Latent Variable Models of Social Networks and Text with Application to Political, Legal and Bibliometric Data

With Andrew McCallum

Friday, December 1st, 2006

Seminar: 11:00 am – 12:00 pm
Lunch: 12:00 pm – 1:00 pm
Thompson Hall, Room 620

Abstract: The field of social network analysis studies mathematical models of patterns in the interactions between people or other entities. In this talk I will present several recent advances in generative, probabilistic modeling of networks and their per-edge attributes. The Author-Recipient-Topic model discovers role-similarity between entities by examining not only network connectivity, but also the words communicated on those edges; I'll demonstrate this method on a large corpus of email data subpoenaed as part of the Enron investigation. The Group-Topic model discovers groups of entities and the "topical" conditions under which different groupings arise; I'll demonstrate this on coalition discovery from many years worth of voting records in the U.S. Senate and the U.N. I'll conclude with further examples of Bayesian networks successfully applied to relational data, as well as discussion of their applicability to trend analysis, expert-finding and bibliometrics.

Andrew McCallum is an Associate Professor of Computer Science at The University of Massachusetts Amherst

Part of the National Center for Digital Government’s Seminar Series.
www.ncdg.org