ETHlogo

**Lecture with Computer Exercises:**

**Modelling and Simulating Social Systems with MATLAB**

Project Report

|  |
| --- |
| Insert Title Here  … |

Name 1 & Name 2

Zürich

Date

**IMPORTANT**

**You MUST include the ETH declaration of originality here; it is available for download on the course website or at**

**http://www.ethz.ch/faculty/exams/plagiarism/index\_EN;**

**It can be printed as pdf and should be filled out in handwriting.**

**Agreement for free-download**

We hereby agree to make our source code of this project freely available for download from the web pages of the SOMS chair. Furthermore, we assure that all source code is written by ourselves and is not violating any copyright restrictions.

|  |  |
| --- | --- |
| Name 1 | Name 2 |

# Table of content

# Abstract

# Individual contributions

# Introduction and Motivations

**Introduction**

Obesity is a major health concern in the world and particularly in the US. In the US, it is estimated that obesity causes annual mortality of more than 300'000 deaths per year (Flegal, Williamson, Pamuk, & Rosenberg, 2004) despite the difficulty to precisely evaluate death directly linked to obesity. The World Health Organization defines obesity as a Body Mass Index higher than 30. WHO implemented in 2004 the WHO Global Strategy on Diet, Physical Activity and Health in order to cope with obesity. While WHO recognizes the importance of "supportive environments and communities" to fight obesity, most proposed solutions put the emphasis on an individual approach such as limiting the quantity of fat absorbed, increasing the consumption of fruits and vegetables or practising physical activity.

Smith and Christakis (Smith & Christakis, 2008) among others have revealed the importance of the social environment associated with the physical environment as a factor of good health hinting at public health interventions which should be elaborated in harmony with the social network. Christakis and Fowler (Christakis & Fowler, 2007) argue that obesity spread through social networks as they find that contacts with obese persons increase the likelihood to become obese. This controversial result has been debated by many including Shalizi and Thomas (Shalizi & Thomas, 2011) who point out the confusion made between homophily where subjects gather with those who resemble them, social contagion and confounding.

Hill et al. (Hill, Rand, Nowak, & Christakis, 2010) have extended the approach of Christakis and Fowler to allow for a transmission of obesity through social networks but also capture spontaneous non social contagion. They introduce a new model, the SISa model, derived from the classic SIS disease model to evaluate the relative effect of social and non social contagion. Hill et al. extend economic diffusion models which also include automatic diffusion by adding the possibility of recovery whereas economic diffusion are considered to be permanently acquired.

[Be clear on the research question]

It is essential to understand how people become obese to design adequate health interventions and to decide if obesity has to be tackled as a clinical issue on an individual basis or through social network to spread positive behaviour to fight obesity.

Therefore, we want to examine if social infection really plays a role in obesity spread and compare the relative effect of social and automatic infection in obesity. We want to test if we find similar results as Hill et al. applying the SISa model to a different sample.

To answer our question, we apply the SISa model to a dataset collected by Aharony et al. (Aharony, Pan, Ip, Khayal, & Pentland, 2011) which contains one of the largest mobile data experiments done in academia to test the validity of the SISa model and estimate the model parameters in order to evaluate the social and non social aspects of obesity contagion. Aharony et al. have deployed a sensing system over 15 months to follow 130 adult members and collected their physical activity, their weight and their friendship status. We have restricted our study to a subset of [85] adults for which complete corpulence data were available.

**Motivation**

[Cite previous report]

[TO BE DELETED AS THIS MIGHT CONFUSE READER]

They implemented three intervention schemes: (i) Control: people are rewarded according to their own physical activity (ii) Peer-review: people are rewarded according to their own physical activity and can see the physical activity of two "buddies" reciprocally, and (iii) Peer-Reward: people monitor both their buddies and personal physical activity but are rewarded solely according to the cumulative physical activity of their "buddies". Contrary to most fitness-related studies which recruit people who want to increase their physical activity, this study was designed as a non competitive game where a non active person can earn the same reward as an active person.

|  |  |
| --- | --- |
| **Number of subjects in each condition** | |
| Condition | Initial |
| Control | 18 |
| Peer-review | 45 |
| Peer-reward | 45 |

Table 1: Number of subjects in each intervention scheme

.

# Description of the Model

***SIS model***

We use an extension of the classic SIS model proposed by Hill et al. The SIS model is an adaptation of the SIR model developed by Kermack and McKendrick (Kermack & McKendrick, 1932). In the SIR model the population is divided into three groups: susceptible, infected, recovered. The disease is transmitted when a susceptible person enters in contact with an infected person with a so called transmission rate β. Once infected, a person can recover with a recovery rate g and then becomes completely immune to the disease. The SIS model allows to model a disease or a behaviour that can occur repeatedly meaning that recovering from the disease do not confer immunity.

In the standard SIS model, infection can only be transmitted through the contact with an infected person. The SIS model is presented in (1):

(1)

where is the number of infected individuals, and is the number of susceptible individuals, is the transmission rate, is the recovery rate and is the total population.

Hill et al. propose an extension of the SIS model to allow for automatic infection without having social contact. They introduce the rate of automatic infection rate to obtain the model described in (2):

(2)

The automatic infection rate and the transmission rate can be deduced from the transition probabilities from susceptible to infected after a time such that:

(3)

while the recovery rate can be deduced from the probability of transition from infected to susceptible after a time :

(4)

In addition, we follow Hill et al. approach and examine how a contact with an infected persons influence the transition between states.

