



Eidgenössische Technische Hochschule Zürich  
Swiss Federal Institute of Technology Zurich

# Agent-Based Modeling: Social System Simulation

Project Report

## **Spread of Obesity in Social Networks**

Chio Ge, Marvin Jarju & Giulia Zheng

Zurich  
December 2019

## **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.

Chio Ge

Marvin Jarju

Giulia Zheng

# Contents

<b>1</b>	<b>Abstract</b>	<b>4</b>
<b>2</b>	<b>Introduction and Motivations</b>	<b>4</b>
<b>3</b>	<b>Description of the Model</b>	<b>5</b>
3.1	The SIS-model . . . . .	5
3.2	The SISa-model . . . . .	6
<b>4</b>	<b>Implementation</b>	<b>6</b>
4.1	Generating a Population . . . . .	7
4.2	Generating the Social Network . . . . .	7
4.3	Time Steps and Iterations . . . . .	8
4.4	Difficulties and Possible Extensions . . . . .	9
<b>5</b>	<b>Simulations, Results and Discussion</b>	<b>10</b>
5.1	Parameter Dependence and Reconstruction of the Evolution of Obesity in Switzerland . . . . .	10
5.2	Predicting the Future Rates of Obesity in Switzerland . . . . .	12
<b>6</b>	<b>Conclusion and Outlook</b>	<b>13</b>
<b>7</b>	<b>Bibliography</b>	<b>14</b>

# 1 Abstract

The prevalence of obesity in our society has experienced a strong increase across the globe. Within the last 25 years, the fraction of obese people in the Swiss population has more than doubled [1] and studies show that genetic factors have no bearings on whether an individual may contract obesity or not [4]. Thus, in our project we would like to explore the possibility that obesity may be viewed as an epidemic that has the ability to spread via social networks whereby "infected" individuals have the capability to pass on the condition to others they come in contact with. Towards this end, we employ the SISa-model as presented in [4] which we use to model the spread of obesity in a social network we construct. This model accepts two rates as parameters, the "spontaneous infection" and the "recovery" rate, and fixes a transmission rate of 0.005. Using the known obesity rates in Switzerland from 1992 to 2017, we use our model to interpolate between these two years and discover that in order to best reconstruct the development of obesity rates in Switzerland starting at 5% in 1992 to 10% in 2017, our model uses an average spontaneous infection rate of 0.006 and a recovery rate of 0.023. In comparison, similar studies pertaining the US arrive at rates that lie in about the same order of magnitude. Using the rates we find in our interpolation, we then predict the development of the obesity epidemic and discover that our model predicts an obesity rate of 16% in 2042 should the hypothesis that obesity spreads like an epidemic be true and considering the simplifications we made to the model.

# 2 Introduction and Motivations

Human behaviour is chiefly governed by societal and environmental influences. Social interactions, nurture and upbringing account for an individual's identity, their choices and way of life, which gives rise to the notion that sociological and behavioural phenomena could be transmitted via social networks in a matter comparable to a viral contagion.

Traditionally, the term "contagion" is associated with strictly epidemiological factors, i.e. infectious diseases or viruses that spread through networks by human contact. In a similar vein, we can extend this connotation by considering a "social contagion" as a behavioural pattern that propagates inter-personally through a *social* network via interaction. Thus, social contagions have the ability to become quintessential determinants for an individual's health and mental well-being. Based on this premise, this hypothesis enables us to model behavioural phenomena using network analytics and classic epidemiological models, albeit most of those will be

subject to slight modifications in order to be compatible with the problems we are concerned with.

