Agent Based Modelling Software V0.3: Simulating the Influence of a Social Network on the Smoking Prevalence of the Age 16+ Population in 2012 Health Survey England

David Tian

University of Sheffield

Contents

1. Introduction of ABM Software V0.3

Agent based modelling (ABM) Software V0.3 simulates the influence of a social network which can represent the friendship of people on the smoking prevalence of the age 16+ population in England over 1 year period. V0.3 is the next version of V0.1 [1]. At the beginning of the simulation, an agent is created for each individual of the Health Survey England (HSE) 2012 data. An agent has 5 characteristics: age, sex, IMD quintile, a state (smoker, quitter, ex-smoker or never smoker) and the proportion of smoking friends. A time step represents a month. At each time step, the following three influences between macro entities and micro entities are modelled:

1. An intervention (macro entity) influences an agent's capability (micro entity) to quit smoking.
2. The proportion (macro entity) of the smoking friends of the agent's social network influences the threshold of quitting smoking (micro entity).
3. A smoking agent (micro entity) quits smoking or continues smoking based on the agent's intention to quit smoking (i.e. probability of quitting smoking), the agent's capability to quit smoking and the threshold of quitting smoking.

The above influences 1 and 2 are modelled as situational mechanisms in the MBSSM frame (Mechanism-based Social Systems Modelling) (Figure 1) [2]. The influence 3 is modelled as an action mechanism.

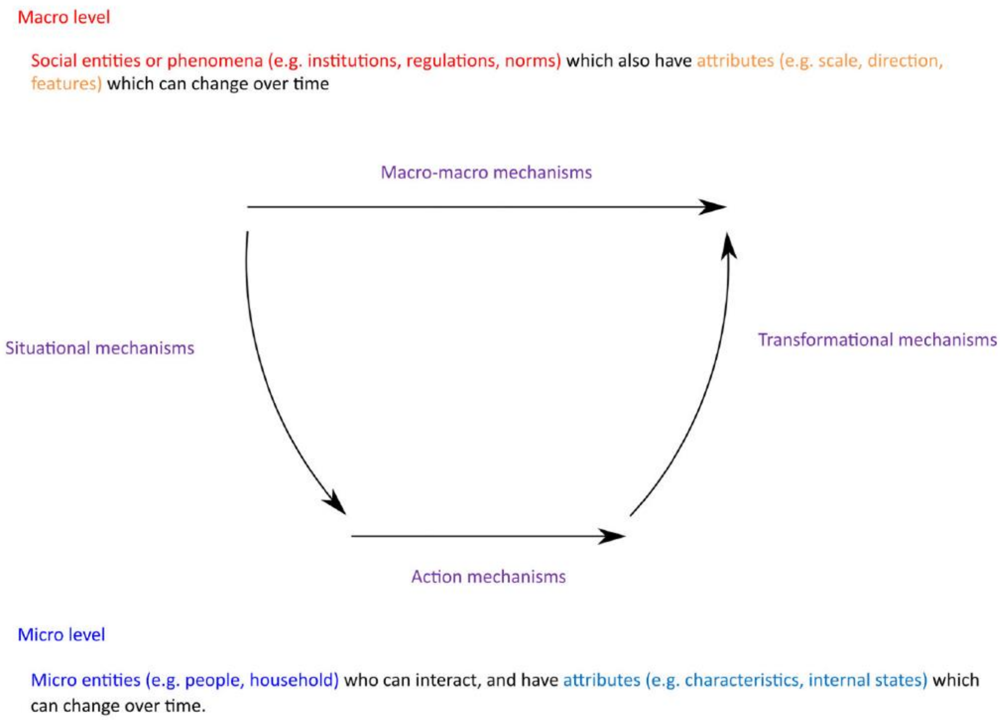


Figure 1: The MBSSM framework [2]

2. The Description of the Simulation

The steps of the simulation are as follows:

1. Create a synthetic social network of agents as follows:
    a. Generate a synthetic network of nodes using a graph algorithm e.g. Watts–Strogatz small-world graph algorithm.
    b. Initialize a population of agents to the STAPM 2012 data.
    c. Associate each agent to a different node of the network.
2. At each time step T (T=0,1,…,12) where T=0 is the beginning of the simulation; T=1 is the end of 1st month:
    a) Run the situational mechanism of each agent as follows:

        If intervention is used
            capability = effect size of intervention
        Else
            capability=1
        End
        calculate_proportion_of_smoking_friends(agent)

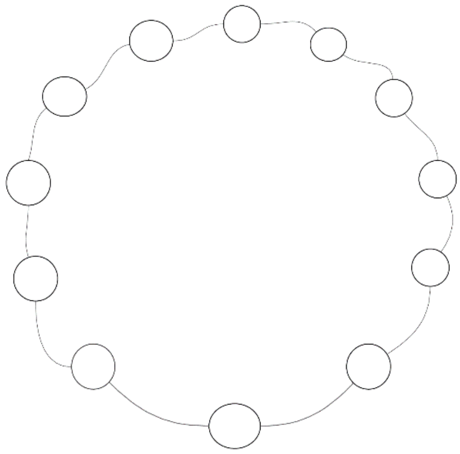
    b) Run the action mechanism of each agent as follows:
        I. Set the variable intention_to_quit to a random number drawn from the Normal(0,1).
        II. Set the variable threshold to the proportion of smoking friends of the agent
        III.
            If agent is a smoker at T-1 and capability x intention_to_quit >= threshold
                The agent is a quitter at T
            Else
                The state of agent is the same as its state at T-1
            End
    c) Calculate and display the smoking prevalence (percentage of smokers) of the current population of agents.

