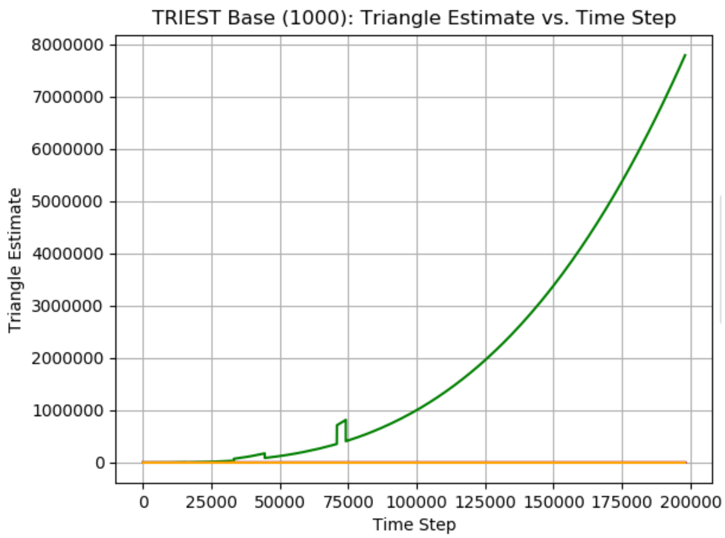
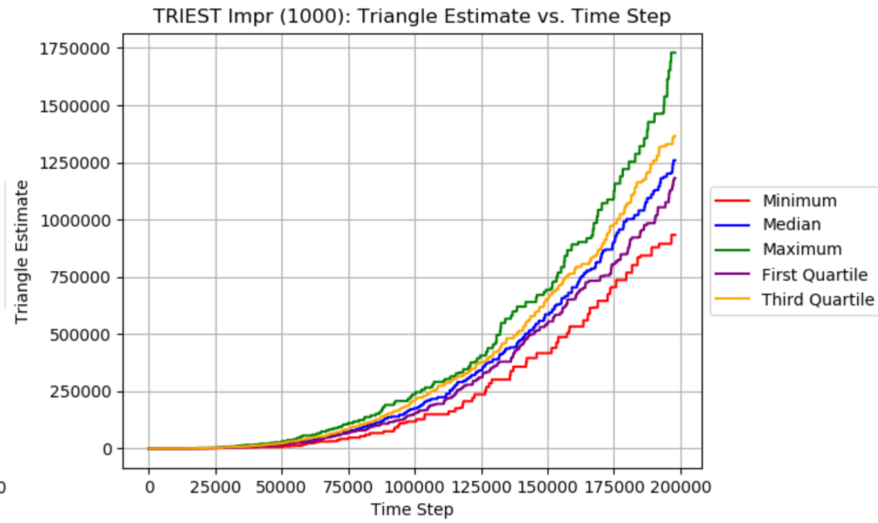
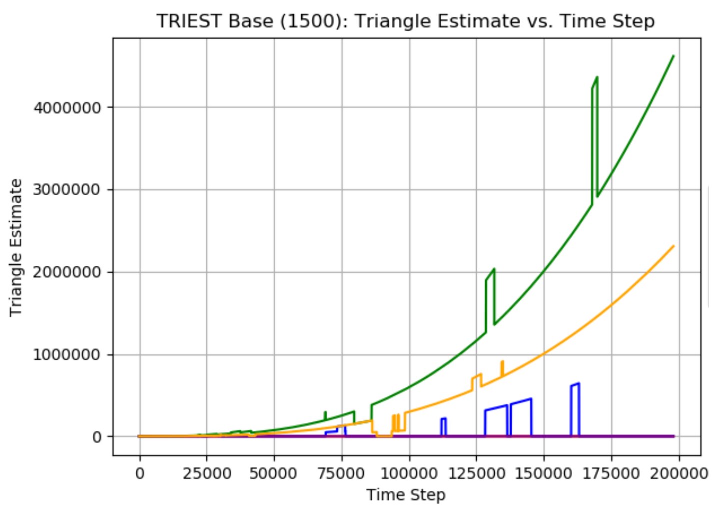
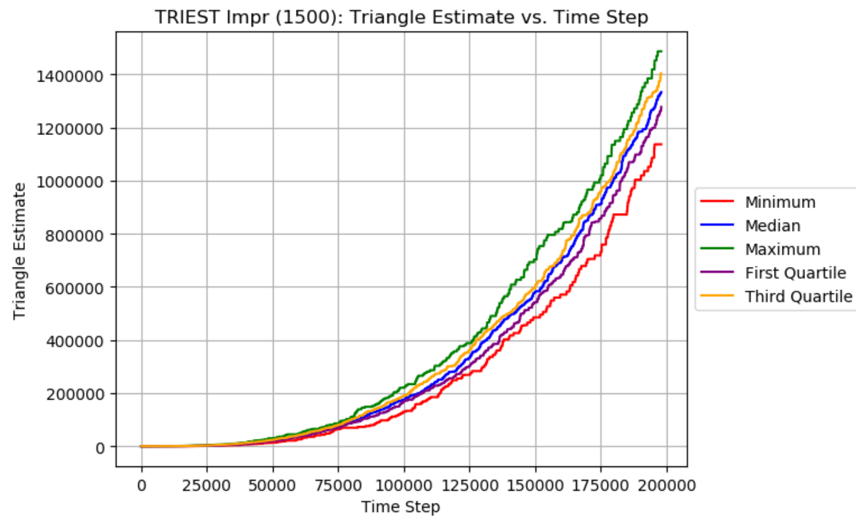
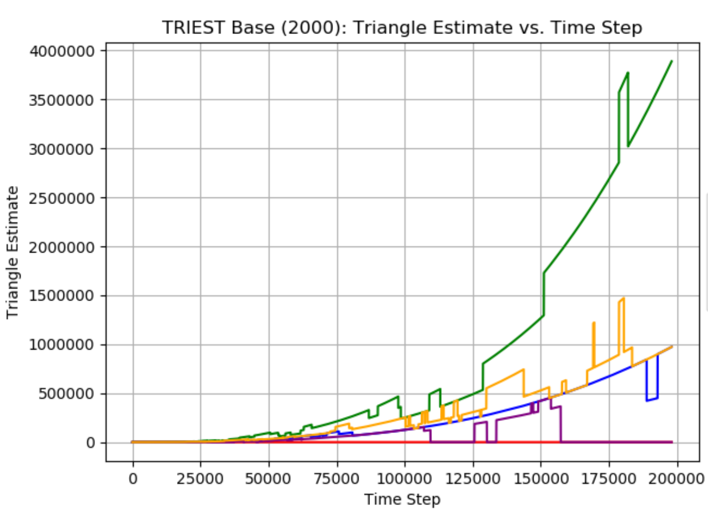
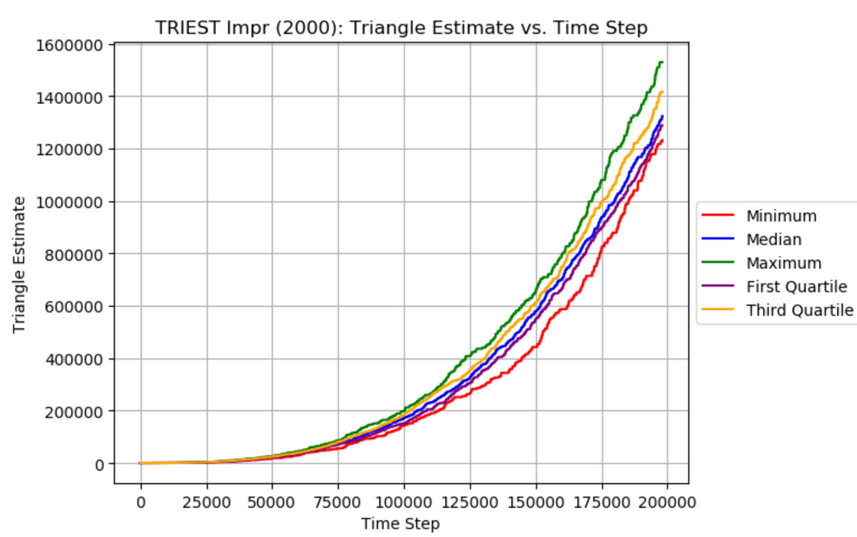

TRIEST Base (1000): Triangle Estimate vs. Time Step
Triangle Estimate
Time Step

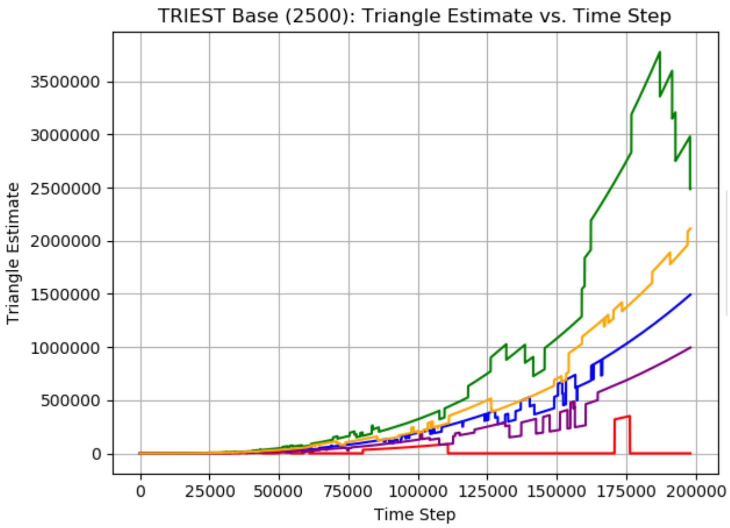

TRIEST Impr (1000): Triangle Estimate vs. Time Step
Triangle Estimate
Time Step
Minimum
Median
Maximum
First Quartile
Third Quartile

