



Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Lecture with Computer Exercises: Modelling and Simulating Social Systems with MATLAB

Project Report

The social contagion of obesity

Amr Ahmed, Elise Ledieu, Emmanuel Munich & François
Wirz

Zurich
May 2014

Agreement for free-download

We hereby agree to make our source code for 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.

Amr Ahmed, Elise Ledieu, Emmanuel Munich, François Wirz



Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zürich

Declaration of originality

The signed declaration of originality is a component of every semester paper, Bachelor's thesis, Master's thesis and any other degree paper undertaken during the course of studies, including the respective electronic versions.

Lecturers may also require a declaration of originality for other written papers compiled for their courses.

I hereby confirm that I am the sole author of the written work here enclosed and that I have compiled it in my own words. Parts excepted are corrections of form and content by the supervisor.

Title of work (in block letters):

MSST/FS14 project report: The social contagion of obesity

Authored by (in block letters):

For papers written by groups the names of all authors are required.

Name(s):

AMMÉ

LEDIEU

NUÑEZ

WIRZ

First name(s):

Amn

Elise

Emmanuel

François

With my signature I confirm that

- I have committed none of the forms of plagiarism described in the 'Citation éthiquette' information sheet.
- I have documented all methods, data and processes truthfully.
- I have not manipulated any data.
- I have mentioned all persons who were significant facilitators of the work.

I am aware that the work may be screened electronically for plagiarism.

Place, date

12/05/2014

Signature(s)

Amn

Elise

Emmanuel

François

For papers written by groups the names of all authors are required. Their signatures collectively guarantee the entire content of the written paper.

Contents

1	Abstract	5
2	Individual contributions	6
3	Introduction and Motivations	7
3.1	Introduction	7
3.2	Motivations	8
4	Description of the Model	9
4.1	SIS model	9
4.2	SISa model	9
5	Implementation	10
5.1	Data description	10
5.1.1	Measure timing	10
5.1.2	Network description and analysis	10
5.1.3	Ego change of state	11
5.2	Regression	11
6	Simulation Results and Discussion	12
6.1	Network characteristic	12
6.2	Regression results	14
6.3	Simulation results	15
6.4	Importance of the α parameter in a SISa model	15
6.5	Society implication	15
6.6	Limitations	16
7	Summary and Outlook	18
7.1	Summary	18
7.2	Outlook	19
8	References	20
9	Annexes	22
9.1	Code listing	22

1 Abstract

Obesity is a major health concern in the US causing annual death of more than 300'000 deaths per year. World Health Organization has implemented intervention program to fight obesity. Understanding how people become obese is key to design adequate intervention program. In 2007, Christakis and Fowler claimed that obesity spread through social networks and argued that contacts with obese persons increase the likelihood of becoming obese. This claim has been subject to much controversy and was refuted by Shalizi and Thomas who highlighted the confusion made between social contagion, homophily and confounding. Hill et al. have extended the findings of Christakis and Fowler and proposed a derived version of the SIS model, the SISa model, to allow for both social transmission of obesity and automatic infection. Using a network of 85 subjects living next to MIT for whom longitudinal weight data were available, we try to fit a SISa model. However, we do not find any evidence of social contagion for obesity using conventional significance level keeping in mind the relatively limited size of our sample.

2 Individual contributions

We initially familiarized ourselves with the current literature and the SIS/SISa model. We agreed on the implementation and divided it into tasks. Emmanuel Munich wrote the code to read the data and perform the necessary preprocessing to it. Amr Ahmed extracted useful information from the data (like adjacency matrices). Elise Ledieu coded the regression model, while Francois plotted the Gephi plots and was responsible for integrating different pieces of code produced by different group members together. Finally, we all participated in plotting the results, discussing them and drawing conclusions from them.

3 Introduction and Motivations

3.1 Introduction

Obesity is a major health concern in the world and particularly in the US. In the US, it is estimated that obesity causes more than 300'000 deaths annually (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. In 2008, 507 Million adults were obese, see Figure 1.

