Facets of Probabilistic Databases

Benny Kimelfeld

Technion – Israel Institute of Technology, Haifa, Israel
bennyk@cs.technion.ac.il

─── **Abstract** ───

Probabilistic databases are commonly known in the form of the tuple-independent model, where the validity of every tuple is an independent random event. Conceptually, the notion is more general, as a probabilistic database refers to any probability distribution over ordinary databases. A central computational problem is that of marginal inference for database queries: what is the probability that a given tuple is a query answer? In this talk, I will discuss recent developments in several research directions that, collectively, position probabilistic databases as the common and natural foundation of various challenges at the core of data analytics. Examples include reasoning about uncertain preferences from conventional distributions such as the Mallows model, data cleaning and repairing in probabilistic paradigms such as the HoloClean system, and the explanation of query answers through concepts from cooperative game theory such as the Shapley value and the Banzhaf Power Index. While these challenges manifest different facets of probabilistic databases, I will show how they interrelate and, moreover, how they relate to the basic theory of inference over tuple-independent databases.