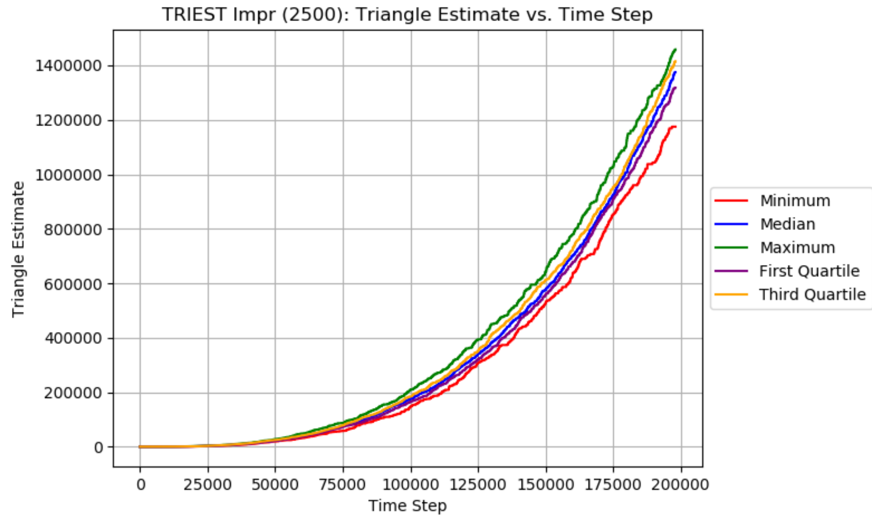

TRIEST Base (1500): Triangle Estimate vs. Time Step
Triangle Estimate
Time Step

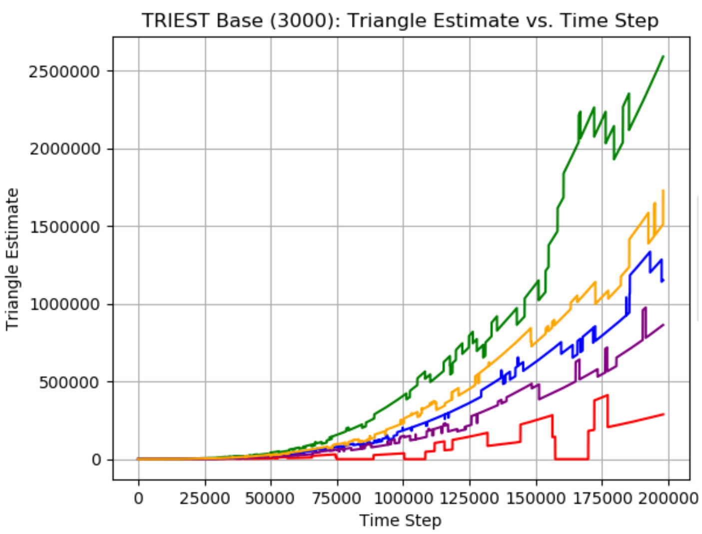

TRIEST Impr (1500): Triangle Estimate vs. Time Step
Triangle Estimate
Time Step
Minimum
Median
Maximum
First Quartile
Third Quartile

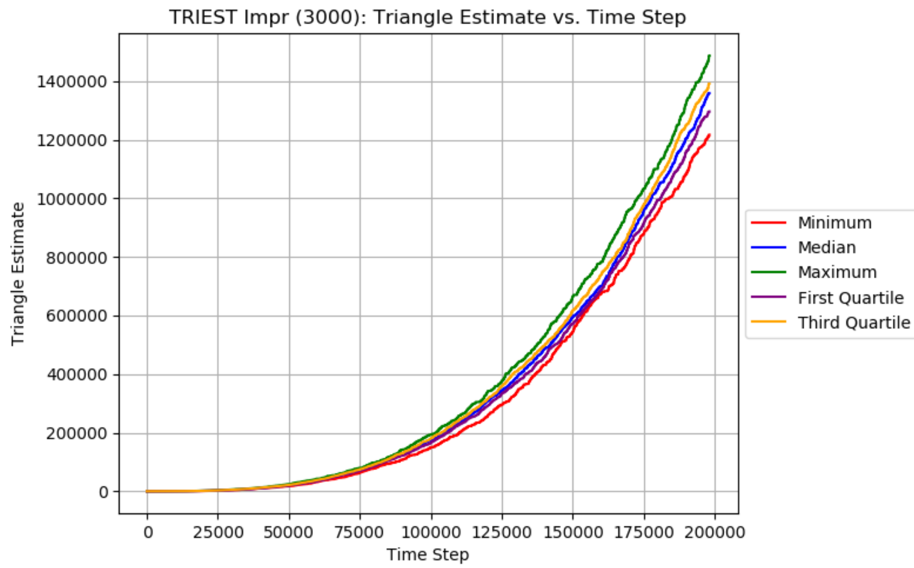

TRIEST Base (2000): Triangle Estimate vs. Time Step
Triangle Estimate
Time Step

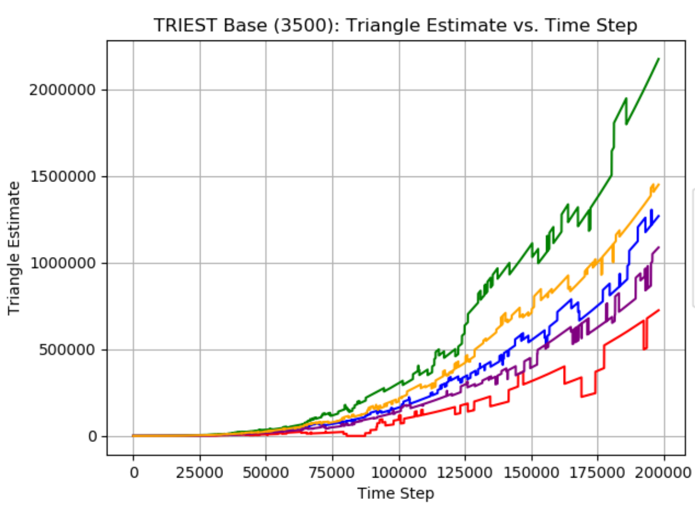

TRIEST Impr (2000): Triangle Estimate vs. Time Step
Triangle Estimate
Time Step
Minimum
Median
Maximum
First Quartile
Third Quartile

