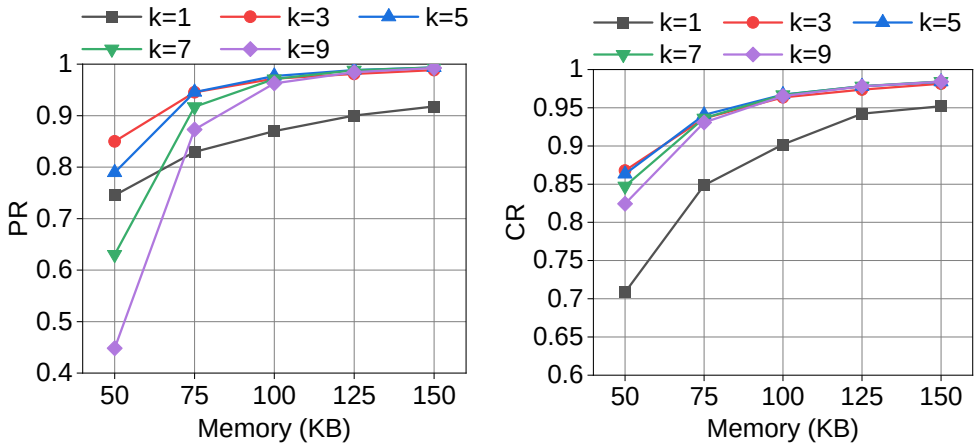
SteadySketch: Finding Steady Flows in Data Streams (Appendix)

Anonymous Author(s)

February 9, 2023

A Experiments on Parameter Settings

In this section, we measure the effects of some key parameters of SteadySketch, namely, the number of hash functions k in SteadyFilter, the ratio r of the memory size of SteadyFilter to the memory size of the whole SteadySketch, the number of hash functions d in GroupSketch, the the variance threshold H for the steady items, and the threshold p for time window period. We use the CAIDA dataset in these experiments, and PR and CR to evaluate the effects.

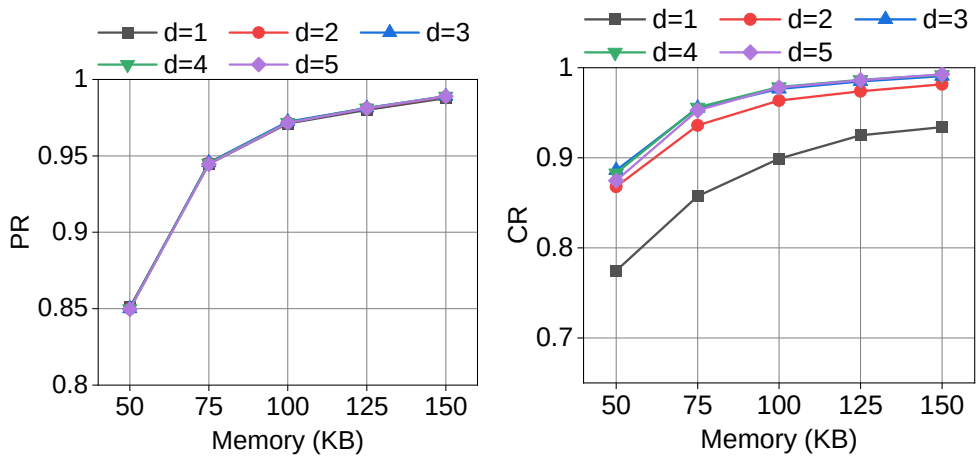


PR of k CR of k

Figure 1: Effect of the parameter: k

1) Effect of k (Figure 1(a) - 1(b)): *The experimental results show that the best value for k is 3.* In this experiment, we fix the d to 2. Under the same memory, as k becomes larger, PR increases first and then decreases, while CR increases first and then does not change significantly. Thus, we set $k = 3$.

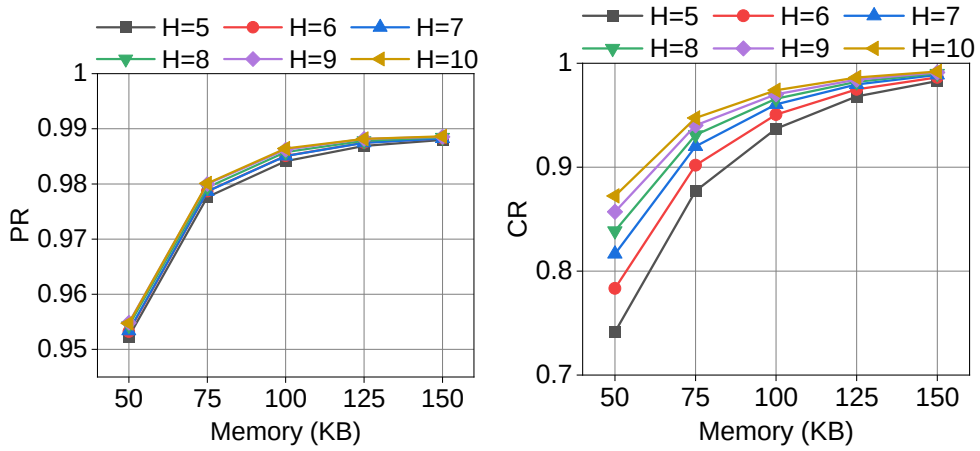
2) Effect of d (Figure 2(a) - 2(b)): *The experimental results show that the best value for d is 2.* In this experiment, we fix the k to 3. Under the same memory, as d becomes



PR of d CR of d

Figure 2: Effect of the parameter: d

larger, CR increases first and then decreases, while PR does not change significantly. Thus, we set $d = 2$.

