university of groningen / faculty of science and engineering / artificial intelligence

An Exploration of Threshold Model Updates

Bachelor's Project Thesis

Jirmaine Tan, s3965090, j.tan.9@student.rug.nl,
Supervisors: Edoardo Baccini - e.baccini@rug.nl

Abstract: Threshold models are used to study how information, behaviours, new technologies, fashions, viruses, etc. diffuse through a social network. These models visualise diffusion through networks; a collection of nodes (vertices) and edges that connect nodes together. A threshold value is assigned to each node. An agent will adopt a behaviour/technology/fashion when the proportion of their neighbours who also have this behaviour meets the threshold. We define threshold models as a Social Networks Model for discussing properties of structural/distance-based social selection and threshold based social influence. We then introduce extended models for similarity-based social selection as well as social abandonment of behaviours for both selection and influence. This paper studies the properties and states of the social networks that these updates create. We first explore properties of the updates using the theoretical framework. We then build a simulation to explore the extended updates empirically.

1 Introduction

Our behaviour is shaped by our surroundings. The people we interact with (unconsciously or consciously) influence our decisions, behaviours and beliefs. We can be influenced into conforming to a new technology or specific diction, dressing a certain way or behaving differently. This is social influence, which determines the way a diffusing novel (or renewed) behaviour gets adopted by a population. Social influence can be explored in networks through the use of threshold models. Threshold models are primarily visualised using graph networks, which are a collection of agent nodes and edges that connect the nodes. They are used to study the dynamics of diffusion by exploiting conformity in the form of threshold-based influence (Easley & Kleinberg, 2010); A threshold model assigns to each node (person) a threshold that determines the proportion of neighbours that must adopt a behaviour for that node to do so as well. These models of social influence have especially gained popularity in logic-based studies of diffusion Baltag et al. (2019), (Granovetter & Soong, 1983), Christoff et al. (2022).

As in Dodds & Watts (2011) we limit social influence to binary decisions, where there are exactly two distinct and alternate options to choose from. There is much hidden complexity in the behaviour of binary decision making. For example, Schelling (1969) modelled binary decisions as a question of race i.e. you are either of a certain ethnicity or you are not. In this demographic abstraction, Schelling showed ethnic communities tend towards segregation by race. Individuals tend towards others of the same race. This shows race is a shared similarity factor, and those who are the same more readily connect. Not only restricted to race, people can be segregated by their acquired traits e.g. hobbies, professions, etc in a binary manner; either you can play the guitar or you can't. Social media platforms allow this expression, where a user curates the information they see online based on their traits (acquired or innate). The accounts one follows, their friends and the groups they join all determine their social profile in a binary manner. Platforms such as Linkedin, Facebook and Instagram exploit these profiles to recommend to each other individuals with similar social profiles (Aiello et al., 2012).

Furthermore, many logic-based accounts of social influence and diffusion permit nodes to adopt behaviours and abandon them (Smets & Velázquez-Quesada, 2019). These are known as non-monotonic updates i.e. a monotonic update prevents abandoning a behaviour. In a monotonic world, I would forever use a fax machine had I adopted the use of one. However, simply put, people change.

We have described here not only social influence but social selection. Social selection determines when two people connect based on how similar they are to one another, be it race or hobbies. This is encapsulated by homophily; the principle that similarity breeds connection McPherson et al. (2001). It has been studied in many disciplines, for example Hebbian learning in neuroscience; Cells that fire together wire together. Computer science makes use of homophily in systems through data clustering. Recommender systems for example exploit homophily to group users who display similar purchase patterns (Sarwar et al., 2002). By grouping these individuals, recommender systems can provide suggestions that are most pertinent. For example a person who purchases with an interest in fashion would be clustered with other 'fashionable' individuals all being suggested online fashion stores instead of music stores.

Homophily has been studied in many ways by varying how one defines similarity. In one, similarity between two people in a population can be approached structurally. Two people are more similar when they share relationships to others. In laymans terms, connection is brought about based on the structure of ones social network (Jeh & Widom, 2002). In another, homophily is brought about through trait similarity (Smets & Velázquez-Quesada, 2020). Two people are more similar when they share more similar traits i.e. race and hobbies. There are other ways in which to define similarity, however these are two of the most prominent and this paper will focus solely on these two forms of homophily.

The goal of this paper is to explore different forms social influence and social selection through the use of threshold models. Social influence and selection are interrelated processes that constantly change the shape of a population (Smets & Velázquez-Quesada, 2019). A result of social influence directly affects the way in which social selection takes place. If one abandons a behaviour those the relationships they formed based on this behaviour may wither, and as the connections one has changes they are less susceptible to the conformity effect of certain behaviours.

We will construct a theoretical framework of a threshold model and compare both theoretically and experimentally diffusion of behaviour(s) through a social network with different forms of social influence and selection.

This paper proceeds as follows: Section 2 constructs the threshold model framework in the form of a Social Networks Model (SNM). It will provide the basis for exploring updates on threshold models. Section 3 defines social selection and influence in addition to defining apparatus that allows us to study the evolution of an SNM such as the states or the relationship between nodes in the network. Section 4 defines further model updates to describe more complex network evolution. Section 5 describes the simulation used to explore the further updates. Section 6 evaluates the experimental results of the simulation and discusses limitations. In section 7 the paper is summarised and some conclusions drawn.