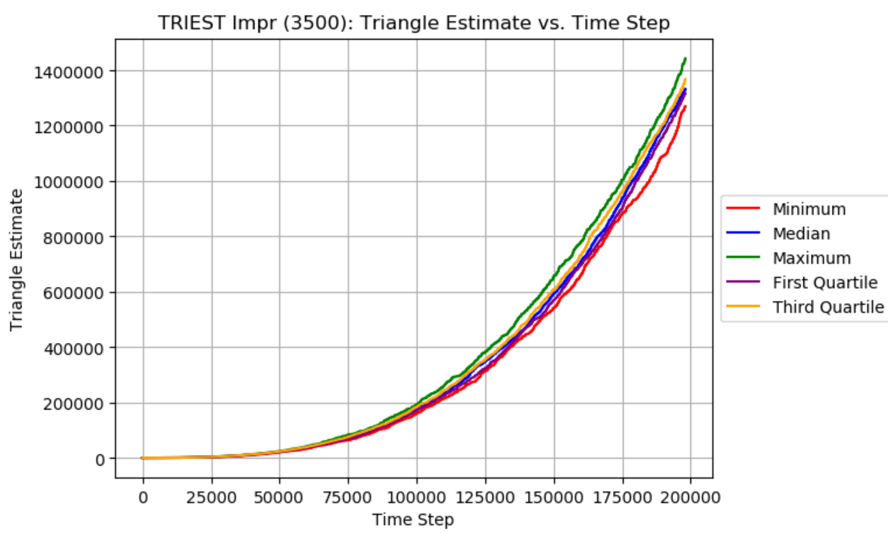

TRIEST Base (2500): Triangle Estimate vs. Time Step
Triangle Estimate
Time Step
0
500000
1000000
1500000
2000000
2500000
3000000
3500000
0 25000 50000 75000 100000 125000 150000 175000 200000

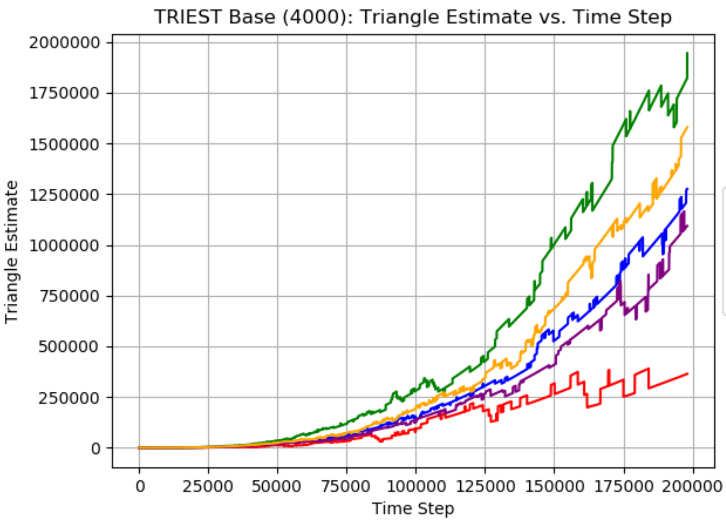

TRIEST Impr (2500): Triangle Estimate vs. Time Step
Triangle Estimate
Time Step
0
200000
400000
600000
800000
1000000
1200000
1400000
0 25000 50000 75000 100000 125000 150000 175000 200000
Minimum
Median
Maximum
First Quartile
Third Quartile

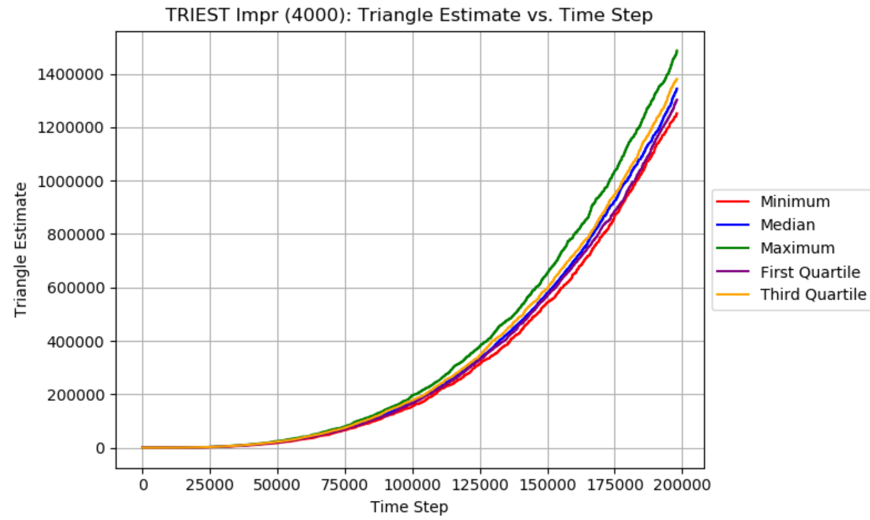

TRIEST Base (3000): Triangle Estimate vs. Time Step
Triangle Estimate
Time Step
0
500000
1000000
1500000
2000000
2500000
0 25000 50000 75000 100000 125000 150000 175000 200000

