Derivation-Graph-Based Characterizations of Decidable Existential Rule Sets

Tim S. Lyon and Sebastian Rudolph

Computational Logic Group, Institute of Artificial Intelligence,

Technische Universität Dresden











JELIA 2023

- Introduction
- Derivation Graphs
- Greedy Derivations
- Showing gbts ⊂ wgbts
- Conclusion

- Introduction
- Derivation Graphs
- Greedy Derivations
- 4 Showing gbts ⊂ wgbts
- Conclusion

Existential Rules: Preliminaries

Instance: set $\mathcal{I} = \{p_1(\vec{t_1}), \dots, p_n(\vec{t_n}), \dots\}$ of atoms

Database: finite set $\mathcal{D} = \{p_1(\vec{a}_1), \dots, p_n(\vec{a}_n)\}$ of ground atoms

Existential Rule: $\forall \vec{x} \vec{y} \beta(\vec{x}, \vec{y}) \rightarrow \exists \vec{z} \alpha(\vec{y}, \vec{z})$ \vec{y} is called "frontier"

Example: $\rho = male(x) \land human(x) \rightarrow \exists z (female(z) \land parent(z, x))$

Knowledge Base: $\mathcal{K} = (\mathcal{D}, \mathcal{R})$ with \mathcal{R} finite set of existential rules

Derivation: $\delta = \mathcal{D}, (\rho_0, h_0, \mathcal{I}_1), \dots, (\rho_{n-1}, h_{n-1}, \mathcal{I}_n)$ (forward chaining)

Notation: $\mathcal{D} \xrightarrow{\delta} \mathcal{I}_n$

Boolean Conjunctive Query (BCQ): $Q = \exists \vec{x}(p_1(\vec{t}_1) \land \cdots \land p_n(\vec{t}_n))$

Example: $Q = \exists z (female(z) \land parent(z, Joe))$

BCQ Entailment Problem: Does $(\mathcal{D}, \mathcal{R}) \models \mathcal{Q}$ hold?

This holds iff some $\mathcal I$ with $\mathcal D \stackrel{\delta}{\longrightarrow} \mathcal I$ is a model of $\mathcal Q$

Goals of This Paper

Issue: BCQ Entailment is undecidable in general!

Solution: Restrict $\mathcal{R} \leadsto \mathsf{Decidability}$

Our Goals:

- Find decidability criteria via proof-theoretic analysis
- Using and adapt the existing tool of derivation graphs
- Relate established notions to existing concept of greediness

Treewidth and Decidability

Bounded Treewidth Sets: A rule set \mathcal{R} is (semantic) **bts** iff $(\mathcal{D}, \mathcal{R})$ has a <u>universal model of finite treewidth</u> for any database \mathcal{D} . not explained, irrelevant for this talk.

Theorem (Baget et al. 2011, Feller et al. 2023)

If R is **bts**, then BCQ entailment is decidable.

Note: All rule sets we consider are **bts** → BCQ entailment is decidable

Problem: Hard to establish whether given rule set is **bts** (undecidable in general) – search for tools and sufficient criteria establishing "easier" subclasses.

- Introduction
- 2 Derivation Graphs
- Greedy Derivations
- 4 Showing gbts ⊂ wgbts
- Conclusion

Derivation Graph (Baget et al. 2011): A DAG that keeps track of how facts are derived and newly introduced terms are propagated.

Derivation Graph (Baget et al. 2011): A DAG that keeps track of how facts are derived and newly introduced terms are propagated.

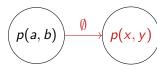
Derivation Graph (Baget et al. 2011): A DAG that keeps track of how facts are derived and newly introduced terms are propagated.

$$\mathcal{R} = \{$$

$$q(x,x) \wedge p(y,z) \rightarrow r(x,y,z).$$

Derivation Graph (Baget et al. 2011): A DAG that keeps track of how facts are derived and newly introduced terms are propagated.