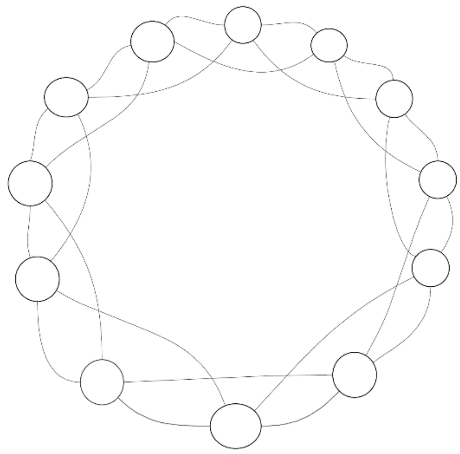
2.1 Watts–Strogatz Graph Algorithm

Watts–Strogatz generates a graph of nodes as follows:

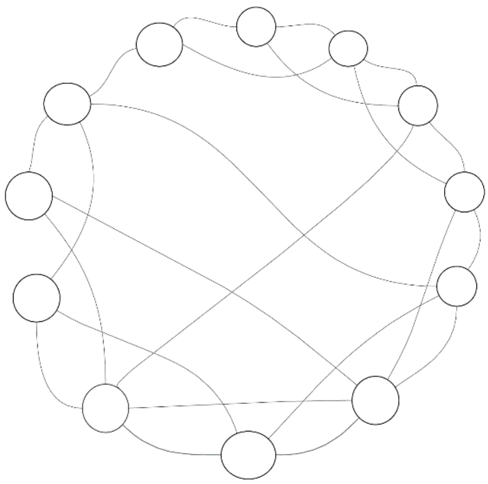
1. Create a ring over all nodes (Figure 2a).
2. Join each node in the ring to its k/2 nearest neighbours on each side. This gives a lattice network (Figure 2b).
3. Randomly rewire each edge (u,v) of the lattice as follows:
    - Choose a new node w uniformly and randomly.
    - With probability p, replace (u,v) with the edge (u,w) (Figure 2c).

The random rewiring does not increase the number of edges. The rewired graph is not guaranteed to be connected. The graph is called a small-world graph and resembles a social network e.g. friendship network. In this simulation, a friend of an agent A is an agent B (a nearest neighbour) who is linked to A by an edge in the network; k is the number of friends of each agent in the lattice network before rewiring of the edges.





a) Ring network b) Lattice network



c) Small-world network

Figure 2: Watts–Strogatz graph algorithm: a) firstly creates a ring network, then, b) transform the ring to a lattice network and c) finally transform the lattice to a small world network by rewiring some edges.

2.2 Parameters of the Simulation

The input data is the STAPM 2012 data of age >=16 people. STAPM 2012 data is a pre-processed data of the Health Survey England (HSE) 2012 data.

The parameters of the simulation are:

- K, the number of friends of each agent
- p, the probability of replacing an edge (u,v) with a new edge (u,w)
- intervention: either 1 (used) or 0 (not used)
- effect size of intervention: a value > 0 and <10
    - 1 represents that intervention is OFF.
    - A value >1 represents that the intervention increases the intention to quit.
    - A value < 1 represents that the intervention decreases the intention to quit.

3. Designing the Architecture of Software V0.3 using Object-oriented Design Approach

Vu et al. [1] proposed a software architecture for the MBSSM (Figure 2). The MBSSM software architecture is an object-oriented software structure [2] which consists of the 7 classes: Model, Micro-Agent, TheoryMediator, Theory, StructuralEntity, StructuralMediator and Regulator. Model is responsible for initializing a population of agents and run the simulation. The MicroAgent class represents agents. Theory consists of different types of mechanisms of an agent such as situation mechanism and action mechanism. TheoryMediator decides which mechanism(s) to run. A MicroAgent instance has a TheoryMediator instance which has 1 or more Theory instances. Each Theory instance is associated with a MicroAgent instance. StructuralEntity represents a macro entity such as social network. Regulator performs macro-macro mechanism. StructuralMediator decides which mechanism to run.

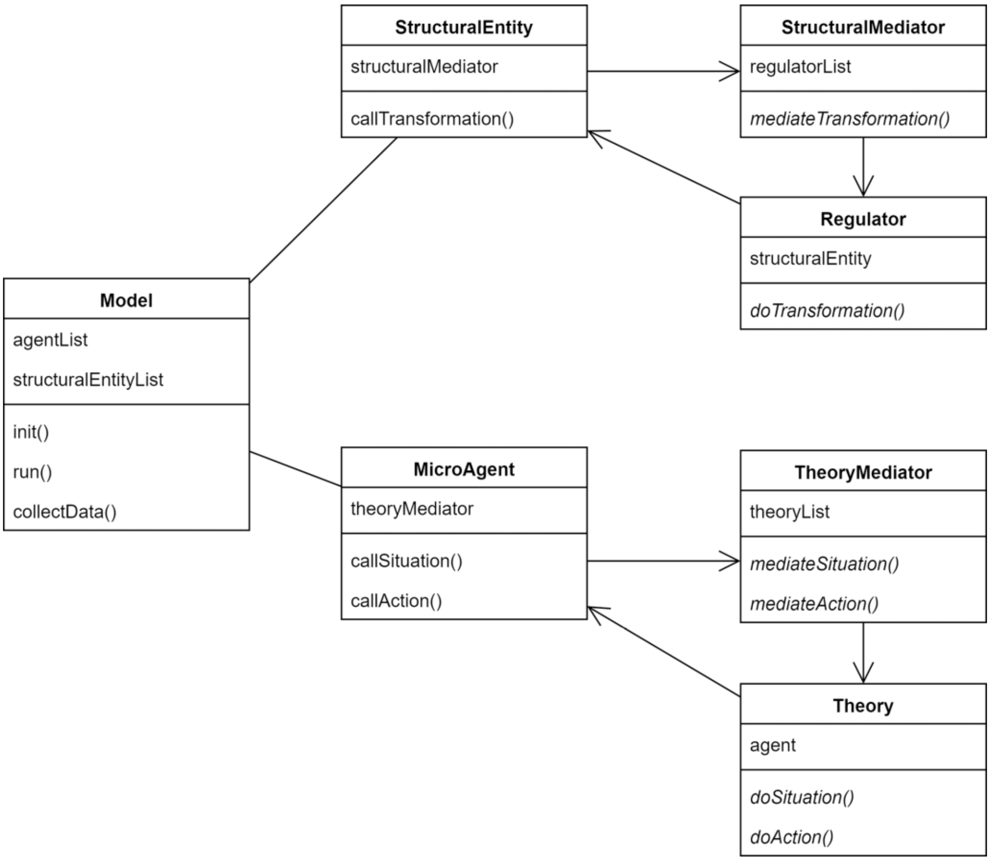


Figure 2: UML class diagram [3, 4] of the MBSSM software architecture

3.1 Integrating Repast4Py Classes Into The MBSSM Software Architecture

In order to run the mechanisms at each time step, a scheduler of execution of events must be integrated into the MBSSM software architecture. Repast4py [5] is an agent-based modelling library in Python. Three of the classes of Repast4py are SharedScheduleRunner, SharedContext and Agent classes. SharedScheduleRunner schedules execution of events at each time step. SharedContext represents a population of agents. Agent class represents a Repast4py agent. The id of a Repast4py agent consists of 3 components: id, type of agent and the process rank of the agent. The id is an integer unique to the agent; type is an integer representing the type of the agent; rank is the id of the process creating the agent. In this tutorial, there is a single type of agents and the type can be set to 0; the simulation is run on 1 process and the rank is set to 0. SharedScheduleRunner and SharedContext are added to Model as attributes (Figure 3). MicroAgent is defined a subclass of Repast4py agent so that MicroAgent inherits the id of a Repast4py agent. SharedContext has numerous MicroAgents.

