Biomedical and Health Informatics Lecture Series
Course Website: Link

Thursday, October 25, 2012
1:30 – 2:20 p.m.
Health Sciences, Room T-359

SPECIAL GUEST PRESENTER:
Eithon Cadag, PhD
Ayasdi Inc.

"Exploration and analysis of complex data using topological data analysis"

The health and life science domains are producing data in larger quantities and with increasing complexity; the task of analyzing the information to obtain actionable knowledge is rapidly becoming a fundamental challenge of our time. Model- and query-based algorithms are useful in systems where complexity is limited, but real-world data are often non-linear and with a high degree of complexity. We make a case for a new and radical approach to analyze complex data problems. Our approach, Topological Data Analysis (TDA), uses higher dimensional versions of distance functions to understand data through shape, suitably defined, and is grounded in the mathematical branch of study called topology. Until recently, this branch of mathematics has only been used to study abstract shapes and surfaces. However, over the last 15 years, there has been a concerted effort to utilize the ability of topological methods to extract patterns to study large and high-dimensional data sets. To date, we have applied TDA to data collected from high-throughput compound screenings, flow cytometry, mass spectrometry, genomics and proteomics platforms, personal health devices, electronic medical records, text and many other diverse sources. We are able to identify sub-populations among cancer patients for better biomarker selection, uncover insights into disease biology and pinpoint outliers and anomalies. TDA is a new approach capable of tackling data that are both large and complex, and we have built software that allows researchers to access these topological algorithms in real time through an interactive user interface.

Eithon Cadag graduated from the BHI Graduate Program in the winter quarter of 2009. He recently started a new job at a small start-up, Ayasdi, Inc. and has been working on using applied topology to analyze all sorts of data, from biological to energy to financial, etc. More recently, he has been working with sequencing and exome mutation data in conjunction with clinical outcomes.