



A Dichotomy for Homomorphism-Closed Queries on Probabilistic Graphs

Antoine Amarilli¹ and Ismail İlkan Ceylan²

September 16, 2020

¹Télécom Paris

²University of Oxford

Uncertain data management

In this talk, we manage **data** represented as a **labeled graph**

Uncertain data management

In this talk, we manage **data** represented as a **labeled graph**

WorksAt

Antoine	Télécom Paris
Antoine	Paris Sud
Benny	Paris Sud
Benny	Technion
İsmail	U. Oxford

Uncertain data management

In this talk, we manage **data** represented as a **labeled graph**

WorksAt

Antoine	Télécom Paris
Antoine	Paris Sud
Benny	Paris Sud
Benny	Technion
İsmail	U. Oxford

MemberOf

Télécom Paris	ParisTech
Télécom Paris	IP Paris
Paris Sud	IP Paris
Paris Sud	Paris-Saclay
Technion	CESAER

Uncertain data management

In this talk, we manage **data** represented as a **labeled graph**

WorksAt		Antoine	Télécom Paris	ParisTech
Antoine	Télécom Paris			
Antoine	Paris Sud			
Benny	Paris Sud			
Benny	Technion		Paris Sud	IP Paris
İsmail	U. Oxford			
MemberOf		Benny	Technion	Paris-Saclay
Télécom Paris	ParisTech			
Télécom Paris	IP Paris			
Paris Sud	IP Paris			
Paris Sud	Paris-Saclay			
Technion	CESAER	İsmail	U. Oxford	CESAER

Uncertain data management

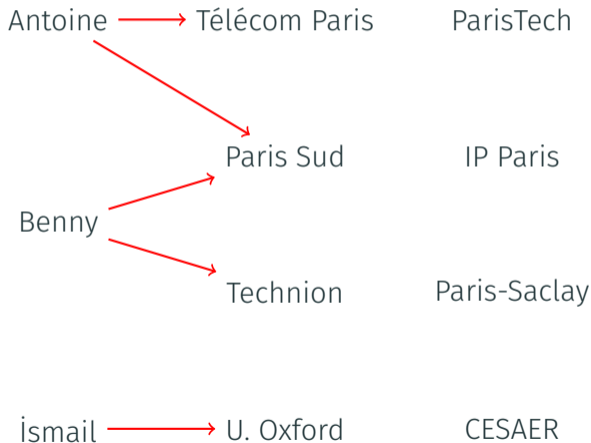
In this talk, we manage **data** represented as a **labeled graph**

WorksAt

Antoine	Télécom Paris
Antoine	Paris Sud
Benny	Paris Sud
Benny	Technion
İsmail	U. Oxford

MemberOf

Télécom Paris	ParisTech
Télécom Paris	IP Paris
Paris Sud	IP Paris
Paris Sud	Paris-Saclay
Technion	CESAER



Uncertain data management

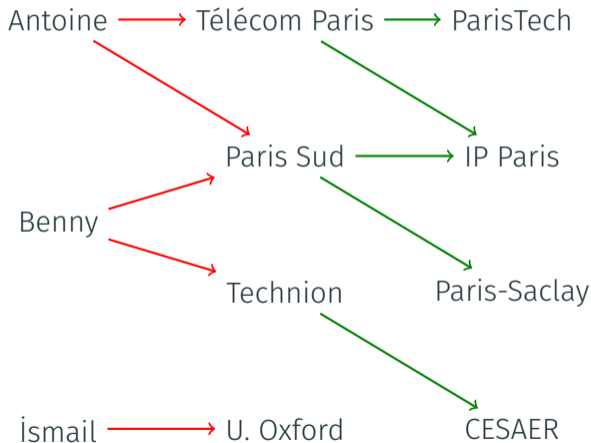
In this talk, we manage **data** represented as a **labeled graph**

WorksAt

Antoine	Télécom Paris
Antoine	Paris Sud
Benny	Paris Sud
Benny	Technion
İsmail	U. Oxford

MemberOf

Télécom Paris	ParisTech
Télécom Paris	IP Paris
Paris Sud	IP Paris
Paris Sud	Paris-Saclay
Technion	CESAER



