1106-60-2402 **Karl Rohe*** (karlrohe@stat.wisc.edu), **Mohammad Khabbazian** and **Haoyang Liu**.

Statistical properties of average linkage hierarchical clustering under the Stochastic Blockmodel.

Agglomerative hierarchical clustering outputs a dendrogram that can estimate hierarchical structure (e.g. Clauset-Moore-Newman 2008, Ahn-Bagrow-Lehmann 2010). It was first proposed in Anderson 1962. Various forms of this algorithm have found wide applicability across various domains. The algorithm is agglomerative; at the first step, every data point is a unique set and at every successive step, the two closest sets are joined together as a new set. The algorithm stops when there is only one set. This requires (1) pairwise measurements between all points (either similarities or distances) and (2) a "linkage" that combine these pairwise measurements into measurements between sets of points.

Taking the arithmetic mean of pairwise measurements in (2) yields one of the most popular forms of this algorithm called "average linkage." Despite its popularity, this form of the algorithm has no known statistical estimation results. This talk will relate average linkage to spectral graph theory. Using the spectral properties of the Stochastic Blockmodel, we show that this classical algorithm is weakly consistent under the Stochastic Blockmodel with pairwise similarity "# of common neighbors". We provide a fast implementation for sparse graphs. (Received September 16, 2014)