Simulation Results & Analysis

To understand whether the developed model provides useful results and meets the initial specification, two separate methods were used. Firstly, using the model to run simulations provides a number of pieces of information. In regards to the problem domain, finding out new information about the impact of social networking upon smoking is a key aim, but at the same time, these simulations should help to demonstrate whether the model is balanced, efficient and a good base for future work. Following this, analysis of the model from a qualitative viewpoint will form the second part of this section. As a part of the specification concerns itself with understanding whether the model is a suitable proof of concept for a commercial model, this is combined with general thoughts on the quality of the model and any future improvements. The basis of this analysis comes from both the design decisions and the results of the simulations carried out.

Simulation Analysis

Using the developed model, a wide range of possible situations can be simulated through using generated or sampled graphs along with configuring certain percentages of the population to smoke, give up or not smoke at all. Whilst running all of these simulations is beyond the scope of this project, ones from prominent categories were run instead. These categories were *simulation parameters*, *large networks*, *sampled networks*, and *composite networks*.

For all simulations, 30,000 simulation steps were carried out with graph and attribute files being exported every 1000 steps. A simulation step is a difficult concept to assign a unit of time to since it is discrete and involves making decisions, assessing influence and rearranging social ties all in one movement. In light of this, the simulation length, whilst not given a specific time period, should be considered to represent months and years rather than hours and days.

Simulation Parameters & Agent Attributes

Given the number of agent attributes and system parameters, there are many initial set-ups to try to model. Since a number of these do not provide useful information in relation to the model or the problem domain, they will not be covered. A high-level investigation into the percentage of smokers and of those giving up was conducted using a base graph as seen in **FIGURE X.** The aim of this test was to get a better understanding of how starting combinations of composition of the graph, relative to percentages of smokers and non-smokers, related to the number of non-smokers left at end of the simulation. A number of runs were executed with the results visible in **FIG Y.**

From **FIGURE Y2**, it can be seen that as the ratio of those giving up compared to those smoking (*num smoking/num giving up)* at the start of the simulation increases, the final percentage of those who do not smoke at the end becomes much more unpredictable. One possible explanation for this is that these people, since they have already given up once, will find it easier to relapse to smoking; if many people start to do this, the influence will then persuade others to do the same causing the final split of smokers and non-smokers to be largely dependent on how connections have formed over the earlier stages of the simulations. Continuing from this, **FIGURE Y3** shows the percentage of smokers in the network at the start against those who are not smoking at the end. For lower starting percentages of smoking such as 20%, the percentage not smoking at the end is around

70%, which is as expected from a stable simulation. When the percentage of smokers increases though, the range of ending percentages increases to the point where at 80% to start, the amount giving up at the end is between 30% and 75%. This points to the possible effect of a core of people giving up having an effect on the smoking population – at low percentages of smokers, those giving up are less likely to interact with smokers so have a lower chance of influencing them. For higher percentages, the people giving up are much more likely to be surrounded by smokers and as such be affected by the influence. As described above, the impact of this influence largely depends on how the network has changed since in the case where around 80% of the smokers start and the simulation ends at 75%, wide-spread cessation occurred but on another run with the same starting smokers but different amount giving up (22%), significantly fewer were non-smokers at the end.

