Relative Expressiveness Within The Calculus of Relations

Jan Van den Bussche
Hasselt University, Belgium

joint work with George Fletcher, Marc Gyssens, Dirk Leinders, Dirk Van Gucht, Stijn Vansummeren, Yuqing Wu

Importance of Graph Databases

Semistructured data, dataspaces,
personal information management

Linked Data, RDF, Semantic Web

Network Data (social, biological, . . .)

GIS Data

Query Languages for Graphs

Graph patterns (conjunctive queries)

First-order logic (FO)

Transitive-closure logic FO(TC)

(Extended) Regular path queries
[Abiteboul&Vianu, Libkin et al.]

Monadic second-order logic [Courcelle]

Navigational Languages

Program logic, dynamic logic

Trees: XPath

Tarski's Calculus of Relations [1941, 1980s]

The Calculus of Relations

A set of operations on binary relations (graphs) over some domain V

- union $\cup$, intersection $\cap$, set difference $-$

- composition

$$r \circ s = \{(x, z) \mid \exists y : (x, y) \in r \ \& \ (y, z) \in s\}$$

- converse

$$r^{-1} = \{(y, x) \mid (x, y) \in r\}$$

- identity

$$\mathsf{id} = \{(x, x) \mid x \in V\}$$

Additional Operations

- diversity di $= \{(x,y) \in V^2 \mid x \neq y\}$

  Allows all $=$ id $\cup$ di and complementation $r^c =$ all $- r$

- Projection

$$\pi_1(r) = \{(x,x) \mid \exists y : (x,y) \in r\}$$
$$\pi_2(r) = \{(y,y) \mid \exists x : (x,y) \in r\}$$

- Coprojection $(i = 1, 2)$

$$\overline{\pi}_1(r) = \{(x,x) \mid x \in V \ \& \ \neg\exists y : (x,y) \in r\}$$
$$\overline{\pi}_2(r) = \{(y,y) \mid y \in V \ \& \ \neg\exists x : (x,y) \in r\}$$

- Transitive closure r^+

Expressions

Fix a binary relational vocabulary Γ

Structures over Γ = edge-labeled graphs

For a set $\mathcal{F}$ of operations, $\mathcal{F}$-expressions are built up from relation names in Γ using the operations in $\mathcal{F}$

E.g. $(R \circ (\mathrm{id} \cup \mathrm{di})) \cap \mathrm{id}$

$\equiv \pi_1(R)$

E.g. $(R^c \circ S^{-1})^c$

$\equiv \{(x, y) \mid \neg\exists z : \neg R(x, z) \wedge S(y, z)\}$

Queries

Binary queries: result is a binary relation

Boolean queries (graph properties): test nonemptiness of result

E.g. $(R \circ R) - R \neq \emptyset \quad \Leftrightarrow \quad$ graph is not transitive

Binary queries expressible in the calculus of relations (without transitive closure) = binary queries expressible in FO^3

Relative expressiveness

Compare different fragments $\mathcal{F}$

- $\cup$, $\circ$, id always present
- add other operations to taste

$\mathcal{F}_1 \preceq \mathcal{F}_2$ if every binary query expressible by an $\mathcal{F}_1$-expression is also expressible by an $\mathcal{F}_2$-expression