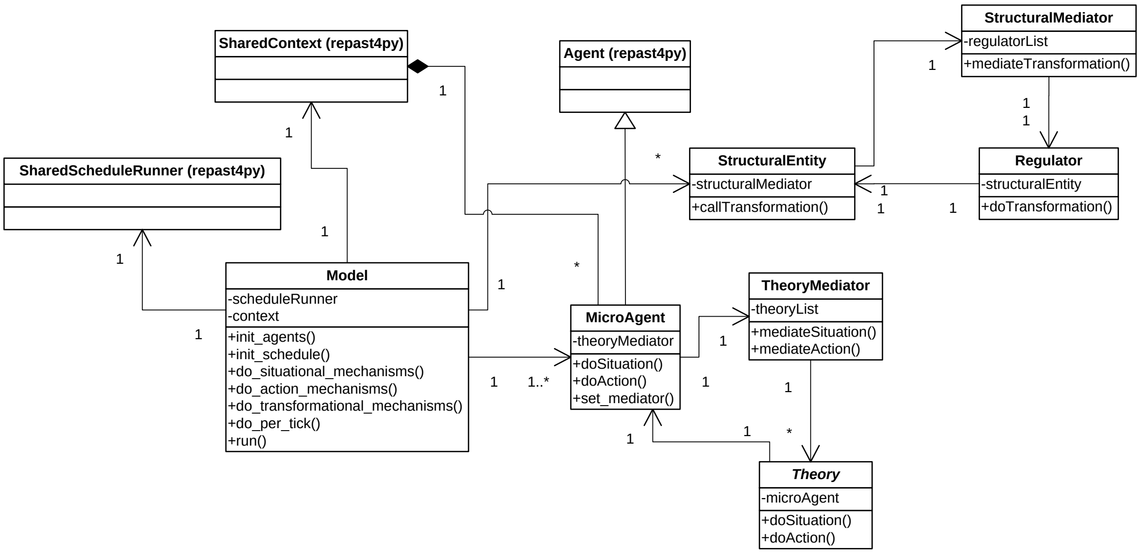


Figure 3: Class diagram of the software v0.3 with Repast4py classes integrated

3.2 Defining the Classes SocialNetwork, SmokingModel, Person, SocialNetworkQuitTheory and SmokingTheoryMediator

To simulate the influence of a social network on the smoking prevalence of the population, SocialNetwork, SmokingModel, Person, SocialNetworkQuitTheory and SmokingTheoryMediator classes are defined as follows:

- SocialNetwork is a subclass of StructuralEntity so that SocialNetwork inherits the attributes and operations of StructuralEntity. Additionally, generate_network_file operation is defined in SocialNetwork to generate a synthetic social network using Watts–Strogatz algorithm of NetworkX python library. Each node of the social network represents an individual in the STAPM 2012 data. The generated social network is saved to a file ‘network.txt’ using the network module of Repast4py.
- SmokingModel_v0_3 is a subclass of Model so that SmokingModel inherits the attributes and operations of Model. The operation add_network_to_context_as_projection creates a Repast4py network object from the file ‘network.txt’ by calling create_person and restore_person operations and returns a Graph object. The operation smoking_prevalence calculates the percentage of smokers at the current time step. The operation generate_Gephi_network_files creates a network file in Gephi visualization tool format. An effectSize attribute and getEffectSize operation are defined in SmokingModel to model the effect of the intervention on intention to quit.
- Person is a subclass of MicroAgent so that Person inherits the attributes and operations of MicroAgent. Additionally, Person has a list of states representing the states of the Person at different time steps and the attributes sex, age, IMD quintiles and proportion of smoking friends.

- SmokingTheoryMediator is a subclass of TheoryMediator so that SmokingTheoryMediator inherits the attributes and operations of TheoryMediator.
- SocialNetworkQuitTheory is a subclass of Theory so that QuitTheory inherits the attributes and operations of Theory. Additionally, SocialNetworkQuitTheory has attributes probability_of_quit, capability, opportunity, motivation and smokingModel.

A main module is defined to run the simulation by running generate_network_file operation of SocialNetwork class, then SmokingModel_v0_3 class.

The class diagram of software v0.3 is shown in Figure 3.