FIG X, base graph

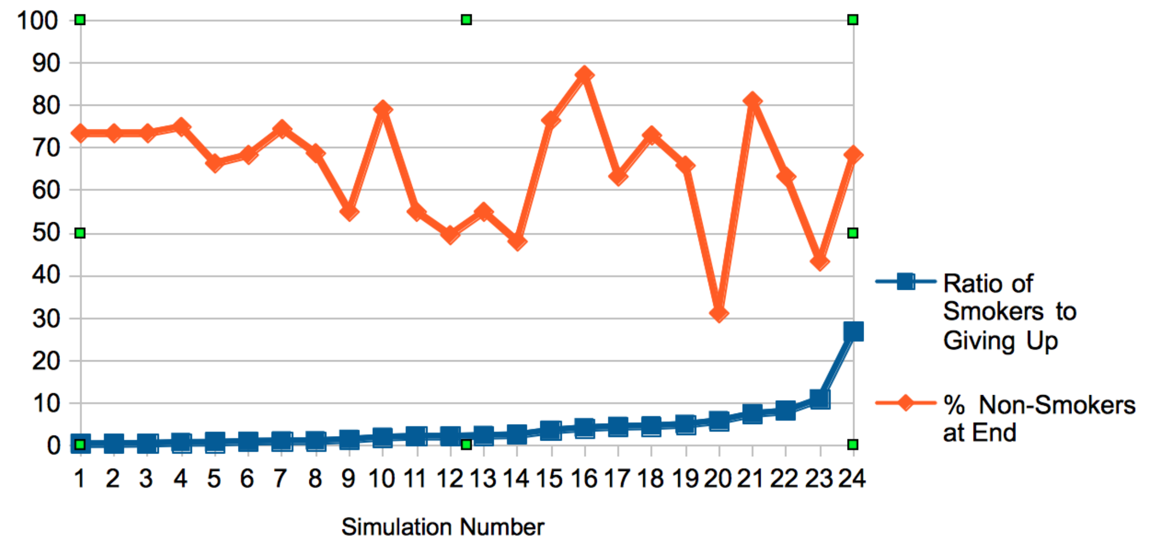

FIG Y Results table

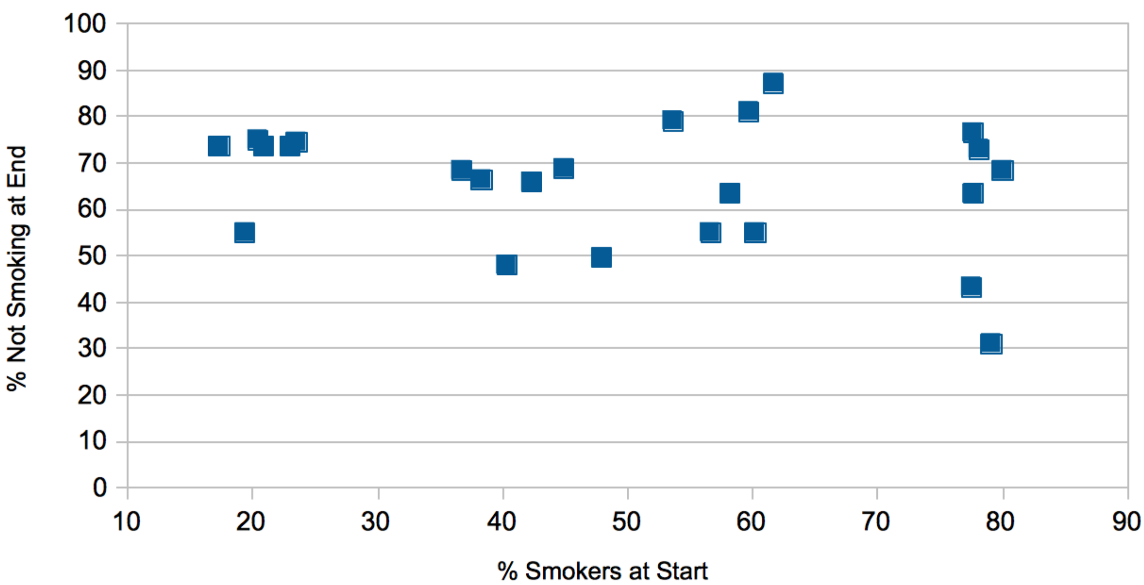

FIG Y2 – ratio against NS at end

FIG Y3 -

Whilst there is no clear conclusion from these tests, it does provide insight into the model itself. It appears that having a larger proportion of those giving up or not smoking as part of the network to begin with destabilises the results somewhat. It indicates that, due to the higher quantity of smokers surrounding those giving up or not smoking, relapses are more common. Additionally, a high percentage of smokers takes a longer time to persuade to give up via influence. As a result, it may be the case that agents attempt but fail to give up smoking regularly, never providing a constant source of influence to others but instead dissipating inconsistent influence to others. This inconsistency then can spread throughout the network to others who are trying to give up and given that there is a larger portion of the network able to do this when the majority of the graph smokers, this could explain the unpredictability in the above results. In regards to real-world behaviour, the main discovery is that trying to get networks saturated with smokers to give up is very difficult and simply introducing some non-smokers is not enough. Instead, it would seem that the optimal way for this to be tackled is that large chunks of the network would need to give up simultaneously, driving down the percentage of smokers overall.