One such instance of a recent issue we are facing is the prevalence of obesity in modern society, which may be regarded as an epidemic in its own right. As of 2017, approximately 40% of the Swiss population is overweight and a further 10% is obese. In comparison to figures of 1992, this represents a full 100% increase in the last 25 years [3]. This issue is not limited to Switzerland: we observe similar trends, if not stronger ones, throughout the rest of the globe. Hitherto, the obesity epidemic cannot be adequately explained by genetic factors (genetics may merely affect one's predisposition towards obesity) [1] and the epidemic permeates through all ages and socioeconomic groups. Hence, as we are lacking congenital or medical explanations to expound this sudden upsurge, this allows the theory of a social contagion to step to the fore.

The hypothesis is that an individual's attitude towards obesity and one's habits change as those around them do. For instance, a person's tolerance for weight-gain may be increased if they converse with multiple contacts that are obese, and group activities like eating-out or smoking may be forced upon an individual by means of peer pressure. Furthermore, if a child has obese parents, this might reflect in the child's upbringing and diet.

In the following, we implement a model to simulate these aforementioned considerations on the basis of a social network that will be representative of the demographics of Switzerland. We interpolate between the figures we observe from 1992 and 2017 and moreover, attempt to make a prediction for the "disease" process in the years to come. Ideally, we want to confirm the correlation between the spread of obesity and social influence, or in the very least, showcase social contagions and the modelling of those as a viable consideration for a plethora of behavioural phenomena and pseudo-epidemics.

## 3 Description of the Model

### 3.1 The SIS-model

The model conventionally used in the mathematical modelling of epidemiology is the SIS-model. Similar to a regular infection, our population can generally enter two different states, "susceptible" and "infectious", respectively. The reason for choosing this model as opposed to alternatives like the SIR-model is due to the fact that our individuals cannot develop immunity and enter a "recovered" state upon recupera-

tion - an individual recently recovered from obesity is still vulnerable to relapse. [4]

The SIS-model acts upon a network which will be modelled by a connected graph of agents, where every connection represents a social connection, e.g. "friendship" or "family", with the caveat that for simplicity, we do not differentiate between different types of relations. Contracting obesity from a friend is thus equally as likely as contracting it from a relative which is not the most realistic assumption - studies do show that transmissions vary depending on the type of relation between two individuals [1].

### 3.2 The SISa-model

The paper we studied presents the SISa-model, which is a modification of the SIS-model, that allows for an additional spreading process. The main modification that is made, is that the spread is not strictly limited to transmission. Instead, beside the "transmission" rate  $\beta$  and "recovery" rate  $\gamma$  used in the traditional SIS-model, an additional "spontaneous infection" rate  $\alpha$  is introduced [4]. As the name suggests, at any given time, a single agent may develop obesity spontaneously independent of the state of those around them. By introducing this rate, it is ensured that the disease will never die out within our population, which is in line with reality.

In the SISa-model, the spontaneous rate serves as a base rate for the probability of contracting obesity. Depending on the transmission rate and number of infectious contacts an individual has, the overall likelihood of contracting obesity will then increase; Hence the probability is cumulative over the number of infected contacts an agent possesses. The more connections an individual has, the more the impact of one single obese contact will be diluted. This means for instance that an agent with two friends which are both obese is more likely to become obese than one with 100 friends out of which two are obese. The rationale behind that decision is that the fewer social contacts one agent has, the more time is spent with each one. The recovery, spontaneous and transmission rates remain static in our model and are only set at the beginning of each simulation.

## 4 Implementation

We use the programming language Python for our implementation. The code relevant to this project is contained in the *contagion.py* file. Within this file, the functions necessary to create, modify, and evaluate a network, along with the relevant plots and simulations are implemented.

## 4.1 Generating a Population

Individual agents are implemented in the Python class *Agent*, containing multiple attributes that describe an individual in the population. The main attributes are *age*, *sex*, and *obese*. The age of an agent is stored as a number, the sex is either 'f' or 'm', and the obesity of an agent is a Boolean value. Additionally, there is a variable called *next\_state*, that is used in each step of the iteration as described in section 4.3

To create a population of agents, we implement the functions *createSwissAgents1992* and *createSwissAgents2017*. Given a number  $n$ , these functions return a list of  $n$  agents, whose rates of obesity are distributed according to the obesity rates [3] for females and males by age group in Switzerland in 1992/2017. Due to lack of sufficient data, the ages of agents for both functions are distributed according to the age structure of Switzerland in 2018 [2].