2 Threshold Models

We are concerned with studying social influence and social selection. We start by defining the notion of threshold models. In this paper a threshold model is an SNM defined below. This model framework will be used in the following section as a basis for describing social influence and selection.

Definition 1 (Social Network Model) *A Social Network Model (SNM) is a tuple* $M = \langle A, O, \beta, F, \theta_a \rangle$ *where:*

1. $A \neq \emptyset$ *is a finite set of agents*
2. $O \neq \emptyset$ *is a finite set of behaviours*
3. $\beta(\sigma) \subseteq A$ *is the set of agents that are adopters of behaviour* $\sigma \in O$
4. $F \subseteq AxA$ *is an irreflexive, symmetrical and serial binary relation with* $(a, b) \in F$ *showing that agents a and b are connected.*
5. θ_a *is an adoption threshold*

To clarify how an SNM is defined, below is an example of an SNM according to definition 1.

Example 1 *Fix the set of agents* $A = \{a, b, c, d\}$*. The behaviour* $O = \{o\}$ *be an arbitrary spreading behaviour.* $\theta_a \in [0, 1]$*. Consider a SNM* $M = \langle A, O, \beta, F, \theta_a \rangle$ *where:*

$$\beta = \{a, b\} \quad \begin{matrix} (a,b) \in F & (b,c) \in F & (c,d) \in F \\ (b,a) \in F & (c,b) \in F & (d,c) \in F \end{matrix}$$

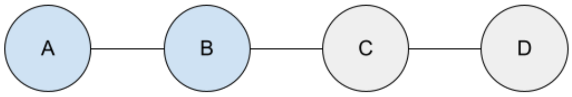

Figure 2.1: SNM presented in example 1

This SNM should be interpreted as follows. There are 4 unique agent nodes a, b, c and d. There is a single diffusing behaviour o. Nodes a and b are adopters of this behaviour. Links exist between agents a and b, b and c and c and d. There are 3 edges in the network however due to the symmetric nature of these edges, formally there are 6. θ_a is the adoption threshold that determines the proportion of a nodes neighbours that must adopt a behaviour before that node does so as well; it is any value chosen between 0 and 1. You could interpret the edges as social influence. An edge between two nodes shows that these two nodes are able to influence one another i.e. they share information. A behaviour can be interpreted as such; perhaps the tendency to play a sport, use specific diction, etc. There can be one or more behaviours diffusing through the network at any time.

3 Social Network Image Update

One has the freedom to share information. For example a passerby can look at a protest and actively engage with those who are protesting. This is a direct exchange of influence where the protesters (adopters) attempt to convert the passerby (non-adopters). An anti-war protest assumes such behaviour and attempts to convert those who are pro-war. Whether the passerby is anti-war or not, by interacting with the protesters they have opened themselves up to influence i.e. formed a connection. Given an SNM we propose updates that allow nodes in the network to behave as agents sharing information and forming connections. We will define here conditions that determines when and why an agent would adopt a behaviour from their neighbours and how two agents become neighbours. Further, we introduce definitions that help describe formally the states in the temporal evolution of a SNM.

3.1 Social Influence of Surroundings

People will choose to adopt information from their surroundings if their surroundings motivate said adoption. Consider again the war protest example. A passerby who is pro-war seeing the anti-war protest may interact and not change their stance i.e. they remain non-adopters. Note that the connection has been made either way (the edge has been formed) as they have shared information. However, in returning home they may find that their family has adopted an anti-war stance. The accumulated influence of their surroundings may be sufficient to encourage the adoption of an anti-war behaviour. In other words, we could describe a threshold for each person that defines how many people must adopt anti-war behaviour them to adopt it as well. Agents are similarly influenced by their surroundings. They can adopt a behaviour, technology, opinion, etc when the proportion of their neighbours who also adopted it is greater than a given threshold.

Definition 2 (Adoption) *For a given SNM* $M = \langle A, O, \beta, F, \theta_a \rangle$, *let* $\theta \in [0, 1]$ *be the adoption threshold. For* $a \in A$ *let* $F[a] := \{i \in A \mid (a, i) \in F\}$ *denote all the neighbours of a. We define the updated model* $M^{\triangle} = \langle A, O, \beta^{\triangle}, F, \theta_a \rangle$ *to be such that it is a SNM after an adoption step. As such adoption is as follows:*

$$a \in \beta^{\triangle} \textit{ iff } \left(\frac{F[a] \cup \beta}{F[a]} \geq \theta_a \;\; \textbf{\textit{OR}} \;\; a \in \beta\right)$$

A node a will adopt when one of two conditions is met. Either the proportion of $F[a]$ who also have this behaviour must be greater than the adoption threshold. Alternatively a must already have adopted this behaviour. Note that this definition of adoption is monotonic. A node will only accumulate new behaviours. For example once the agent has adopted the anti-war behaviour (or any behaviour) it cannot be abandoned. Furthermore, this shows that *adoption* is not idempotent as adoption relies on the behaviour of a node's neighbours, which is altered by this operation. Therefore $(M^{\triangle})^{\triangle} \neq M^{\triangle}$

for any model M, unless M has reached a stable state (a stable state is rigorously defined later).

