Kae-Yang Hsieh(001092745)

Program Structures & Algorithms

Fall 2021 Assignment No. 3

Tasks

* Experiments

The experiment code located at "java/edu/neu/coe/info6205/union_find/UF_HWQUPC_experiment.java".

The experiments were conducted from 10000 to 2560000 sites using doubling method, and each number of sites will be executed for 10 times and then get the average.

Short summary: $c = n - 1$, where c is connections and n is the sites. The result turns out that count function matched what I expected. The number of pairs need to get all objects connected are fixed, which is the number of all sites minus one.

```
Average 46248 paris generated with 9999 connections of 10000 sites
Average 100876 paris generated with 19999 connections of 20000 sites
Average 229355 paris generated with 39999 connections of 40000 sites
Average 494430 paris generated with 79999 connections of 80000 sites
Average 999909 paris generated with 159999 connections of 160000 sites
Average 2246999 paris generated with 319999 connections of 320000 sites
Average 4402904 paris generated with 639999 connections of 640000 sites
Average 9111114 paris generated with 1279999 connections of 1280000 sites
Average 19226965 paris generated with 2559999 connections of 2560000 sites
```

*** Results of relationship between the number of objects (n) and the number of pairs (m)**

Short summary: $m > \frac{1}{2}nln(n)$, where m is the pairs generated and n is the number of objects.

The deduction:

There is a graph has n nodes, and all possible edges will be

$$\binom{n}{2} = \frac{n(n-1)}{2}$$

The randomly generated m pairs can be described as choosing from all possible edges randomly. So the probability of choosing a pair is:

$$p = \left(\begin{matrix} \binom{n}{2} \\ m \end{matrix}\right) = \frac{2m}{n(n-1)}$$

The probability of one isolated node can be described as a node which doesn't have any pair be generated, and it will be:

$$(1-p)^{n-1} \sim (1-p)^n \sim e^{-pn}$$

The final goal is to become one component, that is no isolated node.
So the number of isolated nodes should smaller than 1.
The total number of isolation graph is $n * e^{-pn}$, so:

$$ne^{-pn} < 1 \qquad => \qquad ne^{-\frac{2m}{n-1}} < 1 \sim ne^{-\frac{2m}{n}} < 1 \qquad \text{when } n -> \infty$$

$$e^{-\frac{2m}{n}} < \frac{1}{n} \qquad => \qquad -\frac{2m}{n} < -lnn$$

$$m > \frac{1}{2}nlnn$$

So the generated pairs m should be larger than $\frac{1}{2}nlnn$ to make sure there is no isolated node.

number of sites to random pairs generated

n	m	ln(n)	ln(m)	1/2 n ln(n)
10000	46248	9.21	10.74	46051.70
20000	100876	9.90	11.52	99034.88
40000	229355	10.60	12.34	211932.69
80000	494430	11.29	13.11	451591.28
160000	999909	11.98	13.82	958634.33
320000	2246999	12.68	14.63	2028172.20
640000	4402904	13.37	15.30	4278151.51
1280000	9111114	14.06	16.03	8999917.21
2560000	19226965	14.76	16.77	18887062.81

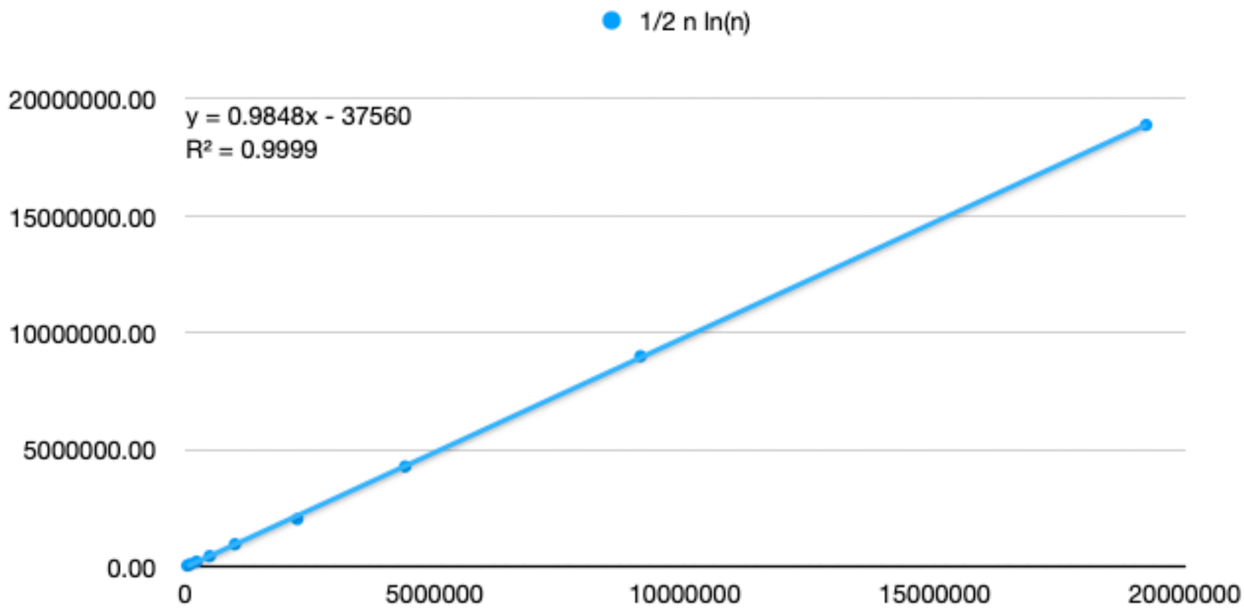

Figure. 1 numbers of pairs generated to the $1/2 * sites * ln(sites)$

The observation:
The chart(Figure. 1) showed a linear relationship of the pairs to the $1/2 * sites * ln(sites)$, which matched the math deduction.

* Unit test results

UF_HWQUPC_Test