



Eidgenössische Technische Hochschule Zürich  
Swiss Federal Institute of Technology Zurich

# Lecture with Computer Exercises: Modelling and Simulating Social Systems

Project Report

---

## Dynamics of Smoking

---

Konstantin Baune, Georg Engin-Deniz,  
Max Glantschnig, Roman Wixinger

Zürich  
December 2018

**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):

Dynamics of Smoking, for the course "Modeling and Simulating Social systems".

**Authored by** (in block letters):

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

Name(s):

Baune

Engin-Deniz

Glaatschwig

Wixinger

First name(s):

Konstantin

Georg

Max

Roman

With my signature I confirm that

- I have committed none of the forms of plagiarism described in the 'Citation etiquette' 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

Zürich, 09/12/2018

Signature(s)

Konstantin

Roman Wixinger

Max Glaatschwig

Georg Engin-Deniz

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

## **Agreement for free-download**

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

Georg Engin-Deniz

Roman Wixinger

Max Glantschnig

Konstantin Baune

# Contents

<b>1 Abstract</b>	<b>3</b>
<b>2 Introduction and research question</b>	<b>4</b>
<b>3 Methods</b>	<b>4</b>
3.1 Model of society . . . . .	4
3.2 Time transition and iteration . . . . .	6
3.3 Possible extensions . . . . .	7
3.4 Outputs . . . . .	8
<b>4 Simulation results and discussion</b>	<b>9</b>
4.1 Visualization of society and formation of groups . . . . .	9
4.2 Stability . . . . .	10
4.3 Influence analysis . . . . .	11
4.4 Investigation of impact parameters . . . . .	12
4.5 Time evolution . . . . .	12
4.6 Difficulties . . . . .	13
<b>5 Conclusion</b>	<b>14</b>
<b>References</b>	<b>15</b>

# 1 Abstract

With this project we tried to predict the evolution of smoking in Switzerland based on data gathered in the US over a span of 30 years in a paper [1] regarding the results of the "Framingham Heart Study" (FHS). We did this by creating a network of friends and family, assigning them influence factors and observing the development over time, where in each time step a person has impact on the people's smoking behavior that he or she is connected to. Although we could not reproduce the full data from our base study, we ended up with a stable system that had certain characteristics that were also mentioned in our base study [1]. For example, at the end of our simulation groups of smokers and groups of non-smokers had been formed in our society.

## 2 Introduction and research question

The goal of our research was to simulate and predict the evolution of smoking in a society. We were inspired by a paper from Christakis and Fowler [1] in which they described the smoking cessation dynamics in the network defined by the data taken from the so-called "Framingham Heart Study" (FHS).

During the "Framingham Heart Study" more than 12,000 people's health statuses in the US were collected for about three decades to gather meaningful data about the evolution of smoking between the 1970s and 2000s. One highly interesting result of this study was the formation of networks between smokers and non-smokers. It was shown how widely connected people quit together rather than independently. The study investigated the impact that friends, family and colleagues have on one's smoking habits. For example it showed that the smoking cessation by a spouse decreased a person's chances of smoking by 67%, compared to 25% after the cessation of a sibling and 36% after the cessation of a close friend. [1]

The main goal of our simulation was to create a social network which we can apply an agent-based model to and that reproduces results that were found in the paper by Christakis and Fowler [1]. We especially wanted to investigate the formation of groups in our society, how our model behaves over time and if our produced results would be realistic for a real society.

## 3 Methods

In order to establish a model to simulate the dynamics of smoking in a society in our code, we needed to produce a realistic model of society and a function that produces the transition of this society over time, regarding each agents smoking habit. As a programming language we used Python.

### 3.1 Model of society

We wanted to create a group of agents with connections that correspond to their friendships and familiar bonds and that are distributed as in a real society. Furthermore, we wanted the initial smoking habits of these agents to match the ones of the American respectively the Swiss society.

