

STANFORD UNIVERSITY
DEPARTMENT OF STATISTICS
DEPARTMENT SEMINAR

4:15 p.m., Tuesday, June 4, 2002
Sequoia Hall Room 200
(Cookies at 3:45 in 1st Floor Lounge)

Daphne Koller
Stanford University

Learning Probabilistic Models from Relational Data

Bayesian networks are a compact and natural representation for complex probabilistic models. They use graphical notation to encode domain structure: the direct probabilistic dependencies between variables in the domain. However, many real-world domains are best described by relational in which instances of multiple types are related to each other in complex ways. For example, in a scientific paper domain, papers are related to each other via citation, and are also related to their authors. Bayesian networks are attribute-based, making it difficult to represent the rich relational structure of complex domains involving multiple entities that interact with each other. The talk will describe probabilistic relational models (PRMs), a new probabilistic modeling language suitable for relational domains. PRMs extend the language of Bayesian networks with the expressive power of object-relational languages. They model the uncertainty over the attributes of objects in the domain as well as uncertainty over the existence of relations between objects. The talk will present techniques for automatically inducing PRMs directly from a relational data set, and applications of these techniques to pattern discovery, including relational classification and cluster. It will demonstrate the applicability of the techniques on complex real-world data sets, including web data and genomic data.

Joint work with Pieter Abbeel, Nir Friedman, Lise Getoor, Avi Pfeffer, Eran Segal, and Ben Taskar.