## 4.2 Generating the Social Network

To model the spread of obesity between agents, we generate a random network whose nodes and edges represent agents in a social network and the social connections between agents respectively. We use the popular library *NetworkX* to create and manipulate network structures in Python.

Given a population of agents, the method *createNetwork* adds nodes to an empty graph, with each node containing the information pertaining to one agent in the population. Next, the social connections (edges) between agents are generated. As we want our network to resemble a real-world social network, we expect to generate a network with a reasonable average degree  $k$ , clustering coefficient (transitivity)  $\Phi$  and average shortest path length  $l$ . Additionally, we expect a power law degree distribution. The graph generator we use is the *LFR\_benchmark\_graph* generator of *NetworkX*. The Lancichinetti–Fortunato–Radicchi benchmark algorithm is designed to generate graphs that resemble real-world social networks [5]. We chose this method of generating a network because it facilitates both the creation of a variety of community sizes and degree distributions, as well as adjusting many of the parameters we are interested in. Using the LFR generator, we generate networks with 500 nodes that have an average  $\Phi \sim 0.22$ ,  $k \sim 15$ , and  $l \sim 3$ . We found these to be sufficiently reasonable parameters for a social network. The edges of the generated network are finally added to the network containing the information of the agents.

The *exportNetwork* method exports the supplied network along with data describing the obesity of a node to a *.gexf* file. This allows us to use the *Gephi* software

to visualise and study the networks we work with. The *obesityRateNetwork* method simply determines the current rate of obesity in a supplied network.

### 4.3 Time Steps and Iterations

After initialising a population and generating the social network, the *simulate* method takes such a network, along with the global rates of transmission  $\beta$ , recovery  $\gamma$  and spontaneous infection  $\alpha$  that are described in section 3.2. Using these rates, the spreading-dynamics of obesity in the network are then simulated for a given number of time steps. This is done by calling the *step* method in each time step. The *step* method iterates over all agents in the network and calls each agent’s *interact* method. Finally, it iterates over all agents again, and calls each agent’s *update* method.

In the *interact* method of an agent, if the agent is currently obese, they will recover (*next\_state* becomes *False*) with probability  $\mu \cdot \gamma$ . Otherwise, if the agent is not currently obese, their initial probability to become obese is  $p = \mu \cdot \alpha$ . For each obese friend,  $p$  increases by  $\delta \cdot \frac{\beta}{k}$ . Finally, they become obese (*next\_state* becomes *True*) with probability  $p$ . Since the rates  $\beta, \gamma$  and  $\alpha$  found in the paper describing the modeling of infectious diseases [4] are global rates, we multiply the rates used in our simulations by some factors  $\mu, \delta$ , which are normally distributed around 1 with a variance of 0.25. This allows us to factor in an agent’s personal tendencies, while still maintaining the expected global rates. The *update* function of an agent simply changes the agent’s current state of obesity to the value of *next\_state* determined in the *interact* function.

The *produceClosestGraph* function takes an initial network, two numbers  $k$  and *timesteps*, along with the rates  $\beta, \gamma$ , and  $\alpha$ . It then performs  $k$  simulations of *timestep* steps on the initial network, saving the final networks of each simulation. Finally, it returns the network whose final obesity rate is closest to the average rate of obesity across all simulations.

The concrete semantics of the *interact* function are shown more accurately in



Algorithm 1.