To create agents we used classes in Python, where each agent has different attributes. These main attributes are age, gender and smoking habit (the so-called `state` of the agent). While the age is simply stored as an integer value and the gender is either "male" or "female", the variable that determines an agent's smoking habit consists of three related

variables: One variable, which is called `state` determines the state of an agent and has either the value 1 or -1, where 1 refers to a non-smoker and -1 to a smoker. Similar to the FHS we only used these two states for the simplicity of the model and used the term smoker for someone who smokes at least one cigarette per day. The other variable `state_con` can be modified continuously between -1 and 1. This variable can change during each time step and determines the state of the agent. If `state_con` is smaller than zero, the agents state changes to smoker, if it is greater than zero, the state becomes non-smoker. Last, there is the variable `next_state` which stores the following state of the agent during each iteration, so that the agent's state can be updated after the iteration is done.

The initial setup of the agents is done in the function `InitializeAgentPopulation` in our code.

Next, the connections between the agents are made. Here, a graph is created, where each agent is a node and a connection between two agents is an edge between the nodes of these two agents. Because we wanted our network to resemble real networks of social connections, we created it to have a reasonable clustering coefficient  $C$  and a mean geodesic distance  $l$ . As a reference for these two values, we used a network of film actors with  $C = 0.2$ ,  $l = 3.48$  and email address books  $C = 0.17$ ,  $l = 5.22$ . [4]

In real societies, however, not every social contact of an agent has the same influence on it. For example, a family member should have a higher influence on the agent than a friend. This distinction of the social contacts was also drawn in the paper by Christakis and Fowler [1], so we decided to add a weight (a real number) to the edges of our network. A weight of 1.0 represents a friendship and a weight of 1.2 represents a family bond.

To implement such a network in Python, we used the library NetworkX which provides functionality for creating, manipulating and analyzing graphs.

The creation is done in the function `GenerateFriendshipGraph` as follows: First, we created a graph containing the agents' "gid" (a member variable denoting the "name" of the agent). The class instance of `GenericAgent` itself is stored in the "data" section of the node. Then, we added edges generated by the Erdős-Rényi model, which is a method to create a random graph. We heuristically chose a Erdős-Rényi-probability between 0.005 and 0.05 for our simulation. However, it turns out that the thereby simulated network has a clustering coefficient of about  $C \approx 0.01$ , which is far from the value we would expect for a real social network. Therefore, we additionally added edges generated by the Block Two-Level Erdős-Rényi (BTER) model, which was first proposed by Seshadhri, Kolda, and Pinar in 2012 [6]. BTER is capable of generating graphs whose characteristics (degree distribution/clustering coefficient) match those of real social networks [6]. The implementation of this model is described by the same scientists (and Plantenga) in [3]. They also provided MATLAB<sup>®</sup> code, the "FEASTPACK Distribution, Version 1.2"<sup>©</sup> [5], which we used to generate the files `300nodes.mat`, `500nodes.mat`, `1000nodes.mat` containing the information about the edges. In the function `GenerateFriendshipGraph`, these files were imported to our Python script and the edges added to the graph.

With these additions, the clustering coefficient then turned out to be between 0.23 and 0.29 where every agent has in average  $\approx 12$  friends and  $\approx 1.5$  family members. We found this values to be reasonable for an approximation of a real social network.

Finally, we made half of our society female and the other half male and gave each agent an age between 15 and 100 years, respecting the age structure in Switzerland from 2017 [8]. This assumes, that people with age lower than 15 years are non-smokers and they are not considered in our model. The smoking habit was then assigned to every agent regarding their sex and age, taking data from the Federal Statistical Office of Switzerland [7] which shows the smoking habits in the Swiss society from 2012. For example, if an agent is a 36 year old man, the probability that the agent smokes is 38.8%. We implemented a similar function for the data from the US, with information from [2].

## 3.2 Time transition and iteration

After the initialization of the agents' network and after giving each agent age, gender and a state (smoking habit), we built a function that does an iteration over a certain number of time steps. In our code, this function is called `simulate`, as it simulates the dynamics of the smoking habits in the society that we set up. In each time step the agents' states can be changed, depending on the connections that the agent has. In our model we wanted that one time step corresponds to one year as this would lead to a similar time scale as in the "Framingham Heart Study", where the results are shown over a period of about thirty years. This way, setting the number of iterations to 30 would mean that we observe our model over a period of 30 years.