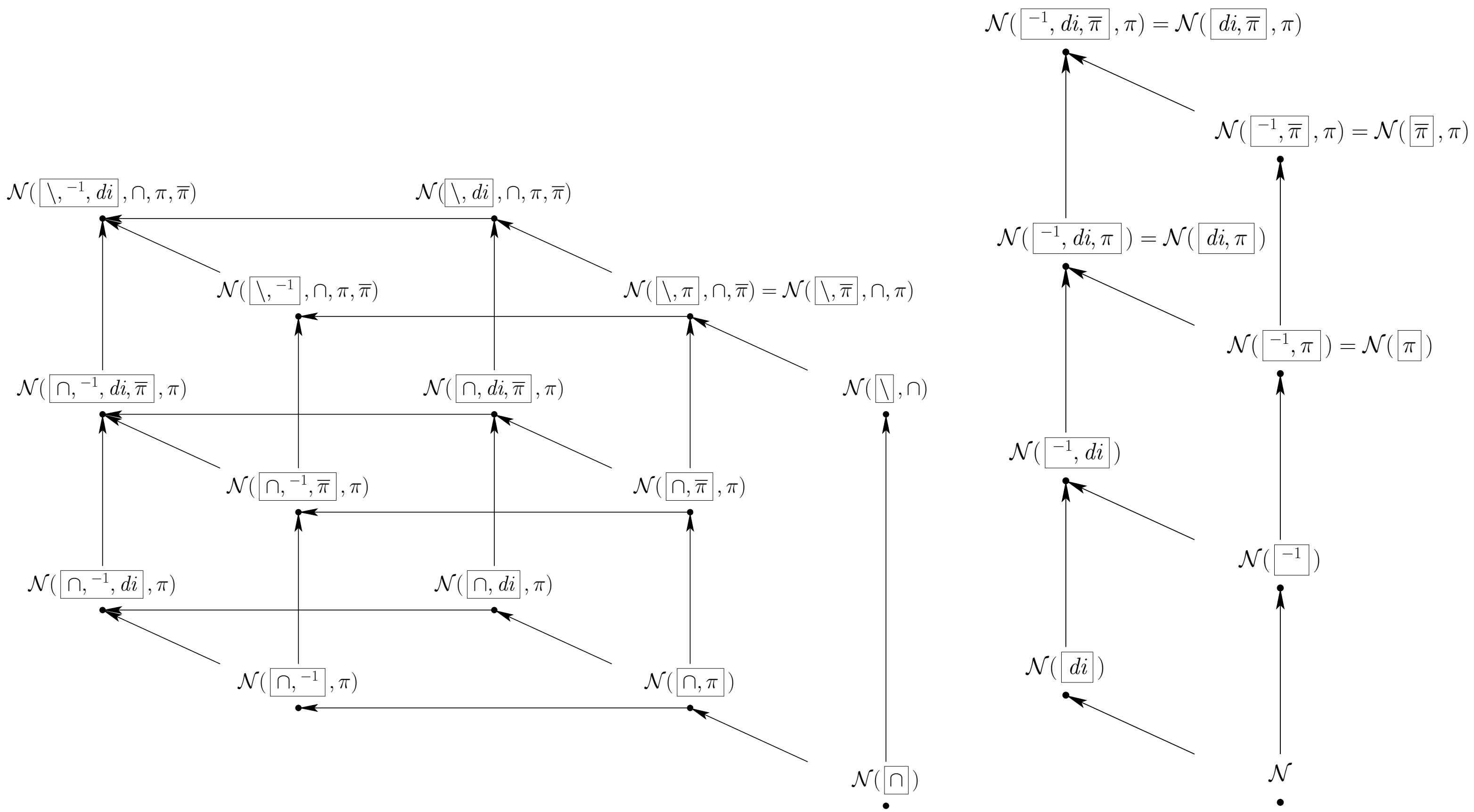
E.g. $(\pi) \preceq (\cap, \mathrm{di})$

$\boxed{\mathcal{F}_1 \preceq^{\mathrm{bool}} \mathcal{F}_2}$ if for every $\mathcal{F}_1$-expression e_1 there is an $\mathcal{F}_2$-expression e_2 such that for all graphs G:

$$e_1(G) \neq \emptyset \quad \Leftrightarrow \quad e_2(G) \neq \emptyset$$

E.g. $(^{-1}) \preceq^{\mathrm{bool}} (\pi)$ but $(^{+}, {}^{-1}) \not\preceq^{\mathrm{bool}} (^{+}, \pi)$

We have a complete picture of $\preceq^{\text{bool}}$



Projection

If fragment $\mathcal{F}$ contains at least one of

- projection
- coprojection
- converse and (intersection or set difference)
- diversity and (intersection or set difference)

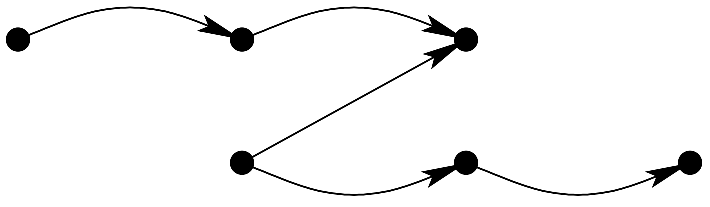
then projection is already expressible in $\mathcal{F}$.

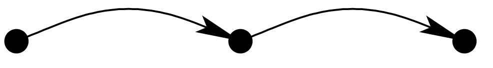
Moreover, for any two fragments $\mathcal{F}_1$ and $\mathcal{F}_2$ *not* like that, we have $(\mathcal{F}_1, \pi) \not\leq^{\text{bool}} \mathcal{F}_2$.