3.2 Agents Shaping their Surroundings

People will choose to connect with other people in many different situations. In one, we meet people through a shared friend i.e. a friend introducing you to their friend(s). Agents are also able to shape their surroundings in this way. They form edges to other agents with whom they share a common connection. This mirrors how we form connections with people who are introduced to us by our existing network of connections. We propose here a distance-based (or structure-based) policy for edge updates (Jeh & Widom, 2002). It is not the same has having a metric distance between two points. The notion of distance-based similarity connects two nodes that are connected to similar nodes.

Definition 3 (Friendship Selection) *For a given SNM $M = \langle A, O, \beta, F, \theta_a \rangle$, let $a, b, c \in A$. We define $M^{\Pi} = \langle A, O, \beta, F^{\Pi} \rangle$ to be such that it is a SNM after a friendship selection step. As such, friendship selection is as follows:*

$$(a, b) \in F^{\Pi} \textit{ iff } ((a, c) \in F \wedge (b, c) \in F) \; \textbf{\textit{OR}} \; (a, b) \in F$$

The definition for friendship selection in Def 3 states that two nodes will have a link in the network if they have a common/shared neighbour or if they were already linked in the previous step. Similar to *adoption*, this operation is monotonic. A node can only accumulate new edges. This operation is also idempotent since it updates the edges in the graph, which is what determines the formation of new edges. As such, $(M^{\Pi})^{\Pi} \neq M^{\Pi}$ for all M unless it has reached a stable state.

3.3 A Holistic Update

We have previously defined two distinct updates; Def 2 defines a way in which an agent is influenced by their surroundings and Def 3 a way in which an agent influences their surroundings. Applying each of these updates to a network leads to distinct model states namely $M^{\triangle}$ and M^{Π} respectively. We have presented them as two distinct processes however in reality these two can happen simultaneously. People are not restricted to performing these behaviours in sequence, in a single conversation one could form a new connection and adopt a new behaviour. We must define formally an update that incorporates both Def 2 and Def 3.

Definition 4 (Holistic Update) *For a given SNM $M = \langle A, O, \beta, F, \theta_a \rangle$ we define $M' = \langle A, O, \beta', F' \rangle$ such that it is a SNM after a holistic update step. β' is as defined in Def 2 and F' as defined by Def 3.*

These two updates occur simultaneously in a single time step. This operation satisfies the same properties as adoption and friendship selection. $(M')' \neq M'$ for all M unless it has reached a stable state.

3.4 Network Descriptions

Now we have constructed a model that describes what network changes occur but not yet a way in which to study them. We intend to explore the speed at which a behaviour cascades through a network with this model. Returning once again to our protest example, it is interesting to see how quickly a behaviour such as an anti-war stance spreads through a community. However, studying the spread of behaviour extends to other phenomena such as new technologies, fashion or deadly viruses. Studying diffusion requires evaluating the sequences of updates of models as the agents interact according to Def 2 and Def 3. These sequences are a collection of discrete time steps. The following definitions will aid evaluating these sequences.

Foremost it is important to define what a sequence of updates entails formally.

Definition 5 (Sequence of Updates) *Let $M = \langle A, O, \beta, F, \theta_a \rangle$ be an SNM. Furthermore, let $M_0 = M$. Then*

$$M_{n+1} = M_n^{?}$$

where $? \in \{\triangle, \Pi, '\}$.

A sequence of updates is a collection of updates on a SNM. Each model is assigned a discrete numerical value n that defines its place in the collection. For example, M_2 is an update that occurs after M_1.

Next, for the evaluation of speed there must be terminal states. A sprinter, marathon runner, swimmer, etc is not faster than their competition unless they win the race. However it is still true that a sprinter can be faster than the others even if they lose i.e. their top-speed. Sequences allow us to determine the same with models as we can directly compare equal time steps. Like a finish line, terminal states define when a behaviour has finished cascading through a network. There are two that we will consider; stable state and complete cascade.

Definition 6 (Stable State) *Let $M = \langle A, O, \beta, F, \theta_a \rangle$ be an SNM. We say that a model is stable iff.*

$$M = M^?$$

where $? \in \{\triangle, \Pi, '\}$.

A stable state occurs when two successive models in the sequence of updates are equal. This means applying another update to the current model M_n will lead to the same model.

Definition 7 (Complete Cascade) *Let $M = \langle A, O, \beta, F, \theta_a \rangle$ be an SNM that has reached a stable state. A complete cascade occurs when all nodes in M have adopted behaviour O.*

$$A = \beta$$

A graph has an incomplete cascade when not all nodes in M have adopted behaviour O.

$$A \neq \beta$$

There are two ways in describing the terminal state of a cascading behaviour. The first is a complete cascade wherein all nodes in the network have adopted. The second is the opposite; an incomplete cascade where all nodes have not adopted.

The term 'speed' in this paper is a direct comparison of the terminal states of two complete cascaded models. As such, 'speed' could be thought of as how long it takes for a sequence of updates to cause a complete cascade in a network in which a complete cascade can happen.

Definition 8 (Speed to Cascade) *Let $M, N = \langle A, O, \beta, F, \theta_a \rangle$ be two SNMs that have reached a full cascade. Model M has a faster speed to cascade compared to N if*

$$M_n^? \text{ and } N_m^? \text{ and } n < m$$

where $? \in \{\triangle, \Pi, '\}$.

M is faster than N for some sequence of updates if M reaches a complete cascade in less steps than N.

