

RESEARCH

Open Access



Improving tobacco social contagion models using agent-based simulations on networks

Adarsh Prabhakaran^{1*}, Valerio Restocchi^{1†} and Benjamin D. Goddard^{2†}

[†]Valerio Restocchi and Benjamin D. Goddard have contributed equally to this work.

*Correspondence: a.prabhakaran@sms.ed.ac.uk

¹ Artificial Intelligence and Its Applications Institute, School of Informatics, The University of Edinburgh, 10 Chricton Street, Edinburgh EH8 9AB, United Kingdom

² School of Mathematics and Maxwell Institute for Mathematical Sciences, The University of Edinburgh, EH9 3FD Edinburgh, United Kingdom

Abstract

Tobacco use is the leading cause of preventable deaths in developed countries. Many interventions and policies have been implemented to reduce the levels of smoking but these policies rarely rely on models that capture the full complexity of the phenomenon. For instance, one feature usually neglected is the long-term effect of social contagion, although empirical research shows that this is a key driver of both tobacco initiation and cessation. One reason why social contagion is often dismissed is that existing models of smoking dynamics tend to be based on ordinary differential equation (ODE), which are not fit to study the impact of network effects on smoking dynamics. These models are also not flexible enough to consider all the interactions between individuals that may lead to initiation or cessation. To address this issue, we develop an agent-based model (ABM) that captures the complexity of social contagion in smoking dynamics. We validate our model with real-world data on historical prevalence of tobacco use in the US and UK. Importantly, our ABM follows empirical evidence and allows for both initiation and cessation to be either spontaneous or a consequence of social contagion. Additionally, we explore in detail the effect of the underlying network topology on smoking dynamics. We achieve this by testing our ABM on six different networks, both synthetic and real-world, including a fully-connected network to mimic ODE models. Our results suggest that a fully-connected network is not well-suited to replicate real data, highlighting the need for network models of smoking dynamics. Moreover, we show that when a real network is not available, good alternatives are networks generated by the Lancichinetti–Fortunato–Radicchi and Erdős–Rényi algorithms. Finally, we argue that, in light of these results, our ABM can be used to better study the long-term effects of tobacco control policies.

Keywords: Agent-based model, Smoking dynamics, Social contagion, Tobacco model, Networks

Introduction

Smoking is one of the leading preventable causes of death, disability and disease across the world (US Department of Health and Human Services and others 2014, 2020; Office for National Statistics 2019) and it is one of the most significant avoidable hazard factors for cancer (Banks et al. 2015) and respiratory diseases (Ferkol and Schraufnagel 2014). Not only is smoking a global health burden, but it is also an economic burden, which

significantly outweighs the economic benefits from tobacco production and sales (Drope et al. 2018).

Acknowledging the need for active efforts on tobacco control, in 2005 182 countries ratified the first international public health treaty, the WHO Framework Convention on Tobacco Control (FCTC) (World Health Organization and others 2004). This involved cigarette taxation, smoke-free zones, public media campaigns (Vallone et al. 2009), restrictions on advertisement for cigarettes (Gilpin and Pierce 1997; Rose et al. 2013), health warnings, cessation support (Brown et al. 2014) and control on access to tobacco products (Millett et al. 2011; Schneider et al. 2011). Effective implementation of such tobacco control policies has decreased deaths, extended both the lifespan and life expectancy of the population (van Meijgaard and Fielding 2012; Holford et al. 2014), and is also associated with a predicted decrease in healthcare expenditure (Lightwood and Glantz 2016). However, despite the success of these tobacco control policies, the rate of decline of smoking prevalence has slowed down, and the world's economies still spend more than one trillion USD per year on smoking-related health expenditures and loss of productivity (Acharya et al. 2016; Goodchild et al. 2018).

An important factor that is commonly overlooked by tobacco-control policy making is the long-term effect of social contagion on both smoking levels and socio-economic inequalities, especially given that the latter can be unexpectedly exacerbated by policies that only optimise short-term effects (Caryl et al. 2021).

In fact, despite the large body of evidence that suggests that tobacco initiation and cessation largely depend on social ties (Christakis and Fowler 2008; Blok et al. 2017; Ennett et al. 2008; Go et al. 2010; Mercken et al. 2009), there is currently no model that fully captures the complexity of social contagion in smoking dynamics.

Consequently, such a model is crucially needed to develop policies that not only accurately take into account long-term effects, but also exploit social contagion to enhance tobacco control.

This is not to say models of social contagion for smoking dynamics do not exist. In fact, many models have been proposed in which smoking is compared to a disease spreading in a population due to its contagion-like behaviour (Sharomi and Gumel 2008; Zaman 2011a, b; Zaman et al. 2017). This similarity allows the use of tools from epidemiology to model the propagation of smoking behaviour, leading most models to be variations of compartmental models such as SIR and SIER that use ordinary differential equations (ODEs) to describe the smoking dynamics. However, epidemiology has successfully moved on to the more flexible agent-based models (ABM) (Turner et al. 2020; Aleta et al. 2022; Hunter et al. 2018), but tobacco control has not. Therefore, existing models of social contagion for tobacco control cannot accurately reproduce the empirically observed complexity of this phenomenon, due to the following reasons.

First, most of these models do not account for the topology of the social ties. When modelling any type of social contagion, it is known that the social interaction between individuals plays an important role (Hodas and Lerman 2014; Shin 2022). Therefore such models can only be accurate if they account for the underlying social network of the population. Even though compartmental models try to incorporate social interactions between groups of individuals, there is no network structure involved. To reduce the complexity and analytical tractability, these models ignore the structure

of the social ties and instead assume a homogeneous well-mixed population, which means that any individual can infect others in the system (Anderson et al. 1992; Kermack and McKendrick 1927, 1932, 1933). This means that vital information from the actual social network is not taken into consideration (Moore and Newman 2000).

Second, when these models do consider network structure, they arbitrarily fix the structure. The models of smoking behaviour which consider the structure of these social ties are agent-based models (Chao et al. 2015; Schaefer et al. 2012, 2013). In these cases, the topology is usually arbitrarily fixed as a scale-free network or on small scale school-network. Since real-world social contact networks of adults can be very different from synthetic and school networks, there is a need for careful characterisation of the smoking behaviour of these models on different network topologies.

Finally, even though empirical research shows that smoking initiation and cessation largely depend on social ties (Christakis and Fowler 2008), not all of these interactions have been considered for modelling the spread of smoking. Although one of the first theoretical studies which modelled smoking cessation advocated the use of interactions between smokers and quitters (Castillo-Garsow et al. 1997), interactions which lead to smoking cessation and relapse are still not used (Schaefer et al. 2013). Consequently, in these models smoking cessation and relapse are usually determined only by spontaneous terms, which causes an underestimation of social contagion effects (Sharomi and Gumel 2008; Zaman 2011a, b; Zaman et al. 2017). As previously discussed, this underestimation can lead to unintended consequences of tobacco control policies such as an increase in socio-economic inequalities.

A closely related field to the spread of smoking behaviour is opinion dynamics, wherein mathematical and computational models are used to study the spread of opinions in a population by considering social influence. Like opinions, smoking behaviour can be influenced by the attitudes and behaviours of others, such as peers, family members, and by external influences like media (Mueller and Tan 2018; Colaiori and Castellano 2015). However, unlike opinions, smoking is a health hazard. This makes the spread of the smoking behaviour also similar to that of an infectious disease where instead of the infection, smoking behaviour is the contagion. Therefore, modelling smoking behaviour can be seen through a hybrid lens of epidemiology and opinion dynamics. Such an approach will help us combine the opinion dynamics perspective which considers the social and psychological factors that influence the adoption and maintenance of smoking behaviour (such as peer pressure, familial attitude towards smoking, etc.) and the epidemiological perspective capturing the mechanisms contributing to the spread of smoking-related health hazards.