Decision Tree Analysis

As the decision tree is the major factor in selecting the behaviours of agents, the number of 'hits' upon each decision reveals a lot about how individuals are guided to their final choice. To get a dataset for this, a *scale-free* graph with 483 nodes, 20% smokers and 10% giving up in order to be close to the UK smoking statistics **[REF],** was simulated over a period of 30000 steps with every decision made by an agent being recorded. The final percentages of smokers was around 9%, a moderate decrease in quantity but by no means significant due to the ratio of smokers to non-smokers. **FIGURE X** shows the decision tree labelled with both the percentage of total 'hits' and the percentage share of hits for the parent decision.

FIGURE X – labelled up decision tree

In regards to overall percentage share of hits, the values appear as expected; the 'non-smoking' arm of the tree encounters more hits than the smoking arm since a greater portion of the network does not smoke than does. Within the smoking arm, 70% of agents fall in the middle branch, i.e. the 'moderate' smoker, meaning that they do not have a high chance of giving up in that decision route. Furthermore, 20% of agent hits were to the 'heavy' category, where although they do have a direct chance of quitting it is unlikely due to the attributes used in that subtree. From this it may be deduced that to get a greater proportion of these smokers to have a chance to give up, they would need to be surrounded by people who smoke less than they do so that their influenced attribute for the quantity of cigarettes smoked would decrease. In turn, this would let them fall into the 'light' smoker category which has a higher chance of giving up. Another explanation for the underrepresented section of light smokers is that few hit this decision regularly because instead, they give up smoking. Interestingly, if the end-states of the light smoker sub-tree is looked at, both choices left a greater proportion of the agents going to the option which required a higher local percentage of non-smokers to give up, which implies that there is a core of light smokers who do not give up. To have the previously described scenario of lighter smokers encouraging heavier ones to reduce consumption, a greater percentage of lighter smokers would be required over the course of the simulation whilst avoiding skewing the tree such that it is not made too difficult to give up.

Of the medium and heavy smoker groups, looking at deeper levels of the tree points to which areas could be targeted to encourage cessation. For the former group, the majority of hits resulted in the agent reassessing their smoking consumption, with the majority of this sub-group opting to reduce consumption. Some, however, were given the chance to quit with the larger portion taking the marginally easier route of 70% of the neighbourhood needing to be not smoking. This matches with the above idea of getting smokers down to lower consumptions before giving up, which could again be improved by planting more, high influence, low consumption smokers in the area. For the 'heavy' group, there is a fairly even split across all sub-groups. It should be noted that there is an 'easy' chance for heavier smokers to give up if their health is lower than those they are related to, though this was not any more popular than any of the other decision routes. This indicates that to get this group to stand more chance of quitting, they would need to be surrounded by people who smoke either less than they do or not at all, and also are of better health. Doing this would guide them to either the light smoker options or the 'easy' option within the heavy smoker tree.

