



STATISTICS
COLORADO STATE UNIVERSITY

Random Partition Models using Prior Information

Spring 2026 Departmental Seminar

Thursday, Sept 17, 2026

9:00 am

Engineering Rm E105

Dr. Sally Paganin

Assistant Professor, Statistics
The Ohio State University



Abstract

A rich literature on Bayesian approaches to clustering is built on random partition models, most commonly through Exchangeable Partition Probability Functions (EPPF) and their Gibbs-type generalizations. In many applications, however, we have concrete information (i.e., expert knowledge, covariates) that we wish to incorporate into the model. In this talk, I will introduce two constructions that refine random partition models to better incorporate such information.

The first considers a prior guess on the partition, such as one based on domain knowledge from experts or earlier studies. We propose a partition model that is linked to the prior guess via auxiliary variables. For each data point, these variables determine whether it stays in its initial cluster or is reassigned under a specified random partition model. Our informed partition model provides flexibility to include varying levels of uncertainty to any subset of the partition (i.e., locally weighted prior information). The second construction incorporates covariate information. Existing methods, such as product partition models with covariates (PPMx), incorporate covariates into Bayesian random partition models, but calibrating how much the covariates should matter is not straightforward: overweighting them can override structure supported by the data, while underweighting wastes useful information. We introduce a calibration mechanism that identifies important covariates and tunes their effect on clustering. We demonstrate this with network data, where node covariates can inform community detection.