Uncertain data management

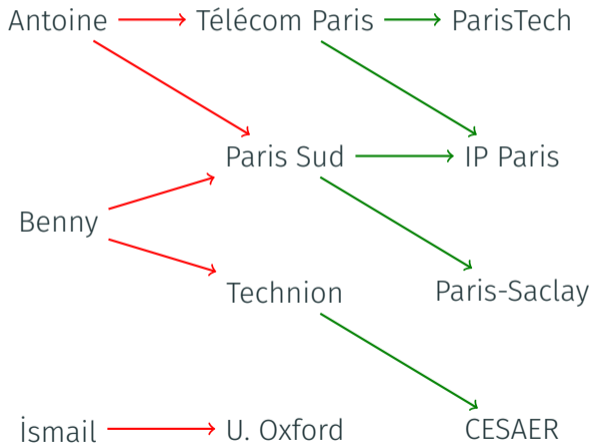
In this talk, we manage **data** represented as a **labeled graph**

WorksAt

Antoine	Télécom Paris
Antoine	Paris Sud
Benny	Paris Sud
Benny	Technion
İsmail	U. Oxford

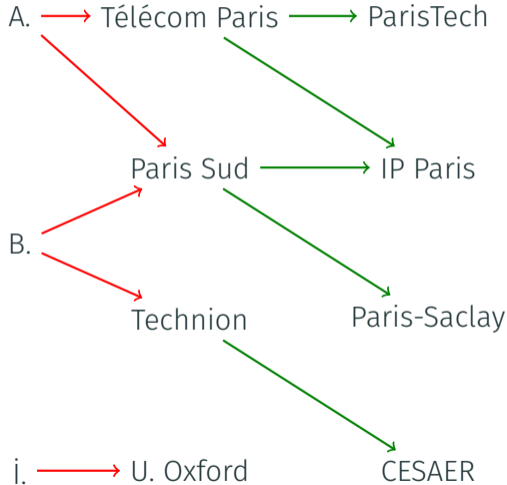
MemberOf

Télécom Paris	ParisTech
Télécom Paris	IP Paris
Paris Sud	IP Paris
Paris Sud	Paris-Saclay
Technion	CESAER



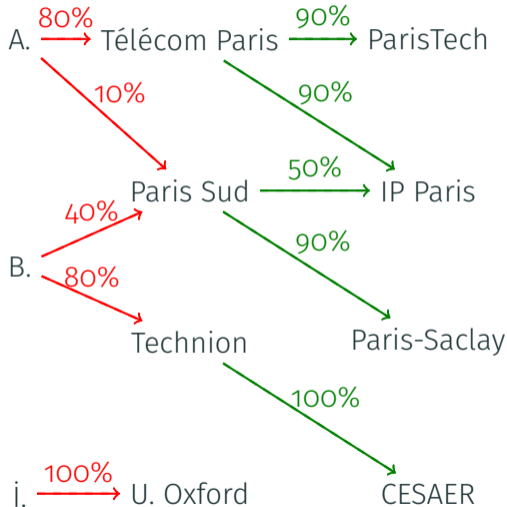
→ **Problem:** we are not **certain** about the true state of the data

Uncertain data model



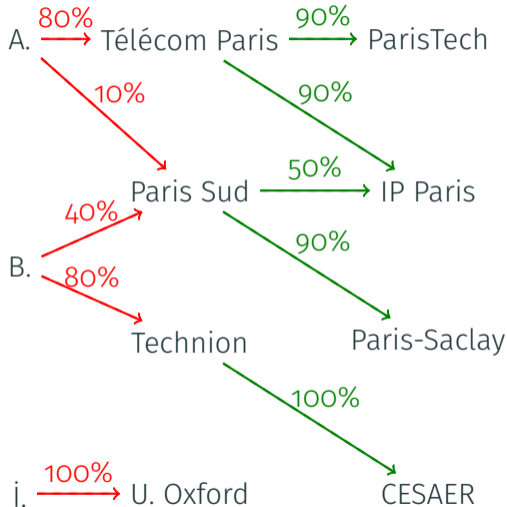
- Uncertain data model: **TID**, for **tuple-independent database**
- Each fact (edge) carries a **probability**

