Repairing Noisy Graphs

Divesh Srivastava
AT&T Labs-Research
divesh@research.att.com

ABSTRACT

Graphs are a flexible way to represent data in a variety of applications, with nodes representing domain-specific entities (e.g., records in record linkage, products and types in an ontology) and edges capturing a variety of relationships between these entities (e.g., an equivalence relationship between records in record linkage, a type-subtype relationship between types in an ontology). Often, the edges in this graph are noisy, in that some edges are missing (i.e., real-world relationships that do not have corresponding edges in the graph) and some edges are spurious (i.e., edges in the graph that do not have corresponding real-world relationships). Directly analyzing noisy graphs can lead to undesirable outcomes, making it important to repair noisy graphs. In this talk, we describe an approach that takes advantage of properties of real-world relationships and their estimated probabilities to ask oracle queries (an abstraction of crowdsourcing) to efficiently repair the noisy graphs. We illustrate this approach for the case of graphs that are unions of cliques (which is the case for record linkage) and graphs that are trees (which is the case for ontologies), and present theoretical and empirical results for these cases.

This is joint work with Donatella Firmani, Sainyam Galhotra and Barna Saha.

About the Speaker

Divesh Srivastava is the head of Database Research at AT&T Labs-Research. He is a Fellow of the Association for Computing Machinery (ACM) and the managing editor of the Proceedings of the VLDB Endowment (PVLDB). He has served as a trustee of the VLDB Endowment, as an associate editor of the ACM Transactions on Database Systems (TODS), and as an associate Editor-in-Chief of the IEEE Transactions on Knowledge and Data Engineering (TKDE). He has presented keynote talks at several international conferences, and his research interests and publications span a variety of topics in data management. He received his Ph.D. from the University of Wisconsin, Madison, USA, and his Bachelor of Technology from the Indian Institute of Technology, Bombay, India.