Definition 9 (Cluster of density d) *Let $M = \langle A, O, \beta, F, \theta_a \rangle$ be an SNM. A cluster of density d is a set of nodes $N \subseteq A$ in M s.t. each node in the set N has at least d proportion of neighbours in the set. Let $\theta_a \in [0, 1]$ be the adoption threshold, if N has density $d > 1 - \theta$ then we say that n is a cluster of critical density.*

It is well known that a graph with clusters of critical density will have an incomplete cascade. This has been proved in Easley & Kleinberg (2010).

Nodes are connected by edges in the network. The existence of these edges implies a distance between nodes. We exploit this with friendship selection (3), allowing nodes to connect if they share a neighbour. In a network a shared neighbour indicates that two nodes have an edge converging to the same neighbour. In other words, these two nodes are separated by two edges. We can think of distance in this way and define it inductively as in Baltag et al. (2019).

Definition 10 (Distance between nodes) *Let $M = \langle A, O, \beta, F, \theta_a \rangle$ be an SNM with agents $A = \{a, b, c\}$. Let $N[a]$ denote the set of neighbours of a, let $d^0(a) = \{a\}$.:*

- $d^{n+1}(a) = d^n(a) \cup \{b \in A : \exists c \in N[a]$ *and* $b \in N[c]\}$

The distance between nodes is the minimum number of links connecting two nodes; the distance between node a and the rest in example 1

$$d^0(a) = \{a\}$$
$$d^1(a) = \{a, b\}$$
$$d^2(a) = \{a, b, c\}$$

for nodes where no such path exists, then

$$d^\infty(a) = \emptyset$$

However, this distance is less exact in that the set contains nodes that are within distance n not exactly at distance n. As such the nodes that are exactly n distant are those in the set $d^{n+1}(a)$ and not in the set $d^n(a)$; or more precisely $d^{n+1}(a)/d^n(a)$. The distance between two nodes is the path containing the minimum number of edges between them. In complex networks it is plausible that there can be several paths connecting two nodes and so we must emphasize that in this paper it is the minimum path.

The following definitions describe the properties of a sequence of updates; heterogeneous and homogeneous.

Definition 11 (Heterogeneous Sequence) *Given an SNM $M = \langle A, O, \beta, F, \theta_a \rangle$ a heterogeneous sequence of updates is defined as follows.*

$$M_0^?,\ M_1^?,\ M_2^? \ldots\ M_n^?$$

where $? \in \{\triangle, \Pi, '\}$.

A heterogeneous sequence of updates is a sequence of model updates where each discrete time step is not equal to the rest. In other words, more than one update is applied in the sequence.

Definition 12 (Homogeneous Sequence) *Given an SNM $M = \langle A, O, \beta, F, \theta_a \rangle$ a heterogeneous sequence of updates is defined as follows.*

$$M_0^?,\ M_1^?,\ M_2^? \ldots\ M_n^?$$

where ? is exactly one element of the set $\{\triangle, \Pi, '\}$.

A homogeneous sequence of updates is the opposite. In the sequence of model updates exactly one update occurs.

Fact 1 *There exist models $M = \langle A, O, \beta, F, \theta_a \rangle$ with behaviour $o \in O$ diffusing with adoption threshold θ from an initial set of adopters s.t. after any application of update 4, no cluster of critical density is formed and M' is faster than $M^\triangle$.*

If a node adopts a behaviour at some time step n then we know they had a sufficient proportion of neighbours who also adopted at n. This is the case for adoption updates in Def 2. However the holistic update in Def 4 has simultaneous link additions meaning there exists a situation for a node wherein the proportion of neighbours meeting the threshold is only true for this update and not adoption; the new edges may sway the proportion in favour of adoption. Consider the following example that illustrates the fact above:

Example 2 *Let $M = \langle A, O, \beta, F, \theta_a \rangle$ be an SNM with agents $A = \{a, b, c, d\}$ and a spreading topic $o \in O$ from initial adopter $\beta = a$ with threshold $\theta_a = \frac{1}{3}$. The connections in F are as follows:*

$$\begin{array}{ccc} (a,b) \in F & (b,c) \in F & (c,d) \in F \\ (b,a) \in F & (c,b) \in F & (d,c) \in F \end{array}$$

At n=0, M is a straight line graph with each node successively connected to the next. Applying only adoption the model reaches a complete cascade at $M_3^\triangle$, whereas the holistic update does so at M_2'. This is due to the formation of links $(b, d) \in F'$ and $(d, b) \in F'$ at n=1.
Based on these findings, we conjecture that if M' is faster than $M^\triangle$ then M has properties s.t.

1. the application of the rule defined in 4 doesn't create clusters of critical density for M'
2. there are nodes $a, z \in A$ s.t. $a \in \beta$ and $z \notin \beta$ and $z \in d^n(a)$ and $n \geq 3$

If applying Def 4 to some network leads to a cluster of critical density then it is obvious that it will be slower than applying Def 2 to that network. Networks with a cluster of critical density never reach a full cascade. Since we have defined speed as the steps necessary for a network to reach a full cascade, networks that do not are infinitely slow. For edge formation in Def 15 to affect a network, the longest distance between an adopter and non-adopter node must be larger than the edge formation distance. The formation of a new edge must create a shorter path for the behaviour to reach a non-adopter node. In sufficiently small networks, any edges do not create shorter paths than those that already exist. As such, new edges will not make cascades faster.