For non-smokers, it is most important to direct them away from end-states where there is a chance to either begin smoking or relapse whilst giving up. For group of agents giving up, the most common branch to follow was for those who have tried to give up between 1 and 5 times. As a side note, this is encouraging for the model balance as it shows that the majority of individuals do not fall into a cycle of giving up and relapsing, causing a very high number of give-up attempts. Within this group, the vast majority had more than 30% of their neighbours as non-smokers so had a lower chance of relapsing. Due to the fact that willpower was the next factor to decide upon, there was a mostly even split. This is the key part of the branch since it shows that by having a contingent of not only non-smokers, but also agents of lower willpower with high influence will lead to the individual finding it easier to avoid smoking again. The section of agents with more than 5 give up attempts were most likely to be surrounded by fewer others giving up – intended to emulate the effect of group support to those who have failed multiple times before - meaning that they had a higher chance of relapsing. In the simulation it can be seen that all of the agents had lower than the *influenced attribute* for health, so stood a much higher relapse chance. To avoid this situation, individuals who have failed multiple times before need to be surrounded by a high quantity of other giving up, ideally of similar health to their own to reduce their chance of relapse. The final section of the 'giving up' subtree is that of the people with only one giving up attempt. Although most agents were of willpower higher than the threshold, the split in health reveals that to increase their chances of not beginning smoking, others of similar or lower health need to be in their neighbourhood. This also holds true for those of lower willpower.

The final section of the tree is that of non-smokers who are not currently giving up. As implied by the final smoker/non-smoker percentage, the majority of the decision hits of the graph were in this area. On the next layer of the tree, a slight skew towards the agents being of higher than their *influenced health* and to maintain a higher chance of avoiding smoking being taken up is to ensure that the agent is surrounded by people of similar health. At the next decision down, willpower is used; a higher willpower here results in a more desirable chance of not relapsing. Since this is not affected by the willpower of those in the neighbourhood, the most important decision here is that of the health. To be more specific, those of low willpower should be in a neighbourhood of lower health but high influence agents, whereas higher willpower individuals can afford to be around those of similar or higher health.

In general, the decision tree reveals a number of pointers for how to drive the network in the direction of giving up smoking then avoiding relapsing. Having smokers be convinced to either smoke less per day or be surrounded by healthy, non-smokers helps in getting them to quit whilst positioning those who are giving up in the correct neighbourhood has a large effect on their chance of relapse, the exact composition of this neighbourhood depending on how many times the person has tried to quit before.

Sampled Networks

In order to consider how the model functions with real networks, a number of datasets were sampled. These are from the Stanford SNAP [**REF**] which provides a number of large datasets sampled from various internet sources. To maintain consistency with the system as a whole, only directed non-multi-graphs were used; undirected graphs would lack the required level relationship information whilst multi-graphs could have many of the same edge which is not catered for in this model. Generally, these datasets were very large and due the timescale of the project, the simulations could not take place on a *cluster* or *high-performance computer* (HPC) so instead, samples using the previously mentioned *snowball sampler* were taken of different sizes and had simulations run upon them. All of the simulations were run for 30,000 steps.

The first dataset was of an email network **[REF]**, with the starting and ending states shown in figures **X AND Y.** It was started with 50% of the network smoking, and of those not smoking, 50% giving up whereas by the end, 80% of the network was not smoking. Over the course of the simulation, the average degree of the graph increased slightly from 2.198 to 2.431 with the network diameter decreasing from 34 to 25 and the average clustering coefficient going from 0.02 to 0.009. Although the connectivity of the graph increased, with both degree increasing and diameter decreasing, the average path length increases from 8.998 to 9.372. Generally, this implies that the graph has become slightly more disconnected over the course of the simulation which is not necessarily desirable – higher connectivity allows greater spread of influence.

More interestingly, if the smokers and non-smokers are separated out into separate graphs, it can be seen that the remaining smokers at the end of the simulation are, mainly, highly connected. This explains the percentage of smokers remaining after such a large proportion of the graph not smoking as they are influencing each other in continuing to avoid giving up. Finding this kind of group is very important in the attempt to encourage cessation as without direct intervention, their connectivity means the influence sways towards keeping smoking. Having a cluster of non-smokers inside this group could break this influence and cause further cessation.

FIGURE X1 – smokers separated out

FIGURE X2 – nonsmokers separated out

Another graph to come here, simulation is taking a while though

<u>Model Analysis</u>

Commercial Analysis

DO I NEED TO TALK ABOUT SANDTABLE NOT GIVING MUCH HERE?

