Title: Continuum Limits of Semi-Supervised Learning on Graphs

Given a data set $X_n=\{x_i\}_{i=1}^n$ and a subset of training labels $\{y_i\}_{i \in Z_n}$ where $Z_n \subseteq \{1, \ldots, n\}$ the goal of semi-supervised is to infer labels on the unlabelled data points $\{x_i\}_{i \not\in Z_n}$. In this talk we use a random geometric graph model with connection radius $\epsilon_n$. The framework is to consider objective functionals which reward the regularity of the estimated labels and impose or reward the agreement with the training data, more specifically we will consider discrete $p$-Laplacian regularization.

The talk concerns the asymptotic behaviour in the limit where the number of unlabelled points increases while the number of training labels becomes asymptotically small. The results are to uncover a delicate interplay between the regularizing nature of the functionals considered and the nonlocality inherent to the graph constructions. To establish asymptotic consistency we make use of a discrete-to-continuum topology that is based on optimal transport and variational methods such as $\Gamma$-convergence. I will give almost optimal ranges on the scaling of $\epsilon_n$ for asymptotic consistency to hold.

This is joint work with Jeff Calder (Minnesota) and Dejan Slepcev (CMU).