1 Framingham Heart Study

The data was collected over 32 years and Christakis and Fowler tried to answer spread of the following traits in that social network :

1. Obesity [2]

2. Smoking [3]

3. Loneliness [1]

4. Happiness [4]

In this and following lecture, we will discuss about:

- the data set used in the Framingham Heart Study (FHS)

- arguments given in the paper [2]

- reviews about these papers by other researchers

1.1 Motivation

The idea of collecting public health data was proposed in 1942 - when penicillin came into existence. (So now TB and infection doesn't kill many people as before). But now cardiovascular disease (CVD) kill many people. In 1950, $\frac{1}{3}^{rd}$ of US males had CVD who were < 60 years of age. Also note the fact that CVD is 2 times prevalence of cancer. The presumptions about the cause of CVD are obesity, seductary life, etc. But there are some exception. For example, Ötzi had CVD but he was not fat, he did not lead a seductary life, he died at the age of 45 years. CVD is the leading cause of death, average life expectancy once you are > 45 years of age did not change.

In this particular data set lots of methodological assumptions are made.It is also called as data set of convenience.

1.2 Framingham Heart Study

Epidemiological study to isolate cancer from CVD is done to quantify risk factor associated with it. There were 2 studies conducted based on the data collected from Department of Public Health at the following places :

1. Newton, MA

2. Framingham, MA (Heart Disease Epidemiological Study in 1947)

The Framingham study was carried out as "Longitudinal cohort study". First physical exam is carried out in 1948 and protocol was established in 1950 and the period is 3 years approximately between each "exam waves ". The aims of the study are :

1. Epidemiology data on arteriosclerosis and Hypertension CVD

2. Data on prevalence of CVD in new population

3. Test diagnostic standards (Cholesterol, obesity,alcohol, smoking, etc.)

Why Framingham?   The choice of collecting public health data at Framingham is made because of the following reasons :

- 1,00,000 people

- middle class

- convenient to hospitals

- have hosted a TB study

- different ethinicities

1.3   FHS Data

How big a sample is?   There were 6000 slots, every 2 years for physicals. Persons were selected from 30-60 year age group who did not have CVD. For this age group there were ~10,000 people in Framingham. There are 3 cohorts of data :

Original cohort 5209 volunteers out of which 5127 had no CVD (therefore, not random because 1/50 had CVD but we expected $1/6^{th}$ having CVD)

1971 offspring cohort 5124 children and spouses of children of original cohort

2002 3rd generation cohort 4095 children of offspring cohort

More on cohorts   Only 10 of 5124 in the offspring cohort left the study for the reason other than death. For this no network information was present. In 2002, the participants were asked to provide their ID, family (1st order relatives such as parents, spouse, siblings, children, etc), specify 1 close friend (only one). These were collected from each of 7 waves in 1971-2003. Also added home address (geo-coded). Some statistics on cohort :

- people in off spring cohort : 5124 (termed as *egos* or *Focus Person (FP)*)

- adding friends and family : 12607 (over 1972-2003)

- 38,611 social and familiar ties (degree average: 7.5 ties/ego)

- 87% of FP had 1 sibling in the network

- 10% of FP had geographical neighbors in the networks

Geodesic distance	% increase
1	45%
2	20%
3	10%

Table 1: Testing null hypothesis for various geodesic distance

- 45% of FP had a friend in network (3604 friendships, 0.7 ties/ego)[1]

- 39% of FP had a co-worker in network (data from only after 2008)

On different types of FP and LP There may be 3 kinds of relations between FP and LP[2]

FP →LP (LP may not be in the network) samples of the friends graph

LP →FP (Both in network) samples of LP's friend only in LP's network

FP →LP (Both in network) samples of real friends but only within the network

1.4 Study on Obesity

Definition 1 *Obesity is defined in terms of body mass index(BMI). If $BMI > 30(Kg/m^2)$ then the person is called a obese person.*

The study is carried out by visualizing the network and following questions were answered.

- Degree Distribution : Not small world, Scale free

- Do Obese people cluster?

    - Homophily - "birds of feather "
    - Confounding- friends are likely to experience a shared context or event leading to same outcome $\Rightarrow$ by observational studies
    - Induction/Contraction: peer effect
    - Resonance (?)

- For each static wave in the probability of the trait occurring between linked node is higher than expected ?

    - Preserve prevalence of the trait
    - Preserve Topology of the network
    - Randomly shuffle trait assignment and get distribution of the occurrence of co-occurrences
        * ie. risk ratio $= \frac{P(trait|neighbor\_trait)}{P(trait|neighbor\_no\_trait)}$
        * test null hypothesis that the observed value till this distribution
        * Repeat this for various Geodesic distance(> 1 hop) (see Table 1)

---

[1] if the contact person is outside the network so the data is missing

[2] LP : Alter Egos or Linked person

Centrality How *important* is the node?

- What is it's Degree of centrality ? Higher degree is more central

- Betweenness: Number of short path go through the node

- Eigen Vector Centrality (explained below)

Eigen Vector Centrality Consider all friendship to be symmetric. Define $a_{ij} = 1$ if i and j are friends. For each person, the centrality is the sum of their neighbors centrality. ie.

$$\lambda X_i = a_{1i}X_1 + a_{2i}X_2 + \ldots + a_{ni}X_n$$

Rewrite such n equations as $\lambda \bar{X} = A^T \bar{X}$. The resulting principle Eigen vector yields Centrality of nodes

References

[1] J.T. Cacioppo, J.H. Fowler, and N.A. Christakis. Alone in the crowd: The structure and spread of loneliness in a large social network. *Journal of Personality and Social Psychology*, 97(6):977, 2009.

[2] N.A. Christakis and J.H. Fowler. The spread of obesity in a large social network over 32 years. *New England Journal of Medicine*, 357(4):370–379, 2007.

[3] N.A. Christakis and J.H. Fowler. The collective dynamics of smoking in a large social network. *New England journal of medicine*, 358(21):2249–2258, 2008.

[4] J.H. Fowler and N.A. Christakis. The dynamic spread of happiness in a large social network. *BMJ: British medical journal*, 337:a2338, 2008.