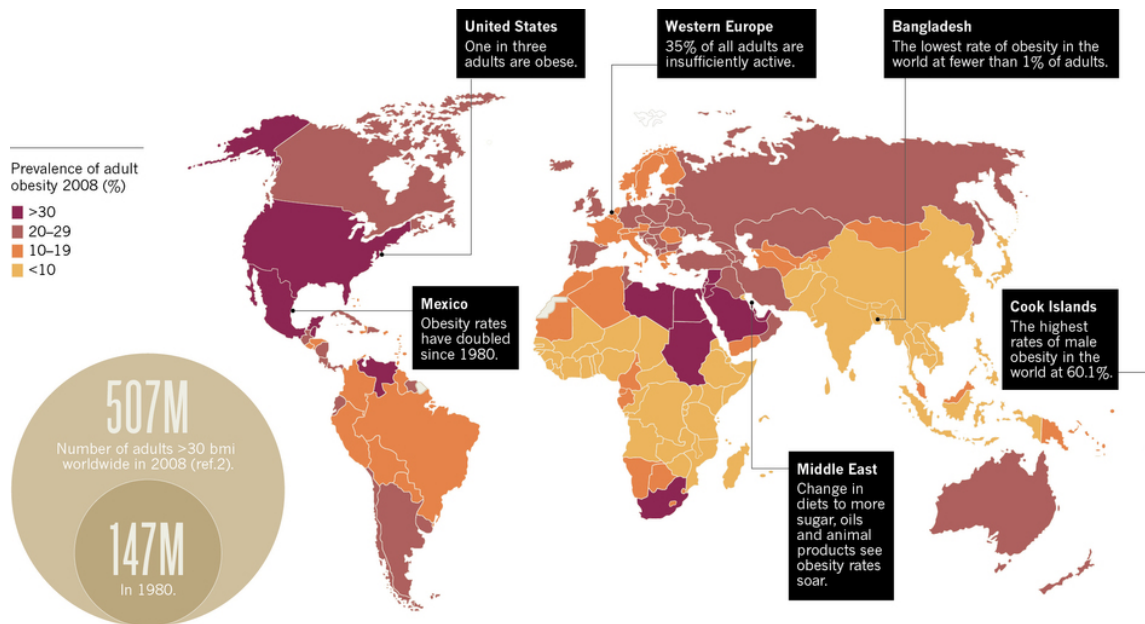


Figure 1: Map illustrating global spread of obesity.

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) highlight 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) among others 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.

3.2 Motivations

It is essential to understand how people become obese to design adequate health interventions and 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 over a period of five months was available.

4 Description of the Model

4.1 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):

$$\begin{aligned}dI/dt &= \beta SI - gI \\dS/dt &= -\beta SI + gI \\I + S &= N\end{aligned}\tag{1}$$

where I is the number of infected individuals, and S is the number of susceptible individuals, β is the transmission rate, g is the recovery rate and N is the total population.

4.2 SISa model

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

$$\begin{aligned}dI/dt &= \beta SI - gI + aS \\dS/dt &= -\beta SI + gI - aS \\I + S &= N\end{aligned}\tag{2}$$

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

$$P(S \rightarrow I, \delta t) \sim (a + \beta n)\delta t\tag{3}$$

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

$$P(I \rightarrow S, \delta t) \sim g\delta t \quad (4)$$

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

5 Implementation

5.1 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 to avoid any bias and have therefore excluded the period running from October 2010 to December 2010.

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

5.1.1 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 the data to solve this problem.

5.1.2 Network description and analysis

Subjects reported their self-perceived closeness to other egos 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 where a 1 stands for close friendship.

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 (see Figure 2).

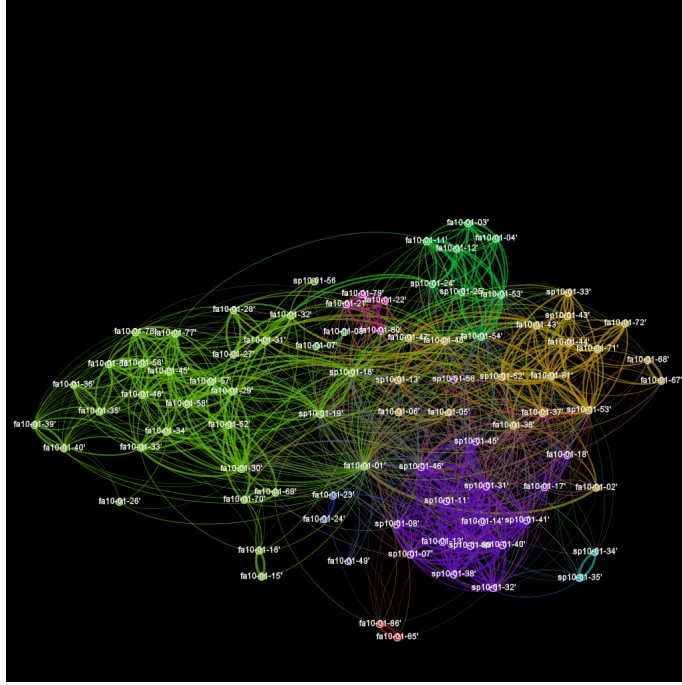


Figure 2: Initial state of the social network of the study participants. The thickness of the links translate the importance of the relation and the Atlas2 clustering algorithm was run.

5.1.3 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 85. This percentage of obese egos is much lower than the percentage of obese adult in the US which was higher than 35% (Flegal, 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.

5.2 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.

6 Simulation Results and Discussion

6.1 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 the Figure 3. The average degree distribution is quite stable over the measurement period standing at 6.7, 6.1, 7.4 for $t=1\dots3$. In comparison, the average degree amounted to 3 at the end of the Framingham Heart Study.

