Carnegie Mellon University
15-826 – Multimedia Databases and Data Mining
Fall 2012, C. Faloutsos
Homework 2 Solution

Prepared by: Kai Ren

1 Q1 – Dynamic time warping [25 pts]

Solution: The answer to the given dataset is:

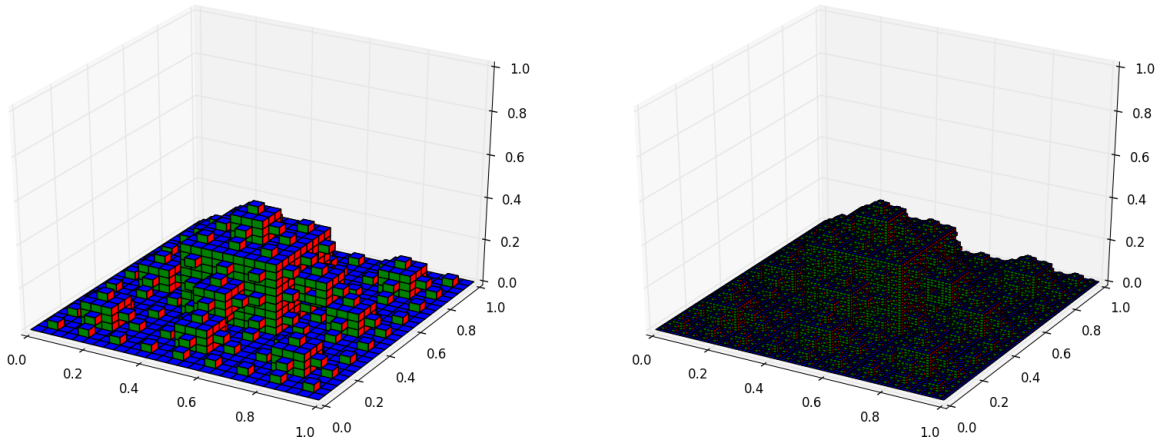
```
0002.dat 0007.dat 18510844
```

See c++ code (*link*).

2 Q2 – Fractals [25 pts]

Solution:

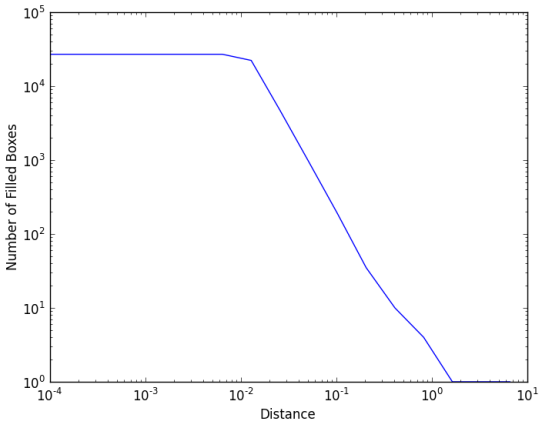
1. See python script (*link*).



(a) K_3 (b) K_4

Figure 1: Example Koch surface for K_3 and K_4

2. The Hausdorff plot for the points of K_4 is shown in Figure 2. This is plotted by using the package provided in the handout. The slope is -2.090, and therefore the Hausdorff or D_0 fractal dimension is 2.090. The slope of the correlation integral of Koch surface is $\log(13)/\log(3) = 2.3347$ in theory.



3 Q3 – Separability [25 pts]

Solution:

1. Figure 2 shows the correlation integral of dataset S, and of dataset H.

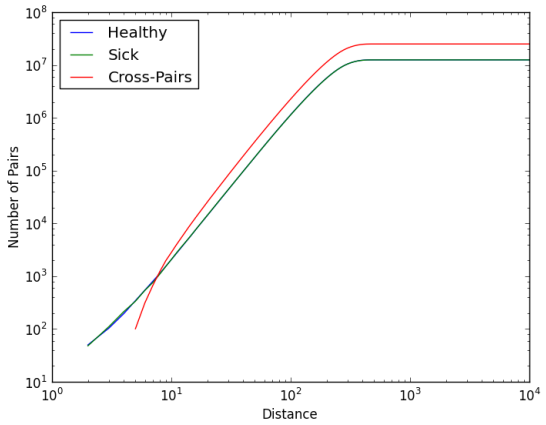


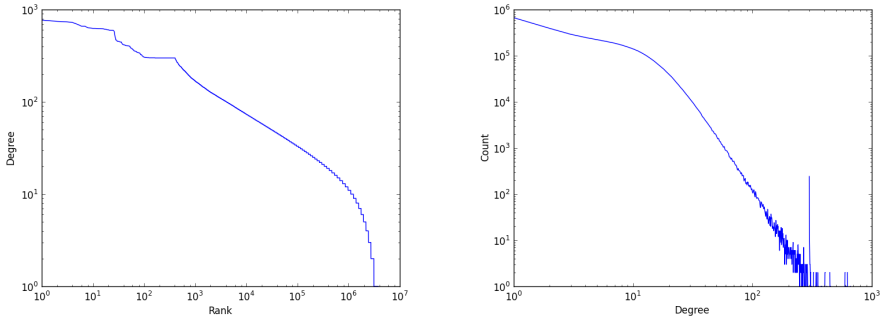
Figure 2: Correlation integral of dataset S and H, and also distribution of distance of pairs from $S \times H$.

2. Both dataset H and S are not uniformly distributed in 4D space since their fractal dimension (= slope of correlation integral) is less than 4.
3. H and S are separable.
4. Figure 2 also draws the distribution of distances of cross-pairs in $S \times H$. It drops at the point where the distance is about 10, and there are no cross-pairs below distance ≈ 5. H and S are actually hyper-planes in 4D space.

4 Q4 – Power Law in Graphs [25 pts]

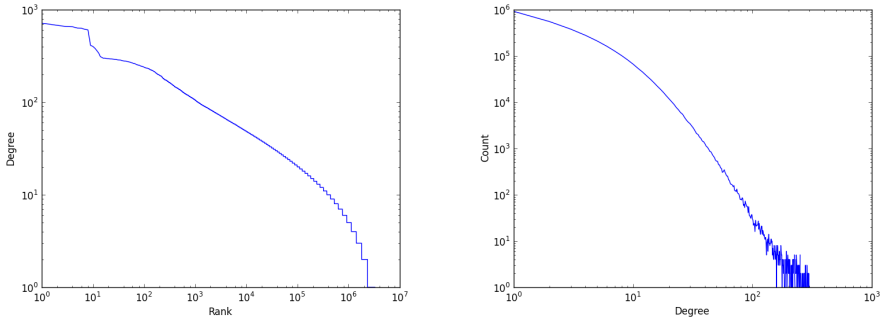
Solution:

1. The six degree plots are listed as below, in Figures 3,4,5.



(a) Rank-Degree (b) Degree-Count

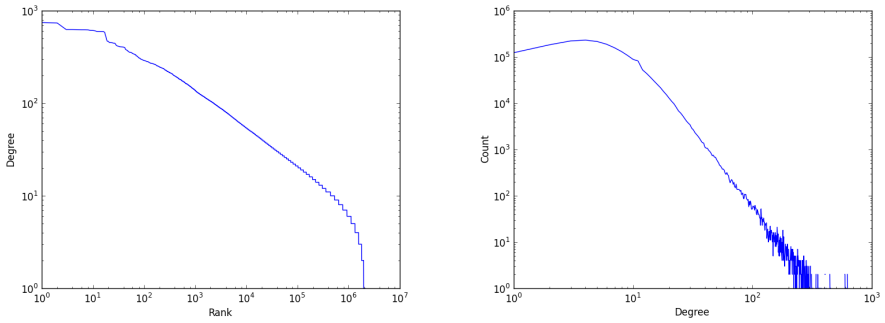
Figure 3: The rank-degree and degree-count plot for total-degree



(a) Rank-Degree (b) Degree-Count

Figure 4: The rank-degree and degree-count plot for in-degree

2. There are 300 fake patents.
3. There is a spike at degree of 300 in Figure 3 (b). This spike is the one responsible for the plateau in Figure 3(a), at rank≈100-1,000 and degree (of course) $d \approx 300$. Because the fake patents form a nearly perfect clique, and there is at most one edge between two patents, the correst guess is 300. And their in-degree or out-degree will be (300, 299, . . . , 1).
4. The injected fake patents are numbered from 1000335 to 1000634. Kudos to you guys who found those fake patents!



(a) Rank-Degree (b) Degree-Count

Figure 5: The rank-degree and degree-count plot for out-degree