COSC 4335 *“Data Mining”*

Review Sheet for 2018 Final Exam

The final exam has been scheduled for Tu., May 8, 11a in 110 AH and will be “open everything” and will take about 115 minutes. The use of computers is not allowed! The exam counts approx. 20% towards the final course grade. The exam will cover the following topics:

1. \*\* Hierarchical Clustering
   1. Transparencies
   2. Book pages 515-523, 524-526 (Key Issues in Hierarchical Clustering Section)
2. \*\*\*\*\*\* Classification Techniques
   1. all transparencies covered in the lecture which cover decision trees, overfitting, k-nearest neighbors, ensembles, neural networks and n-fold cross-validation.
   2. book pages: 147-165, 168-183 186-187, 223-227, 251-256, 276-280, 285-290
3. \*\*\*\* Association Analysis (just centering on association rule mining and interestingness measures such as confidence, support and lift)
   1. book pages: 327-341, 349-353, 370-376 and all transparencies that are associated with those book pages)
4. \*\* PageRank (centering on the algorithm itself)
   1. Class transparencies excluding *Gleich Dissertation transparencies*
   2. Wikipedia PageRank page

Comment: will only ask very basic questions about this topic, but you should know the PageRank algorithm well!

1. \*\*\*\*\* Outlier Detection
   1. Transparencies covered in the lecture
   2. Textbook pages 651-662; 666-670 (subsection 10.2.3 is **not** relevant for the final exam!)
2. \* Preprocessing
   1. Preprocessing transparencies
   2. Textbook pages 44-59
3. \* First 17 slides of the Data Science Talk (will only ask basic questions)

Algorithms and techniques you should know well include: decision tree induction, Adaboost, how k-NN classifiers work; APRIORI algorithm for association rule mining, k-nearest neighbor based and model-based outlier detection.