Commercial viability analysis of this project mainly focuses on whether the current system could be used as a starting point in building a more in-depth model so looks at aspects such as scalability, efficiency and extensibility. This analysis assumes that a commercial-level model should be:

- Efficient – to save excess computing cycles on large and long-running simulations
- Scalable – to be able to run both small and very large models, allowing for analysis on a wide range of social situations such as small friendship groups to populations of cities or towns.
- Extensible & adaptable – should new work influence the configuration of the model, it should not need entirely rebuilding in order to include these changes.

The efficiency of the model in places is good – the decision tree, for example, primarily uses *if* statements for decisions with only basic arithmetic comparisons or precomputed values rather than more complex values computer for each stage. Furthermore, the neighbourhood and influenced attributes are only calculated once, since these are both intensive operations and in a similar manner, the agent tries to calculate datasets such as this once per turn at most. Due to this system being a proof-of-concept, efficiency was not always the biggest concern during development with the favour instead being given to a balanced model. This means that some inefficient sections are present, but are not so intertwined with other components in such a way that makes them difficult to fix.

The first example is due to the fact that the network structure can change from one agent's turn to another, neighbourhoods have to be generated for every agent every turn which is not very efficient. This requires a lot of graph traversal and in larger graphs, a lot of processing time. Unfortunately, this is difficult as caching neighbourhoods introduces the chance of nodes who are newly removed from this set playing a role in calculations and verifying the neighbourhood requires the same amount of computation as creating it. One possible way to get around this is to cache the neighbourhood at the end of each agent's turn. Changes could then be handled by, again at the end of each turn, having each agent broadcast its one-hop neighbours to its one-hop neighbourhood. These nodes can then, between the end of their turn and the start of the next, check for the existence of neighbourhood nodes – new ones can be added, whilst if they are unaccounted for then they can be removed.

A good level of efficiency in the agents lends itself to scalability of the model. This is because the majority of the runtime is spent within agent actions so if these are made to be as efficiently as possible, then the model will perform better when more agents are added. Because this model is not developed with the intention of running simulations in excess of 10,000 nodes, little emphasis has been given to developing with the intention of running large scale tests though in a commercial solution, this would be very important. Developing in this manner requires attention to be split between individual agents and the system as a whole; for example, removing any actions carried out by every agent that could be done once for the whole network would save a lot of compute cycles. If performance was of paramount importance, then the stage at which network reconfiguration happens could be separated out from within the agent to for the whole network at once. This means that rather than agents would work on a fixed network each turn rather than one which changes between separate agent turns meaning steps such as scoring between agents would need to be carried out fewer times.

With regards to scalability, some changes would need to be made in order to provide good performance on large systems. Agent-based systems are difficult to scale due to the amount of processing power required for even smaller ones as even in this project, a basic representation of humans and their connections performed many different operations per step. As more agents are added, on top of the internal processing for this new addition, each existing agent may now have even more nodes in its neighbourhood, resulting in more calculations per agent. The exact effect of more agents being added can depend on a number of factors, such as its degree, size of neighbourhood and more, making it difficult to quantify.

To better analyse this, simulations and graph generations were carried out on a variety of graph sizes on both generated small-world and scale-free graphs. Each simulation ran for 30,000 steps and was timed, with the results being plotted. For reference, the tests were carried out on a 3.4GHz quad-core, Intel i5 processor, using 1.3GB of RAM for the Java Virtual Machine. Graph generation is a one-off step in a simulation that, in its most complete sense, will create a graph of size *n* agents each with a set of attributes and social connections. It should be noted that if a sampled graph is used, the time taken for a *n* nodes to be prepared for running will be less than that of a *scale-free* graph being generated since the sampled graph simply uses the attribute creation step that *scale-free* graphs also use.