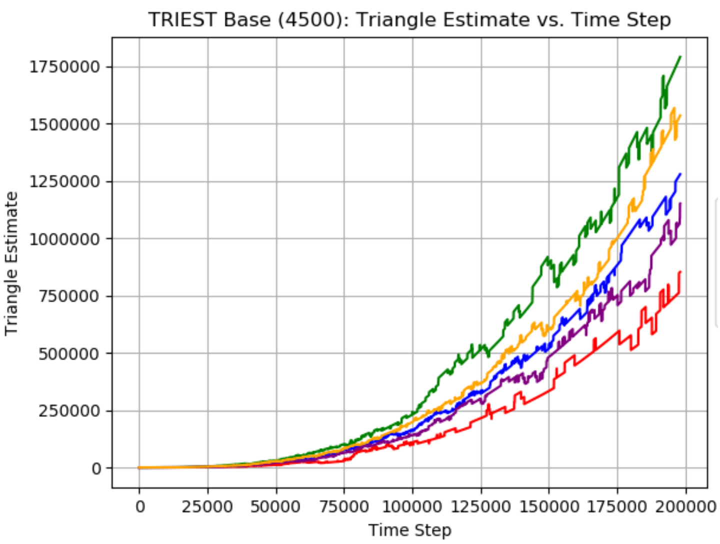

TRIEST Impr (3000): Triangle Estimate vs. Time Step
Triangle Estimate
Time Step
0
200000
400000
600000
800000
1000000
1200000
1400000
0 25000 50000 75000 100000 125000 150000 175000 200000
Minimum
Median
Maximum
First Quartile
Third Quartile

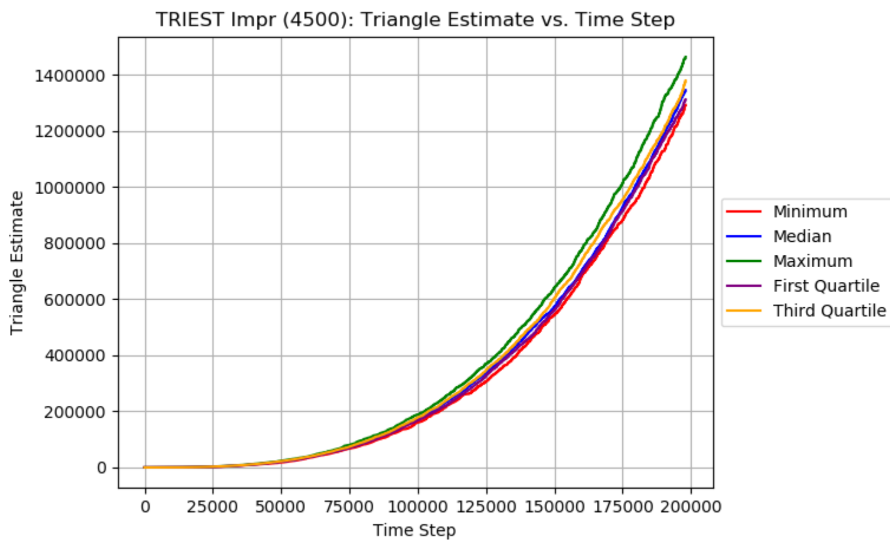

TRIEST Base (3500): Triangle Estimate vs. Time Step
Triangle Estimate
Time Step
0
500000
1000000
1500000
2000000
0 25000 50000 75000 100000 125000 150000 175000 200000

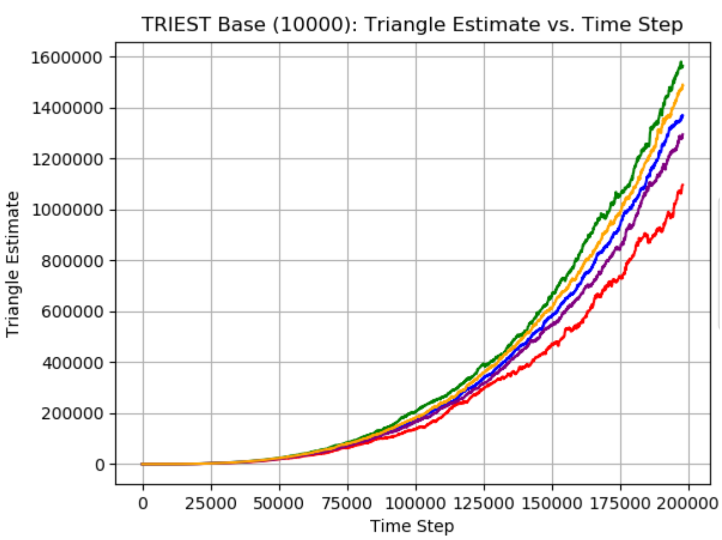

TRIEST Impr (3500): Triangle Estimate vs. Time Step
Triangle Estimate
Time Step
0
200000
400000
600000
800000
1000000
1200000
1400000
0 25000 50000 75000 100000 125000 150000 175000 200000
Minimum
Median
Maximum
First Quartile
Third Quartile

