

# Uncertain Data Management Class Structure

**Antoine Amarilli<sup>1</sup>, Silviu Maniu<sup>2</sup>**

<sup>1</sup>Télécom ParisTech

<sup>2</sup>LRI

November 23rd, 2015



# Class 1 (today): Introduction and Open-World Databases

- **Motivations** and **sources** of uncertain data
- Class **structure** (these slides!)
- **Reminder**: relational algebra and relational calculus
- **Open-world databases**: modeling **incompleteness**

## Class 2 (Nov 30): Closed-World Databases

- Missing values (NULLs)
- c-tables
- Representation size for the various models

## Class 3 (Dec 7): Relational Probabilistic Databases

- **Semantics** of probabilistic databases
- **Models:**
  - Tuple-independent databases
  - Block-independent disjoint databases
  - pc-tables
- **Extensions**

## Class 4 (Dec 14): Non-Relational Probabilistic Databases

- Uncertain XML
  - Semantics of documents
  - Models and compactness
  - Query languages and complexity
- Uncertain graphs
- Other models

## Class 5 (Jan 4): Querying Relational Prob. Databases

- **Semantics** of queries on uncertain data
- **Extensional** approach: evaluating the query **directly**
- **Intensional** approach: computing the query **lineage**
- Tractable **rules** for query evaluation
- Tractable **lineage** classes
- Computational **hardness** results

## Class 6 (Jan 11): Practical Applications

- Practical implementation
- The MayBMS engine
- Updates and repairs
- Crawling the Web
- Learning probabilistic distributions
- Crowdsourcing

## Class 7 (Jan 18): Other Applications and Lab Session

- Other applications
- Lab session