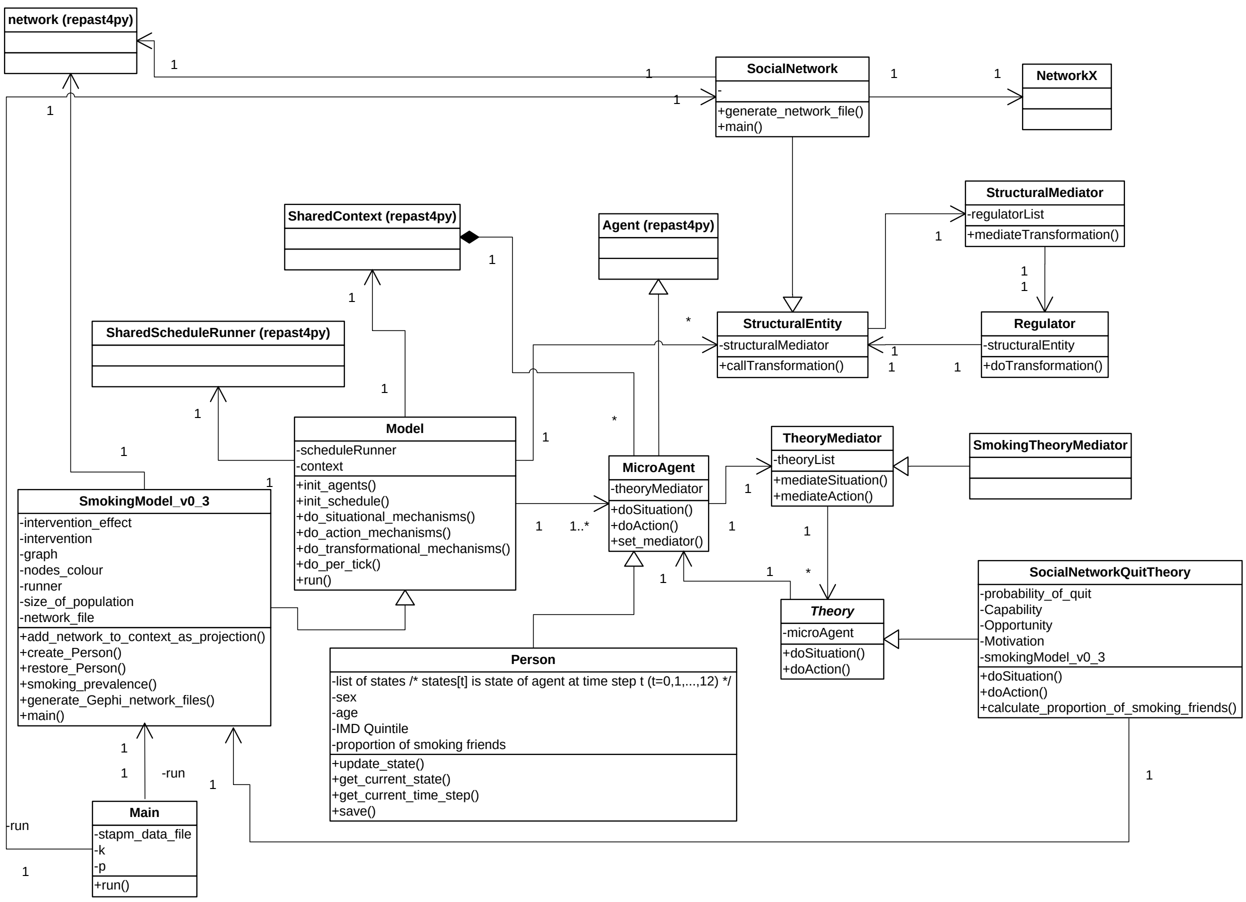


Figure 3: Class diagram of the software v0.3

4. Installation of ABM Software V0.3

ABM software v0.3 was implemented in Python 3.8.10. The Bitbucket repository of ABM software v0.3 is at:

https://bitbucket.org/co1dtx/abm-software-v0.3/src/master/

Download the repository to a directory on your computer e.g. D:/downloads.

4.1 Installation of Python 3, pip, MPI, venv and Repast4py

Repast4py can only be installed on linux or macOS operating systems. To install Repast4py on Windows 10 operating system, Windows Subsystem for Linux (WSL) can be firstly installed on Windows 10, then Repast4py can be installed on the WSL in the same way as installing it on Ubuntu linux (described below).

https://en.wikipedia.org/wiki/Windows_Subsystem_for_Linux

https://learn.microsoft.com/en-us/windows/wsl/about

Install python 3, pip, MPI, venv and Repast4py on Ubuntu linux in the following order:

1. Keep your Ubuntu linux with the latest packages by running the following two commands:

   ```
   sudo apt update
   ```

   ```
   sudo apt upgrade
   ```

2. Install python 3 if it is not available on your platform e.g. Ubuntu linux:

   https://www.makeuseof.com/install-python-ubuntu/

3. Install pip for python 3:

   ```
   apt install python3-pip
   ```

4. Install MPI on your platform (linux or macOS):

   https://repast.github.io/repast4py.site/guide/user_guide.html

5. Install venv to create virtual environment for Python 3:

   ```
   apt install python3-venv
   ```

   Then, create a virtual environment:

   ```
   python3 -m venv my_env
   ```

   More details on creating virtual environment is at:

   https://linuxopsys.com/topics/create-python-virtual-environment-on-ubuntu

6. Install Repast4py in the virtual environment that you created:

   1) Activate the virtual environment:

      ```
      source my_env/bin/activate
      ```

   2) Install Repast4py in the virtual environment:

      ```
      env CC=mpicxx pip install repast4py
      ```

5. Example Runs of ABM Software v0.3

hse2012_stapm.csv is the STAPM 2012 data:

- "age","sex","qimd","state"
- 58,"female",2,"never_smoker"
- 47,"male",2,"never_smoker"
- 39,"male",3,"never_smoker"
- 41,"male",1,"ex-smoker"
- 37,"female",1,"ex-smoker"
- 51,"male",1,"ex-smoker"
- 48,"female",1,"ex-smoker"
- 19,"male",1,"never_smoker"
- …

Run 1

Open model.yaml under props directory using a text editor, then enter and save the following settings of the parameters:

- stop.at: 12
- intervention: 0
- intervention_effect: 2

Run the following command on the command line with the STAPM data 'hse2012_stapm.csv', k=2 and p=0.25:

python main.py hse2012_stapm.csv 2 0.25

The results (smoking prevalence and average proportion of smoking friends of each agent) are displayed on screen:

```
(my_env_pc) uos@acse12349:/mnt/u/smoking cessation/ABM software/code_v0_3$ python main.py hse2012_stapm.csv 2 0.25
STAPM data: hse2012_stapm.csv
parameters of Watts-Strogatz Graph Algorithm for generating a social network:
k=2
p=0.25
stop at time step: 12
intervention: 0
intervention effect: 2
size of population: 8215
At time step: 0 (beginning of simulation), smoking prevalence=18.94%.
At time step: 1 (after doing situation mechanism and action mechanism): smoking prevalence=5.22%.
At time step: 2 (after doing situation mechanism and action mechanism): smoking prevalence=2.24%.
At time step: 3 (after doing situation mechanism and action mechanism): smoking prevalence=1.12%.
At time step: 4 (after doing situation mechanism and action mechanism): smoking prevalence=0.6%.
At time step: 5 (after doing situation mechanism and action mechanism): smoking prevalence=0.3%.
At time step: 6 (after doing situation mechanism and action mechanism): smoking prevalence=0.19%.
At time step: 7 (after doing situation mechanism and action mechanism): smoking prevalence=0.1%.
At time step: 8 (after doing situation mechanism and action mechanism): smoking prevalence=0.04%.
At time step: 9 (after doing situation mechanism and action mechanism): smoking prevalence=0.01%.
At time step: 10 (after doing situation mechanism and action mechanism): smoking prevalence=0.0%.
At time step: 11 (after doing situation mechanism and action mechanism): smoking prevalence=0.0%.
At time step: 12 (after doing situation mechanism and action mechanism): smoking prevalence=0.0%.
At time step 0 (before doing situation mechanism and action mechanism): average proportion of smoking friends of each agent=18.73%, std=29.6%, min=0.0%, max=100.0%
At time step 1 (before doing situation mechanism and action mechanism): average proportion of smoking friends of each agent=18.73%, std=29.6%, min=0.0%, max=100.0%
At time step 2 (before doing situation mechanism and action mechanism): average proportion of smoking friends of each agent=5.04%, std=16.93%, min=0.0%, max=100.0%
At time step 3 (before doing situation mechanism and action mechanism): average proportion of smoking friends of each agent=2.16%, std=11.47%, min=0.0%, max=100.0%
At time step 4 (before doing situation mechanism and action mechanism): average proportion of smoking friends of each agent=1.07%, std=8.27%, min=0.0%, max=100.0%
At time step 5 (before doing situation mechanism and action mechanism): average proportion of smoking friends of each agent=0.53%, std=5.94%, min=0.0%, max=100.0%
At time step 6 (before doing situation mechanism and action mechanism): average proportion of smoking friends of each agent=0.28%, std=4.38%, min=0.0%, max=100.0%
At time step 7 (before doing situation mechanism and action mechanism): average proportion of smoking friends of each agent=0.18%, std=3.51%, min=0.0%, max=100.0%
At time step 8 (before doing situation mechanism and action mechanism): average proportion of smoking friends of each agent=0.07%, std=2.18%, min=0.0%, max=100.0%
At time step 9 (before doing situation mechanism and action mechanism): average proportion of smoking friends of each agent=0.03%, std=1.23%, min=0.0%, max=50.0%
At time step 10 (before doing situation mechanism and action mechanism): average proportion of smoking friends of each agent=0.01%, std=0.78%, min=0.0%, max=50.0%
At time step 11 (before doing situation mechanism and action mechanism): average proportion of smoking friends of each agent=0.0%, std=0.0%, min=0.0%, max=0.0%
At time step 12 (before doing situation mechanism and action mechanism): average proportion of smoking friends of each agent=0.0%, std=0.0%, min=0.0%, max=0.0%
```

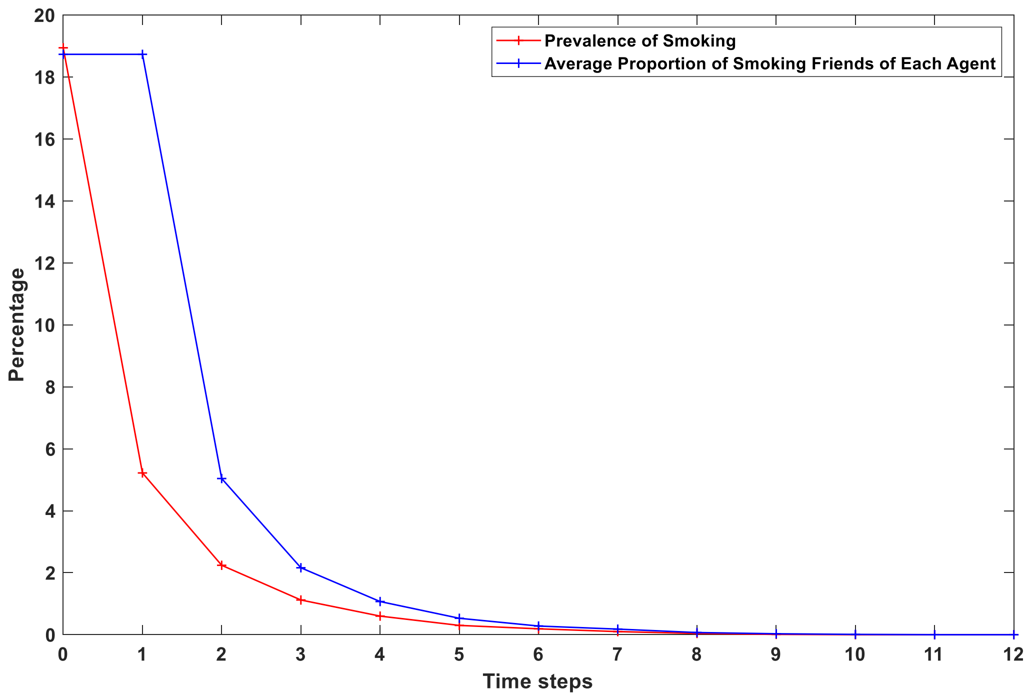


Figure 4: Relationship of the smoking prevalence (%) and average proportion of smoking friends of each agent (%) at each time step (with no intervention used)

The plots.m Matlab script can be used to plot the results of the simulation i.e. smoking prevalence and average proportion of smoking friends of each agent at each time step. Figure 4 was generated using plots.m with the results of no intervention used. As the average proportion of smoking friends of each agent decreases across the duration of simulation, the prevalence of smoking decreases. This result is realistic because the smaller the proportion of smoking friends of a smoker, the more likely that smoker quits smoking resulting in a lower smoking prevalence of the population and vice versa.

<u>Run2</u>

Open model.yaml under props directory using an editor, then enter and save the following settings of the parameters:

- stop.at: 12
- intervention: 1
- intervention_effect: 2

Run the following command on the command line with the STAPM data 'hse2012_stapm.csv', k=2 and p=0.25:

python main.py hse2012_stapm.csv 2 0.25

The results (smoking prevalence and average proportion of smoking friends of each agent) are displayed on screen:

```
(my_env_pc) uos@acse12349:/mnt/u/smoking cessation/ABM software/code_v0_3$ python main.py hse2012_stapm.csv 2 0.25
STAPM data: hse2012_stapm.csv
parameters of Watts-Strogatz Graph Algorithm for generating a social network:
k=2
p=0.25
stop at time step: 12
intervention: 1
intervention effect: 2
size of population: 8215
At time step: 0 (beginning of simulation), smoking prevalence=18.94%.
At time step: 1 (after doing situation mechanism and action mechanism): smoking prevalence=3.03%.
At time step: 2 (after doing situation mechanism and action mechanism): smoking prevalence=0.44%.
At time step: 3 (after doing situation mechanism and action mechanism): smoking prevalence=0.04%.
At time step: 4 (after doing situation mechanism and action mechanism): smoking prevalence=0.0%.
At time step: 5 (after doing situation mechanism and action mechanism): smoking prevalence=0.0%.
At time step: 6 (after doing situation mechanism and action mechanism): smoking prevalence=0.0%.
At time step: 7 (after doing situation mechanism and action mechanism): smoking prevalence=0.0%.
At time step: 8 (after doing situation mechanism and action mechanism): smoking prevalence=0.0%.
At time step: 9 (after doing situation mechanism and action mechanism): smoking prevalence=0.0%.
At time step: 10 (after doing situation mechanism and action mechanism): smoking prevalence=0.0%.
At time step: 11 (after doing situation mechanism and action mechanism): smoking prevalence=0.0%.
At time step: 12 (after doing situation mechanism and action mechanism): smoking prevalence=0.0%.
At time step 0 (before doing situation mechanism and action mechanism): average proportion of smoking friends of each agent=18.97%, std=29.86%, min=0.0%, max=100.0%
At time step 1 (before doing situation mechanism and action mechanism): average proportion of smoking friends of each agent=18.97%, std=29.86%, min=0.0%, max=100.0%
At time step 2 (before doing situation mechanism and action mechanism): average proportion of smoking friends of each agent=2.96%, std=12.88%, min=0.0%, max=100.0%
At time step 3 (before doing situation mechanism and action mechanism): average proportion of smoking friends of each agent=0.41%, std=4.88%, min=0.0%, max=100.0%
At time step 4 (before doing situation mechanism and action mechanism): average proportion of smoking friends of each agent=0.02%, std=1.1%, min=0.0%, max=50.0%
At time step 5 (before doing situation mechanism and action mechanism): average proportion of smoking friends of each agent=0.0%, std=0.0%, min=0.0%, max=0.0%
At time step 6 (before doing situation mechanism and action mechanism): average proportion of smoking friends of each agent=0.0%, std=0.0%, min=0.0%, max=0.0%
At time step 7 (before doing situation mechanism and action mechanism): average proportion of smoking friends of each agent=0.0%, std=0.0%, min=0.0%, max=0.0%
At time step 8 (before doing situation mechanism and action mechanism): average proportion of smoking friends of each agent=0.0%, std=0.0%, min=0.0%, max=0.0%
At time step 9 (before doing situation mechanism and action mechanism): average proportion of smoking friends of each agent=0.0%, std=0.0%, min=0.0%, max=0.0%
At time step 10 (before doing situation mechanism and action mechanism): average proportion of smoking friends of each agent=0.0%, std=0.0%, min=0.0%, max=0.0%
At time step 11 (before doing situation mechanism and action mechanism): average proportion of smoking friends of each agent=0.0%, std=0.0%, min=0.0%, max=0.0%
At time step 12 (before doing situation mechanism and action mechanism): average proportion of smoking friends of each agent=0.0%, std=0.0%, min=0.0%, max=0.0%
```

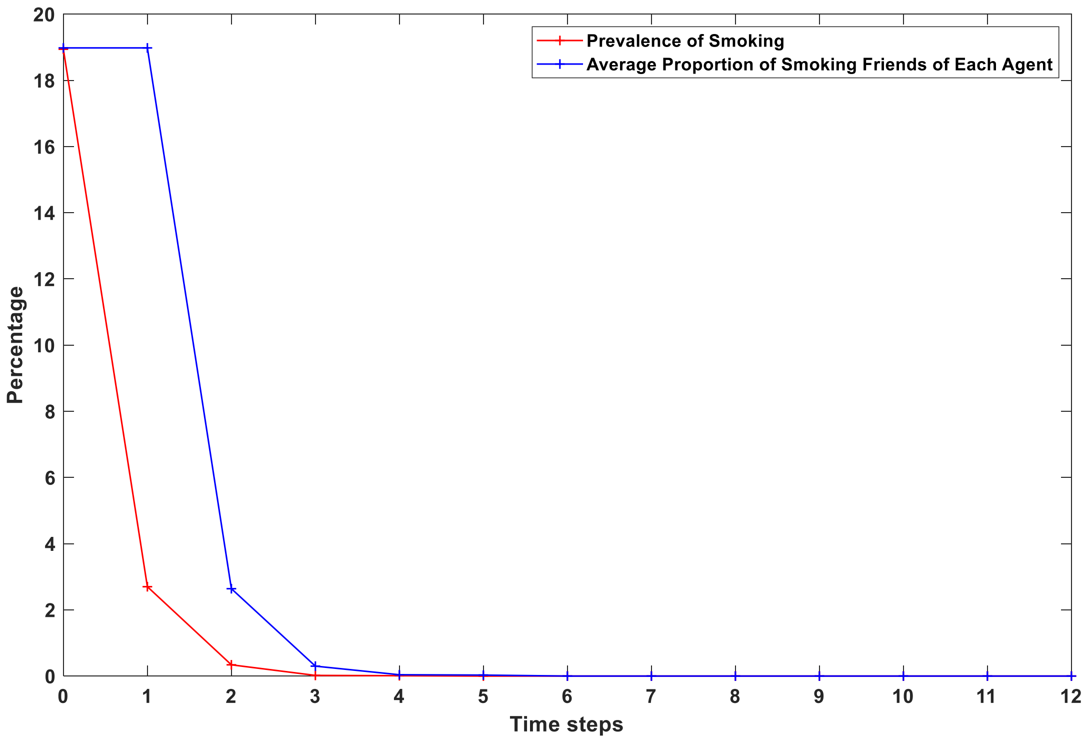


Figure 5: Relationship of the smoking prevalence (%) and average proportion of smoking friends of each agent (%) at each time step (with intervention been used)

Figure 5 was generated using plots.m with the results of intervention been used. As the average proportion of smoking friends of each agent decreases across the duration of simulation, the prevalence of smoking decreases. This relationship is realistic because in reality the smaller the proportion of smoking friends of a smoker, the better chance the smoker quits smoking resulting in a lower smoking prevalence of the population and vice versa. Using intervention causes a faster decrease in prevalence of smoking compared to when no intervention used (Figure 4).

6. Visualization of Social Networks

6.1 Visualization of Social Networks using Gephi Network Visualization Tool

For a social network created with k=2 or 3, Gephi [7] can be used to visualize it as follows.

1. Run main.py which generates the following Gephi files during ABM simulation:
   Gephi nodes files: nodes0.csv,…,nodes12.csv where nodes0.csv corresponds to time step 0.
   Gephi edges files: edges0.csv,…,edges12.csv where edeges0.csv corresponds to time step 0.

