Bedshift: Simulating Permutations of Genomic Interval Sets
(Technical Report)

The Choices of High Frequency Traders in the Flash Crash of 2010
(STS Research Paper)

An Undergraduate Thesis Portfolio

Presented to the Faculty of the
School of Engineering and Applied Science
University of Virginia, Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Computer Science

By

Aaron Gu

May 6, 2020

Socio-technical Synthesis: Genomic Region Similarity and the Flash Crash of 2010

My capstone research project focused on genomic regions, which are areas of interest along the DNA such as genes, promoters, and enhancers. My STS project is in a completely different sector, finance, where I research the ethics of high frequency traders in the stock market flash crash of 2010. Although these two areas are not related, working on these two projects simultaneously allowed me to give more thought into the ethics around finding similar genomic regions from different experiments and patients, as well as the technical details behind the flash crash of 2010.

My capstone research project was about assessing the similarity scoring methods of genomic region sets, which are comprised of the significant areas of the genome mentioned before. Different experiments produce different region sets, and researchers use similarity scoring methods to find other region sets similar to theirs, to see if there is any significance among the regions produced. I created a tool called bedshift that could create a testing framework for the many different similarity scoring methods out there. Bedshift can randomly alter a region set to add, drop, or shift regions, and the resulting region set can then be scored against its original region set to see how well the similarity scoring method reflects the actual change specified in bedshift.

My STS research project was about the ethics of high frequency traders in the flash crash of 2010, an incident during which the S&P500 stock index dropped 6% in five minutes, only to recover just as quickly. High frequency traders first contributed to the crash by rapidly selling the same asset over and over again, then made it worse by withdrawing their algorithms from the market, making many stocks mis-priced. I argue that the high frequency traders did not consider

other market participants before committing their actions, leading to the flash crash that would see some trades being made 60% or more away from the normal value and causing financial loss among those who made those trades. I use the framework of utilitarianism to show that this financial loss affected many more people than if the high frequency traders did not withdraw their algorithms. In addition, their inability to anticipate this kind of market activity and code their algorithms to prevent the rapid sell-off was also unethical.

My STS work has allowed me to consider some ethical viewpoints on my research that I had not considered before. For example, what are implications of region set similarity scoring if used outside of lab experiments on patient data? If similarity scoring is discovered to be inaccurate, then should the patient believe the result from the test? Similarly, my technical work has allowed me to see from the point of view of a coder in the high frequency trading companies. Was it possible to predict a flash crash, and if so was it possible to code an algorithm to prevent the flash crash? This played an important role in my ethical consideration of their decision. Throughout both projects, I have learned to consider both ethical and technical viewpoints in tandem.

Table of Contents