To perform each iteration, we set up the function `step` which performs one time step of the simulation. This function is called once per time step by the `simulate` function. The main part of the `step` function is the function `act`, which is shown in pseudo-code in Algorithm 1 and which is used as follows. A list containing all agents is passed to the function `step` which iterates over it. For each agent `agent*` in the list, an array of the continuous states of the agents that are connected to `agent*` is made. Then the `act` function is called with `agent*`'s information and `agent*`'s friend list. Now the `act` function iterates over all of `agent*`'s friends and changes the continuous state of `agent*` depending on the continuous state of each of the friends. Here, `agent*`'s continuous state increases if a friend is non-smoker and decreases if a friend is a smoker. The impact that a smoker or a non-smoker has on its friends is defined by the impact parameters `impact_smoker` and `impact_non-smoker` and a random number `sample` that is normally distributed around 1. The influence of these impact parameters on our model is investigated and discussed further in section 4.4. When the `act` function is done with the iteration over the friends, the next state of `agent*` is defined, depending on its continuous state. If the continuous state is greater than zero, `agent*`'s next state is non-smoker, and otherwise the next state is smoker. The iteration of the `act` function is shown and explained more accurately in Algorithm 1.

**Data:** Agent's information, agent's friends (with states), impact smoker, impact non-smoker

**Result:** Iteration over this agent is done and the state for the next time step is defined

```

begin
  for friend in friends do
    r = random  $\in N(1, 0.3)$ 
    state_con =  $\begin{cases} \text{state\_con} - r * \frac{\text{impact\_smoker}}{\text{number of friends}} & \text{friend} = \text{smoker} \\ \text{state\_con} + r * \frac{\text{impact\_non-smoker}}{\text{number of friends}} & \text{friend} = \text{non-smoker} \end{cases}$ 
    next_state =  $\begin{cases} \text{smoker} & \text{state\_con} < 0 \\ \text{non\_smoker} & \text{state\_con} \geq 0 \end{cases}$ 
    if state  $\neq$  next_state then
      state_con =  $\begin{cases} \text{state\_con} - 0.3 & \text{next\_state} = \text{smoker} \\ \text{state\_con} + 0.2 & \text{next\_state} = \text{non-smoker} \end{cases}$ 
    end
  end
end

```

**Algorithm 1:** Pseudo-code showing the `act` function which iterates over all agents in each time step and changes their continuous states (`state_con`) and defines the state for the following time step (`next_state`). This function is called once per time step and per agent. `impact_smoker` and `impact_non-smoker` are the variables that determine the impact that a friend's state has on the agent. The impact parameters are divided by the number of friends to insure that the agents smoking behaviour does not depend on the amount of friends, but rather on the share of smokers and non-smokers in its environment. If the state of an agent changes in an iteration, the variable `change` is set to `True` and the continuous variable changes even further, as e.g. people will not quit smoking immediately after starting to smoke.

### 3.3 Possible extensions

As we tried to keep simplicity in our model to make our results reproducible and also understandable we left out certain aspects. Some of the extensions that could be included in our model to increase the complexity and the accuracy are mentioned in the following.

A possible extension for the future could be the addition of the connection type "work colleagues" and the subdivision of family into "spouse" and "relatives" in the network. This would make our network more complex and would lead to more distinct weightings of the edges between agents. Also it would be possible to include directed edges, which

would mean that in a connection between two agents one agent could have more influence on the other agent than vice versa. This could be closely connected to the betweenness centrality of one node in our graph, as a person that is more closely connected than others has a higher social status and thus has more influence on others.

The next layer of complexity could be the addition of a function that decreases the likelihood of single individuals changing their smoking habits. This function could also favor changes in the smoking habits of whole groups like observed as mentioned in the paper by Christakis and Fowler [1].

