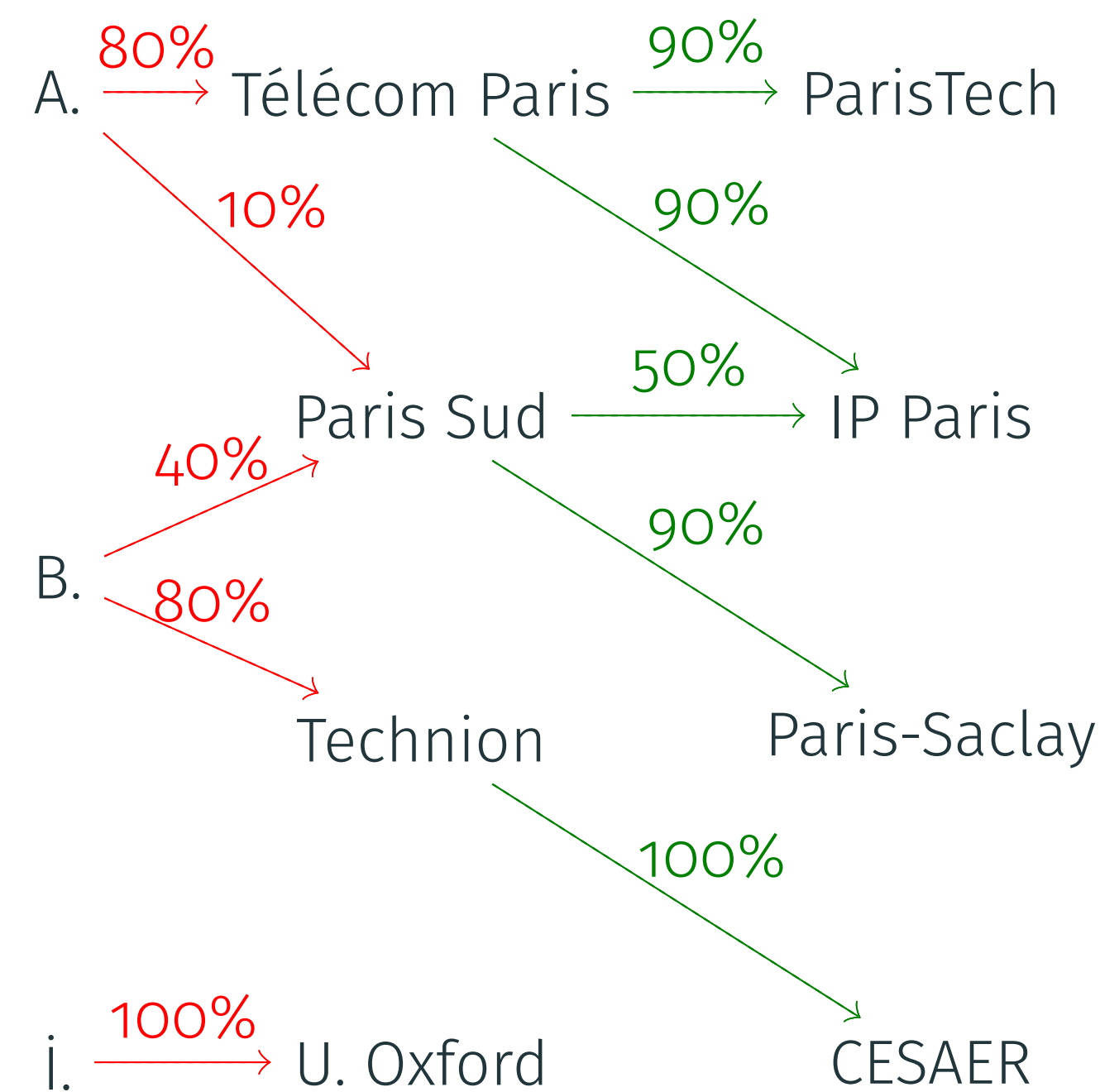# A Dichotomy for Homomorphism-Closed Queries on Probabilistic Graphs

Antoine Amarilli (LTCI, Télécom Paris, Institut Polytechnique de Paris), **İsmail İlkan Ceylan** (University of Oxford)

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
- Each possible world W has a **probability**:

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

Problem statement: Probabilistic query evaluation (PQE)