Example: $\mathcal{D} = \{p(a, b)\}$ and



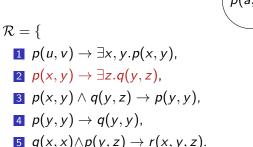
$$\mathcal{R} = \{$$

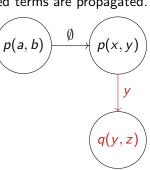
- $p(x,y) \rightarrow \exists z.q(y,z),$

- $q(x,x) \wedge p(y,z) \rightarrow r(x,y,z).$

}

Derivation Graph (Baget et al. 2011): A DAG that keeps track of how facts are derived and newly introduced terms are propagated.





Derivation Graph (Baget et al. 2011): A DAG that keeps track of how facts are derived and newly introduced terms are propagated.

Example:
$$\mathcal{D} = \{p(a, b)\}$$
 and

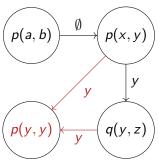
$$\mathcal{R} = \{$$

$$\mathbf{1} \ p(u, v) \to \exists x, y. p(x, y),$$

$$\mathbf{2} \ p(x, y) \to \exists z. q(y, z),$$

$$\mathbf{3} \ p(x, y) \land q(y, z) \to p(y, y),$$

$$\mathbf{4} \ p(y, y) \to q(y, y),$$



Derivation Graph (Baget et al. 2011): A DAG that keeps track of how facts are derived and newly introduced terms are propagated.

Example:
$$\mathcal{D} = \{p(a, b)\}$$
 and

$$\mathcal{R} = \{$$

$$\mathbf{1} \ p(u,v) \to \exists x, y.p(x,y),$$

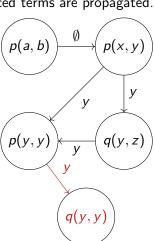
$$\mathbf{2} \ p(x,y) \to \exists z.q(y,z),$$

$$\mathbf{3} \ p(x,y) \land q(y,z) \to p(y,y),$$

$$\mathbf{4} \ p(y,y) \to q(y,y),$$

$$\mathbf{5} \ q(x,x) \land p(y,z) \to r(x,y,z).$$

$$\}$$



Derivation Graph (Baget et al. 2011): A DAG that keeps track of how facts are derived and newly introduced terms are propagated.

$$\mathcal{R} = \{$$

$$\mathbf{1} \ p(u,v) \to \exists x, y.p(x,y),$$

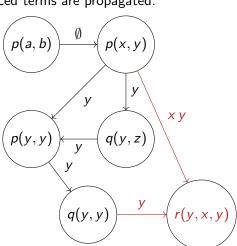
$$\mathbf{2} \ p(x,y) \to \exists z.q(y,z),$$

$$\mathbf{3} \ p(x,y) \land q(y,z) \to p(y,y),$$

$$\mathbf{4} \ p(y,y) \to q(y,y),$$

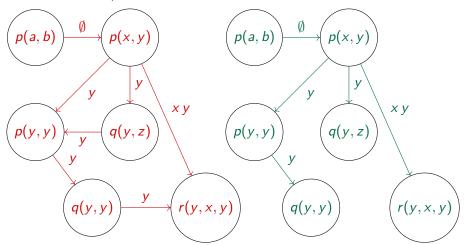
$$\mathbf{5} \ q(x,x) \land p(y,z) \to r(x,y,z).$$

$$\}$$

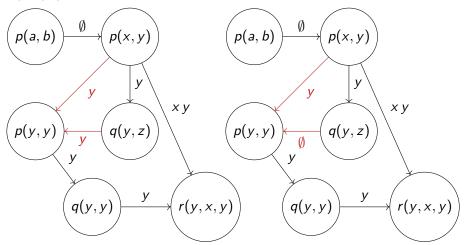


Derivation Graphs and Tree Decompositions

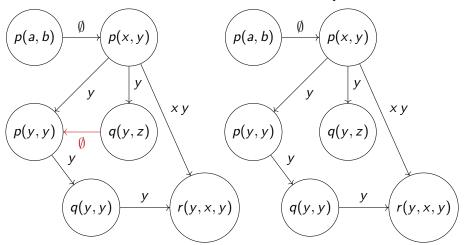
Idea (Baget et al. 2011): If any derivation graph for a rule set \mathcal{R} can be reduced to a tree, then \mathcal{R} is **bts**.