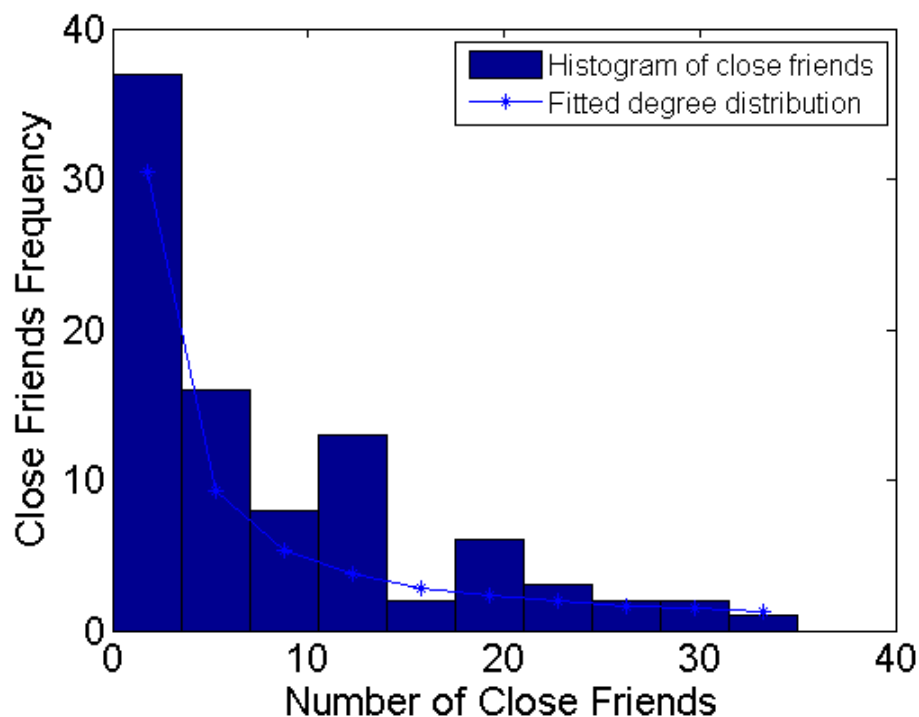


Figure 3: Average degree distribution of close friends

We test a power law to fit the degree distribution of the network as many social networks are considered to be scale free networks (Barabasi, 2009) and find that the

Friends and Family network exhibits characteristics of a scale-free network.

The parameters of the power law fitted function $y=\alpha x^\beta$ are $\alpha = 55$ and $\beta = -1.07$.

We then 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 4 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.

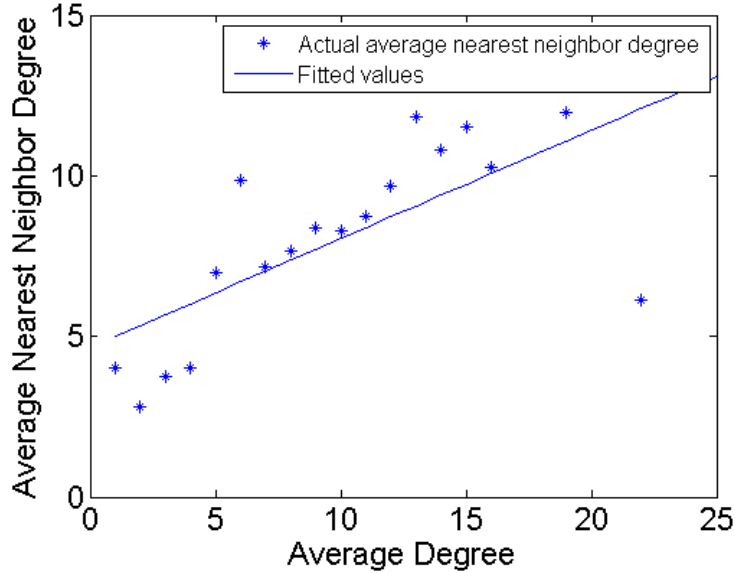


Figure 4: Friends and Family Network Social Assortativity.

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 5). 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.

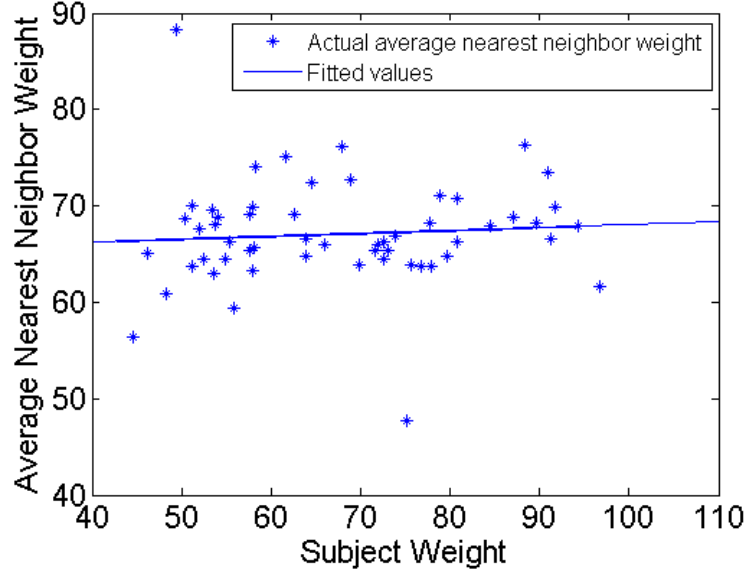


Figure 5: Friends and Family Network Weight Assortativity.

6.2 Regression results

Our findings relate in parts to the outcome of the analysis of Christakis. We solved to some extent the question raised in his talk at Dartmouth College in 2008 concerning the results in a different socio-economic environment, where perception of obesity is different. But when we used the body fat metric which is reportedly a more precise measure for obesity, we surprisingly did not encounter neater proof of the infectious characteristic of obesity, as we find a negative correlation. What's more, we determined in multiple tests that running the regression with even a small shift in the threshold value triggering a change of state, yielded a great impact in the outcome, even negating the correlation when ran against the body fat measure. This surprising mismatch between BMI and body fat, which normally relate to each other in terms of the following approximation formula for an adult (Jackson, 2002):

$$AdultBodyFat\% = (1.20 * BMI) + (0.23 * Age) - (10.8 * gender) - 5.4$$

, is, as far as we can tell, due to the remarkably high skeletal muscle percentage of the total weight reported by the study participants (Janssen, 2000). This reflects the fact that the study participants were drawn from a university setting, where university athletes are overrepresented compared to the general population.

To determine the parameters of the differential systems, we have made a regression over the percentage of change from one state to another given the number of contacts in the target state. The obtained p-values are high. The smaller p-value we obtain is indeed 0.306 for the study of the BMI with a threshold put at 25. The estimated parameters therefore do not seem reliable.

