

Assignment for

Computer Science Theory for the Information Age

Day 3

BY ZEN HUANG

5120309027
2012 ACM class

June 16, 2013

Exercise 1. What is the expected number of squares(4-cycles) in $G\left(n, \frac{d}{n}\right)$? What is the expected number of 4-cliques in $G\left(n, \frac{d}{n}\right)$?

Answer.

Let X be the variable of squares and Y be the variable of 4-cliques in $G\left(n, \frac{d}{n}\right)$.

$$\begin{aligned} E(X) &= \frac{n(n-1)(n-2)(n-3)}{4 \times 2} \left(\frac{d}{n}\right)^4 \\ &= \frac{d^4(n-1)(n-2)(n-3)}{8n^3} \end{aligned} \tag{1}$$

$$\begin{aligned} E(Y) &= C_n^4 \left(\frac{d}{n}\right)^6 \\ &= \frac{n(n-1)(n-2)(n-3)}{4!} \left(\frac{d}{n}\right)^6 \\ &= \frac{d^6(n-1)(n-2)(n-3)}{24n^5} \end{aligned} \tag{2}$$

Exercise 2. Search for WWW for an undirected graph or a data base that can be counted to a graph. Find the connected components and count the number of each size.

Answer.

Here I found the Enron email data¹ from stanford.edu².

Enron email communication network covers all the email communication within a dataset of around half million emails. Nodes of the network are email addresses and if an address i sent at least one email to address j, the graph contains an undirected edge from i to j.

The original graph contains 36692 nodes with 367662 edges. For the convinence of calculating, I removed two edges starting from node 0, which does not influence the results. Here is a piece of mathematical code to deal with the data:

```
rawData = Flatten[Import[#] & /@ FileNames["*.dat"], 1];  
Data = DeleteDuplicates[
```

1. <http://snap.stanford.edu/data/email-Enron.txt.gz>, William Cohen, CMU

2. <http://snap.stanford.edu/data/email-Enron.html>

```

Map[If[#[[1]] < #[[2]], #, {#[[2]], #[[1]]}] &, rawData]];
g = Graph[Data]
Tally[Map[Length, ConnectedComponents[g]]]

```

And here is the output result:

```

{{33695, 1}, {20, 1}, {16, 1}, {14, 1}, {13, 3}, {12, 3}, {11,
2}, {10, 8}, {9, 6}, {8, 7}, {7, 7}, {6, 20}, {5, 44}, {4, 114}, {3,
120}, {2, 727}}

```

which shows the graph contains 1 big component with 33695 edges, 1 component with 20 edges, 1 component with 16 edges and so on.

Also the program comes with a beautiful graph indicating all components:

