Progress Report

Matt Smith

November 13, 2012

1 Project Description

This project, ultimately, aims to provide a simulation of individuals attempting to give up smoking in the situation of a social network. In particular, focus is given to how the network manipulates their behaviour and whether some social organisations and constructs are more suited to triggering the breaking of the smoking habit. To analyse this, a model of both a human and a social network will be created, through which a given period of time can be simulated - the result of this being data that can be analysed to find the previously mentioned social conditions conducive to give up smoking.

2 Progress Summary

On the whole, the project continued to be consistent with the project specification, insofar as rather than large changes having to be made, specific details and implementation methods are now known, and there is a greater understanding about the theoretical basis of the models.

No development towards the final model has been started - to date, experimental code has been written as proof-of-concept programs to examine different approaches. These pieces of work will, however, form the basis for parts of the final model.

3 Model Research Progress

3.1 Human Representation

As modelling a human (also referred to as a *node*) in general is a difficult task, the scope of the project can be reduced significantly by ensuring that only aspects of human behaviour related to the analysis of smoking are used. Furthermore, some assumptions about the way in which nodes behave have to be made to prevent the project encompassing a large part of sociological and psychological theory.

Due to this, it has been decided that the human behaviour will be represented in a probabilistic manner - the reasons for this are twofold. First, is that it provides a simple method of recreating some facets of actions, for example a person could rate, on a scale of one to ten, how likely they are to start smoking in the next few days. Although this is not necessarily an accurate measure, it is enough for a simplistic representation. Example attributes include:

- Susceptibility to peer-pressure

- Likeliness to smoke

- Willpower - how likely they are to continue their giving-up effort.

Some attributes, however, cannot be representing by probabilities. These include items such as how many cigarettes someone smokes, whether someone smokes at all and how likely they are to reach out beyond their local network.

Representing this series of decisions will be done through the use of a decision tree - a series of choices that the person can make, leading them to some state consisting of a collection of attribute values. The

key advantages of this are that it allows a fixed number of choices per tick, with the ability to structure the choices (for example, important ones at the top). The structure and content of this tree is largely a developmental matter and needs to be tuned over time, but a concept tree can be seen below.

Sample Agent Decision Tree to go here

In terms of the model-at-large, it is important for the network to be able to reconfigure itself over time, since people move towards to others with similar interests **REFERENCE** - it should also be noted, however, that friends develop similar interests, thus giving a potential conflict. At this stage, proof-of-concept work has been done in regards to basic reconfiguration, but the actual feature will be created in the development section of the project.

To further balance out this conflict, the system will have Ácts of God, where random behaviours will be introduced with a very small probability. This will be tuned to have the impression of subtly unsettle the network and prevent it falling into equilibrium for long periods of time, whilst also attempting to simulate the human behaviour of occasionally acting out of character. **RESEARCH NEEDED HERE**

Influence propagation will tie the human to the outside world - in the current plans, no formal method of influence propagation will be implemented, however a decision tree can either incorporate or have a standalone tree for the impact of influence from surrounding nodes. Once more, this will be done probabilistically, where the edge weights either dampen or emphasise the influence passing through them (see below). By doing this, influence will still flow through the network without making the project scope too large.

In terms of the items that propagate, attributes are a particular focus - the idea of others affecting your choices is key to this process. As such, the probabilities used in the decision trees will be manipulated by the nodes connected to a given person. The overall aim is that these manipulations will represent the overt and subconscious actions of others that encourage someone to choose a particular action.

3.2   Network Representation

One of the key factors in the success of this project was to be able to generate a realistic social network, as without this, the model would be unable to produce meaningful results. To date, two methods have come out as useful in the research - *small-world* and *scale-free*.

Small-world networks are created using the **INSERT NAME** algorithm, and produce graphs similar to the one shown below **INSERT GRAPH**. In general, the method uses a degree averaging technique to give all members in the graph roughly the same number of edges. As implied by the name, this method is strongly related to the idea that most people know each other, or if not directly, through a small number of connections - as such, this is not entirely useful to this project as it is difficult to realistically model the environment of larger groups of people.

In a larger graph, it is highly unlikely that most people know each other within a few connections, so the creation of a scale-free network (using the **INSERT NAME** method) succeeds here. This uses a *preferential attachment* method - ẗhe rich get richer̈ to generate networks which instead have a number of h́ubńodes, similar to popular figures in friendship circles. Additionally, the graph is significantly less connected, resulting in a seemingly more realistic network.

NEED TO INSERT METRICS OF CLOSENESS, DIAMETER AND AVERAGE DISTANCE HERE INSERT VALUES ABOUT HOW REALISTIC NETWORKS ARE

Whilst using pre-made/sampled networks has been considered for the final data output, at this stage the project will work on a basis of a computer generated network - assuming that the generated network is considered ŕealistic, then this will provide a much wider range of data and thus analysis at the end of the project.

Sampled networks are, however, very useful in the development phase - this is largely because they provide a consistent, realistic network to test new constructs and methods on. As such, when it comes to the development, sampling of some data sets will take place for this purpose.

Moving from the whole network to individual edges, friendships in a social network map well to a directed graph. This is because friendships are not necessarily symmetric - assuming they are removes the concept of role-models, an important feature of social networks, and also blocks another key concept, that of different people being influenced by others to a different degree.

Using this idea, it has been decided that edge-weight should represent the ínfluence factoróf that relationship - if node A has an edge of weight 0.5 to node B, then any behaviours that traverse this edge will

only do so with an impact of half the original. Firstly, this reinforces the idea of role-models (a person is highly influenced by another, but the other is not necessarily influenced, or even knows, the first). It also provides a more realistic approximation of how a friendship works, since although friends, one person may think more of the other and as such tend to be influenced more easily by them. Values between 0 and 1 will be used here to maintain the probabilistic approach.

3.3   User Features

There are two main user features that are to be included in the project, to date. The first is the generation of networks based on loose statistics - this, when combined with both the previously mentioned network generators and the human model, will be a relatively trivial interface between the initialisation of the simulation environment and the user, so should cause no problems in the development phase.

The other feature is the ability to parse and export existing graphs - again, as mentioned above, work has gone into parsing datasets and outputting them as *GraphML* - to work this into the simulation setup should be relatively simple as it is based on code already written. A particularly useful feature would be to export the network to *GraphML* at given simulation ticks, to allow for static analysis of the network at a given time.

3.4   Commercial Model Integration

At this stage, it would not be suitable to consider whether the model will fit into a larger, commercial-level model as not enough development of the model itself has been completed. It should be noted, however, that up to now, there have been no issues that would appear to cause any problems.

3.5   Visualisation

Along with simulating the human and network behaviour, it is important to be able to view snapshots of the model to understand what is happening. So far, this has come via two avenues - *Gephi* and the *Repast GUI*.

Gephi is particularly useful when it comes to generating standalone networks for analysis - examples of this include inspecting algorithm-generated networks for realism and checking if a sampling method is working correctly, producing a reasonable graph.

The Repast GUI provides a similar service but is refreshed on each simulation tick. Due to this, the GUI effectively gives a live view of the network and how it is behaving - of course, this is has some more simplistic tools than Gephi, so works in a complementary way.

3.6   Analysis of Model

Although final analysis methods are yet to be decided, great care has been taken to ensure that there is as larger range as tools as possible available for analysis, and as such stopping any implementation based restrictions on these tools. An example of this include building a *GraphML* serialiser for *JUNG* networks, to allow for *Gephi/Cytoscape* analysis of these graphs.

Considering final analysis methods, statistical methods are likely to form the basis of comparing realistic models to the model produced as part of this project - this will provide a numerical approach to evaluating the ŕealismóf the networks and nodes involved.