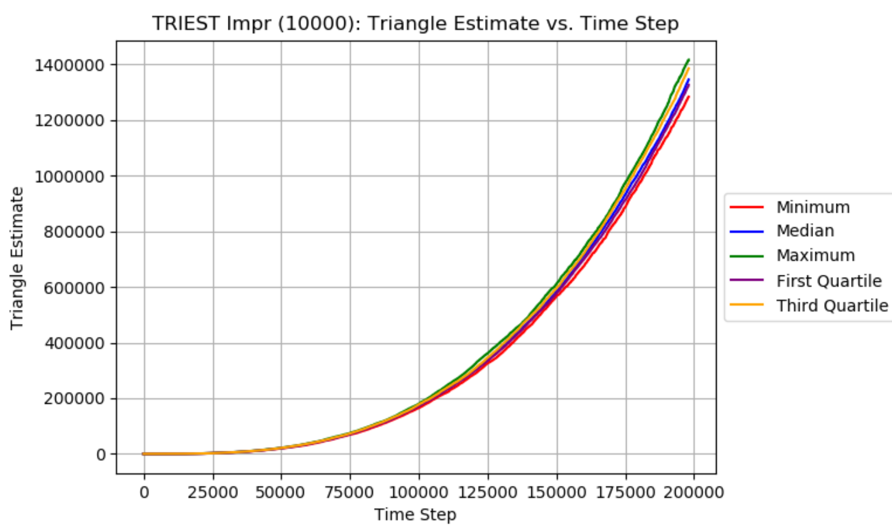

TRIEST Base (4000): Triangle Estimate vs. Time Step
Triangle Estimate
Time Step
0
250000
500000
750000
1000000
1250000
1500000
1750000
2000000
0
25000
50000
75000
100000
125000
150000
175000
200000

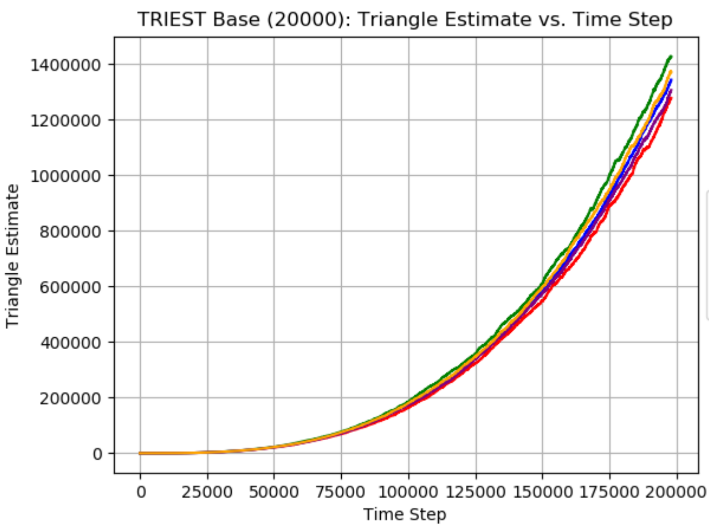

TRIEST Impr (4000): Triangle Estimate vs. Time Step
Triangle Estimate
Time Step
0
200000
400000
600000
800000
1000000
1200000
1400000
0
25000
50000
75000
100000
125000
150000
175000
200000
Minimum
Median
Maximum
First Quartile
Third Quartile

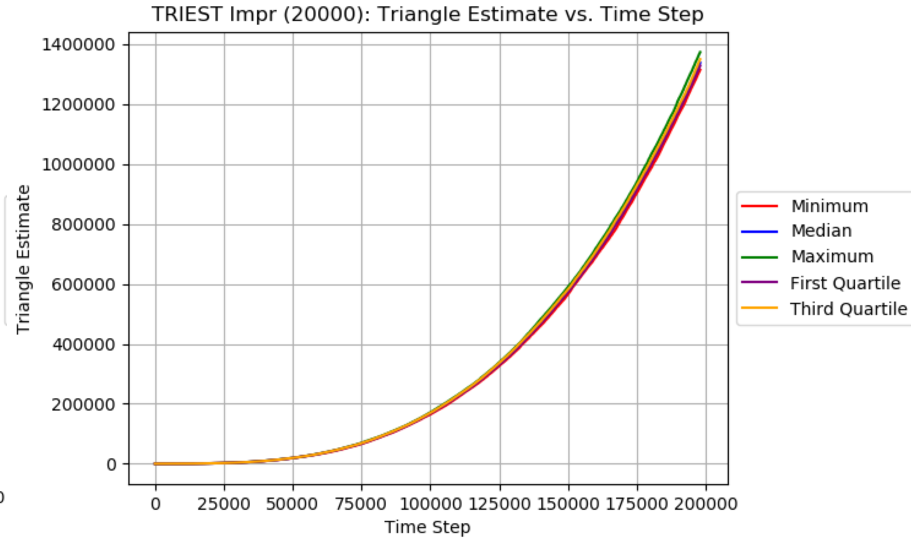

TRIEST Base (4500): Triangle Estimate vs. Time Step
Triangle Estimate
Time Step
0
250000
500000
750000
1000000
1250000
1500000
1750000
0
25000
50000
75000
100000
125000
150000
175000
200000