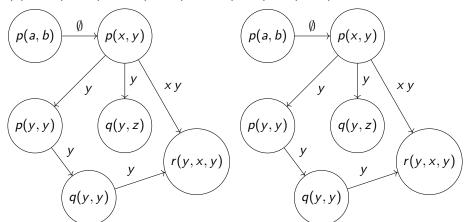
Term Removal: If a term labels two converging arcs, it may be removed from one.



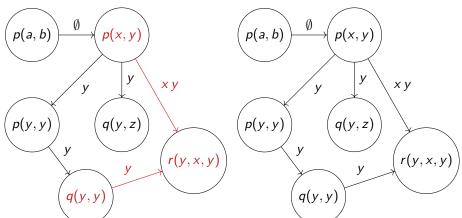
Arc Removal: If an arc is labelled with \emptyset , then it may be deleted.



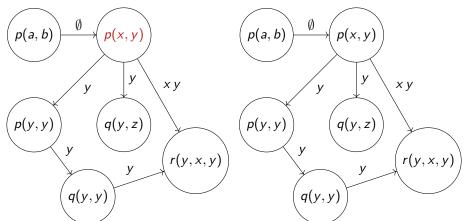
NEW!!!



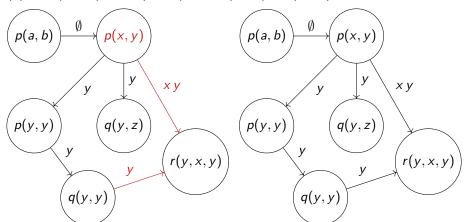
NEW!!!



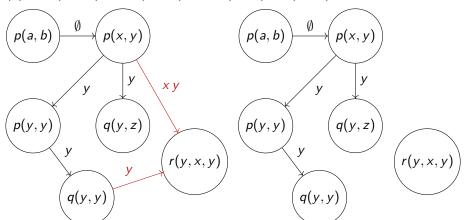
NEW!!!



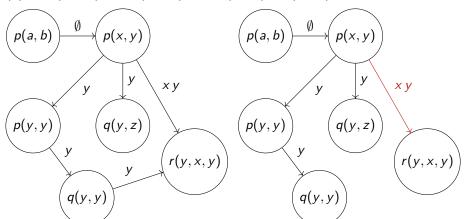
NEW!!!



NEW!!!



NEW!!!



Derivation Graph Sets

 $\mathcal{D} \stackrel{\delta}{\longrightarrow} \mathcal{I} \text{ exhibits cycle-free derivation graph} \Rightarrow \mathcal{I} \text{ has bounded treewidth}.$

- 1 A cycle-free derivation graph corresponds to a tree decomposition.
- 2 Width is bounded by $\max_{\rho \in \mathcal{R}} \{ |terms(\mathcal{D})| + |terms(head(\rho))| \}.$

Cycle-free Derivation Graph Set (cdgs): If $\mathcal{D} \stackrel{\delta}{\longrightarrow} \mathcal{I}$, then the derivation graph of δ is reducible to a cycle-free graph.

Weakly Cycle-free Derivation Graph Set (wcdgs): If $\mathcal{D} \stackrel{\delta}{\longrightarrow} \mathcal{I}$, then there exists a derivation δ' such that the derivation graph of δ' is reducible to a cycle-free graph.

Theorem

- **1** cdgs, wcdgs ⊂ bts.
- 2 If \mathcal{R} is cdgs or wcdgs, then query entailment is decidable.

- Introduction
- Derivation Graphs
- Greedy Derivations
- Showing gbts ⊂ wgbts
- Conclusion

Greedy Derivation (Thomazo et al. 2012) Frontier of any rule application is only mapped to elements which are (1) constants from \mathcal{D} , or (2) from the head of one single previous rule application.

