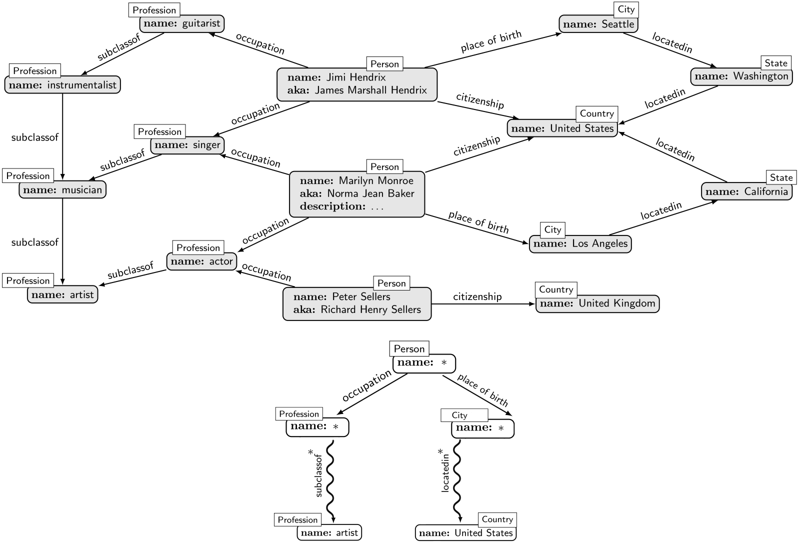
Q_1 is contained/included in Q_2 if for each database B, any match of Q_1 in B is a match of Q_2 in B.

Query containment

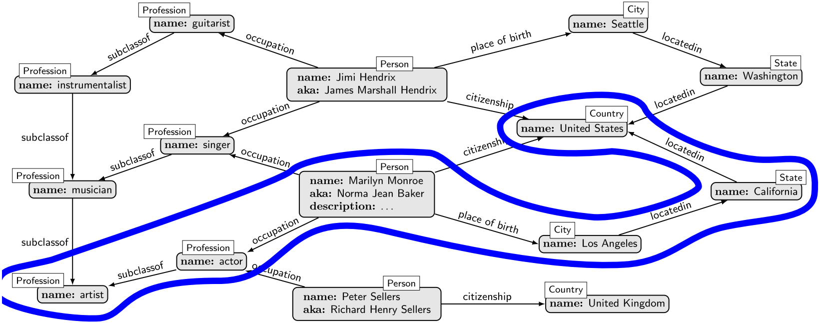
INPUT : two queries Q_1, Q_2.
QUESTION : Is Q_1 contained in Q_2 ?