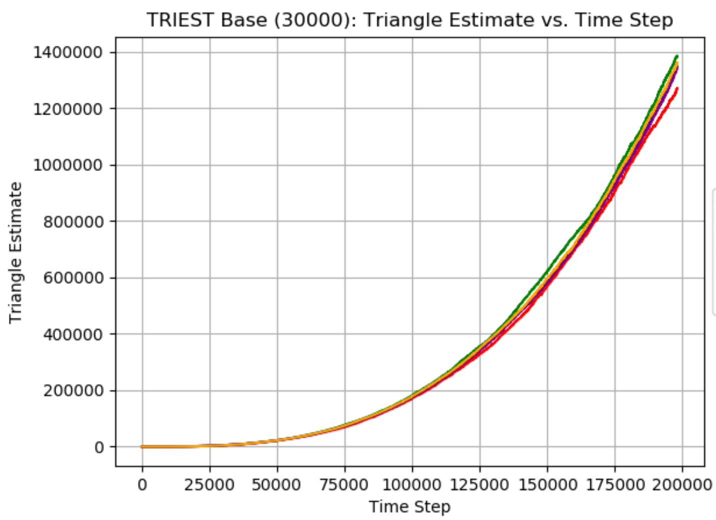

TRIEST Impr (4500): Triangle Estimate vs. Time Step
Triangle Estimate
Time Step
0
200000
400000
600000
800000
1000000
1200000
1400000
0
25000
50000
75000
100000
125000
150000
175000
200000
Minimum
Median
Maximum
First Quartile
Third Quartile

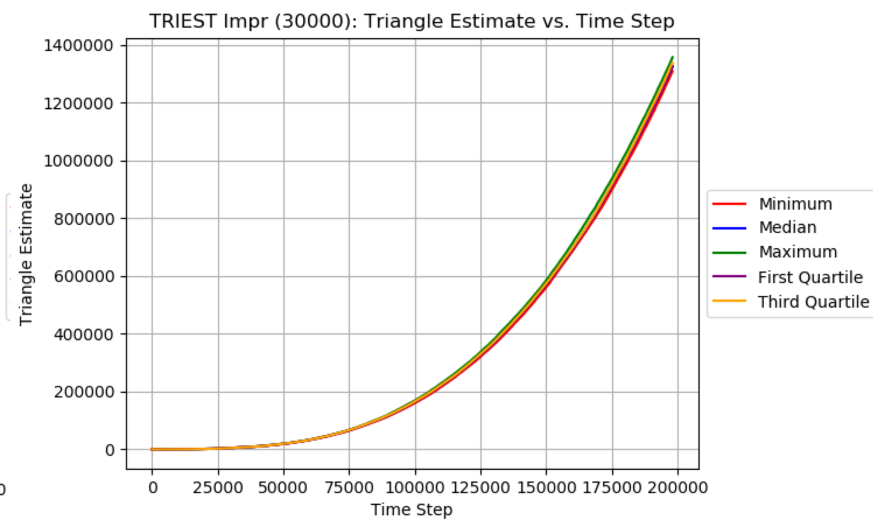

TRIEST Base (10000): Triangle Estimate vs. Time Step
Triangle Estimate
Time Step
0
200000
400000
600000
800000
1000000
1200000
1400000
1600000
0
25000
50000
75000
100000
125000
150000
175000
200000

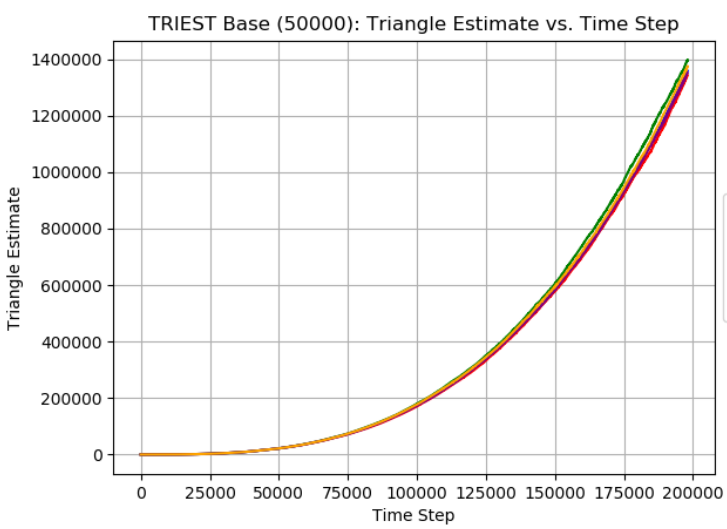

TRIEST Impr (10000): Triangle Estimate vs. Time Step
Triangle Estimate
Time Step
0
200000
400000
600000
800000
1000000
1200000
1400000
0
25000
50000
75000
100000
125000
150000
175000
200000
Minimum
Median
Maximum
First Quartile
Third Quartile