Greedy Derivation (Thomazo et al. 2012) Frontier of any rule application is only mapped to elements which are (1) constants from \mathcal{D} , or (2) from the head of one single previous rule application.

```
\mathcal{D} = \{p(a,b)\}
\mathcal{R} = \{
\mathbf{1} \ p(x,y) \to \exists z. \ q(y,z),
\mathbf{2} \ p(x,y) \land q(y,z) \to q(y,x),
\mathbf{3} \ q(u,v) \to r(v).
\}
```

Greedy Derivation (Thomazo et al. 2012) Frontier of any rule application is only mapped to elements which are (1) constants from \mathcal{D} , or (2) from the head of one single previous rule application.

```
\mathcal{D} = \{p(a,b)\}
\mathcal{R} = \{
\mathbf{1} \ p(x,y) \to \exists z. \ q(y,z),
\mathbf{2} \ p(x,y) \land q(y,z) \to q(y,x),
\mathbf{3} \ q(u,v) \to r(v).
\}
```

$$\mathcal{D} \{p(a, b)\}$$

$$\mathcal{I}_1 \{p(a, b), q(b, x)\}$$

Greedy Derivation (Thomazo et al. 2012) Frontier of any rule application is only mapped to elements which are (1) constants from \mathcal{D} , or (2) from the head of one single previous rule application.

```
\mathcal{D} = \{p(a,b)\} \qquad \mathcal{D} \quad \{p(x,y) \rightarrow \exists z. \ q(y,z), \\ \mathbf{1} \quad p(x,y) \rightarrow \exists z. \ q(y,z), \\ \mathbf{2} \quad p(x,y) \land q(y,z) \rightarrow q(y,x), \\ \mathbf{3} \quad q(u,v) \rightarrow r(v). \\ \}
```

```
\mathcal{D} \{p(a,b)\}
\mathcal{I}_1 \{p(a,b), q(b,x)\}
\mathcal{I}_2 \{p(a,b), q(b,x), q(b,a)\}
```

Greedy Derivation (Thomazo et al. 2012) Frontier of any rule application is only mapped to elements which are (1) constants from \mathcal{D} , or (2) from the head of one single previous rule application.

```
\mathcal{D} \{p(a,b)\} 

\mathcal{I}_{1} \{p(a,b), q(b,x)\} 

\mathcal{I}_{2} \{p(a,b), q(b,x), q(b,a)\} 

\mathcal{I}_{3} \{p(a,b), q(b,x), q(b,x), r(x)\}
```

Greedy Derivations and Decidability

Greedy Bounded Treewidth Set (gbts): if $\mathcal{D} \xrightarrow{\delta} \mathcal{I}$, then δ is greedy.

Weakly Greedy Bounded Treewidth Set (wgbts): if $\mathcal{D} \stackrel{\delta}{\longrightarrow} \mathcal{I}$, then there exists $\mathcal{D} \stackrel{\delta'}{\longrightarrow} \mathcal{I}$ such that δ' is greedy.

Theorem

- \blacktriangleright (w)gbts = (w)cdgs.
- ▶ gbts ⊂ wgbts.
- ▶ cdgs ⊂ wcdgs.
- ▶ All such rule sets have decidable BCQ entailment.

- Introduction
- Derivation Graphs
- Greedy Derivations
- Showing gbts ⊂ wgbts
- Conclusion

Dependency (Baget 2004): A rule ρ' depends on ρ , if ρ can "trigger" ρ' .

Graph of Rule Dependencies (Baget 2004): A directed graph showing all dependencies between rules.

1
$$\rho_1 = p(x) \rightarrow \exists zw. q(x, z, w)$$

2 $\rho_2 = r(y) \rightarrow \exists uv. s(y, u, v)$
3 $\rho_3 = p(x) \land r(y) \rightarrow \exists zwuv. q(x, z, w) \land s(y, u, v)$
4 $\rho_4 = q(x, z, w) \land s(y, u, v) \rightarrow \exists x'. t(x, z, y, u, x')$
}

 $\mathcal{R} = \{$

Dependency (Baget 2004): A rule ρ' depends on ρ , if ρ can "trigger" ρ' .

