Questions 03-28-2018

A Comparison of Graph-Based Synthetic Data Generators for Benchmarking Next-Generation Intrusion Detection Systems:

This article discusses generating synthetic data using graph-based algorithms and measuring the performances of these algorithms both in terms of scalability of data generation and accuracy of data generation to original data used as the seed of the graph. This sets a benchmark to which future studies can compare for the detection of intrusions in cyber security systems. What other algorithms might perform better or worse? What methods of intrusion detection might beat this benchmark? Which ones might not?

Online Machine Learning in Big Data Streams:

This article discusses machine learning of big data that is "in motion." It talks about some considerations regarding fast data as well as some common tools, methods of classification, machine learning algorithms, and evaluations of recommenders. With this article as a reference, how can we compare this to some of the other article we have read that discuss batch processing, or a combination of batch and stream processing? What are the advantages and disadvantages of primarily stream processing versus primarily batch processing or a combination of both?