

Prof. John Stufken

(Director, Informatics and Analytics, Bank of America Excellence Professor,
University of North Carolina Greensboro)

Title:

“Subdata Selection Methods”

Abstract:

These days, the size of data can be enormous. This can create challenges with the storage of the data and, even more so, its analysis. Computations of even the simplest statistics can be challenging. One method to deal with these challenges is based on selecting only some of the data (subdata) from the original full data set. Exploration or analysis would proceed with the subdata. Subdata can be selected through a sampling strategy or through a deterministic method that attempts to optimize a specific criterion. How should the subdata be selected? How much do we lose by only analyzing the subdata? In this talk I will briefly discuss different subdata selection methods.