Graph of Rule Dependencies (Baget 2004): A directed graph showing all dependencies between rules.

```
\mathcal{R} = \{
\mathbf{1} \quad \rho_1 = p(x) \to \exists zw. q(x, z, w)
\mathbf{2} \quad \rho_2 = r(y) \to \exists uv. s(y, u, v)
\mathbf{3} \quad \rho_3 = p(x) \land r(y) \to \exists zwuv. q(x, z, w) \land s(y, u, v)
\mathbf{4} \quad \rho_4 = q(x, z, w) \land s(y, u, v) \to \exists x'. t(x, z, y, u, x')
```

Rudolph (TU Dresden)

Dependency (Baget 2004): A rule ρ' depends on ρ , if ρ can "trigger" ρ' .

Graph of Rule Dependencies (Baget 2004): A directed graph showing all dependencies between rules.

$$\mathcal{R} = \{$$

$$\mathbf{1} \quad \rho_1 = p(x) \to \exists zw. q(x, z, w)$$

$$\mathbf{2} \quad \rho_2 = r(y) \to \exists uv. s(y, u, v)$$

$$\mathbf{3} \quad \rho_3 = p(x) \land r(y) \to \exists zwuv. q(x, z, w) \land s(y, u, v)$$

$$\mathbf{4} \quad \rho_4 = q(x, z, w) \land s(y, u, v) \to \exists x'. t(x, z, y, u, x')$$

Dependency (Baget 2004): A rule ρ' depends on ρ , if ρ can "trigger" ρ' .

Graph of Rule Dependencies (Baget 2004): A directed graph showing all dependencies between rules.

Dependency (Baget 2004): A rule ρ' depends on ρ , if ρ can "trigger" ρ' .

Graph of Rule Dependencies (Baget 2004): A directed graph showing all dependencies between rules.

$$\mathcal{R} = \{$$

$$\mathbf{1} \quad \rho_{1} = p(x) \rightarrow \exists zw. q(x, z, w)$$

$$\mathbf{2} \quad \rho_{2} = r(y) \rightarrow \exists uv. s(y, u, v)$$

$$\mathbf{3} \quad \rho_{3} = p(x) \land r(y) \rightarrow \exists zwuv. q(x, z, w) \land s(y, u, v)$$

$$\mathbf{4} \quad \rho_{4} = q(x, z, w) \land s(y, u, v) \rightarrow \exists x'. t(x, z, y, u, x')$$

An Observation

$$\mathcal{R} = \{$$

$$\mathbf{1} \quad \rho_{1} = p(x) \rightarrow \exists zw. q(x, z, w)$$

$$\mathbf{2} \quad \rho_{2} = r(y) \rightarrow \exists uv. s(y, u, v)$$

$$\mathbf{3} \quad \rho_{3} = p(x) \land r(y) \rightarrow \exists zwuv. q(x, z, w) \land s(y, u, v)$$

$$\mathbf{4} \quad \rho_{4} = q(x, z, w) \land s(y, u, v) \rightarrow \exists x'. t(x, z, y, u, x')$$

A Derivation: Let $\rho \in \{\rho_1, \rho_2, \rho_3\}$.

$$\mathcal{D},\ldots,(\rho_i,h_i,\mathcal{I}_i),(\rho,h_{i+1},\mathcal{I}_{i+1}),\ldots,(\rho_n,h_n,\mathcal{I}_n)$$

An Observation

$$\mathcal{R} = \{$$

$$\mathbf{1} \quad \rho_{1} = p(x) \rightarrow \exists zw. q(x, z, w)$$

$$\mathbf{2} \quad \rho_{2} = r(y) \rightarrow \exists uv. s(y, u, v)$$

$$\mathbf{3} \quad \rho_{3} = p(x) \land r(y) \rightarrow \exists zwuv. q(x, z, w) \land s(y, u, v)$$

$$\mathbf{4} \quad \rho_{4} = q(x, z, w) \land s(y, u, v) \rightarrow \exists x'. t(x, z, y, u, x')$$

