# Read Me

**Introduction**

This is an easy version of Gibbs Sampling for general graphic models.

From the coding level, as the probabilities of each node were not certain, so few probability distributions are pre-defined in the program. The edges are less important, when compared to the vertexes.

**Run the code**

Requirements:

1. Please use the testing dataset *graph-example-2-tp3.json*, and pay attention to the path in Line38.

2. Please import the *commons-math3-3.5.jar* provided by Apache, as the distribution function will be applied.

*Main* function is included, please run directly.

Results:

In the testing code, we assume each vertex has 10 states, values from 0 – 9, with random probability of each.

(x0, x1, x2, x3, x4, x5, x6, x7, x8, x9) 🡪 example: (3,5,1,6,5,3,6,7,0,8)

The result shows (in **Figure 1**):

p(x1=3|x0=5)

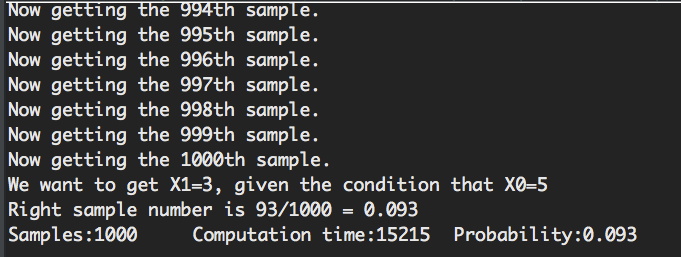


Figure 1: result

**Evaluation**

1. Sample Numbers and Computation Time

Vertex number: 809

State number: 10

Sample number scales from 2, 4,…2048.

Figure 2 shows the plot, Table 1 shows the probabilities.

Figure 2: sample number and computation time

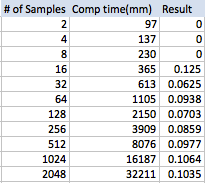


Table 1: sample number and computation time, with probabilities

2. State Numbers and Computation Time

Vertex number: 809

Sample number: 1000

State number scales from 2, 4,…2048.

Figure 3 shows the plot, Table 2 shows the probabilities.

Figure 3: state number and computation time

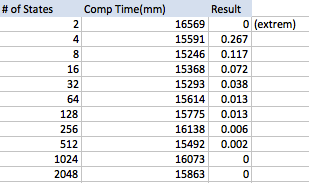


Table 2: state number and computation time, with probabilities

3. Vertex Numbers and Computation Time

State number: 10

Sample number: 1000

Vertex number scales from 2, 4,…2048.

Figure 4 shows the plot, Table 3 shows the probabilities.

Figure 4: vertex number and computation time

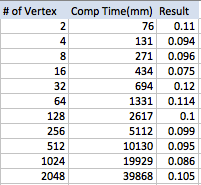


Table 3: vertex number and computation time, with probabilities

**Conclusions**

Edges

For the testing code, the number of vertex is the most important for the general graphic model. The edges are used to defined relationships, especially in Bayesian Model. In this model, only vertexes and edges are not enough for calculate some probabilities. Other probabilities of the vertexes, like conditional probabilities or joint probabilities, are depending on the applications. So here we just predefined as random numbers to simulate the whole process of Gibbs Sampling.

Scalability

For the three variables: number of vertex, number of sample and number of vertex state, computation time depends on the scale of **vertex** and **sample**. When the graph is larger or more samples are needed, computational time increases, which means numbers of vertex and sample are not scalable.

The probability shows the scalability when the number of state is constant.

But only when the state number is increasing, the probability will decrease.