# Department of Mathematics, Statistics and Computer Science
# St. Francis Xavier University
# Presents

## Data Mining in Drug Discovery: Clustering and Recursive Partitioning

### by

Dr. Xu (Sunny) Wang
Post-Doctoral Fellow
Department of Statistics and Actuarial Science
University of Waterloo

March 16<sup>th</sup>, 2009 @ 2:15 in Annex 23a

In the first part of this talk, I will introduce two data mining methods: Clustering and Recursive Partitioning. The second part of this talk, a sequential screening approach, Cluster Structure-Activity Relationship Analysis (CSARA), is explored. Sequential screening integrates High Throughput Screening with mathematical modeling to sequentially select the best compounds as drug candidates. CSARA is a cluster-based and algorithm driven method. To gain further insight into this method, we use three usefully designed experiments to compare predictive accuracy with Recursive Partitioning. The experiments show that CSARA outperforms Recursive Partitioning. Comparisons include problems with many descriptor sets and situations in which many descriptors are not important for biological activity.

Refreshments will be served before the talk in AX24A