Overall, we could not prove the thesis that obesity and overweight spread like a contagious disease nor refute it, and prove that the healthy state spreads, as in both ways the correlation was too small, especially given our small study sample of 85 subjects and the short time scale of the study which lasted less than five months.

6.3 Simulation results

The estimated parameters of the regression yielded a long term prevalence of obesity when inspecting BMI. When we looked at the body fat percentage, the outcome was different, with the healthy state of normal weight becoming dominant.

6.4 Importance of the a parameter in a SISa model

The evolution of the number of persons above a given threshold of BMI is strongly dependant on the threshold given but tend to grow. In order to observe if the SISa model could fit this data with any parameter, we went through a range of parameters, chosen with respect to the ones found by Christakis and Fowler. The SISa model with a maximum correlation corresponds to the one with $\beta=0$ and $g=0$, hinting at a prevalence of automatic contagion over social contagion. Thus, the role played by the parameter a is not to be undermined and eliminates the SIS model as a possible description of the observed behaviour.

To tell things upfront, the SISa model is very sensible to the value of each single value, with the a parameter being the most prevalent. As we inferred from our multiple test, a single decimal change can yield noticeable different outcomes.

6.5 Society implication

The society implication of clearly understanding to what extent and in which manner social spreading of obesity happens are huge. Given that it is morally unacceptable to marginalize obese people and even the small direct correlation is hard to link causation, we prone a more holistic approach where obesity is seen as a risk factor as opposed to a disease. Thus this risk factor has to be contained through environmental reforms on one hand like easier access to sport facilities, bike to work

initiatives, promotion of healthier food alternatives in canteen settings, and social reforms on the other hand where risk reduction due to obesity is rewarded and scheduled physical activity is integrated in the daily work routine (Katz, 2014). Politics also have to play a role as they did with the tobacco advertisement restrictions and warning by translating these efforts to the junk food advertisement restrictions and clearer formulated food declaration, lobbying for "junk food free" as they did for smoke free¹.

6.6 Limitations

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. The BMI distribution of the Friends and Family network does not reflect the BMI distribution of the entire US population (see Figure 6).

¹<http://www.smokefree.ch/en/campaign/>

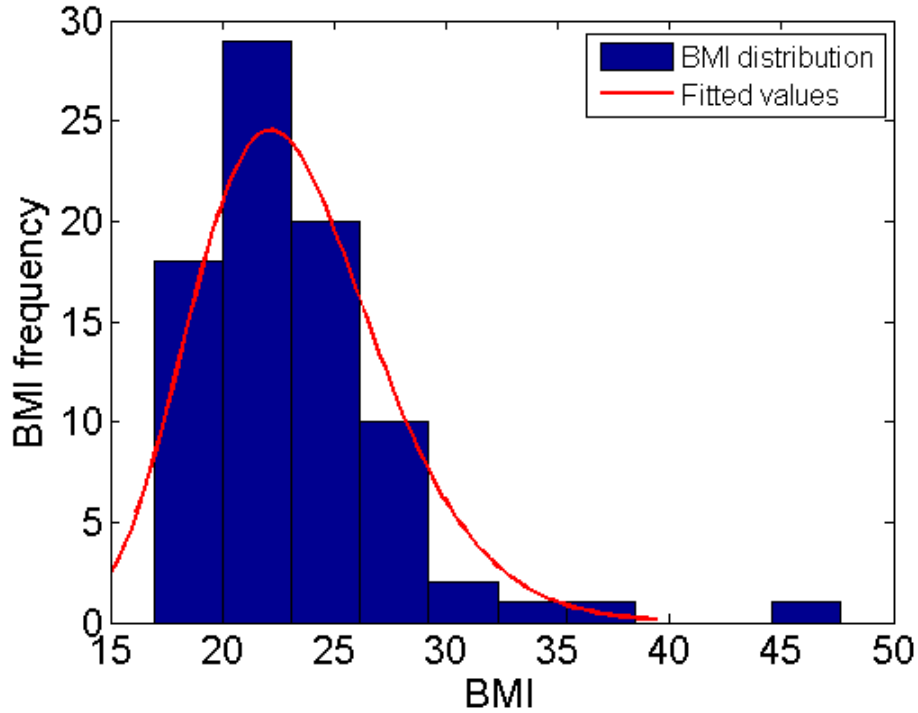


Figure 6: BMI distribution of the Friends and Family network. A lognormal distribution is fitted to the BMI data with a mean of 3.13 and standard deviation of 0.18.

Flegal et al. 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.

Another factor that a model purely based on social contagion does not capture is the alimentary culture that is not directly related to social interaction but largely shared across a social group. Some populations, and therefore the social networks they span, are much more at risk of a deregulation of equilibrium in their macronutrient intake, struggling to cover their protein needs while increasing their overall calories consumption.

We take into account Shalizi and Thomas comments regarding environmental confounding and decided to examine the role of socio-environment in obesity. Using data from the US Department of Agriculture, we run a regression of the 2010 obesity prevalence in 3'137 US counties against the median household income in USD'000

of these same counties. These results highlight the role of confounding and support the argument of Shalizi and Thomas as we find a negative coefficient on the median household income in USD'000 equal to -0.19 and statistically significant at all conventional level (see Table 1). This reveals that a thousand dollar of additional income is associated with a reduction of 0.19% of obesity prevalence. One possible explanation for this is that low income families might consume fatter food of lower quality because of budget constraints.

