Supplemental Material

To illustrate the developed framework can efficiently process the large volume of geo-located tweets, we show the processing time for generating Twitter user trajectories with different time span in Table 1 and Figure 1. The experiment was carried out with Hadoop environment with 11 computing nodes, each nodes is equipped with 256 GB RAM and CPU with Intel Xeon E5-2660 v3, 2.6GHz, 25M Cache, 10 Cores. In particular, with over 1 billion tweets, the corresponding user trajectories can be extracted in approximately 40 minutes.

Table 1: Processing time for generating Twitter user trajectories with different time span

Data	Number of tweets	File size (GB)	Processing time (seconds)
1 month (Jan)	9.95E+07	4.5	260.40
2 months (Jan-Feb)	1.85E+08	8.4	462.57
4 months (Jan-April)	3.70E+08	17	939.16
6 months (Jan -June)	5.67E+08	26	1478.18
8 months (Jan-Aug)	7.86E+08	35	2002.73
10 months (Jan-Oct)	9.23E+08	41	2294.55
12 months (Jan-Dec)	1.04E+09	46	2581.45

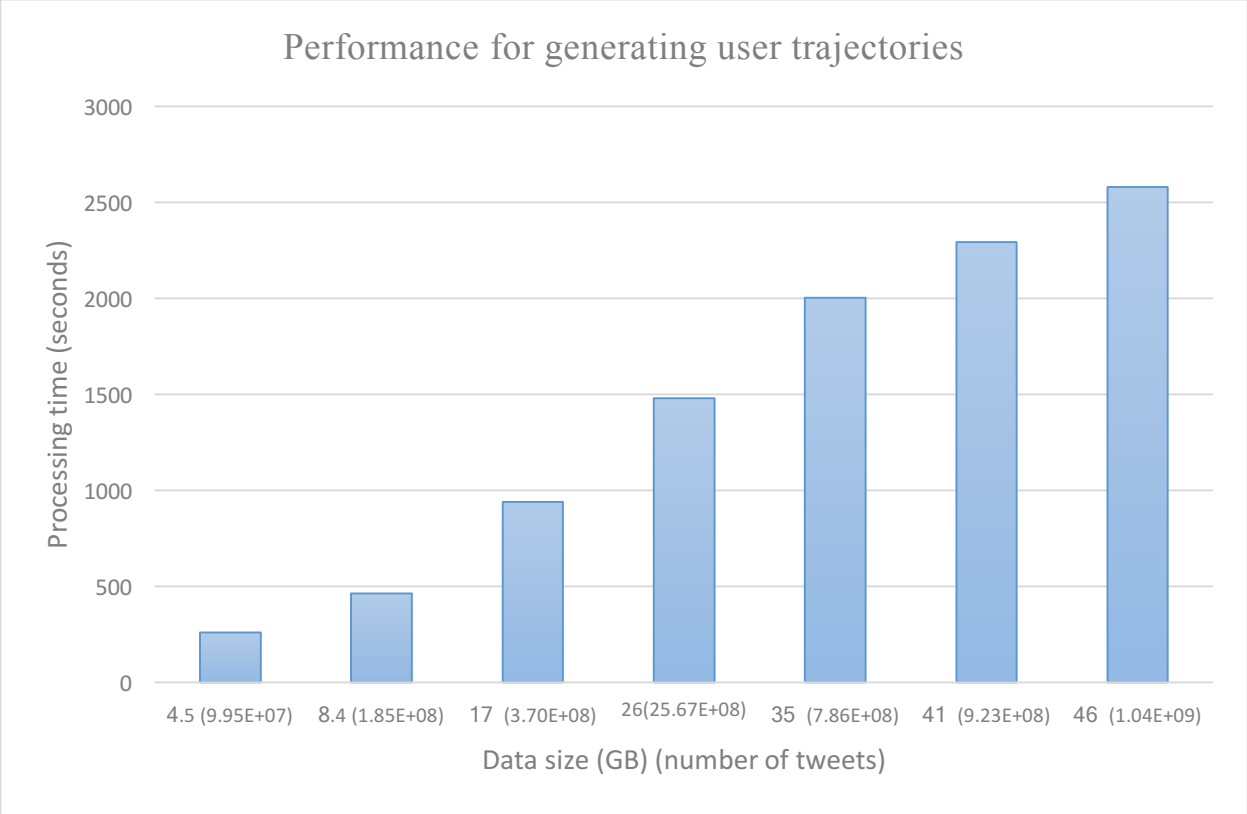

Figure 1: The processing time graph for generating Twitter user trajectories with different time span