Fact 2 *There exist models $M = \langle A, O, \beta, F, \theta_a \rangle$, where behaviour $o \in O$ is spreading with adoption threshold θ_a from an initial set of adopters s.t. after any application of rule in definition 3, no cluster of critical density is formed. Then two unique sequences of $M^\triangle$ and M^Π can generate full cascade models M_n that have the same edges and same distribution of adopters.*

As per link formation 3, agents form a link if their distance $d = 2$. This means that a homogeneous sequence of M^Π updates will always lead to the same final model M_n^{Π} independent of $M^{\triangle}$ updates. However, these link updates change adoption behaviour. Heterogeneous applications of $M^{\triangle}$ and M^{Π} can form a cluster of critical density in a unique sequence, as nodes connect with others that aren't necessarily adopters. This alters the proportion of neighbours who are adopters. However, under the assumption that no critical clusters are formed, two heterogeneous sequences of updates will terminate with identical full cascade models if they have the same proportion of updates. This is not too hard to imagine since any sequence of M^{Π} will lead to the same F^{Π} and a full cascade means $A = \beta$ as defined in 7. If we were to remove the assumption that critical clusters are never formed, then it is possible that two unique sequences don't lead to the same final model. Consider the following example:

Example 3 *Let $M = \langle A, O, \beta, F, \theta_a \rangle$ be an SNM with agents $A = \{a, b, c, d, e\}$ and a spreading topic $o \in O$ from initial adopter $\beta = a$ with threshold $\theta = \frac{1}{3}$. The connections in F at $n = 0$ are as follows:*

$$\begin{array}{cccc} (a,b) \in F & (b,c) \in F & (c,d) \in F & (c,e) \in F \\ (b,a) \in F & (c,b) \in F & (d,c) \in F & (e,c) \in F \end{array}$$

There are sequences that result in both incomplete and complete cascades with this SNM. The two sequences 1 and 2 have an identical final model as they terminate with a complete cascade and have equal proportions of updates.

1. $M_0^{\triangle}$, $M_1^{\triangle}$, M_2^{Π}, M_3^{Π}, $M_4^{\triangle}$

2. $M_0^{\triangle}$, M_1^{Π}, $M_2^{\triangle}$, $M_3^{\triangle}$, M_4^{Π}

3. M_0^{Π}, -

4. $M_0^{\triangle}$, M_1^{Π}, M_2^{Π}, $M_3^{\triangle}$

The same model can have heterogeneous sequences that don't lead to an identical final model. The sequence of updates in 3 leads to an incomplete cascade. The M_0^{Π} update leads to 2 new connections from b to non-adopters $(b,e),(b,d),(e,b),(d,b) \in F^{\Pi}$. At n=1, the proportion of neighbours of b who have adopted is $\frac{1}{4} < \theta_a$. Sequence 4 again leads to a complete cascade. This displays again how the formation of links can increase speed to cascade since 4 reaches a complete cascade at M_4 whereas both 1 and 2 reach a complete cascade at M_5.

4 Extensions

So far we have reasoned about a network that has a single diffusing behaviour which has limited agents' ability to form meaningful connections. In real life there is more information passing through a network which we use to selectively determine our best fit neighbours. We can be introduced to people but not necessarily connect with them. Previously we have presented examples of social influence and selection regarding protesters. However consider now more personal social relationships; close friends or intimate relationships. These such relationships form when people have much in common. You are most likely to share common interests (behaviours) with your close friends or significant others. Note that now we state the plural behaviours, as closer relationships entail sharing multiple behaviours. Following are new definitions regarding social influence and selection that will support reasoning about multiple behaviours diffusing in the network.

Definition 13 *An extended Social Network Model (eSNM) is a tuple $M = \langle A, O, \beta, F, \theta_a, \theta_E \rangle$ where $A, 0, \beta, F\theta_a$ are as defined in Def 1.*

- *θ_E is a friendship threshold*

To clarify how an eSNM is defined, consider again example 1. We extend this SNM to an eSNM as follows:

Example 4 *Fix the set of agents $A = \{a, b, c, d\}$. The behaviours $O = \{w, x, y, z\}$ be arbitrary spreading behaviours. $\theta_a \in [0,1]$ and $\theta_E \in [0,1]$. Consider a SNM $M = \langle A, O, \beta, F, \theta_a, \theta_E \rangle$ where:*

$$\begin{array}{cc} \beta(w) = \{c\} & \beta(x) = \{a,b\} \\ \beta(y) = \{a,b,d\} & \beta(z) = \{a,c\} \end{array}$$

$$\begin{array}{ccc} (a,b) \in F & (b,c) \in F & (c,d) \in F \\ (b,a) \in F & (c,b) \in F & (d,c) \in F \end{array}$$

This eSNM should be interpreted as follows. There are 4 unique agents a, b, c, d. There are 4 diffusing behaviours, w, x, y, z. c is an adopter of w; a, b are adopters of x; a, b, d are adopters of y; a, c are adopters of z. Links exist between a and b, b and c and c and d. As before there are 3 edges however formally there are 6. θ_a is the adoption threshold; any value between 0 and 1. θ_E is the friendship threshold determining the proportion of behaviours that two nodes must coincide for them to connect; any value between 0 and 1. This value will be further explained rigorously later.

