

# Smoothing Dissimilarities for Cluster Analysis: Binary Data and Functional Data

David Hitchcock

University of South Carolina

*E-Mail:* hitchcock@stat.sc.edu

*Abstract:* Cluster analysis attempts to group data objects into homogeneous clusters, often on the basis of the pairwise dissimilarities among the objects. When the data contain noise, we might consider performing a smoothing operation, either on the data themselves or on the dissimilarities, before implementing the clustering algorithm. Possible benefits to such pre-smoothing are discussed in the context of binary data, and we also make connections to previous research on smoothing functional data for cluster analysis. We suggest a method for cluster analysis of binary data based on “smoothed” dissimilarities. The smoothing method presented borrows ideas from shrinkage estimation of cell probabilities. The method is illustrated with an example involving binary item response data.