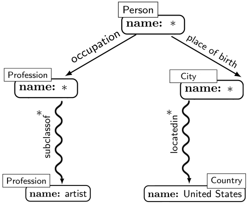
Navigational queries

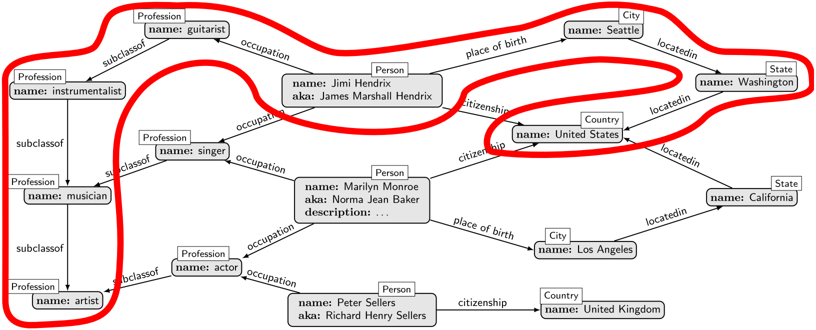


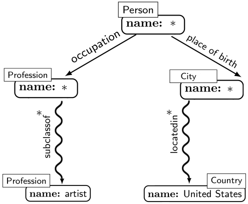

Profession
name: guitarist
City
name: Seattle
subclassof
occupation
place of birth
locatedin
Profession
name: instrumentalist
Person
name: Jimi Hendrix
aka: James Marshall Hendrix
State
name: Washington
citizenship
locatedin
occupation
Country
name: United States
subclassof
Profession
name: singer
citizenship
locatedin
subclassof
occupation
Profession
name: musician
Person
name: Marilyn Monroe
aka: Norma Jean Baker
description: ...
State
name: California
place of birth
locatedin
subclassof
occupation
City
name: Los Angeles
Profession
name: actor
subclassof
occupation
Profession
name: artist
Person
name: Peter Sellers
aka: Richard Henry Sellers
citizenship
Country
name: United Kingdom
Person
name: *
occupation
place of birth
Profession
name: *
City
name: *
subclassof*
locatedin*
Profession
name: artist
Country
name: United States

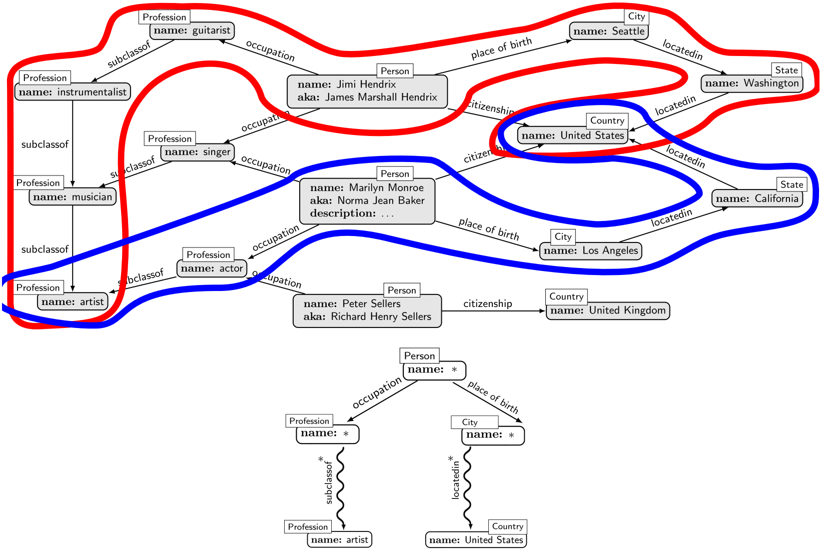
Navigational queries






Profession
name: guitarist
occupation
subclassof
Profession
name: instrumentalist
Person
name: Jimi Hendrix
aka: James Marshall Hendrix
place of birth
City
name: Seattle
locatedin
State
name: Washington
citizenship
locatedin
Country
name: United States
subclassof
occupation
Profession
name: singer
occupation
subclassof
Profession
name: musician
Person
name: Marilyn Monroe
aka: Norma Jean Baker
description: ...
citizenship
locatedin
State
name: California
subclassof
occupation
place of birth
City
name: Los Angeles
locatedin
Profession
name: actor
subclassof
Profession
name: artist
occupation
Person
name: Peter Sellers
aka: Richard Henry Sellers
citizenship
Country
name: United Kingdom


Person
name: *
occupation
place of birth
Profession
name: *
City
name: *
subclassof*
locatedin*
Profession
name: artist
Country
name: United States