Uncertain data model



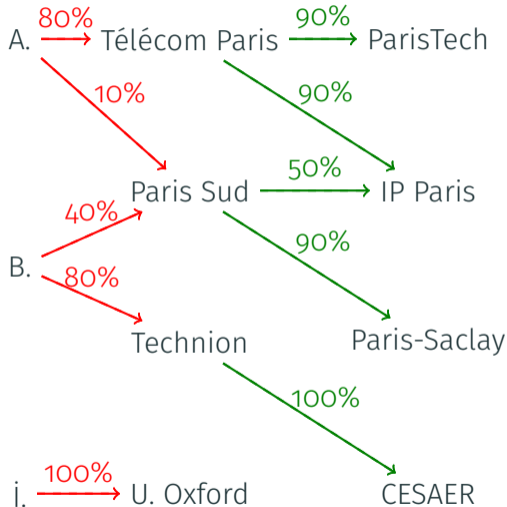
- Uncertain data model: **TID**, for **tuple-independent database**
- Each fact (edge) carries a **probability**

Uncertain data model



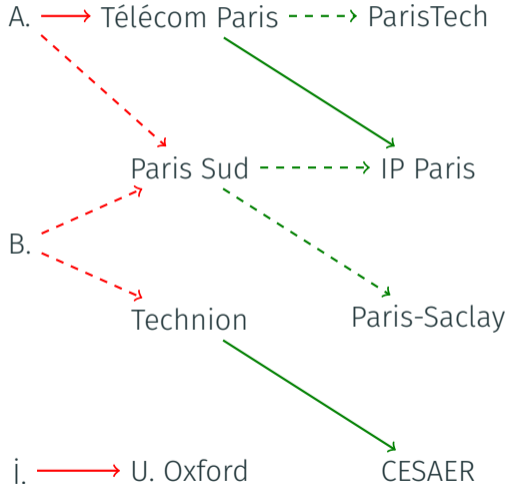
- Uncertain data model: **TID**, for **tuple-independent database**
- Each fact (edge) carries a **probability**
- Each fact exists with its given **probability**
- All facts are **independent**

Uncertain data model



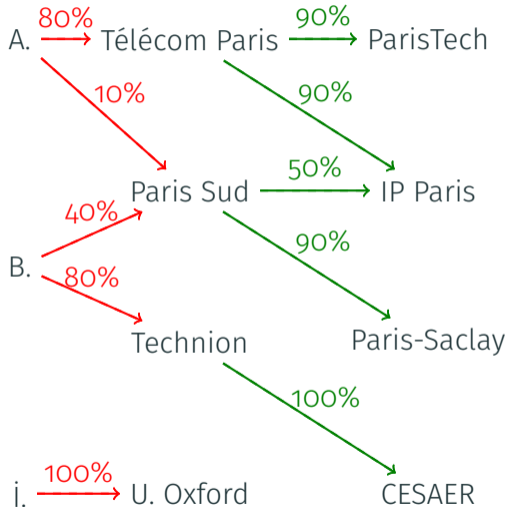
- Uncertain data model: **TID**, for **tuple-independent database**
- Each fact (edge) carries a **probability**
- Each fact exists with its given **probability**
- All facts are **independent**
- **Possible world W** : subset of facts

Uncertain data model



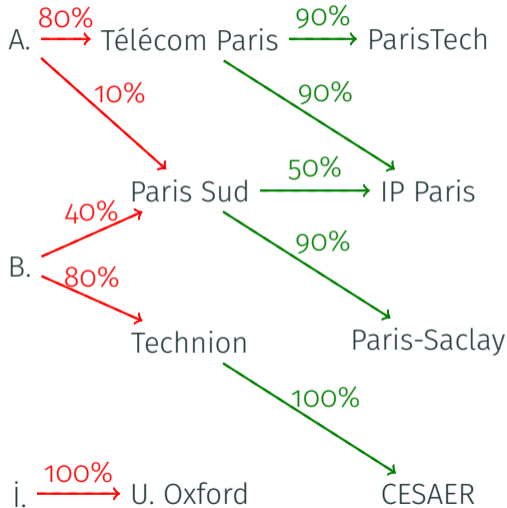
- Uncertain data model: **TID**, for **tuple-independent database**
- Each fact (edge) carries a **probability**
- Each fact exists with its given **probability**
- All facts are **independent**
- **Possible world W** : subset of facts

Uncertain data model



- Uncertain data model: **TID**, for **tuple-independent database**
- Each fact (edge) carries a **probability**
- Each fact exists with its given **probability**
- All facts are **independent**
- **Possible world W** : subset of facts
- **Probability of W** :

