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 especially in the US. In the US, it is estimated that obesity causes annual mortality of around 300'000 deaths per year (Flegal, Williamson, Pamuk, & Rosenberg, 2004) despite the difficulty to precisely evaluate death directly linked to obesity. A Body Mass Index higher than 30 corresponds to obesity according to World Health Organization standards. WHO implemented in 2004 the WHO Global Strategy on Diet, Physical Activity and Health [2]. While WHO recognizes the importance of "supportive environments and communities" to fight obesity, most proposed solutions rather focus on an individual basis such as limiting the quantity of fat absorbed, increasing the consumption of fruits and vegetables or practising physical activity. Few emphasis is put on the potential obesity spreads through social networks. Hill et al. (Hill, Rand, Nowak, & Christakis, 2010) have studied the social contagion of obesity which differs from traditional epidemiological disease. Hill et al. have introduced a new model (SISa) derived from the classic SIS disease model in which they allow for automatic non-social infection. 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 elaborated in harmony with the social network. Before Hill et al., epidemiological were applied to study social contagion and may fail to capture automatic infection. Hill et al. extend economic diffusion models by including the possibility of recovery. To fight obesity, adequate public health policies must be designed to decide if obesity has to be tackled as a clinical issue (i.e. on an individual basis) or as a public health intervention which could better exploit the network phenomena to spread positive behaviour to fight obesity.

[Be clear on the research question]

We want to examine if social infection plays role in obesity spread and if this role is prominent. In addition

We will apply the SISa model to a new 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 model and estimate the model parameters. We follow Hill et al. to determine how contacts with non obese and obese people influence the transition to another state.

We evaluate different social intervention schemes proposed by Aharony et al. We use a subset of the "Friends and Family" dataset where 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. The studied subset consisted of [85] persons for which complete weight data were available.

We created an adjacency matrix based on the self perceived network friendship where each participant rated every other participant on a scale from 0 (not familiar) to 7 (very close) at different step in time

**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 : 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 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 model infection can only be transmitted through the contact with an infected person. The SIS model is presented in equation (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 equation (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 keeping in mind the arguments of Shalizi

et al. (Shalizi & Thomas, 2011) that homophily or covariation of another variable are competing with social influence.

# Implementation

**Data description**

In order to implement the SISa we tried to find a longitudinal dataset with similar properties as the Framingham Heart Study dataset similarly to Hill et al. We used the Friends and Family dataset collected by Aharony et al. The dataset consist 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. In the intervention program subjects were classified into three groups Control, Peer-review, Peer-reward. For each subject, weight, BMI, body fat and skeletal muscle information were collected at different period of time.

Subjects also reported their closeness to other subjects on a scale from 0 to 7 and where a number higher than 2 characterizes close friendship enabling us to build an adjacency matrix at four different time.

One of the major challenges faced with the data was that our dataset on physical condition incorporated the effect of the intervention program thereby introducing a bias which make a direct comparison with Hill et al. results difficult. We have therefore decided to restrict our study to a subsample of the data in which there was no intervention program in place.

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

**Timing**

We have encountered an issue regarding the timing of the examinations as subjects have not been surveyed at the same exact day on the contrary to friendship which was measured at the same time for all subjects. This was a minor obstacle to assess the number of contacts with obese people at a given time.

**Network description and analysis**

We first determined the adjacency matrix of the network where a 1 stands for close self reported friendship whereas 0 stands for an absence of friendship.

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

**Ego change of 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 in our sample is much lower than the percentage of obese adult in the US in 2008 which was higher than 30% (Scully, 2014). Our sample is therefore not representative of the average American person.

For each subject, we track the change in BMI and body fat and define a threshold 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.

Our sample size is not as big as in the Framingham Heart Study and the timescale is reduced as the Friends and Family dataset of larger study containing additional valuable information about subjects.

[**Simulation**]

# Simulation Results and Discussion

**Network characteristic**

The Friends and Family close friends distribution varies across time as the self perceived closeness is measured at four different times. It consists of [85] nodes. The degree distribution are shown in with the average degree distribution is quite stable over the four period of measurement 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 [Add reference]

|  |
| --- |
|  |

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 run the regression of the average nearest neighbor degree against the average degree of a subject in to order to evaluate 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.

However we do not obtain results to support obese 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.



Figure : Friends and Family Network Social Assortativity.



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]
  + [Sample not representative of the society]->[BMI distribution]

# 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.

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.

Scully, T. (2014). Public health: Society at large. *Nature*, *508*(7496), S50–S51. Retrieved from http://dx.doi.org/10.1038/508S50a

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.