Over the years, the effect of social networks on individual and population behaviour have been studied in both opinion dynamics and epidemiology (Rahmandad and Sterman 2008). Multiple approaches have been designed to study the spread of diseases and opinion according to the situation and amount of information available. More recently, these models have been used to study a variety of social contagions including obesity (Hill et al. 2010a), emotions (Hill et al. 2010b), alcoholism (Lee et al. 2010; Sharma and Samanta 2015), substance abuse (White and Comiskey 2007), behaviour change (Badham et al. 2021), and information spreading (Zhou et al. 2020), to name

a few. Due to the similarity of the spread of smoking behaviour to other social contagions, we can use insights from these fields to develop models.

To address these issues with existing smoking dynamics models, we develop an agent-based model (ABM). ABMs are a class of computational techniques which rely on dynamical interactions between autonomous agents to understand the emerging properties of a complex system due to these local interactions. We use ABMs for the following reasons:

First, unlike the ODE models, ABMs can easily be extended to multiple theoretical and real-world networks. This versatile nature allows ABMs to be applied to different population structures and more realistic models of the system. Therefore, the effect of network structure on the smoking dynamics can easily be studied by changing the underlying network topology. Consequently, we can characterise the dynamics of smoking behaviour in multiple different network topologies.

Second, interactions between agents can easily be incorporated into ABMs without significantly increasing the complexity of the model. Conversely, when multiple interactions are considered in an ODE model, the model becomes too complex to be solved analytically. By constructing ODE models which are not solvable (or ones in which the solutions are too complex and lengthy), it becomes difficult to validate and analyse the nature of the solutions analytically. Instead, such ODE systems must be solved using numerical methods to approximate the solutions.

In our ABM for smoking, we include three state change processes: smoking initiation, cessation and relapse. Each of these processes can occur due to both interactions and spontaneously. This accounts for the multiple possible interactions which can lead to a change in smoking behaviour.

Finally, ABMs are flexible enough to become effective test-beds for developing new policies. One of the main applications of studying contagion (both infectious diseases and social contagion) is to develop strategies to contain them. Effective strategies and interventions may prevent smoking initiation, motivate smokers to quit, and stop former smokers from relapsing. Due to the socially contagious nature of smoking, we can potentially use network-based strategies developed from studies on infectious diseases and other social contagions.

First, we develop an agent-based model, which considers multiple possible interactions along with spontaneous terms to study the spread of smoking. Our model can be used to develop network-based intervention strategies and policies for tobacco control. Furthermore, we show the robustness of our ABMs by comparing the dynamics against a traditional ODE model and as expected, our results suggest that our ABM on a fully connected network and the equivalent ODE model provide the same results. Additionally, we show that ABMs on fully connected networks and ODE models should not be used to model smoking behaviour as they replicate the real-world data with poor accuracy compared to the other networks.

Next, we explore the effect of the underlying network topology on smoking dynamics. We find that the underlying network structure affects smoking dynamics considerably. However, synthetic networks with the same average degree reproduced the historic data and showed similar characteristics as that of the real-world networks. Specifically, we show that Lancichinetti–Fortunato–Radicchi benchmark networks and random

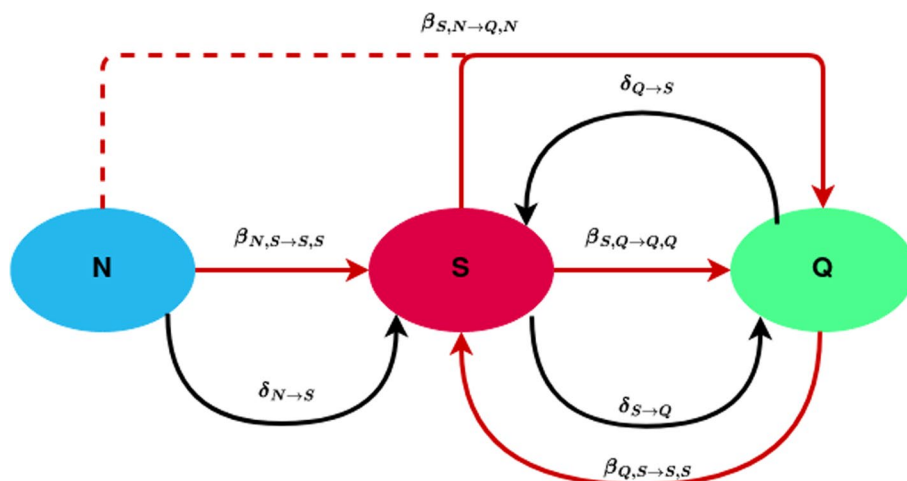


Fig. 1 The figure shows the schematic representation of the state change processes involved in the ABM. The interaction parameters are represented by the red arrows, while the black arrows show the spontaneous terms in the schematic. All three state-change processes are shown in the figure. First, an N-agent can initiate smoking spontaneously ($\delta_{N \rightarrow S}$) or due to the interaction with an S-agent ($\beta_{N,S \rightarrow S,S}$). Similarly, an S-agent can quit spontaneously ($\delta_{S \rightarrow Q}$) or due to interaction with other non-smoker agents (Q-agent: $\beta_{S,Q \rightarrow Q,Q}$ or N-agent: $\beta_{S,N \rightarrow Q,N}$). Like the other processes, Q-agents relapse into smoking spontaneously ($\delta_{Q \rightarrow S}$) or due to interaction ($\beta_{Q,S \rightarrow S,S}$) with an S-agent

networks can be used to develop intervention strategies when complete information on the underlying network topology of a local population is not available.

Methods

This section describes the model structure, data, and the networks used, along with the modelling choices involved in developing the ABM for smoking behaviour.

We highlighted the need for a tobacco control model to incorporate both spontaneous and interaction terms, as well as to consider appropriate network topologies. To achieve this, we use a synthetic population of n (please note the use of small case n , since N will be used to denote never-smokers) agents in an undirected and unweighted network G . To show the effects of the network topology, we make the agents interact on six different networks (described later in the section) and compare the observed dynamics.

Description of the agent-based model

In our model, each agent can be in one of the following smoking states: never-smoker (N), smoker (S) or quitter (Q). An agent is a never-smoker if they have never smoked before, while an agent who smokes any tobacco product daily, or occasionally, falls into the smoker state. Finally, if a smoker quits smoking even temporarily, they are labelled a quitter. We initiate the agents into each of the above states randomly. For a visual representation of the model, please refer to Fig. 1.

To make models more realistic over long time periods, epidemiological models usually include vital dynamics. Traditionally, this is done by including constant mortality and birth rates in the equations. In network-based models, a constant birth rate and mortality rate can lead to older agents having a higher number of social contacts, thus increasing their influence on other agents. However, in reality, the number of social connections

does not increase with age but instead peaks in the mid-twenties and then decreases with age (Bhattacharya et al. 2016). However, tackling these problems would significantly increase the complexity of the model since the model will have to include age-dependent mortality and birth rates, and the network generation process will have to be adapted to ensure that the network retains its properties while adding and removing agents. Hence, we do not include vital dynamics in the model and run our experiments in time frames where the vital dynamics can be ignored. We calibrate and validate our model in time periods of less than 30 years to minimise the effect of vital dynamics.

State change processes

We incorporate three main processes into the model: smoking initiation, smoking cessation and relapse into smoking. The three state change processes involve interaction-based state changes as well as spontaneous state transitions. We assume that each exposure to an agent with a different smoking status is independent of the previous exposure. We then use a binomial approximation to compound the effect of multiple independent interactions on one agent simultaneously to calculate the probability of state change. In Additional file 1: section S1 Appendix, we provide a detailed description and derivation of the expressions used.

