CS 5200 Final Project

Avah Banerjee

Due on: April 26, 11:59 PM

Project Task List

Part-1: Implement the MIS algorithm for interval graphs given in [1].

(a) Before submitting, test your code on the Linux machines in the campus. For instructions on how to use them see – `https://it.mst.edu/services/linux/`.

(b) Permissible languages - `C,C++, Java, Python`.

(c) Your code should be able to take a `.csv` file as an input (from the current directory). If the input has n intervals, then this `.csv` file would contain $2n$ values, where the values at location $2i-1$ and $2i$ ($i \geq 1$) represents the left and the right endpoints of the i^{th} interval. You may assume the name of this `.csv` file is `input.csv`.

(d) You should output your computed MIS as a `.csv` file named `output-#.csv`, where the symbol `#` is a placeholder for the number of intervals in your solution. The output file will contains the set of intervals in MIS in the same format as the input file.

Part-2: Empirically determine the expected size of the MIS for a t-random set of n intervals for different values of t and n. A t-random set ($t \in (0,1]$) of n intervals is defined in the following way. Each interval $[l_i, r_i] \subseteq [0,1]$ and $r_i - l_i \leq t$. First, pick n points independently and uniformly at random from $[0,1)$. These points will constitute the left endpoints $l_1, l_2, l_3, \ldots, l_n$ of the intervals. Then for each i from 1 to n, pick a value v_i in $[0,t]$ uniformly at random and create the interval $[l_i, r_i = \min(l_i + v_i, 1)]$.

(a) Set of t values to used based on n - $\{\frac{1}{n}, \frac{1}{\sqrt{n}}, \frac{1}{\log n}, \frac{1}{4}\}$.

(b) For each t-value, determine the expected MIS as a function of n. You need to computed the expected MIS (denoted by $E(t,n)$) on a sufficient number of distinct n values over a broad range. From which try to infer the asymptotic behavior of $E(t,n)$ as $n \to \infty$.

(c) For each specific t and n, estimate $E(t,n)$ over at least 10 different input instances.

(d) There will be a total of four plots (inside the same figure), one for each t value. The x-axis will indicate the number of intervals (n) and the y-axis will indicate the expected MIS (averaged over at least 10 runs).

(e) You should submit a brief report (no more than 3 pages) describing your understanding of the divide and conquer MIS algorithm and the four plots. The report should in a PDF document.

Important

1. Max group size is 2.
2. All artifacts should be uploaded as a single `.zip` file on Canvas.
3. Late submissions will not be accepted.

References

[1] Snoeyink, J. (2007). Maximum independent set for intervals by divide and conquer with pruning. Networks: An International Journal, 49(2), 158-159.