Projection proof:

case $\mathcal{F}_2$ has intersection or set difference*

Following pattern match is not expressible in $\mathcal{F}_2$:



Indistinguishable from 

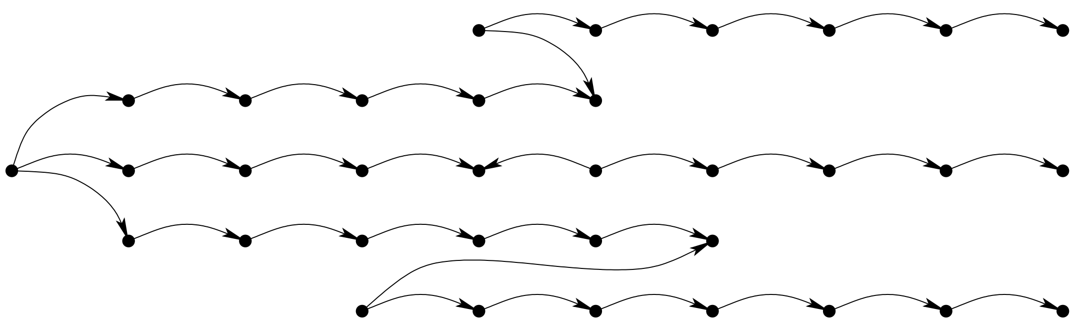
Pattern match expressible as $\pi_1(R^2) \circ R \circ \pi_2(R^2) \neq \emptyset$

*But does not have converse or diversity

Projection proof:

case $\mathcal{F}_2$ does not have intersection or set difference*

Following pattern match is not expressible in $\mathcal{F}_2$:



Expressible as

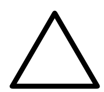
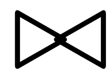
$$\pi_1(R^6 \circ \pi_2(\pi_1(R^6) \circ R)) \circ \pi_1(R^5 \circ \pi_2(\pi_1(R^5) \circ R)) \circ \pi_1(R^4 \circ \pi_2(\pi_1(R^4) \circ R)) \neq \emptyset$$

*But may have converse, diversity, transitive closure

Set difference

If $\mathcal{F}_1$ and $\mathcal{F}_2$ do not have set difference, then $(\mathcal{F}_1, -) \not\preceq^{\text{bool}} \mathcal{F}_2$.

This is not surprising (monotonicity), but we have here a **strong separation**: two finite graphs cannot be distinguished

 versus 

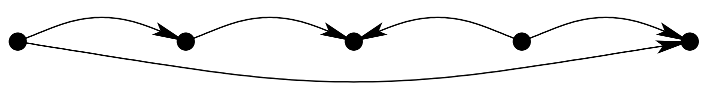
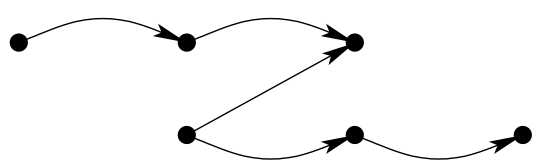
$R^2 - R \neq \emptyset$

Compare to FO where set difference can be expressed using diversity if you know the structure

Converse

For any fragment $\mathcal{F}$ that has neither intersection nor transitive closure, we have $(\mathcal{F}, {}^{-1}) \preceq^{\text{bool}} (\mathcal{F}, \pi)$.

In all other cases, $(\mathcal{F}_1, {}^{-1}) \not\preceq^{\text{bool}} \mathcal{F}_2$ where $\mathcal{F}_2$ is a fragment without converse.

- $\mathcal{F}_1$ has $\cap$: 
- $\mathcal{F}_1$ has TC: $R^2 \circ (R \circ R^{-1})^+ \circ R^2 \neq \emptyset$
- otherwise: 

Transitive closure

$S \circ R^+ \circ T \neq \emptyset$ cannot be expressed in any fragment lacking transitive closure

Do we really need two relations?

- If $\mathcal{F}$ has at most π and di (apart from the default $\circ$, $\cup$, id), then $(\mathcal{F}, {}^+) \preceq^{\text{bool}} \mathcal{F}$ over a single binary relation

- In all other cases $(\mathcal{F}_1, {}^+) \not\preceq^{\text{bool}} \mathcal{F}_2$ where $\mathcal{F}_2$ lacks transitive closure

- $R^+ \cap \text{id} \neq \emptyset$

- $R^2 \circ (R \circ R^{-1})^+ \circ R^2 \neq \emptyset$

- $\overline{\pi}_1((R^+ \circ \overline{\pi}_1(R)) \cup \overline{\pi}_1(R) \neq \emptyset$

“there is a non-sink node from which no sink node can be reached”

Conclusion

Complete understanding of relative expressiveness within fragments considered

Other operations, e.g., residuation

Other modalities for expressing boolean queries, e.g., emptiness instead of nonemptiness

Unary queries