Smoking initiation

First and foremost, we define the transition of a never-smoker into a smoker as smoking initiation. In the model, an N-agent can initiate smoking in two ways. First, through a random probability $\delta_{N \rightarrow S}$ depicting various external influences like advertisements, movies, and the presence of tobacco shops influencing an N-agent to pick up smoking. Second, through interactions with other S-agents in its network neighbourhood with a probability $\beta_{N,S \rightarrow S,S}$. We use the binomial approximation mentioned before to calculate the expression in (1), which gives the probability of smoking initiation due to interaction.

$$P_{N,S \rightarrow S,S} = \frac{n_S}{n} (1 - (1 - \beta_{N,S \rightarrow S,S})^{n_S}) \quad (1)$$

Smoking cessation

Following smoking initiation, we define the process of an S-agent quitting smoking as smoking cessation. Similar to smoking initiation, smoking cessation can also happen in two ways. First, due to various external influences like mass-media campaigns, mandatory warning labels on cigarette boxes, and higher taxes. This external influence is incorporated into the model through the spontaneous term $\delta_{S \rightarrow Q}$. Second, due to interactions with non-smokers. However, both Q-agents as well as N-agents fall under the non-smoker category. Therefore unlike the other state change processes, interactions with both the other states can lead to smoking cessation. $\beta_{S,N \rightarrow Q,N}$ represents the probability of cessation of an S-agent due to N-agents in its network neighbourhood. At the same time, the probability of cessation due to other Q-agents in its network neighbourhood is given by $\beta_{S,Q \rightarrow Q,Q}$. Subsequently, the probability of smoking cessation due to interaction of an S-agent with multiple Q-agents is given by (2) while (3) gives the same probability but due to interaction with multiple N-agents.

$$P_{S,Q \rightarrow Q,Q} = \frac{n_Q}{n} (1 - (1 - \beta_{S,Q \rightarrow Q,Q})^{n_Q}) \quad (2)$$

$$P_{S,N \rightarrow Q,N} = \frac{n_N}{n} (1 - (1 - \beta_{S,N \rightarrow Q,N})^{n_N}) \quad (3)$$

Smoking relapse

Finally, we define picking up smoking after a period of abstinence as a smoking relapse. Smoking relapse is similar to smoking initiation, except that the Q-agent gets influenced instead of an N-agent. Similar to the other two cases, smoking relapse can happen in two ways. $\delta_{Q \rightarrow S}$ represents the probability of a Q-agent relapse into smoking due to external influence. Additionally, $\beta_{Q,S \rightarrow S,S}$ represents the probability of Q-agent relapsing into smoking due to interaction with its immediate network-neighbour S-agents. Here, like the smoking initiation, only interactions with other S-agents can cause a Q-agent to relapse into smoking. The probability of Q-agent relapsing due to its interaction with multiple S-agents is shown in (4).

$$P_{Q,S \rightarrow S,S} = \frac{n_S}{n} (1 - (1 - \beta_{Q,S \rightarrow S,S})^{n_S}) \quad (4)$$

Experiment settings

We run simulations with a total population of $n = 1000$ agents. These agents are connected with each other based on the pre-defined network structure. To understand how this pre-defined network structure affects smoking behaviour, we vary the network structure and study the smoking dynamics observed in each network. This process involves comparing the model on each network with the empirically observed data. Along with this comparison, we also identify the combination of parameter values that best fit observed data and how this combination changes with the underlying network. In addition, we also compare the results of the ABM on different networks with the ODE analogue (described in Additional file 1: section S2 Appendix) of the ABM.

In the ABM, each time step corresponds to a year in the real world. At every timestep, agents follow a three-step procedure sequentially to avoid cascading agent states. First, each agent identifies a potential new state it can transition to based on the agent's state at that time step. Next, each agent calculates the probability of transitioning into the identified new state. Finally, each agent simultaneously transitions into the new state based on the calculated probability. The sequential procedure above removes the chance of cascading agent states in a single step. That is, an N-agent can never change into an S-agent and then a Q-agent in the same time step.

Due to the stochastic nature of ABMs, we iterate each simulation multiple times. We iterate each simulation ten times during the parameter sweeps for calibrating the model due to limits on computational resources. However, we iterate the best-fit combination of parameters 1000 times to validate the model, thus including possibilities of rare events.

For every simulation, a new network is generated, that is, each simulation has a different realisation of the network structure. We then initiate $s_0\%$ and $q_0\%$ (based on the

first data-point in the empirical data) of the total 1000 agents randomly as S-agents and Q-agents, respectively, on the generated network.

The model was built using the modular framework Mesa in Python (Kazil et al. 2020).

Networks

We have run simulations of our ABM on six different network topologies: fully-connected, scale-free (Barabási and Albert 1999), random (Erdős et al. 1960), small world (Watts and Strogatz 1998), Lancichinetti–Fortunato–Radicchi benchmark (Lancichinetti et al. 2008) and a real-world network from the Framingham heart study (FHS) data (Hill et al. 2010a). These networks were chosen due to their unique properties or ubiquitous nature in literature. The scale-free network, random network and small-world networks are standard network topologies in network sciences that are used for spreading phenomenons; hence, we test our model on these networks too. Details of each of the networks used are mentioned below.

1. Fully-connected: The fully-connected network (FC) assumes that every agent is connected to every other agent in the system. This is to replicate the mean-field or perfect-mixing approximation seen in ODE models. However, real-world networks are sparse and seldom fully-connected. Therefore, we generate and explore other network topologies.
2. Random network: Similar to the FC network, every node in a random network tries to form an edge with every other node, but with a probability p_{er} . The situation when $p_{er} = 1$ corresponds to a FC network. We use the Erdős–Rényi (ER) model (Erdős et al. 1960) to generate random networks for our experiments.
3. Scale-free network: These are networks where the degree distribution follows a power law. Many real-world networks have been reported to follow the power-law distribution (Gamermann et al. 2019; Albert et al. 1999). To model this we use the Barabási–Albert (BA) network model (Barabási and Albert 1999).
4. Small-world network (SW): This is a network which is highly clustered with small average shortest paths. These networks are known for local cliques and random long-ranged connections. We use the Watts–Strogatz model to generate the network (Watts and Strogatz 1998).
5. Lancichinetti–Fortunato–Radicchi (LFR) benchmark network: The LFR network encompasses properties of a real-world network like a heterogeneous distribution of degrees and size of communities (Lancichinetti et al. 2008). We use an LFR network due to its unique property of communities embedded into it during the network generations process.
6. Framingham Heart Study (FHS) network: We have also used a real-world network based on the Framingham heart study (FHS) (Dawber 2013) data along with the synthetic networks above. The FHS was a a longitudinal cohort-based study aimed explicitly at studying cardiovascular diseases and identifying the associated factors. However, due to the wide range of documented associated factors, it has become a one-of-a-kind data set on which even detailed network analysis has been carried out. We used a configuration model to generate synthetic networks with the same degree distribution observed in the FHS (Hill et al. 2010a). Unlike the SF and LFR networks, the configuration model allows an arbitrary distribution of degrees and is

therefore not restricted to the power-law distribution (Newman 2003). We estimate two parameters for our model based on the results from the FHS data-related network analysis (Christakis and Fowler 2008).

Each of the networks mentioned above has multiple parameters associated with its generation. We chose these parameters involved in the network generation process such that their average degree is close to the empirically observed one. Another study which does an extensive network analysis using the FHS and tries to model a similar social contagion points out the range of observed average degrees over multiple studies (Hill et al. 2010a). The indicated values of average degree over different years were in the range between 2.8 and 5.3.

Other studies on the composition of social networks in other countries have observed an average degree of 4.45 (Chao et al. 2015). However, the network involved in the spread of smoking behaviour mainly consists of close family members and friends (Christakis and Fowler 2008). Therefore, the number of individuals potentially influencing such behaviour will be less than the average degree of a standard social network. For this reason, we choose the average degree from exam 6 of FHS, $\langle k \rangle = 3$ for the generation of networks in our model. When we are not able to use the value of $\langle k \rangle = 3$ due to algorithmic restrictions (like in the case of small-world network) we use the value of $\langle k \rangle = 4$.

The parameter values for each network and the average expected degree are shown in Table 1 in Additional file 1: S4 Appendix.