Variable	Coefficient	(Std. Err.)
Median Household Income (USD'000)	-0.188 (***)	(0.006)
Intercept	38.7 (***)	(0.27)

Table 1: Results of OLS Model of the Obesity Prevalence in the US

There are also behavioral reasons behind obesity. The increased urbanization of the world tends to affect people's habits, who tend to eat more and exercise less. Also diet has changed significantly compared to the past, as we eat more energy-dense foods containing more fat and sugar while reducing our intake of complex carbohydrates and dietary fibres (Anthes, 2014). Another reason behind obesity is people's sleeping habits. Sleep deprivation can alter the levels of hormones involved in metabolism and appetite (Anthes, 2014). Finally, genetic factors also play a role in obesity. Willyard (Willyard, 2014) suggested that 40–70% of variation in body size is due to genetic factors.

7 Summary and Outlook

7.1 Summary

In order to better understand one of the major plague of our society, we have applied the SISa model proposed by Hill et al. to a sample of subjects living in a community next to MIT. Our aim was to evaluate the extent to which obesity spread through social networks. However when evaluating the SISa model parameters, neither do we find any evidence that obesity spread through social networks nor can we refute this. There are two main reasons accounting for this. The first is that our results are highly influenced by the nature of our sample as there are only few datasets containing large longitudinal personal information on obesity and publicly accessible. The second is that the claim of obesity spread through social network has been subject to virulent controversy and our results may corroborate the doubts raised by Shalizi and Thomas. While we cannot conclude on weight homophily in

our sample, we exhibited the role of social environment, namely household income in our case, in becoming obese.

7.2 Outlook

Given the importance of obesity, we believe it is essential to account for the roles of confounding variables in obesity spread. We do not exclude that obesity may spread through social network and think it is critical to perform a new study based on a dataset comparable to the Framingham Heart Study to clarify the controversy surrounding the potential social contagion of obesity. The number of deaths linked directly or indirectly to obesity as well as the high social costs of obesity (Cawley & Meyerhoefer, 2012) are such that this issue needs to be tackled.

8 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.

Anthes, E. (2014). Treatment: Marginal gains. *Nature*, 508(7496), S54-S56.

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

Cawley, J., & Meyerhoefer, C. (2012). The medical care costs of obesity: an instrumental variables approach. *Journal of health economics*, 31(1), 219-230.

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.

Jackson AS, Stanforth PR, Gagnon J, et al. (2002). The effect of sex, age and race on estimating percentage body fat from body mass index: the Heritage Family Study. *Int J Obes Relat Metab Disord*, 26(6):789-96

Janssen, I., Heymsfield, S. B., Wang, Z., Ross, R. (2000). Skeletal muscle mass and distribution in 468 men and women aged 18-88 yr. *Journal of Applied Physiology*, 89(1), 81-88.

Jordi, Schmid, & Stcheli (2012). Stochastic Simulation of Epidemic Outbreak in a large Network. ETH Course MSSSM FS12

Katz, D. L. (2014). Perspective: Obesity is not a disease. *Nature*, 508(7496), S57S57.

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), 5583.

Shalizi, C. R., & Thomas, A. C. (2011). Homophily and contagion are generically confounded in observational social network studies. *Sociological Methods Research*, 40(2), 211239.

Sotcker, Borer, Pluess, & Buehler (2012). Cholera Epidemic in Haiti 2010. ETH Course MSSSM FS12

Smith, K. P., & Christakis, N. A. (2008). Social networks and health. *Annu. Rev. Sociol*, 34, 405429.

USDA's Economic Research Service (2014). Food Environment Atlas Data Documentation. <http://www.ers.usda.gov/data-products/food-environment-atlas/data-access-and-documentation-downloads.aspx>

Willyard, C. (2014). Heritability: The family roots of obesity. *Nature*, 508(7496), S58-S60.

9 Annexes

9.1 Code listing

File Project_script.m

```
1     load('DATA.mat')
2     prompt = 'What variable do you want to study? Type BMI or body_fat ';
3     studied_variable = input(prompt);
4     prompt = 'Choose a meaningful threshold for the variable: ';
5     threshold = input(prompt);
6     % Generate adjacency matrix of subjects
7     [adj_mat,adj_mat_bin,unique_source,unique_target] = mssmm_adj ...
        (participantID.weight,participantID.weight, ...
        source_friendship,date_friendship.can);
8     % Degree distribution in the network
9     [contacts,avg_degree] = degree_distribution(adj_mat_bin);
10    disp(avg_degree);
11    % Is the ego above the threshold at the exam:
12    PersonID_x_AboveAtExam = persons_above.interpolated( ...
        participantID.weight,studied_variable, date_weigth.can,threshold);
13    %Plots the evolution of the number of person above
14    figure(1)
15    plot(sum(PersonID_x_AboveAtExam)/size(PersonID_x_AboveAtExam,1))
16    % When did a transition from one state to the next happen and which way:
17    PersonID_x_ChangedStateAtDayX = persons_changed.state( ...
        PersonID_x_AboveAtExam );
18    % Number of contacts above the threshold at the exam:
19    [number_of_contacts] = count_obese_contacts(PersonID_x_AboveAtExam, ...
        adj_mat_bin);
20    % Percentage of total transitions happening given a count of contacts
21    % above the threshold:
22    change_under = mssmm_regression(PersonID_x_ChangedStateAtDayX, ...
        number_of_contacts, sum(adj_mat_bin(:, :, 1)), true)
23    change_above = mssmm_regression(PersonID_x_ChangedStateAtDayX, ...
        number_of_contacts, sum(adj_mat_bin(:, :, 1)), false)
24    % regression on the parameters
25    [above b g a] = mssmm_regression_results(change_above, change_under)
26    % plot regression results
27    figure(2);
28    plot(change_above(:, 1), change_above(:, 2))
29    hold on;
30    x = change_above(:, 1);
31    y = b*x+a;
32    plot(x, y*100, 'o');
33    hold off;
34    initial_infected = sum(PersonID_x_AboveAtExam(:, 1)) * (100 / ...
        size(PersonID_x_AboveAtExam, 1));
35    figure(3);
```

```

36     [susceptible,infected] = simulation(a,b,g, initial_infected, ...
        size(PersonID_x_AboveAtExam,2), 0.01);
37
38 % Discussed tests: studied_variable=BMI, threshold=26: slight correlation,
39 % ¬stable state at 75% infected
40 % Second test: studied_variable=body_fat, threshold=25: apparent ...
        negative correlation,
41 % ¬stable state at 15% infected
42 % Third test: studied_variable=BMI, threshold=25: some correlation,
43 % ¬stable state at 90% infected
44 % Fourth test: studied_variable=body_fat, threshold=25: apparent ...
        negative correlation,
45 % ¬stable state at 12% infected

