Name: Jingsi Li USC ID: 1207830860

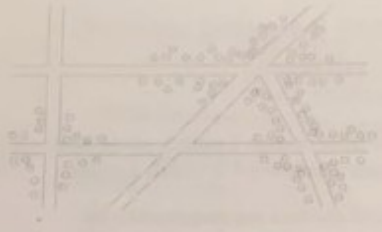

Figure 1: Cholera outbreak

b 6. In the first lecture, we have discussed a number of data mining algorithms and their applications. Figure 1 is the chart about cholera outbreak in London that you have seen in class. The chart was used to discuss which of the following mining algorithms/applications?

a. Association rule mining
b. Clustering
c. Link analysis
d. Product recommendation

7. [4 points] Recall the "evil-doer" example when we talked about the Bonferroni's principle. Suppose that we make the following assumptions. We track 1 million people for 100 days. Each person stays in a hotel 1% of the time. Each hotel holds 100 people and there are 100 hotels.

a. [3 points] What is the expected number of "suspicious" pairs of people (i.e., they went to the same hotel on some two days)?

P and q in same hotel for one day:

$$\frac{1}{100} \times \frac{1}{100} \times \frac{1}{100} = \frac{1}{10^6} = 10^{-6}$$

P and q in same hotel for 2 days:

$$10^{-6} \times 10^{-6} = 10^{-12}$$

2 pairs of people to same hotel on some 2 days.

$$C_2^{100} \times C_2^{10^6} \times 10^{-12} = \frac{10^{12}}{2} \times 10^{-12} \times \frac{1}{2} \times \frac{10^4}{2} = \frac{10^4}{4}$$

$$= 2500$$

b. [1 point] Suppose that we only regard a pair of people as suspect if they were at the same hotel at the same time on three different days. Would the number of such pairs go up or down compared to the number in (a)?

P and q in same hotel for 1 day: $\frac{1}{100} \times \frac{1}{100} \times \frac{1}{100} = 10^{-6}$

P and q in same hotel for 3 days: $(10^{-6})^3 = 10^{-18}$

finally: $C_3^{100} C_2^{10^6} \times 10^{-18} = \frac{100^3}{6} \times \frac{10^{12}}{2} \times 10^{-18}$

$$= \frac{10^6}{6} \times \frac{10^{12}}{2} \times 10^{-18} = \frac{1}{12}$$

down!