A Derivation: Let $\rho \in \{\rho_1, \rho_2, \rho_3\}$.

$$\mathcal{D}, \ldots, (\rho_i, h_i, \mathcal{I}_i), (\rho, h_{i+1}, \mathcal{I}_{i+1}), \ldots, (\rho_n, h_n, \mathcal{I}_n)$$

An Observation

$$\mathcal{R} = \{$$

$$\mathbf{1} \quad \rho_{1} = p(x) \rightarrow \exists zw. q(x, z, w)$$

$$\mathbf{2} \quad \rho_{2} = r(y) \rightarrow \exists uv. s(y, u, v)$$

$$\mathbf{3} \quad \rho_{3} = p(x) \land r(y) \rightarrow \exists zwuv. q(x, z, w) \land s(y, u, v)$$

$$\mathbf{4} \quad \rho_{4} = q(x, z, w) \land s(y, u, v) \rightarrow \exists x'. t(x, z, y, u, x')$$

A Derivation: Let $\rho \in \{\rho_1, \rho_2, \rho_3\}$.

$$\mathcal{D},\ldots,(\rho_i,h_i,\mathcal{I}_i),(\rho,h_{i+1},\mathcal{I}_{i+1}),\ldots,(\rho_n,h_n,\mathcal{I}_n)$$

The Observation: We can permute rule applications!

$$\mathcal{D}, \ldots, (\rho, h_{i+1}, \mathcal{I}_{i-1} \cup (\mathcal{I}_{i+1} \setminus \mathcal{I}_i)), (\rho_i, h_i, \mathcal{I}_{i+1}), \ldots, (\rho_n, h_n, \mathcal{I}_n)$$

Re-writing Derivations via Permutations

Lemma (Permutation Lemma)

Suppose we have a (greedy) derivation of the following form:

$$\delta := \mathcal{D}, \dots, (\rho_i, h_i, \mathcal{I}_{i+1}), (\rho_{i+1}, h_{i+1}, \mathcal{I}_{i+2}), \dots, (\rho_{n-1}, h_{n-1}, \mathcal{I}_n)$$

If ρ_{i+1} does not depend on ρ_i , then δ' is also a (greedy) derivation:

$$\delta' := \mathcal{D}, \ldots, (\rho_{i+1}, h_{i+1}, \mathcal{I}'_i), (\rho_i, h_i, \mathcal{I}_{i+2}), \ldots, (\rho_{n-1}, h_{n-1}, \mathcal{I}_n)$$

where
$$\mathcal{I}_i' = \mathcal{I}_i \cup (\mathcal{I}_{i+2} \setminus \mathcal{I}_{i+1})$$
.

```
\mathcal{R} = \{
  p(x) \rightarrow \exists zw. q(x, z, w)
  r(y) \rightarrow \exists uv.s(y, u, v)
  \exists p(x) \land r(y) \rightarrow \exists zwuv.q(x,z,w) \land s(y,u,v)
   \exists q(x,z,w) \land s(y,u,v) \rightarrow \exists x'.t(x,z,y,u,x') 
  \mathcal{D} \{p(a), r(b)\}
 \mathcal{I}_1 {p(a), r(b), q(a, z, w)}
```

```
\mathcal{R} = \{
  p(x) \rightarrow \exists zw. q(x, z, w)
  r(y) \rightarrow \exists uv.s(y, u, v)
  \exists p(x) \land r(y) \rightarrow \exists zwuv. q(x, z, w) \land s(y, u, v)
   q(x,z,w) \wedge s(y,u,v) \rightarrow \exists x'.t(x,z,y,u,x') 
  \mathcal{D} \{p(a), r(b)\}
 \mathcal{I}_1 {p(a), r(b), q(a, z, w)}
 \mathcal{I}_2 {p(a), r(b), q(a, z, w), s(b, u, v)}
```

```
\mathcal{R} = \{
  p(x) \rightarrow \exists zw. q(x, z, w)
  r(y) \rightarrow \exists uv.s(y, u, v)
  \exists p(x) \land r(y) \rightarrow \exists zwuv. q(x, z, w) \land s(y, u, v)
  4 g(x,z,w) \land s(y,u,v) \rightarrow \exists x'.t(x,z,y,u,x')
  \mathcal{D} \{p(a), r(b)\}
 \mathcal{I}_1 \{ p(a), r(b), q(a, z, w) \}
 \mathcal{I}_2 {p(a), r(b), g(a, z, w), s(b, u, v)}
 \mathcal{I}_3 {p(a), r(b), q(a, z, w), s(b, u, v), t(a, z, b, u, x')}
                               \mathcal{R} is not gbts, but is \mathcal{R} wgbts?!
```