Uncertain data model



- Uncertain data model: **TID**, for **tuple-independent database**
- Each fact (edge) carries a **probability**
- Each fact exists with its given **probability**
- All facts are **independent**
- **Possible world W** : subset of facts
- **Probability** of W :

$$\Pr(W) = \left(\prod_{F \in W} \Pr(F) \right) \times \left(\prod_{F \notin W} (1 - \Pr(F)) \right)$$

Homomorphism-closed queries

- **Query:** maps a graph (**without probabilities**) to YES/NO

Homomorphism-closed queries

- **Query:** maps a graph (**without probabilities**) to YES/NO
- **Conjunctive query** (CQ): can I find a match of a **pattern**? e.g., $x \xrightarrow{\text{red}} y \xrightarrow{\text{green}} z$

Homomorphism-closed queries

- **Query**: maps a graph (**without probabilities**) to YES/NO
- **Conjunctive query** (CQ): can I find a match of a **pattern**? e.g., $x \xrightarrow{\text{red}} y \xrightarrow{\text{green}} z$
- **Union of conjunctive queries** (UCQ): does **one of the CQs** match?

Homomorphism-closed queries

- **Query**: maps a graph (**without probabilities**) to YES/NO
 - **Conjunctive query** (CQ): can I find a match of a **pattern**? e.g., $x \xrightarrow{\text{red}} y \xrightarrow{\text{green}} z$
 - **Union of conjunctive queries** (UCQ): does **one of the CQs** match?
- **Homomorphism-closed query** Q : if G satisfies Q and G has a homomorphism to G' then G' also satisfies Q

Homomorphism-closed queries

- **Query**: maps a graph (**without probabilities**) to YES/NO
 - **Conjunctive query** (CQ): can I find a match of a **pattern**? e.g., $x \xrightarrow{\text{red}} y \xrightarrow{\text{green}} z$
 - **Union of conjunctive queries** (UCQ): does **one of the CQs** match?
- **Homomorphism-closed query** Q : if G satisfies Q and G has a homomorphism to G' then G' also satisfies Q

They generalize **CQs** and **UCQs**, but also **regular path queries** (RPQs), **Datalog**, etc.

Problem statement: Probabilistic query evaluation (PQE)

Here is the problem PQE(Q):

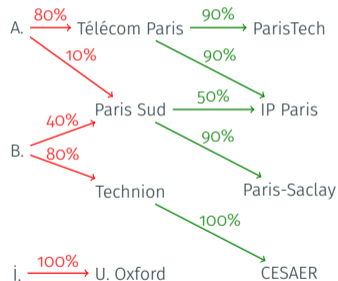
- We **fix** a query Q : $x \xrightarrow{\text{red}} y \xrightarrow{\text{green}} z$

Problem statement: Probabilistic query evaluation (PQE)

Here is the problem PQE(Q):

- We **fix** a query Q : $x \xrightarrow{\text{red}} y \xrightarrow{\text{green}} z$

- The **input** is a TID D :

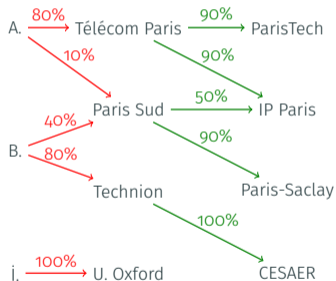


Problem statement: Probabilistic query evaluation (PQE)

Here is the problem PQE(Q):

- We **fix** a query Q : $x \xrightarrow{\text{red}} y \xrightarrow{\text{green}} z$

- The **input** is a TID D :



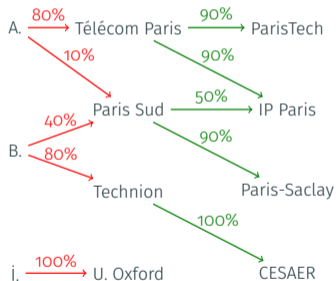
- The **output** is the **probability** that the query is true

Problem statement: Probabilistic query evaluation (PQE)

Here is the problem PQE(Q):

- We **fix** a query Q : $x \xrightarrow{\text{red}} y \xrightarrow{\text{green}} z$

- The **input** is a TID D :



- The **output** is the **probability** that the query is true

→ **Question**: What is the complexity of PQE(Q) depending on the query Q?

