Parallel Solution for Near Repeat

KENNETH MCKANDERS (MOREHOUSE COLLEGE)

MENTOR: DR. HAIHANG YOU

Introduction: What is a Near Repeat?

- A Near Repeat occurs when two or more elements in a system can be related through a set of rules
- Rules can be anything such as a set distance, difference in time, etc.
- Right now, we can consider anything to be an element, such as crimes or cells in the body
- The overall goal is to relate elements to each other
  - derive a definite path between the elements to figure out a starting point
  - Possibly derive elements that can occur after this chain!

Research: Exploring Near Repeat Occurrences

Traditionally, near-repeats are found by comparing each event to every other event in the series.

The complexity for this is $O\left(\frac{n(n-1)}{2}\right)$.

The results of these comparisons are a list of values showing:

- The amount of near-repeats
- The amount of events inside the specified distance, but not the specified time
- The amount of events inside the specified time, but not the specified distance
- The amount of events outside the specified time and distance

Current Progress

Completed functions which can calculate distance and time

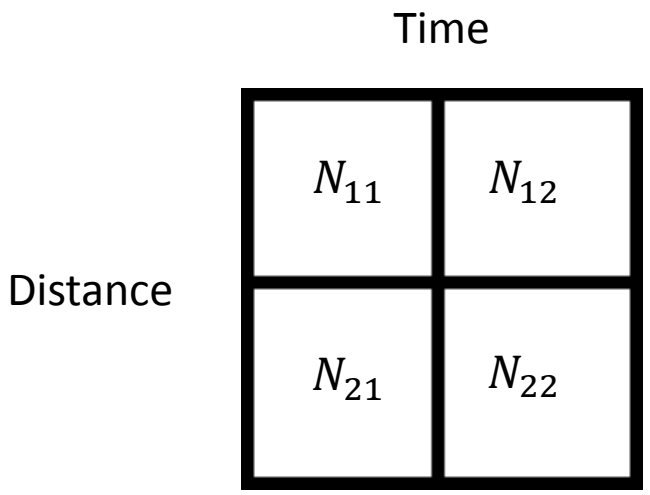
Completed relation generating function

Started testing on large data sets

Future Work

My current goal is to display the following table:

- $N_{11} = |\{(i,j)|d(i,j) \leq d \; and \; t(i,j) \leq t\}|$
- $N_{12} = |\{(i,j)|d(i,j) \leq d \; and \; t(i,j) > t\}|$
- $N_{21} = |\{(i,j)|d(i,j) > d \; and \; t(i,j) \leq t\}|$
- $N_{22} = |\{(i,j)|d(i,j) > d \; and \; t(i,j) > t\}|$



Future Work cont.

Run simulations on Kraken

- Allows for faster table population
- Allows for larger test files

Questions?