Data

To calibrate and validate the model, we use publicly available smoker and quitter prevalence data from the US and UK. The UK data-set (Office for National Statistics 2019) has normalised smoker population and quit ratio (defined as the proportion of smokers who have quit smoking) from 1974 to 2019. Bi-yearly data points are available until 2000, and yearly data from then on. We estimate the quitter population from the quit ratio and use it to calibrate the model.

In the case of the US, we used the data available in the official Surgeon General's report on tobacco (Office of United States Public Health Service and others 2020). This document reports the prevalence of smokers (male and female separately) and quitters (again, male and female separately) between the years 1965 and 2015 (data points every five years).

Additionally, we impose the values of two parameters ($\beta_{N,S \rightarrow S,S}$ and $\beta_{S,Q \rightarrow Q,Q}$) of the model by estimating them from empirical research (Christakis and Fowler 2008). The Additional file 1: section S3 Appendix describes the steps taken to estimate the parameters.

Calibration and validation

To calibrate and validate the model, we use a four-step process. To limit the effects of not including vital dynamics, we calibrate and validate the model on time frames of 25–30 years. Then, we split the time-stamped data into calibration and validation segments in

UK and US scenarios. We use data from 1974 to 2002 (16 data points) in the calibration segment and the remaining in the validation segment in the UK data. Similarly, in the US, we use the data from 1965 to 1990 to calibrate the model and the remaining to validate it.

Step 1: coarse-grained calibration

To identify the combinations of parameter values which best mimic the empirical prevalence data, we run a coarse-grained parameter sweep on all the uncertain parameters of the ABM. In this case, this parameter sweep was carried out for each parameter for ten logarithmically split values between 0 and 1. Further, we iterated each parameter combination ten times to reduce the effects of randomness. We then use these simulation results to identify the range of parameters best fitting the calibration data (top 100 best-fit parameter combinations). Since we have the population sizes time-series data, we use the sum of the Mean Square Error (MSE) of both the S and Q trends to identify the best fitting parameters.

Step 2: sensitivity analysis

After identifying the set of parameters that minimise the MSE, we perform a sensitivity analysis on these values to determine the relative influence of each model parameter on the smoking dynamics. This involves individually varying each parameter across its range while holding other parameters constant and evaluating their impacts on the simulated smoker and quitter prevalence curves over time. Parameters that do not significantly alter the dynamics are identified as potential candidates for removal, pending further testing. The sensitivity analysis aids in interpreting the effects of each parameter and determining options for simplifying the model were supported by the numerical experiments. Details on the sensitivity analysis results, including plots of varying each parameter and discussion of their relative impacts on prevalence, are available in Additional file 1: Section S7 Appendix.

Step 3: fine-grained calibration

To improve the estimated parameters, we re-calibrate the model through a finer grained parameter sweep on each parameter. In this case, this parameter sweep was carried out for each parameter for five equally split values between the range of values identified in step 1. Just as in step 1, we iterate each of the simulations ten times. We use the range of values identified for each parameter from the coarse-grained calibration to run the fine-grained parameter sweep.

Step 4: validation

We validate the calibrated model by comparing the simulated results with the validation data for both US and the UK. Since we used the sum of MSE of the smokers and quitters for calibrating the model, we also use the same sum of MSEs for validation. Along with the MSEs, we also use the unique crossover point of the historical trends of the smoker and quitters populations to improve the validation process.

Results

In this section, we study the ABM for smoking and show its characteristics. First, we compare the ABM with all possible interactions on different networks. Through this comparison, we demonstrate the importance of networks for modelling the spread of smoking behaviour. We further demonstrate the ease with which networks can be

incorporated into ABMs and therefore advocate using them for modelling such a spreading phenomenon. Next, we compare the ABM with an ODE analogue to demonstrate the equivalence of the ABM on an FC network and a traditional ODE model. Finally, we calibrate and validate the model on empirical data observed in the US and the UK. Through the calibration and validation process, we emphasise the need to incorporate networks into such models, which can potentially be used to develop policies. Next, we show that the real-world network (FHS) replicates the empirical data observed in the US and the UK. In addition, we show that in practical situations, when complete information on the actual underlying network is not available, synthetic networks with similar average degrees can be used to develop models. On the other hand, we show that ABMs on FC networks and, therefore, ODE models should not be used for modelling smoking and similar behavioural contagion.

We examine the evolution of total prevalence of smokers (S) and quitters (Q) for each simulation setup. To compare and quantify the temporal dynamics of the populations, we calculate the sum of MSE of the S and Q curves. Since the S and Q curves cross each other in both the UK and US data, we study the unique crossover time-point for the ABM and compare it to the one from empirical data.

Population dynamics in ABMs on networks and its ODE analogue

Figure 2 shows the population dynamics observed from simulations of the ABM for smoking (on all six network topologies) and an ODE analogue (described in Additional file 1: section S2 Appendix) with the same parameter values. We run the ABM and ODE models to simulate a period of 30 years. This time frame is similar to what we used to calibrate the models on US and UK data. We use this observed population dynamics to compare the ABM between different networks and the ABM on these networks with the ODE model.

Our results suggest that the network structure affects the population dynamics of smokers and quitters for the same experimental conditions. Specifically, when the network structure is changed from FC to any other network, the dynamics observed change drastically. In addition, when the average degree of the networks (other than in the FC network) is kept at similar values (close to $\langle K \rangle = 3$), we observe that the deviation in

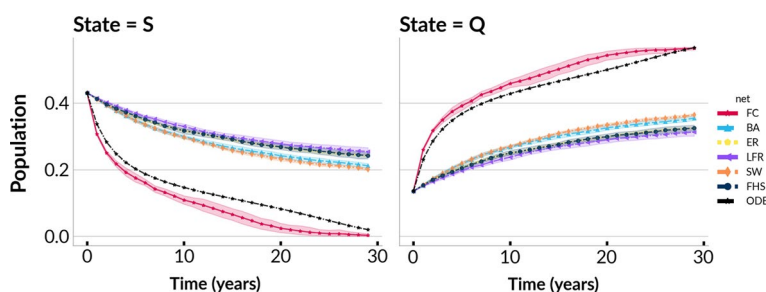


Fig. 2 Dynamics of the S and Q populations from the ODE model and ABM on networks (FC, BA, ER, FHS, LFR and SW networks). We used the best-fit parameters from the coarse-grained parameter sweep of the FHS network ($\beta_{Q \rightarrow S, S}, \beta_{S \rightarrow Q, N}, \delta_{N \rightarrow S}, \delta_{S \rightarrow Q}, \delta_{Q \rightarrow S} = 0.01334, 0.05623, 1e-05, 4e-05, 1e-05$) for these simulations. The black dotted lines in both the plots represent the ODE model results for the respective population for the same parameters mentioned above. The ODE model and ABM on FC show similar temporal dynamics, while the dynamics change drastically when the network structure moves away from FC

the observed population dynamics is minor between different networks. This deviation being minor suggests that the average degree is vital in the dynamics of smoking behaviour.

The ABM on a fully-connected network follows the same qualitative trajectory as the ODE model. Statistical equivalence of a differential equation model, which uses a mean-field approximation to an ABM on a fully-connected network, has been shown before, and our model is consistent with this result (Rahmandad and Sterman 2008). This suggests that a simple traditional ODE model can be used to model smoking behaviour and other similar social contagions only when the local population under study follows a fully-connected network topology. However, real-world networks are sparse and not fully-connected. So ABM on FC, and by extension ODE models, should not be used to model smoking behaviour and other similar behavioural contagion processes.

Validation

When the parameters that best captured the historical trend for the ABM on each network were observed, the variation in the MSE values was small. Therefore to make the parameter selection process more robust, we chose 100 combinations that gave the minimum MSE values instead of choosing only the minimum MSE parameter combination. Additionally, we imposed a condition that each independent parameter in this combination had to fall between the first and third quartile of its values observed in the minimum 100 (the values which fall inside the box in Fig. 5). We then sample this new set of parameters 1000 times to validate the model.

