**COSC 757 Data Mining Assignment 3**

Due date: 04/13/2016 11:59 pm

Instructions: This is an individual assignment. Use Blackboard to submit your answers on the due date (no hard copies please). Late submissions will receive a zero grade.

**Experimentation with Clustering:**  Select a dataset for clustering. To save you some time, this can be one of the datasets that you have worked with in past assignments. The specific requirements for the assignment are as follows:

* Choose a dataset that is of interest to you and is well suited for clustering
* Give a brief description of the dataset
* Test at least 3 clustering algorithms using the software of your choice.
* Use a cluster quality metric to assess the clustering quality of each algorithm. The method can be intrinsic such as the silhouette coefficient or if your data is labeled, you can use the extrinsic methods presented in Chapter 10.6.3 of the Han book.
* Compare the results from each algorithm.
* Write a report that describes your experiment and results. The report should be in either ACM or IEEE conference paper format and should include an introductory section that details the dataset and the objectives of the analysis, a methodology section that explains the approach that you are using to mine the dataset including the algorithms and parameters (e.g. confidence and support) as well as any steps that you had to take to preprocess the data, a results section that shows the results of your analysis and any interesting patterns that you found, and a conclusion section that summarizes your results and discusses the limitations of your approach and any difficulties that you had with your experiment.
  + Links to format templates:
  + <http://www.ieee.org/conferences_events/conferences/publishing/templates.html>
  + <http://www.acm.org/sigs/publications/proceedings-templates>