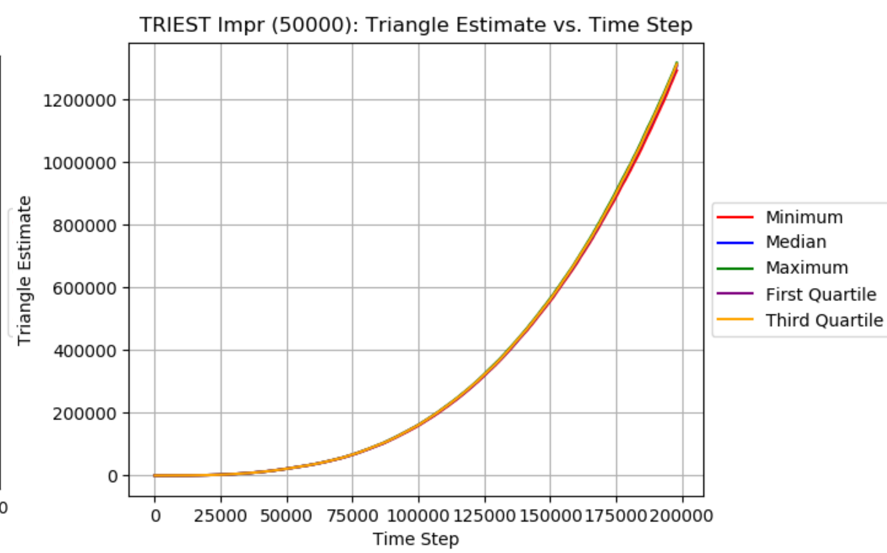

TRIEST Base (20000): Triangle Estimate vs. Time Step
Triangle Estimate
Time Step


TRIEST Impr (20000): Triangle Estimate vs. Time Step
Triangle Estimate
Time Step
Minimum
Median
Maximum
First Quartile
Third Quartile


TRIEST Base (30000): Triangle Estimate vs. Time Step
Triangle Estimate
Time Step


TRIEST Impr (30000): Triangle Estimate vs. Time Step
Triangle Estimate
Time Step
Minimum
Median
Maximum
First Quartile
Third Quartile


TRIEST Base (50000): Triangle Estimate vs. Time Step
Triangle Estimate
Time Step


TRIEST Impr (50000): Triangle Estimate vs. Time Step
Triangle Estimate
Time Step
Minimum
Median
Maximum
First Quartile
Third Quartile

Input file: ca-AstroPh-clean.txt (undirected, processed from ca-AstroPh)

- Nodes: 18,772
- Edges: 198,050
- Triangles: 1,351,441

Because the input file came from a directed graph, additional pre-processing on the original input file (ca-AstroPh.txt) was done to remove duplicated edges and self-loops, generating ca-AstroPh-clean.txt. Previously, TRIEST-BASE and TRIEST-IMPR was coded to handle cases in which duplicate edges and self-edges are present in the original input file (i.e (u,v) = (v,u), and (u,u)). In my prior implementation, all duplicate and self-edges were simply discarded and no changes to the graph or the triangle counter are made. However, realizing that that simply processing these duped/self edges in the algorithms (when they should realistically never come up in an undirected edge stream) may have caused my results to vary from the expected behavior of the algorithms, I decided to opt for pre-processing the data entirely to avoid this issue.

Each algorithm was run 20 times for each of the following 12 values of the sample size parameter: 1,000, 1,500, 2,000, 2,500, 3,000, 3,500, 4,000, 4,500, 10,000, 20,000, 30,000, 50,000. For each of the values, we plotted the minimum, maximum, median, and first and third quartile for each time step across the 20 runs.

From the graphs, it is evident that the improved insertion algorithm (TRIEST-IMPR) results in higher-quality, closer estimations to the actual triangle count compared to BASE as sample size increases. IMPR seems to already give a very good approximation of the global triangle count (~1.3 million) at sample size 4,500 with low variance, in comparison to BASE. Even at a smaller sample size of 2,000, IMPR gives a much better approximation in comparison to BASE. This trend of lower variance in IMPR can be seen for the max, min, median and quartiles measured up to sample size of 20,000 to 30,000, when both algorithms have converged to the actual triangle count with low variance.

Unlike TRIEST-BASE, TRIEST-IMPR will attempt to update its triangle counter for every element on the edge stream before determining whether to add the edge to the local graph. BASE only counts a triangle if all three edges (after an edge has been added essentially) are in the local graph. In IMPR, if two edges are in the local graph made of stored edges, and a third edge that would result in a triangle is present on the stream, the triangles formed from that third edge are counted regardless if that edge is added locally. Additionally, unlike BASE, IMPR will never decrement its local triangle count. This corresponds more closely with how the global triangle count in insertion-only edge streams would behave, as (assuming unlimited memory space) deletion of edges would never occur; thus the global count must be non-strictly increasing. Therefore, IMPR uses the local triangle count to identify new triangles and never decreases its global estimate. Furthermore, the global estimate in BASE is determined by dividing the local triangle count by a constant π_t, while IMPR performs weighted increments to its triangle counter when new triangles are found. By simply rearranging and making small modifications to BASE with how the global estimate is determined, IMPR provides lower variance, higher-quality estimations at smaller sample sizes in comparison to BASE.