To validate the calibrated model, we compare the evolution of simulated population sizes with the historical data. Since we calibrate the model using the MSE values of S and Q curves, we also use the same for validation. For validation, we use periods of the empirical data, which were not used to calibrate the model (validation data, periods: 2003–2019 for the UK and 1995–2015 for the USA).

Case 1: UK

Figure 3 shows the distribution of MSE values, crossover points and the population dynamics of 1000 iterations of the simulations using the best-fit parameters for each network in the UK. To compare each of these MSE distributions, independent two-sample t-tests were carried out. Specifically, we carried out t-tests between the MSE distributions for every pair of networks. The results demonstrate that the fully-connected (FC) network is significantly different from all other networks, with p values of 0 versus the alternatives. This aligns with the poor performance of FC observed during validation. On the other hand, the MSE values for the empirical FHS network and ER network are the most similar, with a p value of 0.049. The BA network showed no significant difference compared to LFR and SW networks, with p values of 0.3 and 0.5 respectively. However, all other pairs of networks exhibit highly significant differences in their MSE distributions, with p values of 0.

The distributions of crossover points in Figs. 3 and 4 (row 2) show the time steps at which the simulated smoker and quitter curves intersect over the 1000 runs (if there is a crossover) for each network topology. The width of these distributions illustrates the variability in when the crossover occurs between different iterations of the model

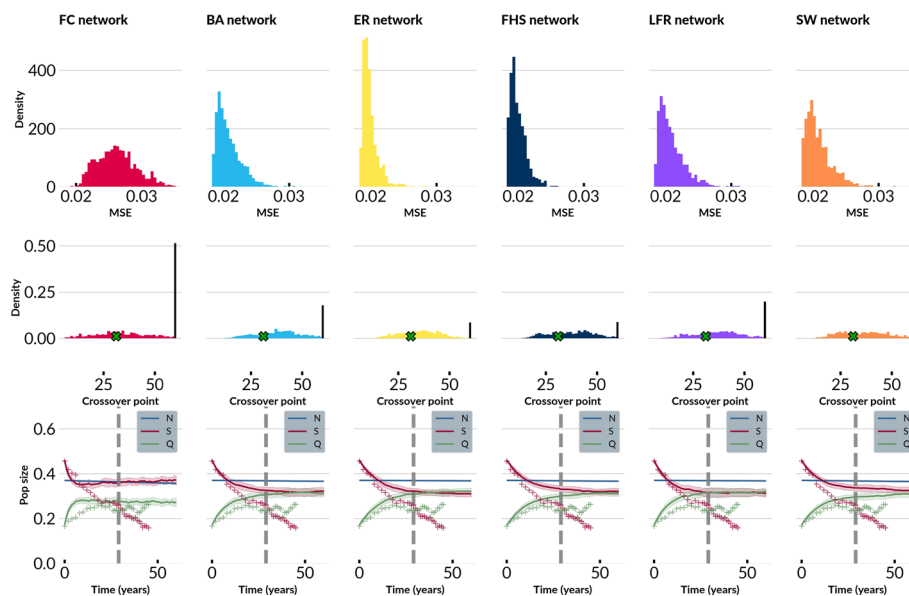


Fig. 3 The figure shows simulated characteristics and population plots from 1000 runs of the ABM for the best-fit parameters on each network for the UK. The bars in the first row represent the MSE value (sum of S and Q) of the ABM with the validation data, the bars in the second row show the crossover point between the smoker and quitter populations, and the third row shows the mean population value with a 95% confidence interval (CI) around the curve, obtained from 1000 simulation runs. The green crosses on the second row on the x-axis represent the actual crossover point in the empirical data. The black bars show the number of times the S and Q curves do not cross each other. The third row shows the mean simulated population curves over time for smokers (S) and quitters (Q) from 1000 runs, with the shaded areas indicating the 95% CI. The lines with the marker + indicate the actual historical prevalence data. The grey dotted line, dividing the plot indicates the time-step till which the model was calibrated. Among the six networks, we see that ER (MSE mean = 0.01984, SD = 0.00069) and the FHS (MSE mean = 0.01998, SD = 0.00071) network reproduces the data most accurately. The BA (MSE mean = 0.02076, SD = 0.00104), SW (MSE mean = 0.02081, SD = 0.00135) and LFR (MSE mean = 0.02082, SD = 0.00114) are very similar to each other in terms how good they replicate the data. While the ABM on FC network (MSE mean = 0.02617, SD = 0.00226) provides the worst fit for the validation data

parameters. The range of crossover points observed in the distribution always contains the actual crossover point observed in the historic trend. However, the number of times the crossover happens are very different when the underlying network is changed. The S and Q curves cross over 485, 822, 915, 912, 802, 836 times for FC, BA, ER, FHS, LFR and SW respectively. The FC network both gives a higher variability in the crossover point distributions and a lower number of successful crossovers, while the ER gives the highest number of successful crossovers. The third row in Fig. 3 shows the mean simulated population curves over time for smokers and quitters, along with the 95% confidence interval (CI) bounds. The CI illustrates the range of dynamics observed across the 1000 iterations of the best-fit model parameters for each network. Additionally, the lines with the '+' marker depict the actual historical smoking prevalence data from the UK for comparison to the model results. From the difference in MSE values and the variability in the crossover-points we can conclude that the network structure does matter when modelling the smoking behaviour. The real-world FHS network and the ER network replicate the historic trend observed very well and also has the highest number of successful crossovers.

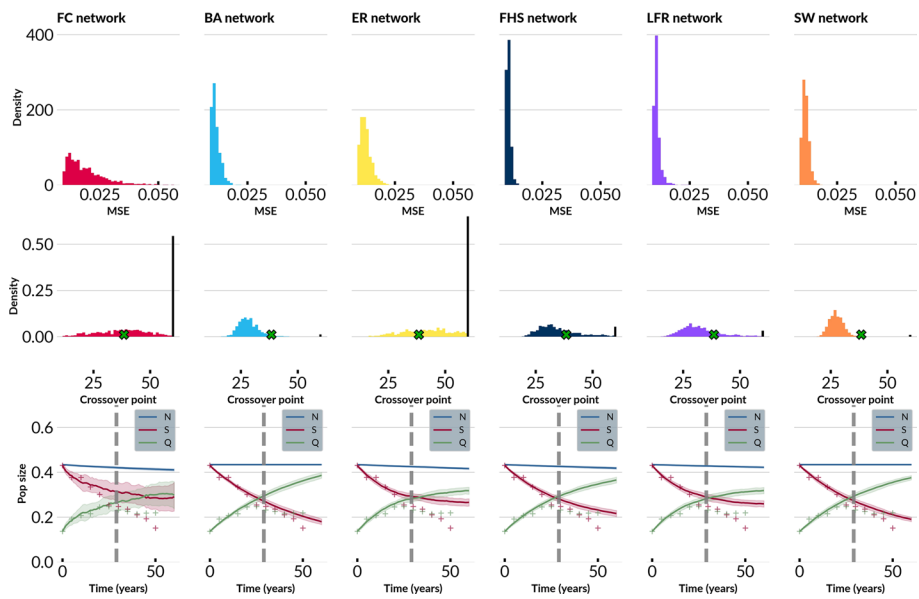


Fig. 4 The figure shows simulated characteristics and population plots from 1000 runs of the ABM for the best-fit parameters on each network for the US. The bars in the first row represent the MSE value (sum of S and Q) of ABM with the validation data, the bars in the second row show the crossover point, and the third row shows the mean population plot with a 95% CI. The green crosses on the second row on the x-axis represent the actual crossover point in the empirical data. The black bars show the number of times the S and Q curves did not cross each other. The third row shows the mean simulated population curves over time for smokers (S) and quitters (Q) from 1000 runs, with the shaded areas indicating the 95% CI. The lines with the marker + indicate the actual historical prevalence data. The grey dotted line, dividing the plot indicates the time-step till which the model was calibrated. Amongst the six networks, we see that the ABM on the FHS network reproduces the data most accurately (MSE mean = 0.01048, SD = 0.00053). The LFR (MSE mean = 0.01102, SD = 0.00091), BA (MSE mean = 0.0114, SD = 0.0013) and SW (MSE mean = 0.01151, SD = 0.0007) are again very similar to each other in terms how good it replicates the data. The ER network (MSE mean = 0.0124, SD = 0.00175) closely follows all the networks except FC. While the ABM on FC network provides the worst fit for the validation data FC (MSE mean = 0.01878, SD = 0.00622)