Another extension would be to include more attributes of the agents. For example the age and the gender of the agents are not yet used in the `act` function, as this has lead to instability in our first tries and as we have no adequate rule that we could use to logically include the agent's gender and age. At the moment, gender and age are only used for the initialization of our society where the smoking behavior is assigned according to the agent's age and gender and the age is assigned according to the age structure in the US respectively in Switzerland, as mentioned in section [3.1].

### 3.4 Outputs

In order to be able to evaluate the results of our simulation, we implemented functions that plot certain quantities of our system, so that we had a visual output of the results which we then investigated further.

The function `ExportGraph` uses a NetworkX-Graph as input and generates a ".gexf" file which contains all the relevant data concerning the nodes and edges of our network. This file can be imported to the program Gephi<sup>©</sup>, which is a tool for visually analyzing graphs. As explained in [3.2], the `simulate` function is responsible for computing the time propagation. It also keeps track of the numbers of smokers for every time step which is stored in the array `numbers`. The function `run_simulation` uses this data to plot the share of smokers in the population as a function of time. This was important for us when initially developing and later improving our simulation because it presents the macroscopic (i.e. the absolute numbers of smokers and non-smokers) properties of the population in a straightforward way. Also, the clustering coefficient of the graph and the initial and final percentage of smokers is printed. The function `average_friends` prints, as the name already suggests, the average number of friends and family members per agent. Similarly, the function `Graph_test` prints the average percentage of smoking neighbors for both smokers and non-smokers in the initial state. It was used to determine the parameters for a reasonable social network as described in section [3.1]. By setting the boolean variables `analyze_inf` and `analyze_quitting_inf`, we would print the results described in [4.3].

The outputs of `run_experiment1` and `run_experiment2` are described in sections [4.4] and [4.5] and the output of `Determinism_test` is discussed in section [4.2].

## 4 Simulation results and discussion

In this section we will present and discuss the outputs of our simulations. Each subsection will look at different attributes of our model and how these attributes are present in our simulation. We tried to visualize most of our results and outputs as plots and graphs in order to make the results easier to understand.

