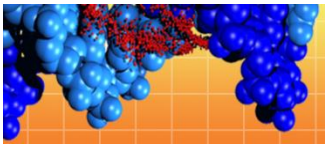

CENTER FOR BIOINFORMATICS & COMPUTATIONAL BIOLOGY

BIOINFORMATICS 2016 FALL SEMINAR SERIES

Hosted by: Department of Computer and Information Sciences,
Department of Electrical and Computer Engineering &
Center for Bioinformatics and Computational Biology
http://bioinformatics.udel.edu/Seminars/Current

MONDAY, September 19, 2016
3:30pm
DBI Room 102

Development of a Scalable Method for Creating Food Groups Using the NHANES Dataset and MapReduce

Michael R. Wyatt II

PhD Student – Computer & Information Sciences

ABSTRACT: In this talk we tackle the need for meaningful food group classifications in dietary datasets such as the National Health and Nutrition Examination Survey (NAHNES) that are less subjective in nature by defining a new objective method of identifying food groups exclusively based on the food's micro- and macro-nutrient content. We first perform extensive preprocessing of the NHANES raw data to mitigate impacts of missing nutrient values, redundancies, and different food intake quantities and scales. We then utilize an unsupervised learning clustering algorithm to create food groups within the preprocessed NHANES data and identify food groups with similar nutrient content. Finally we parallelize our method to benefit from the scalable MapReduce paradigm. Our results show that our method identity food groups with smaller diameter and larger cluster separation distances than the standard, expert-informed, method of grouping food items.