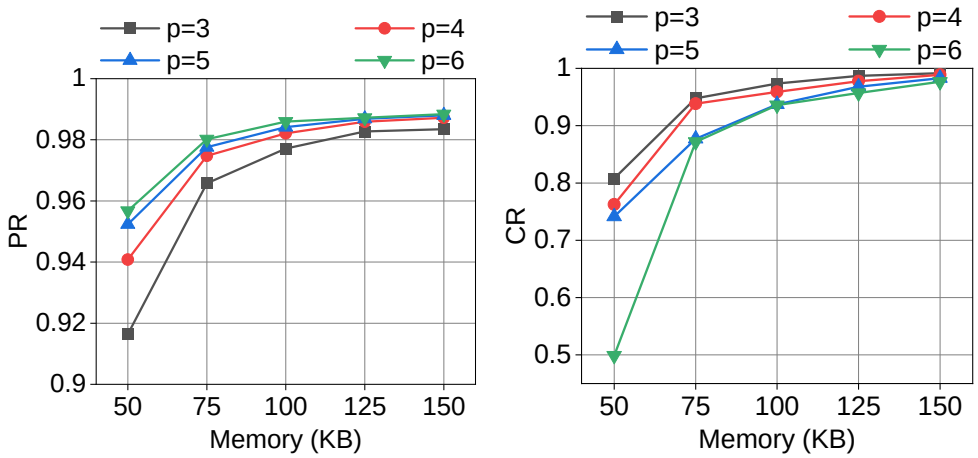


PR of H CR of H

Figure 3: Effect of the parameter: H

3) Effect of H (Figure 3(a) - 3(b)): *Our experimental results show that the PR and CR of different H values are close.* In this experiment, we fix the k to 3 and d to 2. As H becomes larger, CR gradually increases but gradually approaches as the memory becomes larger, while PR does not change significantly. Since this value is as small as possible, we finally choose $H = 5$.

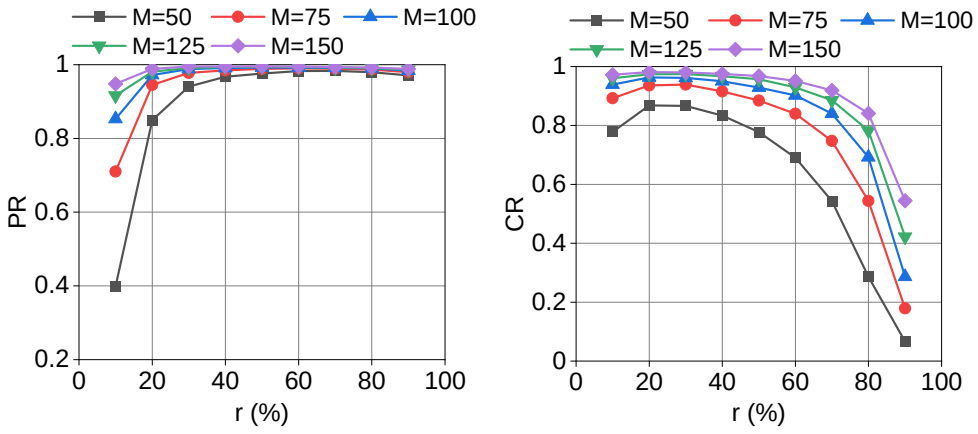
4) Effect of p (Figure 4(a) - 4(b)): *The experimental results show that the optimal value for p is 5.* In this experiment, we fix the k to 3, d to 2 and H to 5. As p becomes larger, PR gradually increases but gradually approaches as the memory becomes larger, while CR gradually decreases but gradually approaches as the memory becomes larger.



PR of p CR of p

Figure 4: Effect of the parameter: p

Therefore, we set $p = 5$.



PR of r CR of r

Figure 5: Effect of the parameter: r

5) Effect of r (Figure 5(a) - 5(b)): *The experimental results show that the optimal value for r is 40%.* In this experiment, we fix the k to 3, d to 2, H to 5 and p to 5. M refers to the total memory of SteadySketch, which consists of the memory of SteadyFilter and RollingSketch. As r becomes larger, PR gradually increases, while CR increases first and then decreases. Therefore, we choose $r = 20\%$ because it can trade off PR and CR well for different values of M.