2. Visualize the network at time step 0 using Gephi as follows:

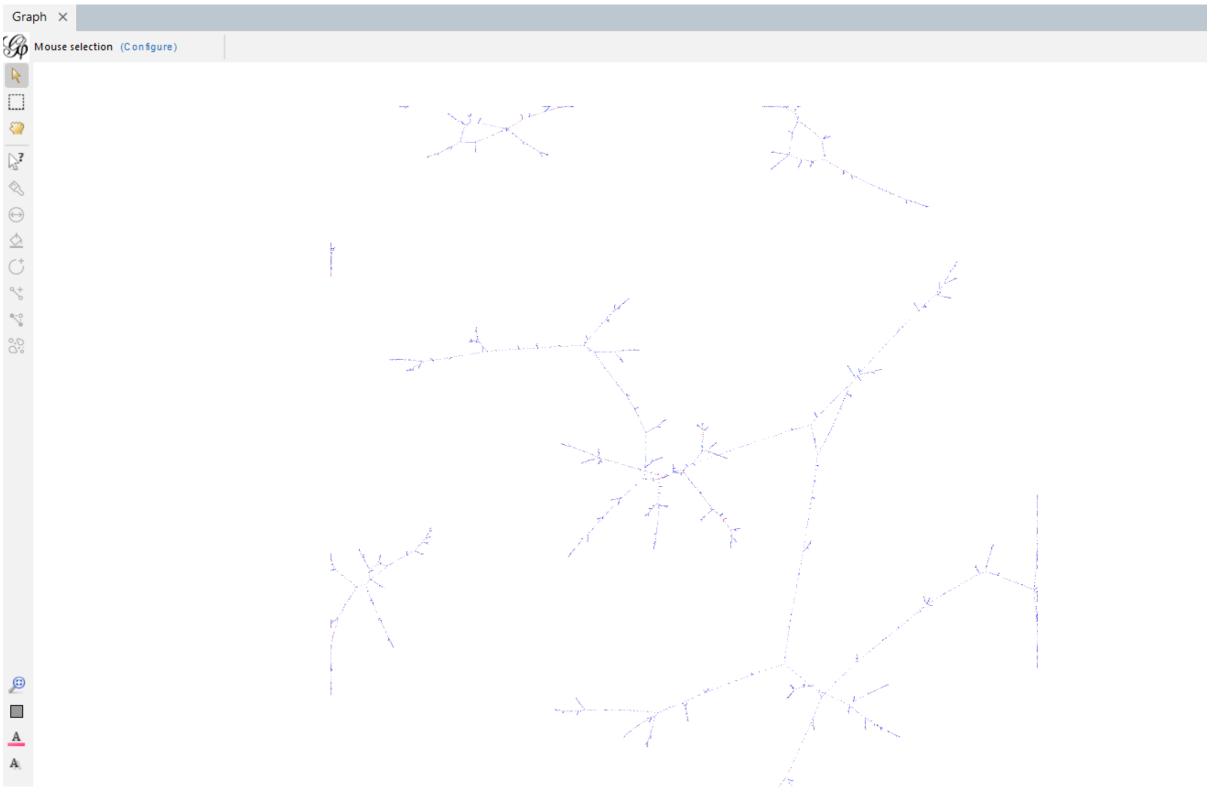
   1) Open nodes0.csv and edges0.csv in Gephi.

   2) Set the nodes of smokers to red and those of non-smokers to blue.

3) Layout the network using ForceAtlas2. For k=2 or 3, ForceAtlas2 algorithm layouts the network quickly (~1 or 2 minutes). The network has 8215 edges.

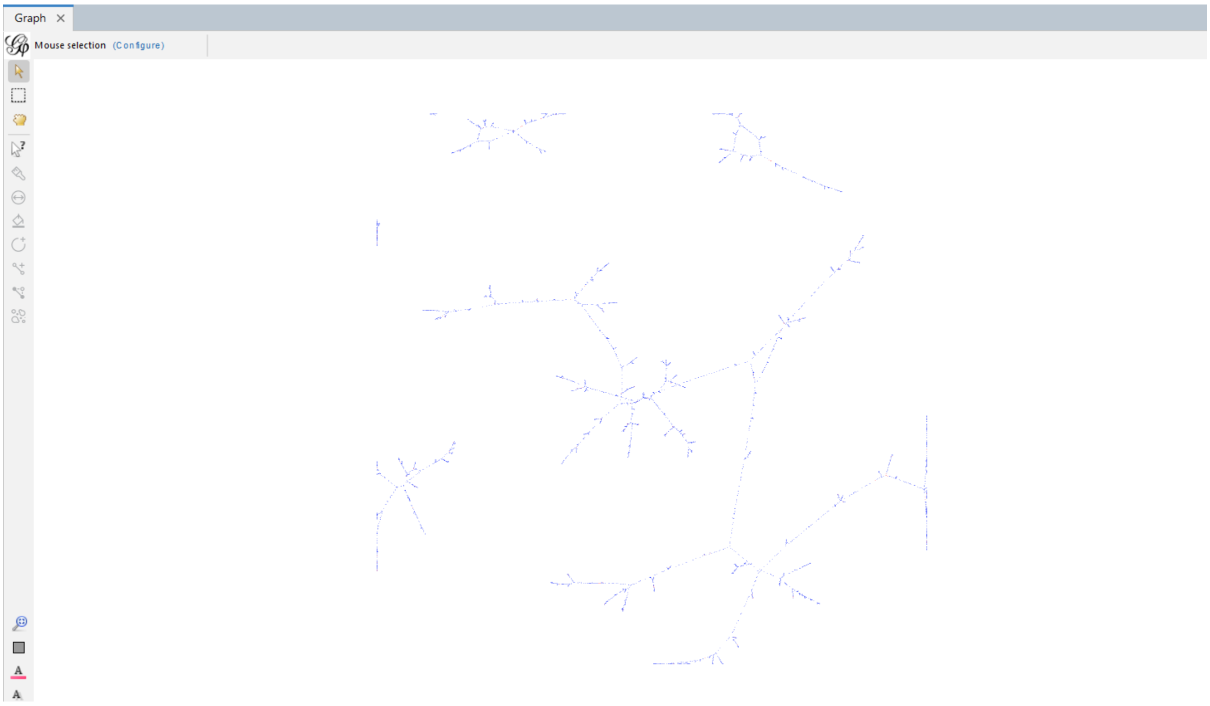
   4) Save the network as a gml file e.g. graph0.gml

   5) Export the network to a PDF file e.g. graph0.pdf.



3. To visualize the network at time step 1, update the states of the nodes to the states of time step 1 as follows.
   - Run update_gml_file.py passing **nodes1.csv** and **graph0.gml** as input and save the updated gml file as **graph1.gml**

4. Visualize the network by opening graph1.gml using Gephi. Then, export the visualization to a PDF file e.g. graph1.pdf.