### 4.1 Visualization of society and formation of groups

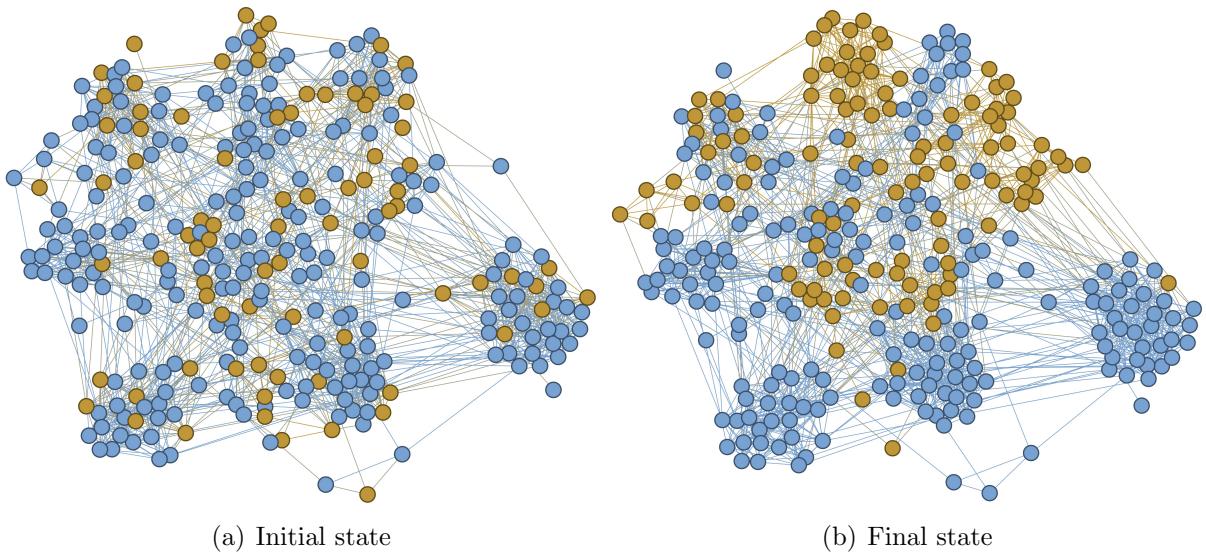


Figure 1: *Visualization of the graph of our society with the initial state (a) and the final state after the iterations (b). Here, smokers are shown in brown while non-smokers are the blue nodes. In this simulation, the initial percentage of smokers is 27% and the final percentage of smokers is 36.33%. One can see that the society started with a random arrangement of smokers and non-smokers. After the iteration groups of smokers and groups of non-smokers have formed, similar to what is described in the paper by Christakis and Fowler [1]. These two graphs were created with the program Gephi<sup>©</sup> using the layout "Force Atlas".*

The first result of our code is the result from the simulation with the start parameters from Switzerland as described above. As the initialization of our population depends on many random factors, every simulation looks different, while some characteristics of the simulation are always the same. Figure 1 shows the initial state and the final state of such a simulation, visualized with the program Gephi<sup>©</sup>. For the initialization of the society we set the Erdős-Rényi-probabilities of both the family and the friend connections to 0.005 for the graph. Here we can see that at the beginning of each simulation smokers and non-smokers are randomly distributed in our society. Furthermore we can see that the connections in

the society are as desired, with smaller communities that are closely connected and then different random connections between the communities. For the simulation we used the impact parameters `impact_smoker` = 0.2 and `impact_non-smoker` = 0.1 and iterated over 30 time steps with a population of 300 agents. After the simulation we can see how groups of smokers and groups of non-smokers have formed and that some communities consist either of only smoking or only non-smoking agents, similar to what is described in the paper of Christakis and Fowler [1].

## 4.2 Stability

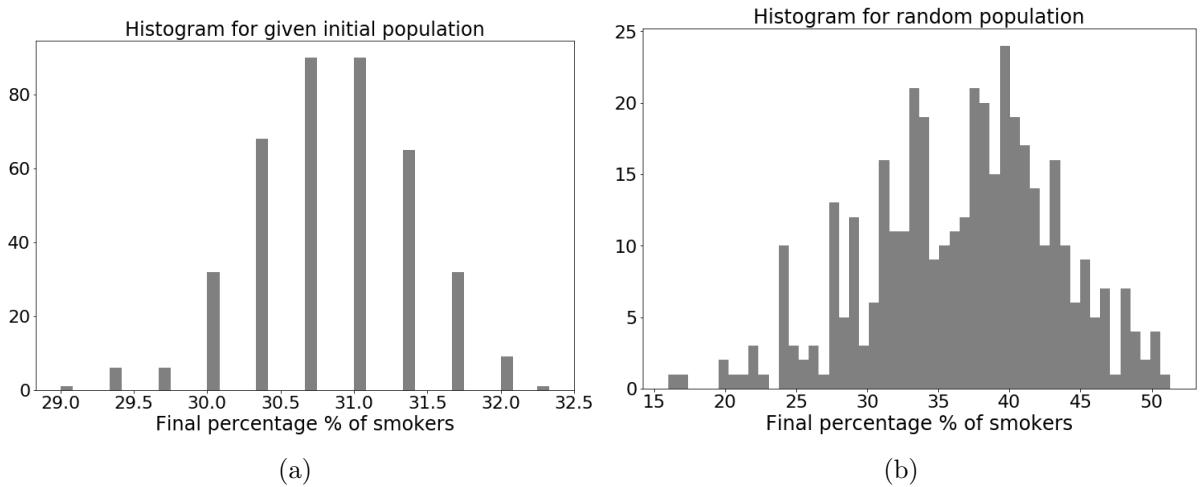


Figure 2: *Distribution of final percentage of smokers for given initial population (a) and for a random population (b).* Both stability test ran with 400 simulations and the impact parameters were set to `impact_smoker` = 0.2 and `impact_non-smoker` = 0.1 and for the initialization we used the data from Switzerland (as described in section 3.1). Especially in this plot a normal distribution of the values is visible with standard deviation 0.56%, while the the standard deviation for the data in Figure 2(b) is 6.6%. Both plots together show that our model is highly dependent on the initial state of the population and that for a random population the final outcome is not predictable.

In order to find out how stable our model is, we first examined how much the final percentage of smokers depends on the random numbers used for the simulation with a fixed initial population. Fixed initial population means, that here all edges between the agents are the same for each simulation and that all initial states of the agents are the same for the beginning of every simulation. For the final percentage of smokers with fixed initial population, we found a standard deviation of 0.6% for 400 runs of the system and so here the function seems stable.

Secondly, we generated different populations and examined the final percentage of smokers.

The additional variation of the initial population resulted in a high deviation of the results, mostly varying between about 20% and about 50% when the initial data from Switzerland (as described in section 3.1) was used.

The resulting plots from these stability tests, generated by the function `Determinism_test`, are shown in Figure 2.

From the stability tests we see that our system is highly dependent on the initial population and is not stable if the initial population is changed, but we also see that it seems stable if the initial population is the same, as then the variation in the final state stays small.

### 4.3 Influence analysis

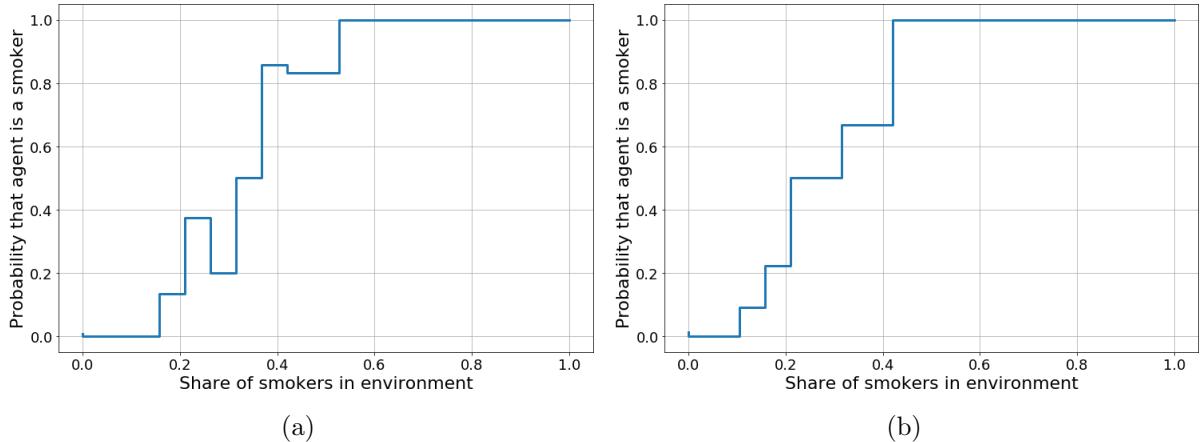


Figure 3: *Probability that an agent is a smoker (at the end of the simulation) as a function of the share of smokers in his environment. Plots are taken from simulations of the Swiss society with the same parameters (`friend_prob = 0.005`, `impact_smoker = 0.18` and `impact_non-smoker = 0.1`). Although the results differ for every simulation (as was discussed in 4.2), there is an overall trend visible: A gradual increase of probability between a share of smokers of 0.1 and 0.5.*

In order to understand how the environment of an agent influences his smoking behaviour, we divided the population into subgroups of agents with similar shares of smokers in their environment. We then approximated the probability of an agent being a smoker with the percentage of smokers in his subgroup.

In Figure 3, generated by the function `analyze_influence`, we see that the probability of an agent being a smoker increases with the share of smokers in his environment. This finding matches the formation of groups of smokers and non-smokers which could be observed in the visualization of our society in Figure 1(b). However, when we look at agents living in extreme environments with a share of smokers below 10% or above 60%, we find

probabilities of this agents being smokers or non-smokers which are too extreme compared to our everyday observations in reality.