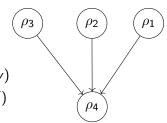
$$\mathcal{R} = \{$$

1
$$p(x) \rightarrow \exists zw. q(x, z, w)$$

$$r(y) \rightarrow \exists uv.s(y, u, v)$$

$$\exists p(x) \land r(y) \rightarrow \exists zwuv.q(x,z,w) \land s(y,u,v)$$

$$q(x,z,w) \wedge s(y,u,v) \rightarrow \exists x'.t(x,z,y,u,x')$$



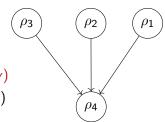
$$\mathcal{R} = \{$$

1
$$p(x) \rightarrow \exists zw. q(x, z, w)$$

$$r(y) \rightarrow \exists uv.s(y, u, v)$$

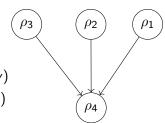
$$\exists p(x) \land r(y) \rightarrow \exists zwuv.q(x,z,w) \land s(y,u,v)$$

$$q(x,z,w) \wedge s(y,u,v) \rightarrow \exists x'.t(x,z,y,u,x')$$



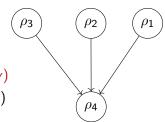
$$\mathcal{R} = \{$$

- 1 $p(x) \rightarrow \exists zw. q(x, z, w)$
- $r(y) \rightarrow \exists uv.s(y, u, v)$
- $p(x) \land r(y) \rightarrow \exists zwuv.q(x,z,w) \land s(y,u,v)$
- $q(x,z,w) \wedge s(y,u,v) \rightarrow \exists x'.t(x,z,y,u,x')$



$$\mathcal{R} = \{$$

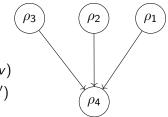
- 1 $p(x) \rightarrow \exists zw. q(x, z, w)$
- $r(y) \rightarrow \exists uv.s(y, u, v)$
- $\exists p(x) \land r(y) \rightarrow \exists zwuv.q(x,z,w) \land s(y,u,v)$
- $q(x,z,w) \wedge s(y,u,v) \rightarrow \exists x'.t(x,z,y,u,x')$



Idea 1: Use 3 to simulate 1 and 2.

$$\mathcal{R} = \{$$

- $r(y) \rightarrow \exists uv.s(y, u, v)$
- $\exists p(x) \land r(y) \rightarrow \exists zwuv.q(x,z,w) \land s(y,u,v)$
- $q(x,z,w) \wedge s(y,u,v) \rightarrow \exists x'.t(x,z,y,u,x')$



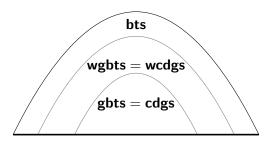
Idea 2: Permutation Lemma →

- (i) Permute instances of 1 or 2 backward to instances of 2 or 1 (resp.)
- (ii) Replace 1,2 or 2,1 instances with 3.

- Introduction
- 2 Derivation Graphs
- Greedy Derivations
- 4 Showing gbts ⊂ wgbts
- Conclusion

Summary and Future Work

Our Results:



Open Questions:

- 1 Are gbts and wgbts recognizable?
- 2 What is the complexity of BCQ entailment in (w)gbts?
- 3 Can other reduction operations generalize (w)cdgs?