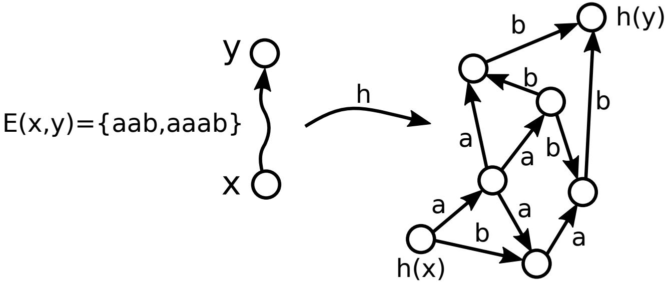

Profession
name: guitarist
Profession
name: instrumentalist
subclassof
occupation
Person
name: Jimi Hendrix
aka: James Marshall Hendrix
place of birth
City
name: Seattle
locatedin
State
name: Washington
citizenship
locatedin
Country
name: United States
Profession
name: singer
occupation
subclassof
occupation
Profession
name: musician
Person
name: Marilyn Monroe
aka: Norma Jean Baker
description: . . .
locatedin
State
name: California
place of birth
City
name: Los Angeles
locatedin
occupation
Profession
name: actor
subclassof
Profession
name: artist
occupation
Person
name: Peter Sellers
aka: Richard Henry Sellers
citizenship
Country
name: United Kingdom
Person
name: *
occupation
place of birth
Profession
name: *
City
name: *
subclassof
*
locatedin
*
Profession
name: artist
Country
name: United States

Graph database : Arc-labeled directed graph, labels are from a fixed alphabet $\Sigma = \{a, b, c, \ldots\}$

Query graph : Directed graph with label $E(x, y)$ for each arc (x, y).
Labels are sets of words over Σ.

Definition - Navigational homomorphism of Q to B

Mapping $h : V(G) \to V(H)$ that preserves labels :
$(x, y) \in A(Q) \implies$ there exists a directed walk from $h(x)$ to $h(y)$ in B whose associated word is in $E(x, y)$.

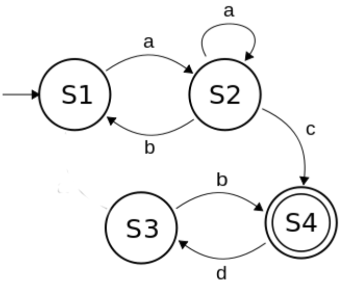


The sets of words labeling the arcs are regular languages.

Regular language : set of words over alphabet Σ that is closed under union $+$, concatenation, and Kleene-star *.

Example : $a(a^*ba)^*c(db)^*$

Correspond to word sets recognized by finite automata :



RPQ : one arc
conjunctive RPQ (CRPQ) : regular graph pattern (“graph of RPQs”)

B : a fixed database

CRPQ evaluation for B

INPUT : a regular graph pattern Q.
QUESTION : Does there exist a match (n-homomorphism) of Q in B ?

→ Barceló-Romero-Vardi (LICS'17) : NP-complete/polynomial dichotomy using the CSP dichotomy theorem (Bulatov'17 + Zhuk'17)

How to model navigational query containment using homomorphisms ?

Query graphs : Directed graphs with label $E(x,y)$ for each arc (x,y).
Labels are sets of words over Σ. Example : $\{a, aa, ab, ba, bcab\}$

Definition - Navigational homomorphism between two CRPQs Q_1 and Q_2

Mapping $h : V(Q_1) \to V(Q_2)$ such that :
$(x,y) \in A(Q_1) \implies$ there exists a directed walk from $h(x)$ to $h(y)$ in Q_2 whose associated concatenation of sets of words is in $E(x,y)$.

CRPQ containment

Homomorphism-based CRPQ containment

INPUT : two CRPQs Q_1 and Q_2.
QUESTION : Does there exist a homomorphism of Q_1 to Q_2 ?

Homomorphism-based CRPQ containment

INPUT : two CRPQs Q_1 and Q_2.
QUESTION : Does there exist a homomorphism of Q_1 to Q_2 ?

Theorem (Beaudou, F., Madelaine, Nourine, Richard, 2019)

Homomorphism-based CRPQ containment is in EXPTIME, but PSPACE-hard.

hardness : trivial reduction from **Regular Language Inclusion**.
$L_{|n}$: n-truncation of L (words of L of length at most n)

Lemma

$A, B_1, \ldots, B_k$ regular languages recognized by automata with $n_A, n_1, \ldots, n_k$ states.
Then, $L(B_1) \cdot \ldots \cdot L(B_k) \subseteq L(A)$ if and only if $L(B_1)_{|n_A n_1} \cdot \ldots \cdot L(B_k)_{|n_A n_k} \subseteq L(A)$.