Furthermore, we tried to analyze the influence of an agent who quitted smoking on his environment. For this, we compared the environments after the simulation of agents that quit to the environments of agents that stayed smokers. The function `analyze_influence_quitting`, which we used to quantify this influence, delivered reasonable results most of the time. The function mainly found a positive impact of quitting on the environment, which means that if an agent quits smoking, smokers in his environment are more likely to quit too. However we did not get the stability of the result that we hoped for and thus did not include the results of this experiment in this paper.

## 4.4 Investigation of impact parameters

The main parameters that determine our system and that have the most influence on the final state are the impact parameters `impact_smoker` and `impact_non-smoker`, that are used in the `act` function of our simulation. In order to understand how the evolution of our population depends on these parameters and how they are correlated, we ran our simulation on a grid of different values of these parameters. This was implemented in the function `run_experiment1`. The final percentage of smokers in the population was plotted on the grid of the two parameters. The plot is shown in Figure 4(a). In this plot one can see that the result mainly depends on the quotient between the two parameters as we can see approximately straight lines in the plot.

## 4.5 Time evolution

To see how our simulation evolves in time, we evaluated the final percentage of smokers on a grid of different simulation times and values of the parameter `impact_smoker`. For this we set `impact_non-smoker` = 0.1 and evaluated the system for 30 time steps with a given initial population. Here a certain parameter `impact_smoker` is taken and the share of smokers in the society for each time step is plotted. In the code this analysis is called `run_experiment2`. In Figure 4(b) the outcome of this analysis is shown. One can see that the whole plot looks continuous without peaks, which also shows the stability of our system for a given initial population, as described in section 4.2. Furthermore we can see the influence of the parameter `impact_smoker` and that our society is rather stable for a value of about `impact_smoker` = 0.2, which is also the value that we chose for most of our simulations in the other sections.

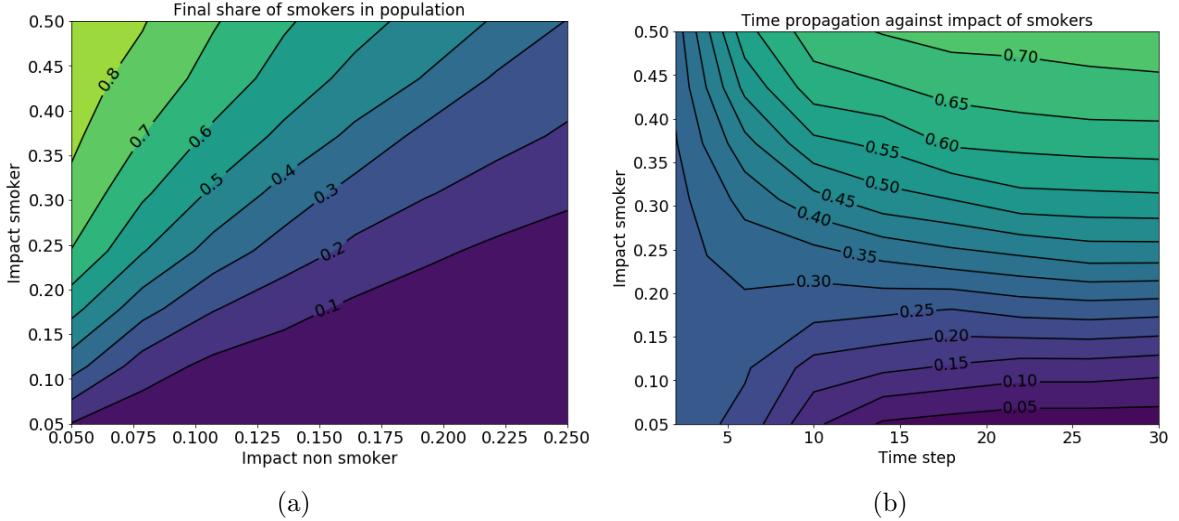


Figure 4: *Parameter dependence of the simulation.* Figure (a) shows the final shares of smokers in a given society for different combinations of the impact parameters `impact_smoker` and `impact_non-smoker`. This plot makes it possible to predict the outcome of our simulation for a given initial population, when changing the impact parameters. This plot is further described in section 4.4. In Figure (b) the results of the second experiment (which is described in section 4.5) are shown. Here the share of smokers in the society in all time steps is shown for different values of `impact_smoker`, while the parameter `impact_non-smoker` is fixed at 0.1 for all of these simulations.