FIG X SCALE FREE gen

FIG Y SMALL-WORLD gen

As visible in **FIGURE X**, the generation time for scale-free graphs increases almost quadratically as the number of nodes increases, reaching around 30 minutes creation time for a 3000 node graph. Adding to this, as the generation method leaves disconnected nodes in the graph, many of these have to be removed causing the operational node count to be often lower. Obviously, this level of performance would not be effective in a commercial environment since large graphs of over 1000 nodes would take too long to create using this method. This, however, may be partly due to the implementation rather than the method; **FIGURE Y** shows the generation times for a small-world graph, using the *Repast* built-in small-world creation algorithm. A very short creation time across the range of node counts tested with a linear increase can be seen, which is a better performance however the method for generating small-world graphs is much less intensive. At least some of the cause of this is the fact that the calculation used to determine if an edge is to be added or removed is simpler for small-world graphs than scale-free. From this, it can be seen that to make this aspect of the model more commercially viable, it needs to offer a more scalable generation method for scale-free graphs.

FIX X2 SF run

FIG Y2 SW Run

Perhaps more important than this is the time taken for simulations to run. This is a difficult area to quantify since two graphs of identical node count can vary in many ways such as edge density, average degree and more, in turn affecting how the graph may behave. Using the graphs generated for the previous tests, the results of the simulations carried out can be seen in figures **X2 and Y2**. A point to note is that due to the scale-free method not producing graphs of exactly the number of nodes required, creating a 3000 node graph to match the upper limit of the small-world tests would have been very difficult. It can be seen that both follow a linear increase as nodes are added, with simulations of around 750 nodes taking around 4 minutes for a small-world graph and just over 5 minutes for a scale-free one. The similarity in the trends of these graphs indicate that the type of graph may be independent of how long a simulation on it takes. If this is the case for all graphs, then the focus for scalability improvements lie within agent efficiency, as detailed above. In regards to the simulation times at the moment, whilst 20 minutes for a 3000 node small-world graph may seem high, if run on a cluster or HPC, the runtime would be much shorter. As such, increasing the number of agents significantly would not cause untenable runtimes, should the requite computing power be available. From this basic investigation, the model appears to be fairly scalable but would benefit from work on agent efficiency if simulations were to take place on personal computers.

On the whole, the system displays a good level of extensibility and ease of maintenance. From the start of development, focus has been on building not only a useful model but also a platform on which makes it easier to improve said model. This was done by ensuring that any tools developed were done so in as general a manner as possible, as well as grouping them into related packages. The agent itself separates out its actions into methods, especially the three steps of neighbourhood finding, running the decision tree and reconfiguring connections. A useful feature that this allows is the easy replacement of any of these components, without having to disentangle them from the rest of the agent implementation – a company could use this to create a number of separate approaches to the decision tree and network reconfigurer and test them by simply changing the method calls.

Furthermore, the Repast framework is a large, public project which causes it to be well organised and ensures that user code is separated from system code. By making user code interface with the system instead of building within it, modifications to said system can be avoided and the efficiency and speed of the underlying simulator can be guaranteed. This is important since it shifts the emphasis for code to be well written and efficient to the model and relating tools thus. Furthermore, keeping this separation means that if the model implementation is maintainable, the developer does not need to worry about the simulation environment code. Although this is very useful, the model is developed within *Repast Simphony* whereas a commercial level model would probably require HPC access and due to this would need to be rewritten in C++. Assuming the same principles would be used in that model as in this, this task would not need as much time or effort to develop and so would not be a major problem.

<u>*Model Analysis*</u>

Overall, the model did provide some insight into the problem domain and from this is in some respects successful. Considering the fact that the intention was for the system to hold a proof-of-concept status, keeping the modelling of both the agent and network to basic levels, it is encouraging to see that the project does seem to exhibit behaviours that would be expected such as being able to run on networks without them collapsing into tight clusters but at the same time with the network actually changing. Furthermore, as described above, it does produce results that indicate how networks can be manipulated in order to aid in giving up smoking. With this in mind, a number of other points have been discovered that are important in the development of this type of model.