```

File mssmm_adj.m

```

1  function ...
    [adj_mat,adj_mat_bin,unique_source,unique_target,unique_funf_group] ...
    = mssmm_adj(source,target,source_friendship,date_friendship_can)
2  %% Returns the adjacency matrix, a binary adjacency matrix, a vector ...
    of unique elements of source and target
3  % Input:
4      % source is the vector source in FunFit Excel file, target is
5      % the vector source in FunFit Excel file, source_friendship is the
6      % vector source_friendship in SurveyFriendship Excel file, ...
        date_friendship is the
7      % vector date_friendship in SurveyFriendship Excel file
8      % self-reported friendship ranging from 0 to 7. The binary
9      % adjacency matrix displays 1 when the the self reported
10     % friendship > 2
11 % Output:
12     % adj_mat is a 3D adjacency matrix, the first two ...
        dimensions corresponds to
13     % the person and the third dimension to the date of friendship
14     % unique_source is the vector of unique source (idem for
15     % unique_target)
16
17     load('DATA.mat')
18     %% Vector of unique source
19     unique_source = unique(participantID.weight);%zeros(1,1);
20     number_of_legos = length(unique_source) % for logging
21     %% Vector of unique target
22     unique_target = unique(participantID.weight);%zeros(1,1);
23     %length(unique_target)
24
25     %% Return the experiment to which a person belongs ...
        (Control,See-Peer,Pay-Peer)
26     % unique_funf_group = cell(length(unique_source),1);

```

```

27     % for i = 1:length(unique_source)
28     %     funf_group_index = find(ismember(source,unique_source(i)));
29     %     unique_funf_group{i} = funf_group{funf_group_index};
30     % end
31
32     %% Creation of the adjacency matrix
33
34     date = unique(date_friendship_can);
35     adj_mat = ...
36         zeros(length(unique_source),length(unique_target),length(date));
37     adj_mat_bin = ...
38         zeros(length(unique_source),length(unique_target),length(date));
39
40     for i = 1:length(source_friendship)
41
42         % source_friendship_index returns index of person in the ...
43         % adjacency matrix
44         source_friendship_index = ...
45             find(ismember(unique_source,source_friendship(i)));
46         target_friendship_index = ...
47             find(ismember(unique_target,target_friendship(i)));
48         date_index = find(ismember(date,date_friendship_can(i)));
49
50         % Adjacency matrix element value equal to weight_friendship
51         % value ranges from 0 to 7
52         adj_mat(source_friendship_index, ...
53             target_friendship_index,date_index) = weight_friendship(i);
54         if weight_friendship(i)>2
55             adj_mat_bin(source_friendship_index, ...
56                 target_friendship_index, date_index) = 1;% binary ...
57                 adjacency matrix 1
58         else
59             end
60     end
61 end

```

File degree_distribution.m

```

1 function [contacts,avg_degree] = degree_distribution(adj_mat_bin)
2     %% Returns the degree distribution (distribution of contact numbers) ...
3     % of the network at each time
4
5     % Input:
6     %     % Binary adjacency matrix: 1 if person are close friends and 0
7     %     % otherwise
8
9     % Output:

```



```

8         % number_of_contacts matrix with the number of contacts of ...
          each person at the
9         % each time
10        % The plots displays the degree distribution histogram
11
12    date = size(adj_mat_bin,3); % number of date
13    len_source = size(adj_mat_bin,1); % number of persons
14
15    contacts = zeros(len_source,date);
16    avg_degree = zeros(date,1);
17
18    for t = 1:date
19        for k = 1:len_source
20            contacts(k,t) = sum(adj_mat_bin(k,:,t)); % Number of close ...
                friends of person k at time t
21        end
22
23        total_contacts = sum(contacts(:,t));
24
25        % figure
26
27        % bar(1:max(number_of_contacts(:,t)), ...
            1/total_contacts*histc(number_of_contacts(:,t), ...
            1:max(number_of_contacts(:,t)),'histc')% Histogram of ...
            degree distribution
28
29        % figure
30
31        avg_degree(t,1) = total_contacts/len_source;
32        % hist(number_of_contacts(:,t))% Histogram of degree distribution
33        % xlabel('Number of close friends')
34        % ylabel('Close friends frequency')
35        % title(['Close friends distribution t = ...
            ',num2str(t)], 'FontSize',10, 'fontweight','bold')
36
37    end
38
39 end