- We **fix** a query Q, for instance the CQ: $x \longrightarrow y \longrightarrow z$
- The **input** is a TID D:
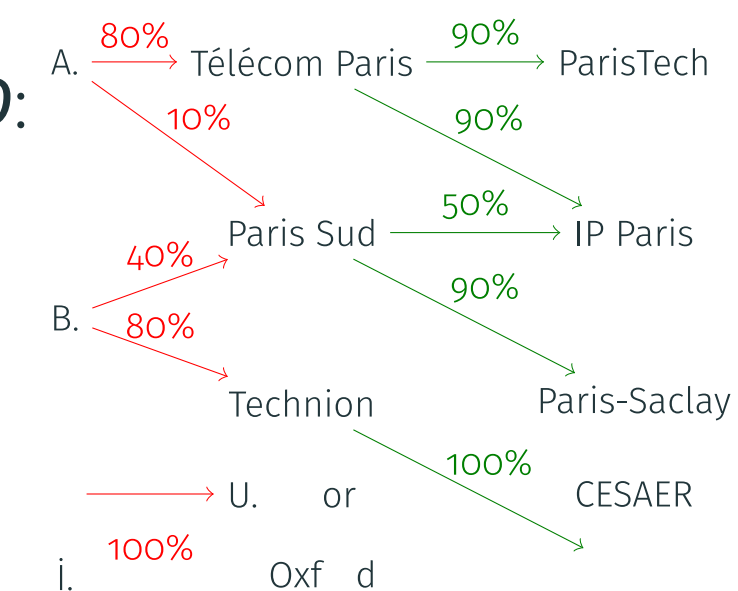
- The **output** is the **total probability** of the worlds which satisfy Q:
  - Formally: $\sum_{W \subseteq D, W \models Q} \Pr(W)$
  - → Intuition: the **probability** that the query is true
- For a fixed query Q, we write the problem $\mathrm{PQE}(Q)$

→ Complexity of $\mathrm{PQE}(Q)$ (in D), depending on Q?

Homomorphism-closed queries

- A **homomorphism** from a graph G to a graph G' maps the vertices of G to those of G' while preserving the edges

 has a homomorphism to

- **Homomorphism-closed query** Q: for any graph G, if G satisfies Q and G has a homomorphism to G' then G' also satisfies Q
- Homomorphism-closed queries include **all CQs**, **all UCQs**, some **recursive queries** like **regular path queries** (RPQs), **Datalog**, etc.
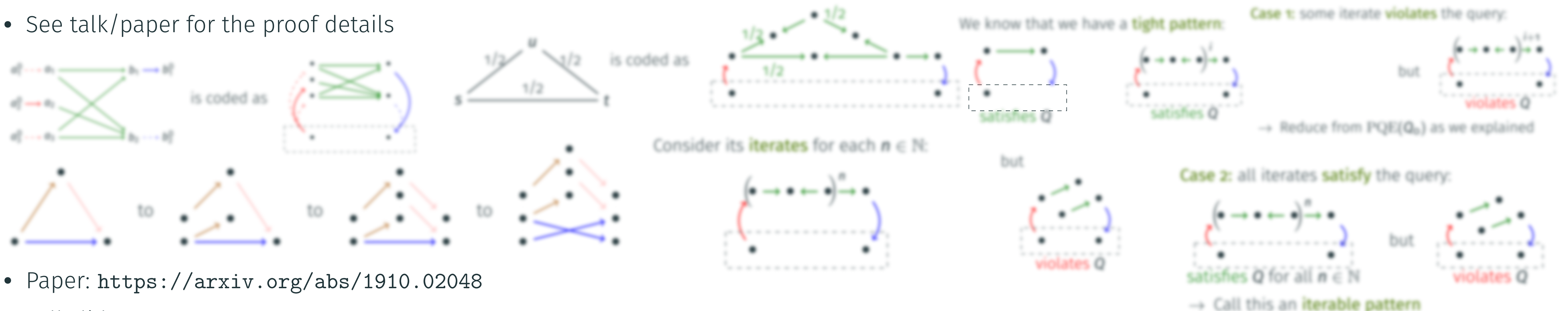
Our result

Theorem (Amarilli and Ceylan 2020)

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

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

Presented at ICDT'20 (best paper award)

Proof?

- See talk/paper for the proof details

- Paper: `https://arxiv.org/abs/1910.02048`
- Full slides: `https://a3nm.net/work/talks/icdt2020/amarilli2020dichotomy_slides.pdf`