A major factor during the development of the project was that of *balancing* the model. Trial and error indicated that balance was indicated by how long the social network could maintain a structure similar to its beginning state; for example, if a scale-free network collapses into a small-world network in fewer than 10,000 simulation steps, then the model would appear to be very imbalanced, whilst a collapse over a longer period of time would indicate a better configuration. It was found that certain aspects of the system had a much greater impact on this balance than others:

- Social connection add/remove boundaries – a lower bound of 0.4 and upper bound of 0.7 was used throughout the simulations though whilst finding this boundary set, many others were tried. As might be expected, raising the lower bound too high caused the graph to become sparse over time, since it would be overzealous with edge removal. On the other hand, dropping the upper bound by just 0.05, the network became saturated with edges very quickly, resulting in many feedback loops and lots of edges becoming high influence.

- Decision tree end-state inputs – the values which are fed into the decisions that choose if a person begins or stops smoking significantly affects the distribution of smokers to non-smokers. Setting the thresholds too low makes it very easy for that decision to cause a change such as smoker to non-smoker, meaning that agents will readily make this change and flick between two states regularly. In contrast, setting the values too high means that extraordinary situations are required for changes in behaviour, which may not happen. This would cause very little change in the graph and could cause it to become stale, reaching a saddle-point very quickly.
- Decision ordering – the order in which an agent reaches decisions changes the significance of attributes since on average, the higher in the decision tree it is the more agents will encounter it. For example, the most important decision is for whether the person is a smoker in the tree in this model, so it is at the top. If, for example, willpower was placed at the top of the tree, then any changes to willpower would have a much greater effect on agent behaviour.
- Influenced attributes – given that these are generated based on the autonomous behaviour of the agent to the point of calculation, there is a chance that influence causes these values to tend towards the extremities of their ranges. Should these attributes be used as the basis of decisions within the tree, this can cause entire branches to be inaccessible to agents, heavily skewing the balance of the tree and by extension, the model.

If the model were to be used for further research, these areas should be the focus when it comes to rebalancing, if necessary.

Behaviour-wise, the model performs generally as expected but with some areas for improvement. The fact that it can be configured to avoid mass clustering indicates that the interaction between agents performs in a similar but basic manner compared to human behaviour. This appears to be done by maintaining a number connections that are of different influence level with others whilst avoiding having a very high number of these connections – much like a person would find it difficult to maintain a lot of friendships simultaneously. Furthermore, the manner in which connections are reconfigured takes uses abstract representation of evaluating likenesses, meaning that similar people aim to connect with each other and, due to this connection, then influence each other. In social networking terminology, this is *homophily*, as described in the *Literature Review* section of the document. Given that this is observed in real-world networks, its occurrence here is useful.

Areas in which some more unexpected behaviours occurred were, on the whole fairly specific apart from the fact that attaining consistent results is difficult. Over a number of runs, the same simulation setups can give a wide range of finishing splits between smokers and non-smokers which means that for results to be determined, many iterations must be run then averaged. Even if this is the case, should the spread of results be uniform across the range, this would imply that the particular parameters used have no effect on the final state. This is hard to identify because the model relies on agent autonomy and in doing so, a certain variance between iterations should be expected; as such, a lack of consistency may be endemic due to the choice of approach. Tackling this involves paying attention not only to the number of iterations used but also the variance and distribution of the results.

Some of the more specific undesirable behaviours, whilst small, have an impact on the model at large. Firstly, it seems that the model makes beginning smoking (particularly relapsing) too easy resulting in a number of give-up attempts over the course of the simulation or, in extreme cases, constant switching between being a smoker and non-smoker. This is not ideal as the spurious changes drive up the number of attempts and giving up, effectively nullifying the impact of previous attempts, as well as them not providing a source of influence consistent with one behaviour. Adding to this, by the end of the simulations there was very rarely any individuals giving up. As agents that do give up and then relapse increase their chances of doing so in the future so if this is repeated a number of times, the length of the runs (30,000 steps) will mean that any attempts to give up are likely to fail rapidly. Ultimately this is a model stability issue and to address it, a better method of having willpower and previous attempts at giving up would need to be developed that could incorporate the time since the last effort and possibly other attributes. Doing this would also contribute towards solving the issue of it being too easy to begin smoking, especially those who are at risk of relapsing.