Results on PQE

Existing dichotomy on the **unions of conjunctive queries** (UCQs):

Theorem [Dalvi and Suciu, 2012]

- Some UCQs Q are **safe** and $\text{PQE}(Q)$ is in **PTIME**
- All others are **unsafe** and $\text{PQE}(Q)$ is **#P-hard**

Results on PQE

Existing dichotomy on the **unions of conjunctive queries** (UCQs):

Theorem [Dalvi and Suciu, 2012]

- Some UCQs Q are **safe** and $\text{PQE}(Q)$ is in **PTIME**
- All others are **unsafe** and $\text{PQE}(Q)$ is **#P-hard**

We study PQE for **homomorphism-closed queries** and show:

Results on PQE

Existing dichotomy on the **unions of conjunctive queries** (UCQs):

Theorem [Dalvi and Suciu, 2012]

- Some UCQs Q are **safe** and $\text{PQE}(Q)$ is in **PTIME**
- All others are **unsafe** and $\text{PQE}(Q)$ is **#P-hard**

We study PQE for **homomorphism-closed queries** and show:

Theorem [Amarilli and Ceylan, 2020]

For any **query Q closed under homomorphisms**:

- Either Q is equivalent to a **safe UCQ** and $\text{PQE}(Q)$ is in **PTIME**

Results on PQE

Existing dichotomy on the **unions of conjunctive queries** (UCQs):

Theorem [Dalvi and Suciu, 2012]

- Some UCQs Q are **safe** and $\text{PQE}(Q)$ is in **PTIME**
- All others are **unsafe** and $\text{PQE}(Q)$ is **#P-hard**

We study PQE for **homomorphism-closed queries** and show:

Theorem [Amarilli and Ceylan, 2020]

For any **query Q closed under homomorphisms**:

- Either Q is equivalent to a **safe UCQ** and $\text{PQE}(Q)$ is in **PTIME**
- In all other cases, $\text{PQE}(Q)$ is **#P-hard**

Results on PQE

Existing dichotomy on the **unions of conjunctive queries** (UCQs):

Theorem [Dalvi and Suciu, 2012]

- Some UCQs Q are **safe** and $\text{PQE}(Q)$ is in **PTIME**
- All others are **unsafe** and $\text{PQE}(Q)$ is **#P-hard**

We study PQE for **homomorphism-closed queries** and show:

Theorem [Amarilli and Ceylan, 2020]

For any **query Q closed under homomorphisms**:

- Either Q is equivalent to a **safe UCQ** and $\text{PQE}(Q)$ is in **PTIME**
- In all other cases, $\text{PQE}(Q)$ is **#P-hard**

So **bad news**: all homomorphism-closed queries are **hard** except safe UCQs

What's next?

- The result only applies to **graphs**, not higher-arity databases
 - We **conjecture** that it holds for arbitrary arity

What's next?

- The result only applies to **graphs**, not higher-arity databases
 - We **conjecture** that it holds for arbitrary arity
- Adapting to **unweighted** PQE, where all probabilities are $1/2$?
 - We have a recent result on **non-hierarchical self-join-free CQs** [Amarilli and Kimelfeld, 2020]
 - Recent paper by Suciu and Kenig [Kenig and Suciu, 2020]

What's next?

- The result only applies to **graphs**, not higher-arity databases
 - We **conjecture** that it holds for arbitrary arity
- Adapting to **unweighted** PQE, where all probabilities are $1/2$?
 - We have a recent result on **non-hierarchical self-join-free CQs** [Amarilli and Kimelfeld, 2020]
 - Recent paper by Suciu and Kenig [Kenig and Suciu, 2020]

Thanks for your attention!

References i



Amarilli, A. and Ceylan, I. I. (2020).

A dichotomy for homomorphism-closed queries on probabilistic graphs.

In *ICDT*.



Amarilli, A. and Kimelfeld, B. (2020).

Uniform reliability of self-join-free conjunctive queries.

arXiv preprint arXiv:1908.07093.



Dalvi, N. and Suciu, D. (2012).

The dichotomy of probabilistic inference for unions of conjunctive queries.

J. ACM, 59(6).



Kenig, B. and Suciu, D. (2020).

A dichotomy for the generalized model counting problem for unions of conjunctive queries.

arXiv preprint arXiv:2008.00896.