4.1 Update Variations

Those who form a community share similar behaviour (Das & Biswas, 2023). Previously in Def 3 we have defined a structural/distance-based approach where two nodes are connected if they are connected to similar nodes. However as agents with behaviour these nodes can have a notion of similarity; the behaviours in which they share and those in which they differ (Smets & Velázquez-Quesada, 2020). Close relationships tend to coincide many behaviours. This provides a further condition for the formation of edges and will affect how nodes shape their surroundings. The similarity between two agents is the set of behaviours that two agents both share and that neither have. In a single behaviour model, homophily is restricted by the binary nature of similarity i.e. two agents can either be fully similar or not at all. Consequentially, similarity in these models would result in non adopted and adopted agents never interacting. With additional behaviours agents can be somewhat similar. Importantly for edge formation, two agents can now connect if they are sufficiently similar. The similarity of agents is derived from the adopter (β) sets. For example, $\beta(w) = \{c\}$ shows that agents a, b and d are similar in that they are all non-adopters of this behaviour (they do not appear in the set). On the other hand, $\beta(x) = \{a, b\}$ shows that a and b are similar in that they are both adopters (they both appear in the set).

Definition 14 (Similarity) *Let $M = \langle A, O, \beta, F\theta_a, \theta_E\rangle$ be an SNM with agents $a, b \in A$. Let V_a denote the set of behaviours a has and $\overline{V_a}$ denote the set of behaviours a does not have. We therefore define similarity as*

$$sim(a, b) = (V_a \cap V_b) \cup (\overline{V_a} \cap \overline{V_b})$$

$$S_{a,b} = |sim(a, b)|$$

Two agents have a similarity equal to the number of behaviours that they coincide. These are the behaviours that both agents agree on and disagree on. Consider example 4. The following table shows the similarity between some pairs of nodes in example 4.

$$\begin{array}{ll} sim(a,a) = \{w,x,y,z\} & sim(a,b) = \{w,x,y\} \\ sim(b,c) = \emptyset & sim(a,c) = \{z\} \\ S_{a,a} = |O| & S_{a,b} = 3 \\ S_{b,c} = 0 & S_{a,c} = 1 \end{array}$$

In example 4 there are 4 behaviours diffusing in the network: w, x, y and z with initial adopters. The similarity between agents a and c is 1. This means there is exactly one behaviour in which both agents either agree on and both agents disagree on. The similarity between agents a and b is 3; there are exactly 3 behaviours that these two nodes both agree and disagree on.

Now that agents have a measure of similarity, they can use this knowledge to form more meaningful connections. Just as in real life, closer relationships can form through shared similarity. As with adoption there must be a certain number of behaviours that makes two agents similar enough to connect (in other words a threshold). Perhaps your closest friends play the same sport and listen to the same music. Extended friendship selection incorporates this and implements a similarity-based policy (Smets & Velázquez-Quesada, 2020).

Definition 15 (Extended Friendship Selection) *Let $M = \langle A, O, \beta, F\theta_a, \theta_E\rangle$ be an SNM. Let $M^E = \langle A, O, \beta, F^E\theta_a, \theta_E\rangle$ be an SNM after an extended friendship selection update. Additionally, let $\theta_E \in [0, 1]$ be the friendship threshold.*

$$(a, b) \in F^E \text{ iff } \{(a, c) \in F \wedge (b, c) \in F \text{ or } (a, b) \in F \text{ and } \frac{S_{a,b}}{|O|} > \theta_E$$

Extended friendship selection allows an edge to form between two nodes if they have a common neighbour as in Def 3 but they must also have a proportion of similar behaviours greater than the friendship threshold θ_E.

Previously in Def 2 and 3 the updates were monotonic. This meant that agents could not abandon a behaviour once learned e.g. once you play a sport you never stop. Adding another layer of complexity to 15 permits agents to disconnect from another agent if their similarity drops below the threshold i.e. non-monotonic. People are not bound to the relationships they have formed, at any point they have the agency to disconnect to those who they have connected with. We provide the same to agents in the network.

Definition 16 (n-Friendship Selection) *Let $M = \langle A, O, \beta, F\theta_a, \theta_E \rangle$ be an SNM. Let $M^Y = \langle A, O, \beta, F^Y\theta_a, \theta_E \rangle$ be an SNM after an extended friendship selection update. Additionally, let $\theta_E \in [0, 1]$ be the friendship threshold.*

$$(a, b) \in F^Y \textit{ iff } \{(a, c) \in F \wedge (b, c) \in F \textit{ and } \frac{S_{a,b}}{|O|} > \theta_E$$

Importantly, *n-friendship selection* in comparison to the extended variant is a non-monotonic update. This allows for nodes to leave social groups if their similarity drops below the friendship threshold. In other words, the neighbour set of a node $F[a]$ can shrink. Otherwise, the same properties hold.

Similarly, we extend adoption by adding non-monotonic updates, allowing for nodes in the graph to abandon a behaviour if their proportion of neighbours with this behaviour drops below the threshold.

Definition 17 (Extended Adoption) *Let $M = \langle A, O, \beta, F\theta_a, \theta_E \rangle$ be an SNM. let $M^N = \langle A, O, \beta^N, F\theta_a, \theta_E \rangle$ be M after an extended adoption step.*

$$a \in \beta^N \textit{ iff } \frac{F[a] \cup \beta(\sigma)}{F[a]} \geq \theta_a$$

Just as in 2, a node will adopt behaviour $\sigma \in O$ if their proportion of neighbours is greater than a threshold θ_a. The operation displays the same behaviour of non-monotonicity as in 15. Nodes can now drop a behaviour σ if the proportion of σ neighbours is less than a threshold. Otherwise the same properties hold.