From the results above, it is clear that the decision tree for cessation and social networking part of the model do have worth since they provide the opportunity to manipulate them in a number of ways or to track their behaviour, allowing for analysis of particular model setups. The decision tree provides a good way to keep the decision-making process simple whilst offering a range of different choices to the agents and because of how the values for said decisions can be changed, the balance of the tree can be adjusted with ease. The network itself proved to be a good representation of real-life networks, allowing the opportunity for both generated and sampled networks to be simulated upon. Influence within the model was built into the decision tree ensuring that the influence did actually affect the choice of behaviours. This was further reinforced by the *end-states* in the tree relying on the behaviour of the neighbourhood; in retrospect, this weighs a little too heavy in favour of the actions of those around the agent so the inclusion of willpower or stubbornness would help to recreate those who will behave in a certain way regardless of their surroundings. Obviously, for a commercial level model, the tree would be greatly expanded with more attributes added to the agent; this would then give further configurability and more insight into the different areas of smoking cessation. Coupled with further integration of influenced attributes into the tree, the decision tree and graph approach is promising.

Further Improvements

Given that this model is a starting point rather than a complete solution to the problem, there are many possible changes that would bolster the performance and accuracy. These range from minor changes to reworking of entire sections and in general were not included due to time constraints or simply being too complex to include.

The agent itself could be adjusted to include more attributes and for those already included, more accurate models. The former is a difficult proposition since to make these additions, it should be the case that the attributes have been shown to have some effect on smoking through other research. Due to how circumstantial smoking behaviours seem to be, acquiring general information about factors in smoking cessation is not straightforward. As such, to undertake this, the best route would be to undertake specialist research in the field with the specific intention of understanding smoking behaviour contributing factors. In spite of this, some attributes such as stress could be included, assuming a reasonable model could be incorporated which would account for both internal and external contributors to the stress. Accurately modelling existing parameters is easier since as shown in the *Research & Literature Review* section, work towards this has already been done. The process for each attribute would involve isolating each concept and investigating it individually, through which a mathematical model could be developed for it. To prove the accuracy, these models would then have to be tested against some real-world measure. Whilst this would aid in the general result quality of the model, it could incur extra computational cost which may be a problem if large simulations are the main focus. Avoiding this would then require either efficient attribute representations or only modelling the most significant ones in such a detailed manner.

In terms of the decision tree, there are a number of design-level changes that, whilst would not impact the operation of the tree, would allow for much faster and easier tree prototyping, in turn making it more straightforward when developing new or adapting existing trees. A modular approach could be used where a decision is represented by some abstract type, allowing the specification of which attributes to decide on and their relevant thresholds for each as well as some list of which decisions lead to others. This would remove the need for extra programming when implementing new decision trees so would make the model more usable for those who do not know how to program. On top of this, it would be easier to build visualisation tools for and get an understanding of the tree at a glance compared to the current implementation which uses nested *if* statements.

Another decision-tree based change would be the inclusion of 'dead-zones', which are tree states that result in no change to behaviours and/or attributes. By defining more 'no change' behaviours, this would provide another aspect of configurability and also, by offering the ability to not change, produce more gradual results. An extension to this would be to separate out physical and mental attributes. This would allow for influence, for example, to act upon mental state which in turn acts upon the individual's decisions and behaviours. Combined with the idea of dead-zones, gradual changes could be made to the mental attributes resulting in more of a 'decision-making over time' behaviour by the agent.

Finally, some extra tools would be very useful in order to produce graphs more accurate to real-world ones. Examples of this would be to generate graphs that could have configurable clusters such as groups of smokers, or of highly connected nodes. This would provide better emulation of smaller, dense social circles within a larger network, similar to groups of friends and their interactions with other groups. This would require elements of both small-world and scale-free graphs, so neither method alone would be suitable to generate them. Additionally, some method of specifying specific attribute changes within agents at certain points of the simulation would be very useful in investigating how changes within a subset of nodes affects the network at large. This would allow a network to be set up and run for an amount of time before some pre-determined changes happen, providing a more organic simulation compared to starting with the network changes required.