# Implementation

**Data description**

In order to implement the SISa we searched for a longitudinal dataset with similar properties as the Framingham Heart Study dataset used by Hill et al. We used the Friends and Family dataset collected by Aharony et al. The dataset consists of 130 subjects who all belong to a young residential community with at least one family member affiliated to MIT. The data were collected from October 2010 to May 2011 and an intervention program was carried out from October to December 2010. We have therefore decided to restrict our study to a subsample of the data in which there was no intervention program in place and have focussed our study on the period from end of December 2010 to May 2011.

Another difficulty was to harmonize the different files of the Friends and Family dataset which had to be cross-checked.

**Measure timing**

We have encountered an issue regarding the timing of the examinations as subjects have not been surveyed at the same exact day as the friendship was measured. We have therefore interpolated data to solve for this problem.

**Network description and analysis**

Subjects reported their self-perceived closeness to other subjects on a scale from 0 to 7 where a number strictly higher than 2 stands for close friendship enabling us to build a binary adjacency matrix over time.

Furthermore, we have analyzed additional properties of the network, namely its degree distribution, its diameter as well as the density. We both use Matlab and Gephi to perform the network analysis and visualization.

**Ego change of state**

We follow Hill et al. to determine how contacts with non obese and obese people influence the transition to another state by performing longitudinal regression.

For each subject, we track the change in BMI and body fat and define a threshold above which a person changes state. While WHO defines obesity as a BMI higher than 30, our sample only contained 5 obese subjects out of a total of [108/90/85]. This percentage of obese egos is much lower than the percentage of obese adult in the US which was higher than 35% (Flegal, Carroll, Kit, & Ogden, 2012) and we have therefore decided to lower the threshold BMI above which a person changes state. As we want to evaluate the impact of having a contact with an obese person on the state of ego, we retrieve the number of obese contacts for each subject.

[**Regression**]

We follow Hill and al. approach and perform a regression of the probability of transitioning against the number of contacts which are in a particular state namely obese or non-obese.

Based on the results of the regression, we estimate the coefficients of the SISa model and run a simulation to predict the spread of obesity.

[**Simulation**]

# Simulation Results and Discussion

**Network characteristic**

The Friends and Family network consists of [85] nodes. The friends distribution varies across time as the self perceived closeness is measured at three different times. The degree distribution at t=1 is shown in Figure 1. The average degree distribution is quite stable over the measurement period standing at 6.7, 6.1, 7.4 for t=1...3. In comparison, the average degree amounted to 3 at the end of the Framingham Heart Study.

We test a power law to fit the degree distribution of the network as many social networks are considered to be scale free networks (Barabási, 2009) and find that the Friends and Family network exhibits characteristics of a scale-free network.

|  |
| --- |
|  |

Figure : Average degree distribution of close friends. The parameters of the power law fitted function y=αxβ are α=55 and β=-1.07.

We test two types of assortativity: social assortativity and weight assortativity. We run the regression of the average nearest neighbor degree against the average degree of a subject to test the social assortativity of the network. The Friends and Family network is socially assortative as shown in Figure 2 as subjects with a higher number of close friends tend to have a higher nearest neighbor degree. We find a positive coefficient on the average degree equal to 0.34 and significant at the 5% level.



Figure : Friends and Family Network Social Assortativity.

However we do not obtain significant results to support weight assortativity when running the regression of the average nearest neighbor weight against a subject's weight (see Figure 3). The coefficient of the regression slope of the average nearest neighbor weight against a subject weight is not significant at any conventional level. We therefore conclude that a subject weight is not associated with its average nearest neighbor weight.



Figure : Friends and Family Network Weight Assortativity

[Regression result]

* be clear on cause
* Have clear figures

[Simulation either based on actual results or not]

[Society implication]

Limitations

* + [Homophily, social influence]

Despite Aharony et al. claims, the Friends and Family network does not appear to be representative of the US population as all subjects were members of a residential living community living next to MIT. Flegal et al. (Flegal et al., 2012) have found a prevalence of obesity of 35% among US adults. In comparison, the mean BMI in the Friends and Family network is 23 and 5 out of the [85] subjects are obese in the sample at the beginning of the study.



Figure : BMI distribution of the Friends and Family network. A lognormal distribution is fitted to the BMI data with a mean of 3.13 and standard deviation of 0.18.

# Summary and Outlook

# References

Aharony, N., Pan, W., Ip, C., Khayal, I., & Pentland, A. (2011). Social fMRI: Investigating and shaping social mechanisms in the real world. *Pervasive and Mobile Computing*, *7*(6), 643–659.

Barabási, A.-L. (2009). Scale-free networks: a decade and beyond. *Science*, *325*(5939), 412–413.

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

Flegal, K. M., Carroll, M. D., Kit, B. K., & Ogden, C. L. (2012). Prevalence of obesity and trends in the distribution of body mass index among US adults, 1999-2010. *Jama*, *307*(5), 491–497.

Flegal, K. M., Williamson, D. F., Pamuk, E. R., & Rosenberg, H. M. (2004). Estimating deaths attributable to obesity in the United States. *American Journal of Public Health*, *94*(9), 1486.

Hill, A. L., Rand, D. G., Nowak, M. A., & Christakis, N. A. (2010). Infectious disease modeling of social contagion in networks. *PLoS Computational Biology*, *6*(11), e1000968.

Kermack, W. O., & McKendrick, A. G. (1932). Contributions to the mathematical theory of epidemics. II. The problem of endemicity. *Proceedings of the Royal Society of London. Series A*, *138*(834), 55–83.

Shalizi, C. R., & Thomas, A. C. (2011). Homophily and contagion are generically confounded in observational social network studies. *Sociological Methods & Research*, *40*(2), 211–239.

Smith, K. P., & Christakis, N. A. (2008). Social networks and health. *Annu. Rev. Sociol*, *34*, 405–429.