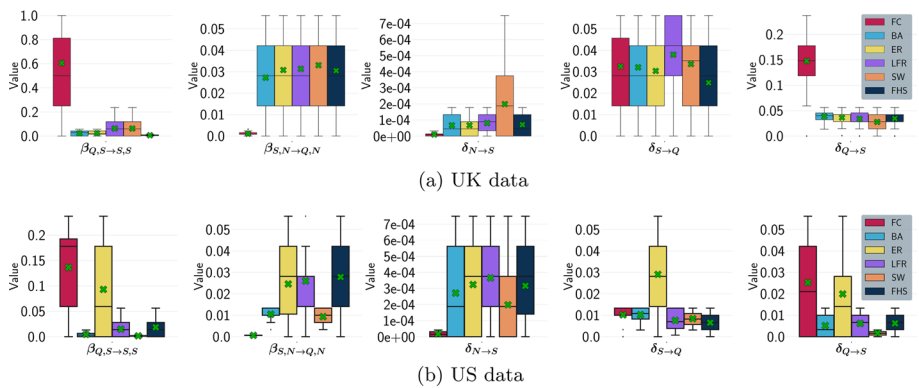


Fig. 5 Box plots representing the range of values for each parameter in the 100 simulations that best fit the calibration data for each network. The green crosses in each box show the mean value of the parameters. From left to right, we have the FC (red), BA (light blue), ER (yellow), FHS (dark blue), LFR (purple) and SW (orange) networks for each plot

For the spread of smoking behaviour, the influence of other individuals is only substantial when they are a close family member or a close friend (Christakis and Fowler 2008). This limits the average degree of the required network for the spread of smoking

behaviour. This network will have a much lower average degree than a standard social network. In such situations, the degree distribution of the ER network can approximate real-world ones, such as that of the FHS network. This can be seen in our results, the ER and FHS network best replicate the validation data out of the six networks. The LFR and SW networks also generate low MSEs. This trend is seen in the number of successful crossovers as well. Following the MSE distributions, the FHS and ER also gives very high successful crossovers, while BA, LFR and SW give similar values of successful crossovers. Additionally, the ABM on FC networks gives the worst fit to the validation data and the least number of successful crossovers showing that the ABM on FC networks and, by extension, ODE models should not be used to model smoking and similar behavioural contagion. We can conclude that ABM on the ER network can be potentially used to model smoking behaviour when information on real-world network is not available.

Additional file 1: section S6 Appendix gives a detailed analysis of the networks used for the simulations.

Case 2: US

Figure 4 shows the distribution of MSE values, crossover points and the population dynamics for the ABM over 1000 iterations using the best-fit parameters calibrated with the US data. As in the case of the UK, we also carried out independent two-sample t-tests for MSE distributions between each network in the US. However, there was a significant difference between the ABM on FHS and all other networks.

Even though the distribution of MSE values were significantly different, the average MSE values of the ABM on each of the networks were lower than that of the ABM on FC and very close to each other. The third row in Fig. 4 shows the mean simulated population curves over time for smokers and quitters, along with the 95% confidence interval (CI) bounds. We find that the ABM calibrated on the FHS network provides the overall best fit to the empirical smoking data for both the US. However, the models using the LFR, BA, ER, and SW networks also match the historic trends. We find that the FHS network fits the real-world data the best in the case of the US, followed by the LFR network. As opposed to the others, the LFR network has a unique property of community structure embedded in the network generation process. This suggests that communities play a role in the spread of smoking behaviour.

Analysis of best-fit parameters

Figure 5 shows a box plot of values seen in the 100 best-fit parameters for each network on both US and UK data. When the 100 best parameter combinations that fit the data best were compared, the FC network consistently gave significantly different parameter combinations in both US and UK data sets. In both US and UK, at the 5% significance level, the values of at least 80% of the parameter values found from calibration on the FC network were significantly different from the other networks. Moreover, 100% of them were significantly different from the FHS network. On the other hand, all other networks return parameter values in which at least 20% of them are not significantly different from

that of the FHS. The only exception is the SW network in the case of the US, where all parameters values returned were significantly different.

The ER and BA networks perform well in the case of the UK (almost 60% of the parameters are not significantly different from the FHS network), but in the case of the US, similarity of the parameters drops (only 40% are not significantly different). However, the LFR network performs decently in the UK data set (60% are not significantly different) and very well in the US case (none of the parameters is significantly different).

Our results thus indicate that when the average degree is kept constant, the parameter values found for the FHS network are somewhat similar to that of the LFR, ER and BA networks.

By comparing the ABM calibrated on different networks between the US and the UK, we see that in the LFR, SW, and FHS networks, 20% of the parameters were not significantly different between the US and UK. While in the ER network, 40% of the parameters were not significantly different. However, moving from the UK to the US in the BA and FC network, all the parameter values found from the calibration process significantly differed. This shows that the ABM on FHS, ER, SW and LFR networks are robust despite geographic variations. However, this is not true for the FC and BA networks.

Implications to policy

When the underlying network structure is changed, the parameters best replicating the empirical data also change. Therefore, when models are used to develop policies, parameter estimation becomes very important to predict the outcome of potential new policies. If the wrong network structure is used for the model, the calibrated parameters will also be different, which would lead to inaccurate strategies being developed. However, our results (robustness of the FHS, ER, SW and LFR networks to changes in geographic regions and the similarity of parameters of LFR, ER and BA within each region) suggest that when the real-world network structure is not available, the LFR and ER networks provide a satisfactory approximation. Thus, LFR or ER networks could potentially be used to develop strategies for controlling smoking behaviour when the local population's underlying network structure is unavailable.

Discussions

Using our ABM, we study the effect of network topology on the dynamics of the spread of smoking and see that the network structure affects it. The effect of network topology can be clearly seen when comparing results from an FC network to those obtained on other networks. However, the difference is minor within other networks when the average degree is similar. This suggests that, in addition to the network structure, the average degree of the underlying network might also play a role in the spread of smoking behaviour.

Our UK and US results suggest that ER and LFR networks replicate the empirical data better than the other synthetic networks (that is, excluding the FHS network). Apart from this, by analysing the parameter values found during calibration, we observe that only ER and LFR networks are robust to changes in the geographic region and also return a combination of parameters of which at least 40% are not significantly different from those obtained by calibrating the model on the FHS network. Upon closer

observation of the network characteristics of LFR, ER and FHS networks (Fig. 7 in Additional file 1), we see that all three of the networks have very similar average degrees and thus also form a similar number of edges. At the same time, the BA and SW have the same average degree as each other, which is a bit higher than the other networks. This suggests that the average degree might play a crucial role in the dynamics of smoking behaviour and possibly, any synthetic network with an average degree similar to that of the real-world network can be used to model smoking behaviour.

We have showed the ABM on networks can be used to reliably model social contagion trends in tobacco use dynamics. Even with the wrong network, these models can predict the trajectory of the populations qualitatively, albeit with a lower accuracy. Nevertheless, when a network changes, the parameter values returned on calibration also change. The differences in the model and its best-fit parameters become important when policymakers use the predictions to develop strategies to curb smoking. Incorrect approximation of network structure and thus the model parameters can potentially lead to the development of ineffective policies.

