Efficient Approximate PageRank 10605 15 Fall

Jingyuan Liu
AndrewId: jingyual
`jingyual@andrew.cmu.edu`

1 Question 1

The probability that the k-th point is the last to be eaten is $\frac{1}{n}$.

$$p(k \ is \ the \ last \ to \ be \ eaten) = \frac{1}{n} \tag{1}$$

Before k is eaten, we must first eat the k-1 or k+1. Consider the symmetry, we can asuume that it is the k-1 that is eaten before k. The k is the last means that the k+1 must be eaten already before k. This could happen if the random walk passes cloclwise from k-1 to k+1 before eating k.

The probability of eats k-1 before k+1 is $\frac{1}{2}$, because the graph is symmetric. To conclude, we could see this result is reasonable. The overall probability of any point be the last one:

$$\sum_{k}^{n} p(k \ is \ the \ last \ one) = 1 \tag{2}$$

2 Question 2

We could use several different criteria to evaluate a static-graph sampling algorithms:

In-degree and out-degree distribution.

The distribution of sizes of weakly connected and strongly connected components.

Hot-plot: The number P (h) of reachable pairs of nodes at distance or less.

Hot-plot on the largest WCC.

The distribution of the first left singular vector of the graph adjacency matrix versus the rank.

The distribution of singular values of the graph adjacency matrix versus the the rank.

The distribution of the defined clustering coefficient.

3 Question 3

Place more outlinks from V would not change the PageRank score. Basically, place more outlinks would not directly increase the inlinks, therefore, would not bring direct increase. On the other hand, the self-reserved flow is not changed by the outlink numbers. In conclusion, placing more outlinks would not change the PageRank score.

Obtain more links to V would increase the PageRank score, since more links to the V website, the PageRank score would increase as defined.

Split V into two Va and Vb without any links between them and copy all out links would not change the PageRank for V. This is because there would not be more links to the V. So the PageRank score would not change.

Sybil Attack would increase the PageRank score. Basically, using Sybil Attack would add more inlinks to V. This would increase the PageRank score.

4 Question 4

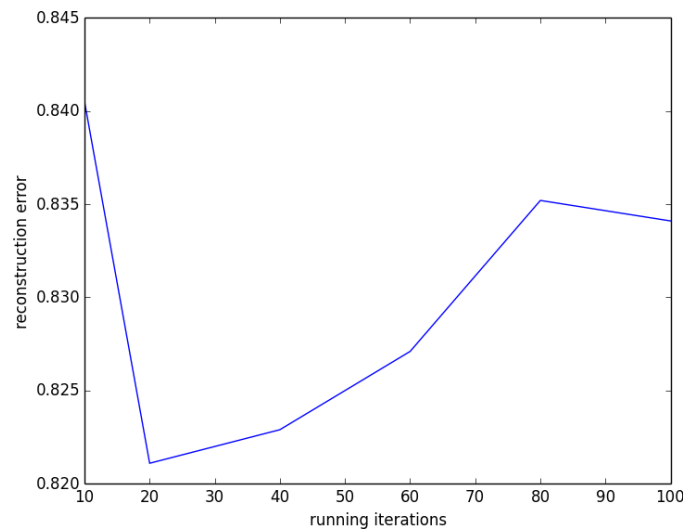## 4.1 Question 4.1

The trend is:

Figure 1: Error over iteration

We could see at first, with the increase of iterations, the reconstruction error would decrease. Then it would slowly increase with more iterations.

4.2 Question 4.2

The trend is:

We could see that as the increase of workers, the error would increase. This is quite reasonable because with increase of workers, the error of approximation like number of iterations would be more influential.

4.3 Question 4.3

The trend is:

We could see at first, with the increase of topics or factors, the error would decrease. Then with the increase of factors after 20, the error would increase very fast. This is because we set too many topics to the dataset.
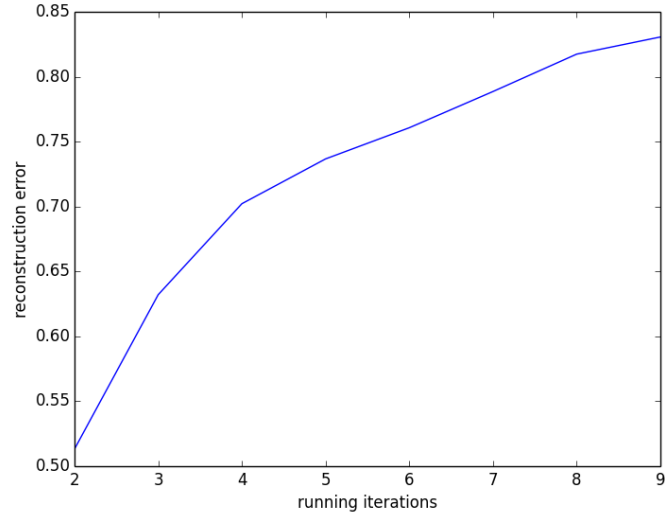
4.4 Question 4.4

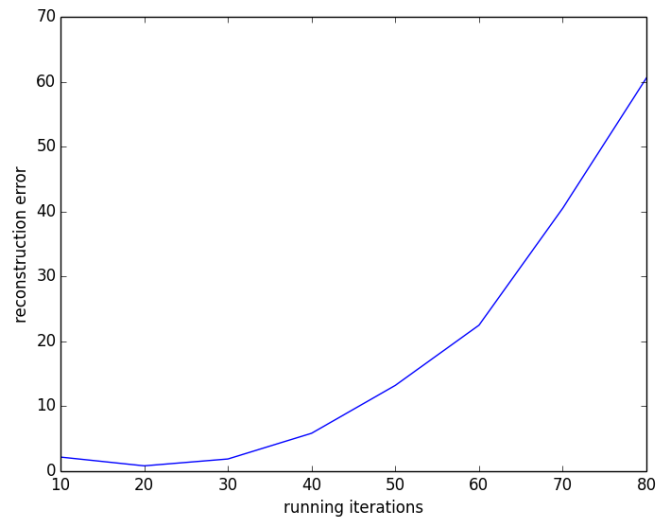The trend is:

Figure 2: Error over worker

Figure 3: Error over factors

We could see the trend is at first decrease, and then is stable, not sensitive to the increase of beta

5 Question 5

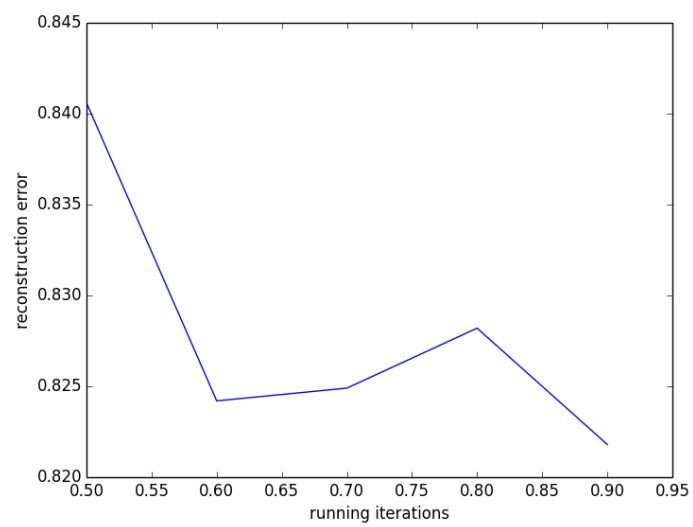I received helps from Chenran Li, whose AndrewId is chenranl. He told me how to design the blockify matrix function and how to understand the perform sgd function.

Figure 4: Error over beta