---

**Algorithm 1:** *interact* function of an agent.

---

**Input:** Agent *self*, Network *G*, rates of transmission  $\beta$ , recovery  $\gamma$ , and spontaneous infection  $\alpha$

**Output:** agents next\_state is updated

```

1  $\mu \sim N(1, 0.25)$ ;
2  $\epsilon \sim U(0, 1)$ ;
3 if self is obese then
4   if  $\epsilon < \mu * \gamma$  then
5     next_state = False;
6 else
7    $p = \mu * \alpha$ ;
8    $k = \text{size}(\text{contacts}(\text{self}))$ ;
9   for contact in contacts(self) do
10    if contact is obese then
11       $\delta \sim N(1, 0.25)$ ;
12       $p += \delta * (\frac{\beta}{k})$ ;
13   if  $\epsilon < p$  then
14     next_state = True;
```

---

#### 4.4 Difficulties and Possible Extensions

One problem we faced when generating the social network was the initial random distribution of obese agents throughout the network. Ideally, to improve the authenticity of the model, we would generate a network that already contains clusters of obese agents, as this would be more representative of the real world. However, realising this proved to be quite difficult. A related difficulty was finding and using publicly available data sets. These often were too big to run simulations on using the hardware we had available. They also tended to have a density, transitivity or degree distribution that was too different from the ones we considered reasonable for our purposes. Using the LFR benchmark graphs [5] allowed us to fine tune the parameters as we saw fit.

A possible future extension that might be included is allowing a variety of social connections. Some of these might include spousal, family or close friendship connections. This would allow us to adapt the weight of edges depending on the type of connection between two agents, which could be used to change the impact an agent

has on the chance of becoming obese of another agent as presented in [1]. Along with this, parameters such as the age and gender of an agent might also have an impact on the spreading dynamics, as currently only the current state of obesity is considered in our model.

## 5 Simulations, Results and Discussion

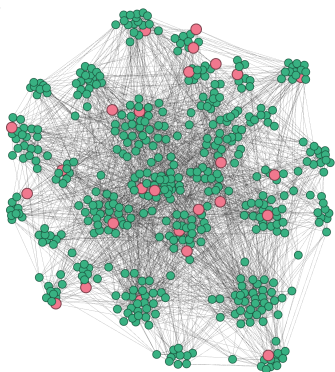


Figure 1: **Network simulation of Swiss obesity rate in 1992.** Nodes represent agent that can be either obese (red) or non-obese (green) and they are connected by edges corresponding to their social bonds. The percentage of obesity is 5.4 % [3] and obese agents are randomly distributed.

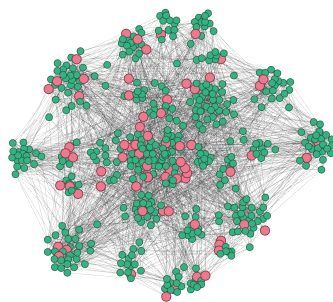


Figure 2: **Network simulation of Swiss obesity rate in 2017.** Representation of the evolution of Figure 1 after 25 time-steps, with transmission rate  $\beta = 0.005$ , recovery rate  $\gamma_0 = 0.023$  and spontaneous rate  $\alpha_0 = 0.006$ . These are the rates for which the evolution of the simulated network best represent the Swiss data [3]

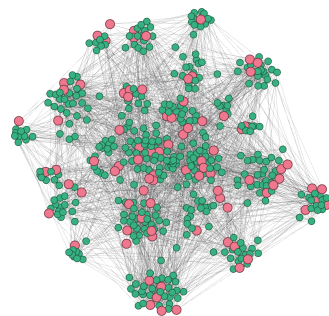


Figure 3: **Network simulation of Swiss obesity rate in 2042.** Representation of the evolution of Figure 1 after 50 time-steps, with  $\beta = 0.005$ ,  $\gamma_0 = 0.023$  and  $\alpha_0 = 0.006$ .

