



Maude Menten Institute and Mathematical Biology Seminar

Monday, November 3, 2025



1 pm MT – Room 4-450 UComm (in person)

Join Zoom Meeting

<https://ualberta-ca.zoom.us/j/97624718507>

Meeting ID: 976 2471 8507



Khanh Dao Duc

Assistant Professor

Department of Mathematics

University of British Columbia

Building new metrics for analyzing large biological shape data

Recent advances in experimental methodologies and community efforts led to a surge in large and heterogeneous biological datasets across all scales, that require the developments of new methods to extract meaningful information. In this context, I will describe our recent efforts to leverage optimal transport theory with the introduction of new metrics and algorithms, to compare data points in high dimensional spaces with applications in structural, molecular, and cell biology. After motivating these methods and briefly introducing the concept of Wasserstein distance, I'll introduce two new frameworks. To quantify heterogeneity arising from large collections of 2D or 3D cell shapes, we define the stratified Wasserstein kernel, which embeds shape data in Euclidean space via ranked local distance profiles. This embedding yields an isometry-invariant Euclidean distance and a positive-definite kernel for population analysis, with a consistent sample-based estimator that supports large datasets in near-quadratic time. By leveraging kernel methods, the framework enables statistically rigorous tasks such as nonparametric hypothesis testing, providing theoretical guarantees as well as interpretability. Second, to improve and automate the fitting of protein subunits into large complexes imaged from cryo-EM, we formulate a new objective function -called Joint Gromov Wasserstein (JGW)-, which extends the mathematical concepts underlying the Gromov-Wasserstein objective, such as metric measure spaces and isomorphisms, to handle collections of objects. We prove theoretical properties of the JGW objective, analyzing its metric properties and asymptotic behavior from point sampling, and adapt existing approximation techniques to produce feasible algorithms to compute it.