



THE UNIVERSITY OF CHICAGO

Department of Statistics

DISSERTATION PRESENTATION AND DEFENSE

JONATHAN ESKREIS-WINKLER

Department of Statistics
The University of Chicago

MULTIRESOLUTION ANALYSIS ON DISCRETE SPACES

WEDNESDAY, December 04, 2019, at 1:00 PM
Jones 303, 5747 S. Ellis Avenue

ABSTRACT

It continues to be much cheaper to store data than to analyze it. This state of data analysis motivates methods that make minimal assumptions in order to reduce the complexity of data. In order to address scalability in certain applied, network-based problems, we bring together three distinct fields: multiresolution analysis (MRA), hyperbolic embeddings, and community detection. Together, the work in these fields provides a path forward for intractable problems in large data analysis. We provide a theoretical background, a description of our contributions, and a number of applications to show the viability of our ideas in a real-world setting.

The primary contributions of this dissertation are threefold. First, we draw a connection between a class of currently-used methods in computational MRA to hyperbolic representations of data. This connection allows us to suggest a rationale for when different MRA methods are most appropriate, and to define related tree-based kernels for wavelet construction.

Second, we broaden the existing work describing the mechanics of how the multiresolution matrix factorization (MMF) summarizes data. We look at this MRA framework from two perspectives. First, we consider how MMF may be viewed as a hyperbolic embedding. Second, we consider the way MMF may be seen as a regularization operator. We show that for certain graphs, the regularization imposed by MMF enables methods to perform more accurate inference than when no regularization is used. This work fits within the existing field of high dimensional statistics where regularization may turn a massive problem into a manageable one.

Third, we show the way that MMF, and other similar multiresolution methods, may be used efficiently in unsupervised and semi-supervised settings. By taking advantage of the localization of wavelets in time and frequency space of graphs, we are able to detect complex, multiscale, overlapping community structures, as well as combine network structure with the small amounts of labeling for semi-supervised learning.