However, our results suggest that in cases where the real-world network information is not entirely available due to practical constraints, policymakers can use LFR networks and ER networks to approximate the real-world network topology, as our social contagion process on these networks replicates the empirical data with good accuracy and, during calibration, returns parameter values which are not significantly different from the real-world network. This potentially opens the way to population-wide tobacco control interventions that exploit social contagion and will require minimal knowledge of real-world parameters, such as average number of close contacts.

ABM is a good technique for incorporating network structure and interactions between individuals to study the macroscopic outcome. Our model shows that the network structure of the population is essential while modelling smoking, which should be taken into consideration while developing policies. However, some limitations due to modelling choices to preserve the model's simplicity should be noted.

First, we have assumed that every individual behaves in the same way within a group. However, this is not the case in a real-world setting. Additionally, many social ties manifest asymmetrically, whereas we only consider undirected networks. Incorporating directionality could improve accuracy. Exploring simplicial complex frameworks constitutes a promising avenue for enhancement. Such higher-order representations can capture empirical group social contagion effects (Iacopini et al. 2019).

Second, we have not considered vital dynamics in the model and the age-dependent nature of reaction to influence. Usually, a constant mortality rate and birth rate are incorporated to model vital dynamics. However, in network models, there is a risk of older agents gaining more centrality just because of the implementation of network growth. A careful understanding of age-dependent mortality rates and social ties should be incorporated into the model to circumvent this problem. However, this is beyond the scope of this paper and will be explored in future work. Therefore, to limit the effects of not including vital dynamics into the model, we calibrate and validate the model in time-periods of 25–30 years.

Third, the degree of influence on smoking behaviour changes with the social tie you have with the other person (Christakis and Fowler 2008). As a starting point in modelling influence on smoking behaviour, we have assumed that every kind of relationship affects the smoking behaviour similarly. Further, many social ties can be one-directional (as perceived by one individual in a social tie). However, as mentioned above, we have not considered any directed graphs in the model.

Since our model assumes that the total population is constant, $N = 1 - Q - S$, no new N-agents are being introduced to the system. Additionally, the S curve is calibrated against a decreasing trend observed in the empirical data, which explains the absence of variation never-smoker population.

Additionally, in a real world setting, smokers tend to be show interesting group behaviours, wherein they are part of smaller subgroups than non-smokers, and groups of smokers tend quit smoking together (Christakis and Fowler 2008). However, due to limitations on data, we have only studied the effect of random initiation of smokers in the model. We leave this task for the future with a hope for more detailed network data on smoking.

Conclusions

We have developed an agent-based model for smoking dynamics that considers the contagious nature of smoking behaviour by including network effects. This model can act as a test-bed for network based policies and strategies to control the spread of smoking behaviour.

Our results suggest that, when interactions between individuals are used to model population-level smoking dynamics, the underlying network of the local population becomes very important. By changing the network topology from a fully-connected network to other theoretical networks and, finally, a real-world network, we show that the dynamics deviate drastically from those of a traditional ODE model.

We show that our model is robust and consistent with the historical trends observed in two countries—US and UK. In both countries, the network based on a real-world local population best replicates the trends observed.

Moreover, our results suggest that the network topology, the average number of close social ties, and the presence of communities in the population improve the accuracy of the model.

Importantly, given the difficulties in collecting data on offline social networks, we find that the LFR and ER network replicates the empirical data with accuracy and also their calibrated parameter values are not significantly different from those of the FHS network, suggesting that the LFR and ER networks can be used for social contagion models of tobacco use.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1007/s41109-023-00580-5>.

Additional file 1. S1–S8 appendices.

Acknowledgements

This work has made use of the resources provided by the Edinburgh Compute and Data Facility (ECDF) (<http://www.ecdf.ed.ac.uk/>).

Author contributions

AP contributed by conceptualising the study, developing the methodology, running the experiments, collecting the results, writing the first draft, and editing the paper. VR contributed by conceptualising the study, developing the methodology, supervising the project, and reviewing and editing the draft. BG contributed by conceptualising of the study, developing the methodology, supervising the project, and reviewing and editing the draft.

Funding

This project has been funded by the Population Health Agent-based Simulation network (PHASE). PHASE is supported by the UK Prevention Research Partnership, an initiative funded by UK Research and Innovation Councils, the Department of Health and Social Care (England) and the UK devolved administrations, and leading health research charities. (Grant Reference MR/S037594/1) and their support is gratefully acknowledged.

Availability of data and materials

The model code and historical data used to validate the model is available at: https://github.com/adprabhak/ABM_net_smoking.

Declarations**Competing interests**

The authors have declared that no competing interests exist.

Received: 3 July 2023 Accepted: 6 August 2023

Published online: 23 August 2023

References

- Acharya A, Angus K, Asma S, Bettcher DW, Blackman K, Blecher E, Borland R, Ciecierski C, Cui M, Silva VL, et al (2016) The economics of tobacco and tobacco control. Technical report
- Albert R, Jeong H, Barabási A-L (1999) Diameter of the world-wide web. *Nature* 401(6749):130–131
- Aleta A, Martín-Corral D, Bakker MA, Piontti A, Ajelli M, Litvinova M, Chinazzi M, Dean NE, Halloran ME, Longini IM Jr et al (2022) Quantifying the importance and location of sars-cov-2 transmission events in large metropolitan areas. *Proc Natl Acad Sci* 119(26):2112182119
- Anderson RM, Anderson B, May RM (1992) *Infectious diseases of humans: dynamics and control*. Oxford University Press, Oxford
- Badham J, Kee F, Hunter RF (2021) Network structure influence on simulated network interventions for behaviour change. *Soc Netw* 64:55–62
- Banks E, Joshy G, Weber MF, Liu B, Grenfell R, Egger S, Paige E, Lopez AD, Sitas F, Beral V (2015) Tobacco smoking and all-cause mortality in a large Australian cohort study: findings from a mature epidemic with current low smoking prevalence. *BMC Med* 13(1):38. <https://doi.org/10.1186/s12916-015-0281-z>
- Barabási A-L, Albert R (1999) Emergence of scaling in random networks. *Science* 286(5439):509–512
- Bhattacharya K, Ghosh A, Monsivais D, Dunbar RI, Kaski K (2016) Sex differences in social focus across the life cycle in humans. *R Soc Open Sci* 3(4):160097
- Blok DJ, Vlas SJ, Empelen P, Lenthe FJ (2017) The role of smoking in social networks on smoking cessation and relapse among adults: a longitudinal study. *Prev Med* 99:105–110
- Brown T, Platt S, Amos A (2014) Equity impact of interventions and policies to reduce smoking in youth: systematic review. *Tob Control* 23(e2):98–105
- Caryl FM, Pearce J, Reid G, Mitchell R, Shortt NK (2021) Simulating the density reduction and equity impact of potential tobacco retail control policies. *Tob Control* 30(e2):138–143
- Castillo-Garsow C, Jordan-Salivia G, Rodriguez-Herrera A et al (1997) Mathematical models for the dynamics of tobacco use, recovery and relapse. Technical report series
- Chao D, Hashimoto H, Kondo N (2015) Dynamic impact of social stratification and social influence on smoking prevalence by gender: an agent-based model. *Soc Sci Med* 147:280–287
- Christakis NA, Fowler JH (2008) The collective dynamics of smoking in a large social network. *N Engl J Med* 358(21):2249–2258
- Colaioni F, Castellano C (2015) Interplay between media and social influence in the collective behavior of opinion dynamics. *Phys Rev E* 92(4):042815
- Dawber TR (2013) *The Framingham study*. In: *The Framingham study*. Harvard University Press
- Drope J, Schluger N, Cahn Z, Drope J, Hamill S, Islami F, Liber A, Nargis N, Stoklosa M (2018) *The Tobacco Atlas, Sixth Edition* Jeffrey Drope and Neil W. Schluger, Editors, pp 21–222526
- Ennett ST, Faris R, Hipp J, Foshee VA, Bauman KE, Hussong A, Cai L (2008) Peer smoking, other peer attributes, and adolescent cigarette smoking: a social network analysis. *Prev Sci* 9(2):88–98
- Erdős P, Rényi A et al (1960) On the evolution of random graphs. *Publ Math Inst Hung Acad Sci* 5(1):17–60
- Ferkol T, Schraufnagel D (2014) The global burden of respiratory disease. *Ann Am Thorac Soc* 11(3):404–406
- Gamerhann D, Triana-Dopico J, Jaime R (2019) A comprehensive statistical study of metabolic and protein-protein interaction network properties. *Physica A* 534:122204