Defined previously in section 3.4 were a number of definitions used to describe network states. These definitions were bound to the network states that were created by applying one of three updates, namely adoption, friendship selection and network adoption update. These definitions must now include the new updates extended friendship selection and adoption and n-friendship selection. Definitions 5, 6, 8, 11 and 12 are now redefined such that the set of possible updates includes those defined in section 4. Formally, $? \in \{\triangle, \Pi, '\}$ is redefined to $? \in \{\triangle, \Pi, ', E, N\}$.

5 Simulation

Due to the complexity of the extended updates we will explore the new models of social influence and selection using a simulation to provide empirical results. The updates Def 15, 16 and 17 will be combined into 3 different rules similar to the holistic update 4 that determine how agents behave. The first rule connects agents based on their similarity and allows them to abandon behaviours. The second rules is as the first, however agents can abandon edges as well. In the third no edge manipulation is permitted i.e. agents can neither form new edges or abandon old ones. With the simulation our objective is to provide empirical evidence regarding; (1) how readily these rules lead to stable state models, (2) how readily these rules lead to full cascade models and (3) the speed at which (1) and (2) occur.

Definition 18 (Simulation Rules) *Let $M = \langle A, O, \beta, F \rangle$ be an SNM. let $M^1 = \langle A, O, \beta^N, F^E \rangle$ be M after applying R1. Let $M^2 = \langle A, O, \beta^N, F^Y \rangle$ be M after applying R2. Let $M^3 = \langle A, O, \beta^N, F \rangle$ be M after applying R3.*

R1 updates the network M by applying extended adoption (17 and extended friendship selection (15) simultaneously. R2 updates the network M by applying extended adoption (17 and n-friendship selection (15) simultaneously. R3 is simply the application of extended adoption in 17. We will apply each of the 3 rules independently to each of the randomly generated networks. We will collect data on the number of steps taken until the network either stabilised or is terminated after 50 steps. The data will be aggregated to compare the difference in average steps between the 3 rules. The code for the simulation was designed in Python and can be found here.

5.1 Random Networks

Networks are generated randomly using the Erdős–Rényi model described in Gilbert (1959). We use the networkx package in python to perform this. It implements the $G(N, P_{edge})$ model that generates a random network G of N nodes with each edge in the graph having a probability P_{edge} of forming in G_0. For $G(N, 0.1)$ each node has an independent 10 percent chance of having an edge with all other nodes. As P_{edge} approaches 1, the probability that G has complete edges approaches 1. Likewise as P_{edge} approaches 0, the probability of G being an empty graph approaches 0. We must highlight that the time it takes to generate a random graph G and subsequently perform updates on G increases exponentially with the number of nodes. For any $n \in N$ there are $N-1$ possible links, meaning for all nodes there are $N*(N-1)$ possible links in the network.

5.2 Network Values

We will generate 1000 random networks $G(N, P_{edge})$ for each size $N \in \{10, 20, 50, 100\}$ with constant $P_{edge} = 0.1$. Each simulation run will have the same identical factors:

1. $P_{adopt} = 0.2$
2. $\theta_a = 0.3$
3. $\theta_E = 0.5$
4. $O = \{A, B, C, D\}$

P_{adopt} declares a 20 percent chance for a node to be an initial adopter for each behaviour. On average for each behaviour 20 percent of the population will be initial adopters. This value does not change the behaviour of the network but rather the generation of them. A higher P_{adopt} will create networks where the diffusion is closer to a terminal state. The opposite it true for lower values. It is important that this value supports the adoption threshold. Small values of P_{adopt} will have small populations of initial adopters. Since they are assigned randomly there is a higher probability that these initial adopters don't have the necessary size to cause a cascade and the behaviour instead dissipates. We are interested in studying full cascade networks. P_{adopt} was sufficiently high enough to allow nodes in the initial network to begin adopting.

This was also the reason for the adoption threshold $\theta_a = 0.3$. This value was the maximum value at which for this distribution of initial adopters, behaviours would only dissipate in rare conditions. It defines that 30 percent of an agent's neighbours must adopt a behaviour before they do so.

$\theta_E = 0.5$, the friendship threshold, defines that two agents must coincide in half of their behaviours for them to connect. Finally, O is simply the set of behaviours diffusing in the network, here denoted arbitrarily as A, B, C and D. For this number of behaviours there are only 3 meaningful intervals of similarity between nodes; $\frac{1}{4}$, $\frac{2}{4}$, $\frac{3}{4}$ i.e. quarter intervals. These intervals are what determines behaviour not the number of behaviours i.e. 8 behaviours will have more intervals (eighths) but the behaviour of the network at the half interval threshold $\frac{4}{8}$ would mirror that of $\frac{2}{4}$. The half threshold was chosen as it was high enough that networks wouldn't form complete networks in too few steps but not high enough that networks would become empty with R2. There is a trade-off in the number of behaviours between computational complexity and data. 4 behaviours was the compromise that provided data in reasonable time. For the same reason we restrict network sizes to a maximum of 100 nodes.