## 4.6 Difficulties

One of the questions we faced was how to cope with varying results. Because we did not want to make our simulation deterministic we had to include random factors. These give each run a distinctive result, however they slightly hinder the reproducibility of our project. As we have seen in section 4.2, the unpredictability of the results originated mostly from the fact that we generated a random social network for every simulation. To cope with this, we could have generated a certain network right from the beginning to run all simulations and tests solely on this network. However, although using US/Swiss data for generating the network, there is still no reason why to favour one certain network, especially when keeping in mind that the results depend so heavily on it. As a result, we refrained from the approach to match the quantitative results of [1] and we also did not want to predict any developments of the number of smokers in Switzerland.

As a consequence, we focused on investigating qualitative properties of our model which lead to meaningful results as described in sections 4.3, 4.4 and 4.5.

## 5 Conclusion

As our main motivation for the simulation was the paper by Christakis and Fowler [1], we wanted to see in which ways we could reproduce the results of this study. As discussed in the section 4.6, we focused on reproducing qualitative aspects stated in [1] as it was not possible for us to reproduce the quantitative results. But qualitatively our model showed the formation of groups of smokers and groups of non-smokers in our society and it was possible for us to show that agents in our model are more likely to stop smoking if one of the connected agents quits smoking, which both are aspects mentioned in [1].

As far as the question whether the model can predict real life developments is concerned, we can only answer it in the negative (at least at the moment). However, as discussed in section 3.3, there are different ways to extend the model and add more complexity to it, which could lead to a better approximation of the reality.

## References

- [1] Nicholas A. Christakis and James H. Fowler. “The Collective Dynamics of Smoking in a Large Social Network”. In: *New England Journal of Medicine* 358.21 (May 2008), pp. 2249–2258. DOI: [10.1056/nejmsa0706154](https://doi.org/10.1056/nejmsa0706154).
- [2] Centers for Disease Control and Prevention. *Population by age groups, race, and sex for 1960-97*. Dec. 5, 2018. URL: <https://www.cdc.gov/nchs/data/statab/pop6097.pdf>.
- [3] Tamara G. Kolda et al. “A Scalable Generative Graph Model with Community Structure”. In: *SIAM Journal on Scientific Computing* 36.5 (Jan. 2014), pp. C424–C452. DOI: [10.1137/130914218](https://doi.org/10.1137/130914218).
- [4] Mark Newman. *Networks: An Introduction*. Oxford University Press, 2010. ISBN: 978-0-19-920665-0.
- [5] Tamara G. Kolda; Ali Pinar; others. *FEASTPACK v1.2*. Sandia National Laboratorie. Nov. 27, 2018. URL: <https://www.sandia.gov/~tgkolda/feastpack/>.
- [6] C. Seshadhri, Tamara G. Kolda, and Ali Pinar. “Community structure and scale-free collections of Erdős-Rényi graphs”. In: *Phys. Rev. E* 85 (5 May 2012), p. 056109. DOI: [10.1103/PhysRevE.85.056109](https://doi.org/10.1103/PhysRevE.85.056109). URL: <https://link.aps.org/doi/10.1103/PhysRevE.85.056109>
- [7] Bundesamt für Statistik. *Tabakkonsum in der Schweiz*. Oct. 7, 2018. URL: <https://www.bfs.admin.ch/bfs/de/home/statistiken/gesundheit/determinanten/tabak.html>.
- [8] *Switzerland Age structure*. Nov. 23, 2018. URL: [https://www.indexmundi.com/switzerland/age\\_structure.html](https://www.indexmundi.com/switzerland/age_structure.html).