### 5.1 Parameter Dependence and Reconstruction of the Evolution of Obesity in Switzerland

We generate a network of 500 agents using the obesity rates of the Swiss population in 1992 [3] as described in section 4.2. Red (green) nodes represent obese (non-obese)

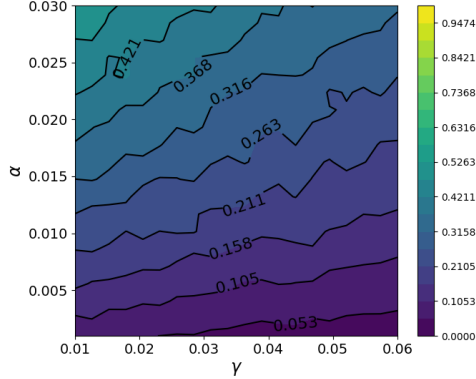


Figure 4: Obesity rate after 25 time-steps as a function of recovery rate  $\gamma$  and spontaneous rate  $\alpha$ . The initial obesity rate is 5.2 % and the transmission rate  $\beta = 0.005$

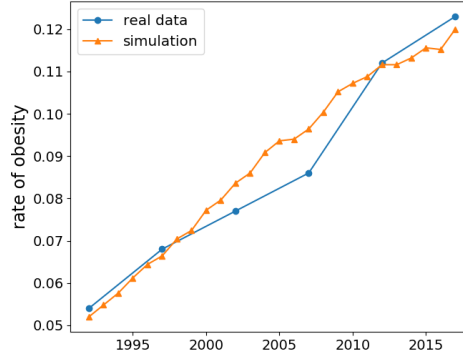


Figure 5: Comparison between the evolution of the obesity rate during 25 years with the real data of the Swiss population between 1992 and 2017. The simulation is run with  $\beta = 0.005$ ,  $\gamma_0 = 0.023$ ,  $\alpha_0 = 0.006$  and an initial obesity rate of 5.2 %

agents. Therefore the ratio between red and green nodes corresponds to the obesity rate of the network year. Note that here the obese agents are initially randomly distributed. The generated network is shown in Figure 1. The age distribution and the gender of individual agents are not represented, as it is not relevant in our simplified model. Additionally, since only one type of social connection between agents is considered, there is no meaning behind the color of edges.

Parameter dependency analysis is performed to examine the impact of varying rates  $\gamma, \alpha$  on the final rate of obesity in simulations. We keep the transmission rate  $\beta$  constant at a value of 0.005 as found in [4]. While this rate was found to fit the population studied in the US, the corresponding rate in Switzerland would most likely be different. Nevertheless, this is a simplification made in our model, since we assume that Switzerland and the US do not differ in that regard as strongly as in the other two rates. We run simulations for pairs of varying rates  $\gamma, \alpha$  and plot them against the rates of obesity produced by simulations using those rates. The recovery and spontaneous infection rates used for this analysis are twenty linearly spaced values between 0.01 to 0.06, and between 0.01 to 0.03 respectively. For each pair of parameters, five simulations of 25 time steps are ran to model the development from

1992 to 2017. The average obesity rates in each time step of these five simulation are recorded. Figure 4 shows the parameter dependence between  $\gamma, \alpha$  and the average final rate of obesity in the network.

A non-linear least squares approach is then used to find the pair of parameters that best fits the development of obesity in Switzerland [3]. Consider for this a fixed pair of parameters  $(\gamma_i, \alpha_j)$ . The residual for every year 1992, 1997, 2002, 2007, 2012, and 2017 is then defined as the difference between the actual rate of obesity in Switzerland in the given year and the average rate recorded in the corresponding time step of the simulations using  $\gamma_i, \alpha_j$  recorded previously. These residuals are then squared and summed over all six years. Finally, we choose the pair of parameters with the lowest square sum of residuals, which is then the least squares solution that best fits our data.

Using this approach, we obtain the values  $(\gamma_0 = 0.023, \alpha_0 = 0.006)$  as the solution, with a sum of squared residuals  $\epsilon = 1.63 \cdot 10^{-3}$ . Figure 5 shows the evolution of the obesity rate in Switzerland along with the simulated evolution obtained using  $\gamma_0$  and  $\alpha_0$ . Indeed, we observe similar trends.

