



統計科學研究所

INSTITUTE OF  
STATISTICAL SCIENCE



S.E.M.I.N.A.R.



S.T.A.T.A.S.

## Postdoc Seminar

Title : Bayesian graph clustering with random spanning trees

Speaker : Mr. Changwoo Lee

(Department of Statistics, Texas A&M University)

Time : 10:30 ~11:30, Tuesday, July 25, 2023

Place : Auditorium, B1F, Institute of Statistical Science, AS

### Abstract

Clustering is one of the most important topics in statistics and data science. In this talk, I will give a general overview of probabilistic clustering, which can provide uncertainty quantification of the clustering estimate based on Bayesian inference. In the first half of the talk, I will introduce random partition models and motivate with several popular application examples such as mixture models, topic models, and stochastic block models. In particular, I will discuss how the choice of random partition prior affects the clustering results. While previous studies primarily focused on the induced number of clusters, I introduce a concept of balancedness which has been largely neglected in the literature. It provides a better understanding of the different behaviors of the probabilistic clustering results and provides insights to practitioners to decide on a suitable model according to specific applications. The second half of the talk will be devoted to probabilistic clustering methods and algorithms with random spanning trees. Data collected from the complex constrained domain, such as geospatial/traffic data, demand clustering methods that respect its own nontrivial geometries. By considering an undirected graph that reflects its structural information, I introduce probabilistic clustering methods and algorithms using random spanning trees that provide contiguous clustering estimates along its uncertainty quantification.

※ Tea reception starts at 10 : 10.

※ Lecture in Mandarin. Online live streaming through Cisco Webex will be available.