GABRIEL RIOUX, Purdue University

Comparing Heterogeneous Data: an Optimal Transport Approach

Modern machine learning pipelines must handle heterogeneous datasets that differ in modality (e.g., images and text) or representation (e.g., when no canonical representation exists or representations vary across sources). Optimal transport (OT) provides a powerful geometric framework for comparing and transforming datasets, but requires a cost function that captures the desired correspondence between data points. For complex heterogeneous data, this cost is often chosen ad hoc and may fail to capture meaningful similarities.

Gromov–Wasserstein (GW) problems provide a quadratic generalization of OT that can naturally handle distributions on different spaces and align datasets, such as graphs, that lack a canonical representation. However, GW generally leads to a nonconvex quadratic optimization problem, complicating its statistical and computational analysis.

This talk will describe recent progress on the structural, statistical, and computational analysis of GW. The key result enabling these developments is a new representation of GW as the optimal value of an optimization problem whose objective depends on the input distributions only through a parameterized OT problem.