



Statistical Sciences
UNIVERSITY OF TORONTO

SEMINAR

December 1, 2016 at 3:30pm

Refreshments will be provided at 3:15pm

Sidney Smith Hall, Room 2108

Speaker: Barbara Engelhardt, Princeton University

Host: Stanislav Volgushev

Structured Factor Models to Find Interpretable Signal In Genomic Data

Latent factor models have been the recent focus of much attention in 'big data' applications because of their ability to quickly allow the user to explore the underlying data in a controlled and interpretable way. In genomics, latent factor models are commonly used to identify population substructure, identify gene clusters, and control noise in large data sets. In this talk I present a general framework for Bayesian structured latent factor models. I will illustrate the power of these models for a broad class of problems in genomics via application to the Genotype-tissue Expression (GTEx) data set. In particular, by using a Bayesian biclustering version of this model, the estimated latent structure may be used to identify gene co-expression networks that co-vary uniquely in one tissue

type (and other conditions). We validate network edges using tissue-specific expression quantitative trait loci.