```

File persons_above_interpolated.m

```

1 function PersonID_x_AboveAtExam = persons_above_interpolated( persons, ...
    vector_measure, date_measure, threshold)
2     % SUMMARY: persons_above computes for each unique person if they are ...
        above the
3     %threshold at a given day.
4     %
5     % OUTPUT:: PersonID_x_AboveAtExam : Returns a unique-persons by

```

```

6         % first_date_of_measure_to_last_date_of_measure matrix with ...
           value 1 if above
7         % the threshold and 0 else.
8     % INPUT::
9         % persons: a vector of cells of personID strings,
10        % vector_measure: a vector of numeric values, where value at row x
11        % corresponds to person at row x.
12        % date_measure: a vector of dates in integer form, where value ...
           at row x
13        % corresponds to measure at row x.
14        % threshold: a threshold above which the state is 1.
15
16    egos = unique(persons); % different persons
17    duration = max(date_measure)-min(date_measure) + 1; % Time elapsed ...
           from first measurement (inclusive) to last measurement (inclusive)
18    date_measure = date_measure - min(date_measure) + 1; % normalize ...
           date.weight
19
20    PersonID_x_AboveAtExam = zeros(length(egos),duration); % make empty
21
22    %warning('off', 'MATLAB:polyfit:RepeatedPointsOrRescale'); % ...
           temporarily turn off those warnings about badly conditioned ...
           polynomials
23
24    for i= 1:length(egos)
25        personInd = find(ismember(persons,egos(i))); % personInd = ...
           person indexes
26
27        measureDateEgo = date_measure(personInd); % We need the indexes ...
           corresponding to a person sorted by date, in order to ...
           compute the final matrix in one pass per person.
28        dates_measures = [measureDateEgo, personInd]; % Prepend date ...
           vector to index vector
29        dates_measures = sortrows(dates_measures); % Sort after date vector
30        mbd = dates_measures(:,2); % extract sorted by date index vector ...
           of the measures i.e. MeasuresByDate
31        measureDateEgo = dates_measures(:,1); % dates sorted
32
33        person_measures = vector_measure(mbd); % all the measures of an ego
34        % remove multiple measurements that were taken on the same day to
35        % avoid badly conditioned polynomial
36        [measureDateEgo, index_measureDateEgo, ~] = ...
           unique(measureDateEgo, 'stable');
37        person_measures = person_measures(index_measureDateEgo);
38        measure_count = length(measureDateEgo);
39        degree = min(measure_count-1, 1); % restrict polynomial to degree ≤ ...
           2 to avoid edge effects
40        p = polyfit(measureDateEgo,person_measures,degree);
41        x = 1:duration;
42        y = polyval(p,x);

```

```

43         plot(x,y);
44
45
46         y(y ≤ threshold) = 0; % binary matrix: 0 -> value below ...
            threshold at this date for this person
47         y(y > threshold) = 1; % binary matrix: 1 -> value above or equal ...
            threshold at this date for this person
48         PersonID_x_AboveAtExam(i,:) = y;
49     end
50 end

```

File persons_changed_state.m

```

1  function PersonID_x_ChangedStateAtDayX = persons_changed_state( ...
    PersonID_x_AboveAtExam )
2  %SUMMARY: persons_changed_state computes for each unique person if they ...
    changed
3  %state with respect to previous day.
4  %
5  % OUTPUT:: PersonID_x_ChangedStateAtDayX : Returns a unique_persons by
6      % first_date_of_mesure_to_last_date_of_mesure matrix with value ...
        0 if the person doesn't changed state with respect to ...
        previous day.
7      % If the person state changed from under the threshold to above the
8      % threshold on the next day, the value is -1.
9      % If the person state changed from above the threshold to under the
10     % threshold on the next day, the value is 1.
11 % INPUT::
12     % PersonID_x_AboveAtExam
13
14 [m,n] = size(PersonID_x_AboveAtExam);
15
16 PersonID_x_ChangedStateAtDayX = ...
    [zeros(m,1),PersonID_x_AboveAtExam(:,1:n-1)] - ...
    [zeros(m,1),PersonID_x_AboveAtExam(:,2:n)];
17
18 end

```

File count_obese_contacts.m

```

1  % function returns a matrix 'number_of_obese_contacts'
2  % that has the number of obese contacts each person has at every timestep
3  % weight_overtime: entry (i,j) represents whether person i is obese at
4  % time j (entry in the matrix is one) or not (entry is zero)
5  % adj matrix, entry (i,j) represents whether person i knows person j or not

```

```

6 function [number_of_obese_contacts] = ...
    count_obese_contacts(weight_over_time, adj_matrix)
7     number_of_persons = size(weight_over_time,1);
8     time_steps = size(weight_over_time,2);
9     number_of_obese_contacts = zeros(number_of_persons, time_steps);
10    for i = 1:number_of_persons
11        for j = 1:number_of_persons
12            if adj_matrix(i,j) ≥ 1      %check if person i is in relation ...
                with person j
13                for k = 1:time_steps    % at any time step if person 'j' ...
                    is obese, increment the number of obese contacts of ...
                    person 'i' by one at that time step
14                    if weight_over_time(j,k) ≥ 1
15                        number_of_obese_contacts(i,k) = ...
                            number_of_obese_contacts(i,k) + 1;
16                    end
17                end
18            end
19        end
20    end
21 end

