



ID de Contribution: 21

Type: **Poster**

Finding missing data by modelling and solving the MCC problem

Missing data is a pervasive issue in empirical datasets, arising when individuals or data-collection devices fail to record observations, resulting in missing attribute values or, in some cases, entire records. Such incompleteness is common in real-world domains, including clinical databases such as Traumabase. A central challenge is whether missing values can be imputed in a manner that preserves the underlying statistical properties of the data, thereby improving the validity and accuracy of downstream analyses without altering their intended scope.

We address this challenge by formalizing the problem at the intersection of incomplete database theory and causal modeling of missingness mechanisms. Incomplete database theory provides the semantic foundation for reasoning over partially observed data through possible worlds semantics. Within this framework, we define a world as a complete instantiation of the database, represented as a set of tuples, and a class as a set of possible worlds sharing identical multiset content, such that all worlds in a class yield identical answers to any query.

To capture the generative process underlying missing data, we incorporate causal assumptions following Pearl's theory of missingness mechanisms. Conditional dependencies among variables and their corresponding missingness indicators are modeled using a Bayesian network, referred to as a Missingness Graph (as introduced by Pearl and Mohan). Given an incomplete dataset, our objective is to identify a class of possible worlds that is consistent with the observed data while minimizing the distance between its induced statistical distribution—defined over joint tuple probabilities—and the target distribution implied by the causal missingness model. This approach unifies incomplete database semantics with causal inference, providing a principled framework for statistically faithful data imputation under explicit assumptions about missingness.

Auteur: BOUKRIA, Ali

Classification de Session: Poster Flash Talks