5. To visualize the network at a time step i e.g. i=2, repeat steps 3 and 4 on nodesi.csv and graph0.gml to create graphi.gml and graphi.pdf.

6.2 Visualization of Social Networks using Matlab

For a social network created with k>=4, Gephi gives a poor layout of the network. For k = 4 or 5, ForceAtlas2 gave poor layout for the network and the process never converged. This may be because the network has a large number of edges (16430 edges). Matlab can be used to visualize social networks generated with k>=4. In a social network, the nodes representing smokers are red and the nodes representing non-smokers are blue.

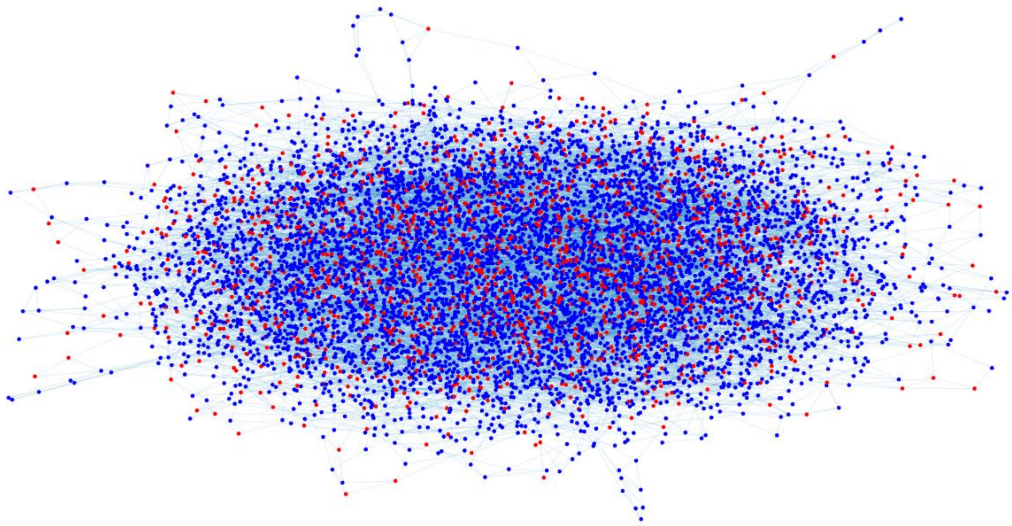
1. Run main.py which generates the following files during ABM simulation:
    - nodes colour files: **nodes_colour0.csv,…,nodes_colour12.csv** where nodes_colour0.csv corresponds to time step 0.
    - A gml file e.g. **graph_k=5.gml**

2. To visualize the network with k=5, open visualize_gml_file.m using Matlab, then pass **nodes_colour0.csv** and **graph_k=5.gml** to visualize_gml_file.m.

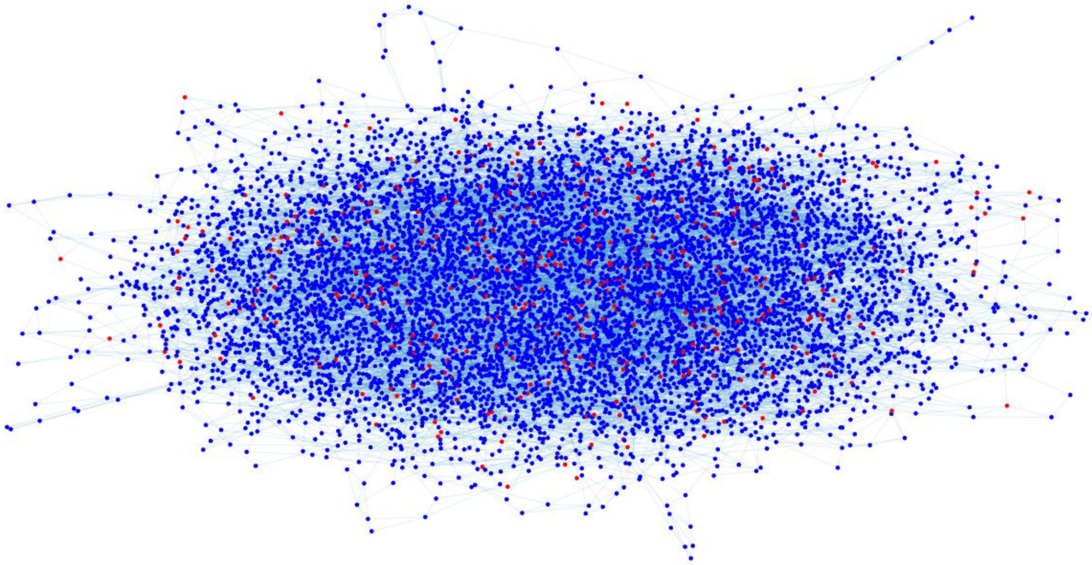
visualize_gml_file.m

```
g=read_gml("./results/k=5_p=0_25/graph_k=5.gml");
t=struct2table(g);
t2=struct2table(t.edge);
source=table2array(t2(:,1));
target=table2array(t2(:,2));
source=source+1;%indices start from 1
target=target+1;
g=graph(source,target);
c=readmatrix("nodes_colour0.csv");
%c=readmatrix("nodes_colour1.csv");
%c=readmatrix("nodes_colour2.csv");
%c=readmatrix("nodes_colour3.csv");
h=figure(1);
p=plot(g,'NodeColor',c,'EdgeAlpha',0.1,'Layout','force');
%savefig(h,'graph0_k=5_intervention=1.fig');
```

3. Run visualize_gml_file.m in Matlab. A visualization of the social network at time step 0 is displayed on screen.



4. To visualize the network with k=5 at time step i e.g. i=1, pass **nodes_colouri.csv** and **graph_k=5.gml** to visualize_gml_file.m and run it. For example, to visualize the network at time step 1, pass nodes_colour1.csv and graph_k=5.gml to visualize_gml_file.m.



References

[1] ABM Software V0.1 https://bitbucket.org/co1dtx/abm-software-v0.1/src/master/

[2] Vu et al. A software architecture for mechanism-based social systems modelling in agent-based simulation models. Journal of Artificial Societies and Social Simulation, 2020; 23(3):1

[3] https://en.wikipedia.org/wiki/Object-oriented_design

[4] https://en.wikipedia.org/wiki/Class_diagram

[5] https://en.wikipedia.org/wiki/Unified_Modeling_Language

[6] https://repast.github.io/repast4py.site/index.html

[7] Gephi Network Visualization Tool https://gephi.org/