```

File mssmm_regression.m

```

1 function changes =mssmm_regression(people_changed_states, ...
    count_contacts_above, number_contacts, under_to_above)
2 % SUMMARY: Compute the number of changes from under to above (or under) ...
    the threshold value
3 % with respect to the number of contacts under (or above) threshold at ...
    the time step
4 % of the change
5
6 %OUTPUT :: A matrix of two columns. The first one corresponds to the ...
    number of contacts observed
7 % the second one to the percentage of the changes that took place in ...
    this particular configuration of contacts
8
9 %INPUT ::
10 % people_changed_state : output of the function people_changed_state
11 % (Returns a unique_people by
12 % first_date_of_mesure_to_last_date_of_mesure matrix with absolute ...
    value 1
13 % if the person changed state with respect to previous day and 0 else.)
14 % count_matrix_above : output of the count_matrix function
15 % number_contacts : column vector with the total number of contacts of
16 % each person
17 % under_to_above : boolean value. if TRUE, then changes from ...
    under_to_above ; else

```

```

18 % then compute the changes from state above to state under threshold
19
20
21 pcs = -persons.changed.states;
22 contacts.new.state = count.contacts.above;
23
24 if ¬(under.to.above)
25     pcs = persons.changed.states;
26     number.contacts = transpose(number.contacts);
27     n = size(count.contacts.above, 2);
28     contacts.new.state = repmat(number.contacts, 1, n)-contacts.new.state;
29 end
30
31 changes.with.counts = tabulate(uint8(contacts.new.state(pcs>0)));
32 changes = changes.with.counts(:, [1 2]);
33 changes(:, 2) = changes(:, 2)/size(persons.changed.states, 1);
34
35 end

```

File mssmm_regression_results.m

```

1 function [above b g a] =mssmm_regression_results(change_above, change_under)
2 % SUMMARY: Compute the parameters to be used in the SIS and SISa model
3
4 %OUTPUT :: The parameters to be used in the SIS model and SISa model, by
5 %determining the state that is infectious ( cf. Hill and al.)
6 % above : boolean variable. 1 if above is the infectious state. 0 if under
7 % is the infectious state
8
9 %INPUT ::
10 % change_above : output of the function mssmm_regression on the changes
11 % from the under threshold state to the above threshold state
12 % change_under : output of the function mssmm_regression on the changes
13 % from the above threshold state to the under threshold state
14
15
16 x_above = change_above(:, 1);
17 y_above = change_above(:, 2)/100;
18
19 x_under = change_under(:, 1);
20 y_under = change_under(:, 2)/100;
21
22 lin_above = fitlm(x_above, y_above, 'linear') % use ...
23     GeneralizedLinearModel.fit instead of firlm in versions prior to ...
24     MatLab R2014a
25 lin_under = fitlm(x_under, y_under, 'linear')
26
27 coef_above = cell2mat(table2cell(lin_above.Coefficients));

```

```

26  coef_under = cell2mat(table2cell(lin_under.Coefficients));
27
28  if double(coef_under(2,4)) > double(coef_above(2,4)) % Corresponds to a ...
    comparison of the p-value
29      above = 1;
30      b = double(coef_above(2,1));
31      a = double(coef_above(1,1));
32      g = change_under(1,2);
33  else
34      above = 0;
35      b = double(coef_under(2,1));
36      a = double(coef_under(1,1));
37      g = change_above(1,2);
38  end
39
40  end

```

File simulation.m

```

1  function [susceptible,infected] = simulation(a,b,g, initial_infect, ...
    duration, step)
2      % SUMMARY: Simulates the SISa model for the automatic (or spontaneous)
3      % infection factor a, the transmission rate b, and the recovery rate g,
4      % for an initially infected percentage of initial_infect, a duration of
5      % 'duration' days, with a resolution in the timesteps of step.
6
7      % OUTPUT::
8          %susceptible vector with the susceptible percentage per timestep
9          % infector vector with the infected percentage per timestep
10     % INPUT::
11         %parameters: a=spontaneous infection rate; b=transmission ...
            rate; g=recovery
12         %rate; duration=duration of the simulation; step=time step
13
14     dt=step;          % time step, initialise manually here when desired
15     d = duration;    % simulated duration, initialise manually here when ...
            desired
16
17     % initial condition
18     S=100-initial_infect;      % Susceptible egos
19     I=initial_infect;    % Infected individualsS
20     % book-keeping variable
21     index = 0;
22
23     % simulation loop
24     for t=1:dt:d+1
25         index = index + 1;
26         dI = (b*S*I-g*I+a*S)*dt;

```

```

27         dS = (-b*S*I+g*I-a*S)*dt;
28         I = I + dI;
29         i(index) = I;
30         S = S + dS;
31         s(index) = S;
32     end
33     format short
34     x = linspace(1,t,(d/dt)+1);
35     plot(x,s)
36     grid on
37     hold on
38     plot(x,i,'o')
39     hold off
40     infected = i;
41     susceptible = s;
42 end

```

Additional regression on the US obesity by county against the median household income in 2010 File `obesity_confounding.m`

```

1 function [ output_args ] = obesity_confounding()
2 % Returns the regression of the US obesity by county against the median
3 % household income in 2010
4 % Source:http://www.ers.usda.gov/data-products/food-environment-atlas/ ...
5 % data-access-and-documentation-downloads.aspx#.U3E4pSg-Vo4
6 % Load input file
7 load('DATA.mat')
8 % Regression
9 fitlm(MEDHHINC10/1000,PCT_OBESE_ADULTS10)
10
11 end

```