Theorem (Beaudou, F., Madelaine, Nourine, Richard, 2019)

Homomorphism-based CRPQ containment is in EXPTIME, but PSPACE-hard.

Remark : General CRPQ containment is EXPSPACE-complete (Florescu et al. PODS'98)

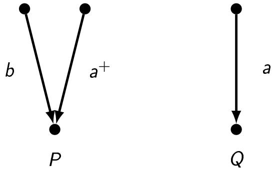


Fig : $Q \not\to P$ but Q contained in P

CRPQ containment

Special case of unary alphabet : $\Sigma = \{a\}$ and walks of type "a" or "a^+" ($a^+ = \{a, aa, aaa, aaaa, \ldots\}$)

Motivation : XPath (XML Path Language), SPARQL ()

XPath operators "" (child node) and "" (descendants or self)

objectsprice : returns all prices that are below the "items"

CRPQ containment

Special case of unary alphabet : $\Sigma = \{a\}$ and walks of type "a" or "a^+" ($a^+ = \{a, aa, aaa, aaaa, \dots\}$)

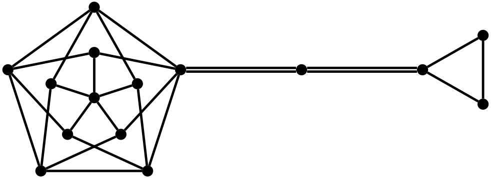
Proposition

Homomorphism-based $\{a, a^+\}$**-CRPQ containment** has a polynomial/NP-complete dichotomy.

Special case of unary alphabet : $\Sigma = \{a\}$ and walks of type "a" or "a^+"
($a^+ = \{a, aa, aaa, aaaa, \ldots\}$)

Theorem (Beaudou, F., Madelaine, Nourine, Richard, 2019)

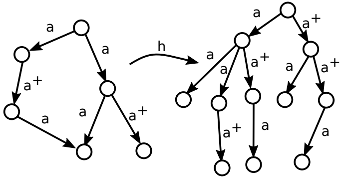
Homomorphism-based $\{a, a^+\}$**-CRPQ containment** for undirected CRPQs is polynomial-time if the core of the target has at most one edge, NP-complete otherwise.



Special case of unary alphabet : $\Sigma = \{a\}$ and walks of type "a" or "a^+" ($a^+ = \{a, aa, aaa, aaaa, \ldots\}$)

Theorem (Beaudou, F., Madelaine, Nourine, Richard, 2019)

Homomorphism-based $\{a, a^+\}$**-CRPQ containment** is polynomial-time if the target is a directed path.



Special case of unary alphabet : $\Sigma = \{a\}$ and walks of type "a" or "a^+" ($a^+ = \{a, aa, aaa, aaaa, \dots\}$)

Theorem (Beaudou, F., Madelaine, Nourine, Richard, 2019)

Homomorphism-based $\{a, a^+\}$**-CRPQ containment** is polynomial-time if the target is a directed path.

Proof. If only a's : parallel scheduling with relative deadlines.

Generally : majority (median) polymorphism.

ternary **majority polymorphism** of T : homomorphism $h : T^3 \to T$ with $h(x,x,y) = h(x,y,x) = h(y,x,x) = x$

Theorem (Feder-Vardi)

If T has a ternary majority polymorphism, then **CSP(T)** can be solved in cubic time by path-consistency.

Interface of rich research areas :
Databases - Graph theory/algorithms - CSP - Language/automata theory

Some selected problems :

• Is **Homomorphism-based CRPQ containment** EXPTIME-hard or in PSPACE ?

• Is **Homomorphism-based** $\{a, a^+\}$**-CRPQ containment** NP-complete on rooted directed trees ?

• Many special cases are interesting !
Examples : $\{a, a^+\}$; $\{a, a^*\}$; $\{a^k, k \in \mathbb{N}\}$; Unary regular languages in general...