- Gilpin EA, Pierce JP (1997) Trends in adolescent smoking initiation in the United States: Is tobacco marketing an influence? *Tob Control* 6(2):122–127
- Go M-H, Green HD Jr, Kennedy DP, Pollard M, Tucker JS (2010) Peer influence and selection effects on adolescent smoking. *Drug Alcohol Depend* 109(1–3):239–242
- Goodchild M, Nargis N, d'Espaignet ET (2018) Global economic cost of smoking-attributable diseases. *Tob Control* 27(1):58–64
- Hill AL, Rand DG, Nowak MA, Christakis NA (2010) Infectious disease modeling of social contagion in networks. *PLOS Comput Biol* 6(11):e1000968
- Hill AL, Rand DG, Nowak MA, Christakis NA (2010) Emotions as infectious diseases in a large social network: the SISa model. *Proc R Soc B Biol Sci* 277(1701):3827–3835
- Hodas NO, Lerman K (2014) The simple rules of social contagion. *Sci Rep* 4(1):1–7
- Holford TR, Meza R, Warner KE, Meernik C, Jeon J, Moolgavkar SH, Levy DT (2014) Tobacco control and the reduction in smoking-related premature deaths in the United States, 1964–2012. *JAMA* 311(2):164–171
- Hunter E, Mac Namee B, Kelleher J (2018) An open-data-driven agent-based model to simulate infectious disease outbreaks. *PLoS ONE* 13(12):0208775
- Iacopini I, Petri G, Barrat A, Latora V (2019) Simplicial models of social contagion. *Nat Commun* 10(1):2485
- Kazil J, Masad D, Crooks A (2020) Utilizing python for agent-based modeling: the mesa framework. In: Thomson R, Bisgin H, Dancy C, Hyder A, Hussain M (eds) *Social, cultural, and behavioral modeling*. Springer, Cham, pp 308–317
- Kermack WO, McKendrick AG (1927) A contribution to the mathematical theory of epidemics. *Proc R Soc Lond Ser A* 115(772):700–721 **(Containing papers of a mathematical and physical character)**
- Kermack WO, McKendrick AG (1932) Contributions to the mathematical theory of epidemics. II—the problem of endemicity. *Proc R Soc Lond Ser A* 138(834):55–83 **(Containing papers of a mathematical and physical character)**
- Kermack WO, McKendrick AG (1933) Contributions to the mathematical theory of epidemics. III—further studies of the problem of endemicity. *Proc R Soc Lond Ser A* 141(843):94–122 **(Containing papers of a mathematical and physical character)**
- Lancichinetti A, Fortunato S, Radicchi F (2008) Benchmark graphs for testing community detection algorithms. *Phys Rev E* 78(4):046110
- Lee S, Jung E, Castillo-Chavez C (2010) Optimal control intervention strategies in low-and high-risk problem drinking populations. *Socioecon Plan Sci* 44(4):258–265
- Lightwood J, Glantz SA (2016) Smoking behavior and healthcare expenditure in the United States, 1992–2009: panel data estimates. *PLoS Med* 13(5):1002020
- Mercken L, Snijders TA, Steglich C, Vries H (2009) Dynamics of adolescent friendship networks and smoking behavior: social network analyses in six European countries. *Soc Sci Med* 69(10):1506–1514
- Millett C, Lee JT, Gibbons DC, Glantz SA (2011) Increasing the age for the legal purchase of tobacco in England: impacts on socio-economic disparities in youth smoking. *Thorax* 66(10):862–865
- Moore C, Newman ME (2000) Epidemics and percolation in small-world networks. *Phys Rev E* 61(5):5678
- Mueller ST, Tan Y-YS (2018) Cognitive perspectives on opinion dynamics: the role of knowledge in consensus formation, opinion divergence, and group polarization. *J Comput Soc Sci* 1:15–48
- Newman ME (2003) The structure and function of complex networks. *SIAM Rev* 45(2):167–256
- Office of National Statistics (2019) Adult smoking habits in the UK: 2019. *Stat Bull* 1–15
- Office of United States Public Health Service and others (2020) Smoking cessation: a report of the surgeon general [internet]. Technical report
- Rahmandad H, Sterman J (2008) Heterogeneity and network structure in the dynamics of diffusion: comparing agent-based and differential equation models. *Manag Sci* 54(5):998–1014
- Rose SW, Myers AE, D'Angelo H, Ribisl KM (2013) Peer reviewed: retailer adherence to family smoking prevention and Tobacco Control Act, North Carolina, 2011. *Prev Chronic Dis* 10:E47
- Schaefer DR, Haas SA, Bishop NJ (2012) A dynamic model of US adolescents' smoking and friendship networks. *Am J Public Health* 102(6):12–18
- Schaefer DR, Adams J, Haas SA (2013) Social networks and smoking: exploring the effects of peer influence and smoker popularity through simulations. *Health Educ Behav* 40(1-suppl):24–32
- Schneider S, Gruber J, Yamamoto S, Weidmann C (2011) What happens after the implementation of electronic locking devices for adolescents at cigarette vending machines? a natural longitudinal experiment from 2005 to 2009 in Germany. *Nicotine Tob Res* 13(8):732–740
- Sharma S, Samanta G (2015) Analysis of a drinking epidemic model. *Int J Dyn Control* 3(3):288–305
- Sharomi O, Gumel AB (2008) Curtailing smoking dynamics: a mathematical modeling approach. *Appl Math Comput* 195(2):475–499
- Shin H (2022) Social contagion of academic behavior: comparing social networks of close friends and admired peers. *PLoS ONE* 17(3):0265385
- Thurner S, Klimek P, Hanel R (2020) A network-based explanation of why most covid-19 infection curves are linear. *Proc Natl Acad Sci* 117(37):22684–22689
- US Department of Health and Human Services and others (2014) The health consequences of smoking—50 years of progress: a report of the Surgeon General. US Department of Health and Human Services, Centers for Disease, Atlanta, GA
- US Department of Health and Human Services and others (2020) Smoking cessation: a report of the surgeon general. US Department of Health and Human Services, Atlanta
- Vallone DM, Allen JA, Xiao H (2009) Is socioeconomic status associated with awareness of and receptivity to the truth campaign? *Drug Alcohol Depend* 104:115–120
- van Meijgaard J, Fielding JE (2012) Estimating benefits of past, current, and future reductions in smoking rates using a comprehensive model with competing causes of death. *Prev Chronic Dis* 9:E122
- Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. *Nature* 393(6684):440–442
- White E, Comiskey C (2007) Heroin epidemics, treatment and ode modelling. *Math Biosci* 208(1):312–324

- World Health Organization and others (2004) Who framework convention on tobacco control. Technical report, WHO Regional Office for South-East Asia
- Zaman G (2011) Optimal campaign in the smoking dynamics. *Comput Math Methods Med* 2011:163834
- Zaman G (2011) Qualitative behavior of giving up smoking models. *Bull Malays Math Sci Soc* 34(2):403–415
- Zaman G, Kang YH, Jung IH (2017) Dynamics of a smoking model with smoking death rate. *Appl Math* 44:281–295
- Zhou B, Pei S, Muchnik L, Meng X, Xu X, Sela A, Havlin S, Stanley HE (2020) Realistic modelling of information spread using peer-to-peer diffusion patterns. *Nat Hum Behav* 4(11):1198–1207

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