The final network obtained by simulations using  $\gamma_0$  and  $\alpha_0$  is shown in Figure 2, with an average obesity rate of 10%. Here, one can observe the formation of clusters with prevalence of either obese or non-obese agents as described in [1].

## 5.2 Predicting the Future Rates of Obesity in Switzerland

Supposing that recovery and spontaneous rates remain constant over the next 25 years, we use our model to make predictions about the future evolution of obesity in Switzerland. Figure 3 shows the final network obtained in section 5.1 after another 25 time steps using the network from Figure 2. The final obesity rate our model produces on average is 16.4%.

## 6 Conclusion and Outlook

Despite the simplifications our model was subjected to, and the limited set of parameters we utilised, we find that we extract spontaneous and recovery rates that in terms of order of magnitude, are highly in line with those found in research papers pertaining population studies in the US [4]. While in our simulations we arrive at a spontaneous rate of 0.006 and a recovery rate of 0.023 (with fixed transmission rate of 0.005), the research conducted in [4] result in a spontaneous rate of 0.02 and a recovery rate of 0.04. For what it's worth, our lower rates seem logical considering the general obesity prevalence in Switzerland is lower than that of the USA. The fact that our rates seem appropriate and reasonably in line with those of other independent research results, hints at the possibility that the theory of a social contagion holds water, indeed.

Furthermore, bearing in mind the number of simplifications we made, we find there are numerous ways our model can be expanded upon: one may begin to differentiate between different kinds of relationship (friendship, family, etc.) [1] or utilise the fact that transmission rates can vary with the sex of a social contagion. Another caveat we had to work under was that our initial distribution of obese agents was completely at random: realistically, clusters should already have formed from the get-go in our baseline network. If we were to include all those considerations and worked these features into our simulations, we may inch even closer to the real rates and achieve an accurate prediction for 25 years down the line, provided this hypothesis of the existence of a social contagion is in fact proven true.

Should this indeed be the case, this would have wide-ranging consequences for society as we know it. We would find ourselves in a downward spiral of infections whereby more and more of the population will become obese and infect the social circles surrounding them which only exacerbates the issue even further. In our simplified model, we already predict obesity to rise as high as 16% in 2042 with no break in this upward trend in sight. The conclusions needed to be drawn from this are that what we call an person's spontaneous/recovery rate must be lowered/increased, respectively, if we ever hope to curb this epidemic. Thus, the implications for society are that obese people must receive increased support in order to facilitate recovery and healthy individuals need to be more discouraged from contracting obesity "spontaneously", i.e. via improved education on the subject and promotion of more healthy lifestyles.

## 7 Bibliography

### References

- [1] Nicholas A. Christakis and James H. Fowler. The spread of obesity in a large social network over 32 years. *New England Journal of Medicine*, 357(4), 2007. PMID: 17652652.
- [2] CIA. Age structure of switzerland 2018. <https://www.cia.gov/library/publications/the-world-factbook/geos/sz.html>.
- [3] Bundesamt für Statistik. Body mass index (bmi) nach geschlecht, alter, bildungsniveau, sprachgebiet - 1992, 1997, 2002, 2007, 2012, 2017: Tabelle. <https://www.bfs.admin.ch/bfs/de/home/statistiken/gesundheit/determinanten/uebergewicht.assetdetail.6466017.html>.
- [4] Alison L. Hill, David G. Rand, Martin A. Nowak, and Nicholas A. Christakis. Infectious disease modeling of social contagion in networks. *PLOS Computational Biology*, 6(11), 11 2010.
- [5] Andrea Lancichinetti, Santo Fortunato, and Filippo Radicchi. Benchmark graphs for testing community detection algorithms. *Physical Review E*, 78(4), Oct 2008.