CIS 630 – Distributed Systems

Project 1

Due: 4-20-14

Adam Martini
martini@cs.uorgeon.edu

Simulating Many Random Walk on a Large Graph

The sample data provided contains roughly 1.5 million nodes and 15 million edges. To load and process this data efficiently we use a C++ program. We use an adjacency list to store edge connections, which is implemented as an unordered map with unordered set values. These data structures provide $O(1)$ insertions and look-up for efficient updates and membership testing. The unordered sets are stored on the heap to facilitate faster data structure updates for data structure initialization.

We have tested the many random walk simulation on data sets of various sized to confirm mixing time convergence of the distribution. Figure 1 demonstrates that the standard deviation of the credit vectors converges to a constant value. Figure 2 shows that the average squared difference between credit vectors C(t-1,i) and C(t,i) converges to 0 with an increasing number of steps.

These two plots together confirm that the distribution of the credit/degree ratio converges to the same value as the number of rounds increases.

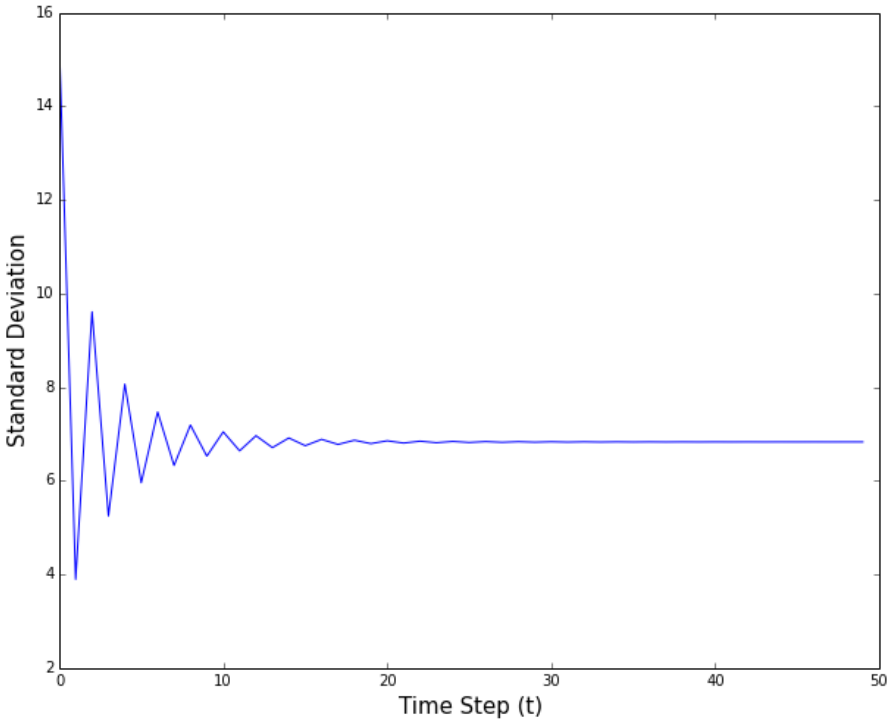

Figure 1: The standard deviation of the normalized credit vector over time.

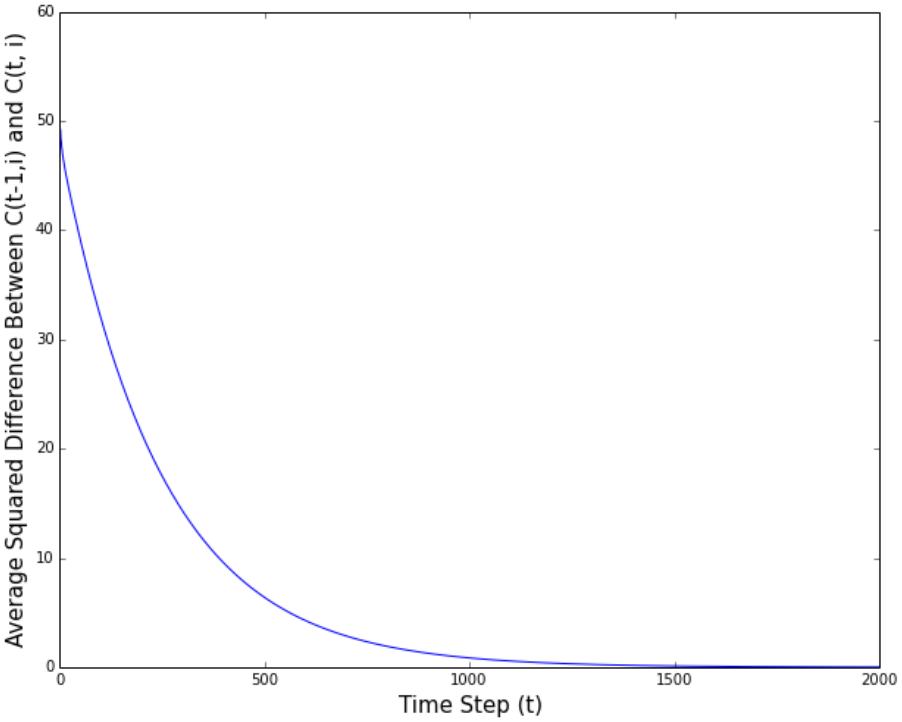

Figure 2: The average squared difference between credit vectors at each update over time.