6 Results

We will first look at the data in aggregate meaning networks either stabilised or stopped simulating after 50 steps. Then restrict the data to only networks that stabilised and finally networks that stabilised and had a full cascade. For each of these conditions we present a pair of graphs that display (1) the average number of steps until stabilisation (or stoppage) and (2) the difference in the average number of steps between each rule. The differences between rules are calculated as follows.

$$\Delta_{orange} = \bar{R}1 - \bar{R}2$$
$$\Delta_{green} = \bar{R}1 - \bar{R}3$$
$$\Delta_{blue} = \bar{R}2 - \bar{R}3$$

Deltas can be negative or positive, the sign indicates which rule had fewer steps until stabilisation or stoppage i.e. $+\Delta_{orange}$ for stable network data indicates that $\bar{R}1 > \bar{R}2$ and therefore R2 took fewer steps to stabilise a network. For example consider $\Delta_{orange} = -1$ in networks that stabilised. This shows that R1 on average takes 1 less step than R2 to stabilise a network. Conversely, $\Delta_{orange} = +1$ shows R1 on average takes 1 more step to stabilise than R2 i.e. R1 is slower than R2 by one step.

6.1 Stable or Stoppage

Figure 6.1 displays the mean number of steps until the network either stabilised or was terminated after 50 steps. Observations worth highlighting in

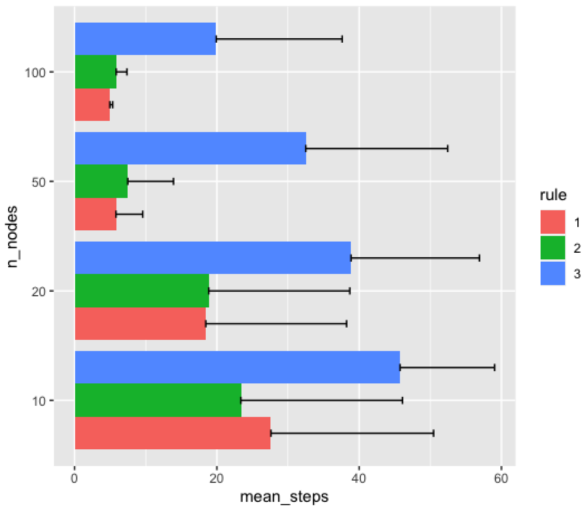

Figure 6.1: Average steps until stabilisation or stoppage after 50 steps for all network sizes. Y-axis displays the network size, the x-axis displays average steps value and the rules are categorised by colour; R1 (orange), R2 (green), R3 (blue). Standard deviations are also shown.

Figure 6.1. First, all rules decrease in steps as network size increases i.e. steps and network size appear inversely related. Significant decreases happen at the boundary between 20 and 50 node networks for R1 and R2. R3 has a consistent decrease in steps across all network sizes. Second, variance in stabilisation decreases significantly with network size for R1 and 2 but not for R3. This suggests that R3 networks do not readily reach a steady state at all sizes and the manipulation of links increases the probability that a network will reach a steady state. Figure 6.2 further clarifies the disparity in stabilisation between rules. Clearly, R1 and 2 differ insignificantly in comparison to how they differ to R3. Only with 10 node networks does there appear to be a significant difference between rules 1 and 2. Note 100 node networks. The step difference between R1-R3 and R2-R3 decreases significantly.

We conjecture that due to the size of the network, the number of edges in the network at $n = 0$ are sufficient to decrease the effect of edge manip-

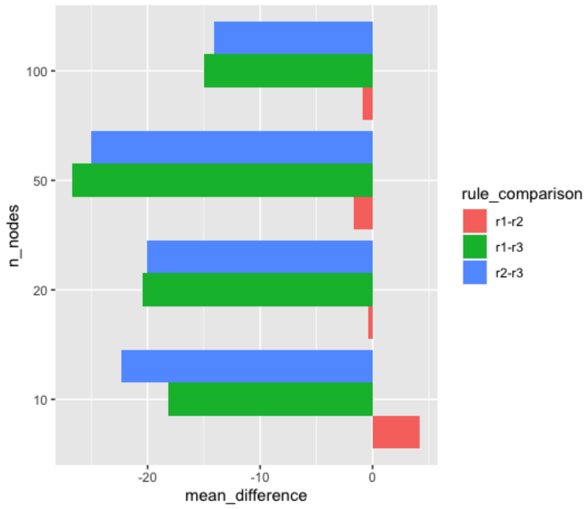

Figure 6.2: Deltas of stabilisation or stoppage. Y-axis displays network size. X-axis displays delta values. Comparisons are categorised by colour; R1-R2 (orange), R1-R3 (green), R2-R3 (blue).

ulation on the speed of stabilisation. Recall that the number of edges in the network increases exponentially as network size increases. This means that larger networks are generally more interconnected than small networks even though there are more nodes. As such, proportionally, new links have a smaller effect on the process of diffusion. We explore these notions further by running pairwise t-tests between rules with the bonferroni correction. The difference in steps between R1 and 2 are insignificant ($p = 0.84 < 0.05$). However, for all comparisons of R3, there are significant differences ($p < 2e - 16 < 0.05$).

6.2 Stable

We are also interested in relationship between rules when the networks stabilise. Figure 6.3 displays the mean number of steps until the network stabilised. What is immediately apparent in contrast to what was found in aggregate is that the mean number of steps does not decrease as network size increases. This suggests that the high number of average steps for small networks in aggregate was skewed by the inclusion of networks that never reached a steady state. In other words, the previously suggested in-

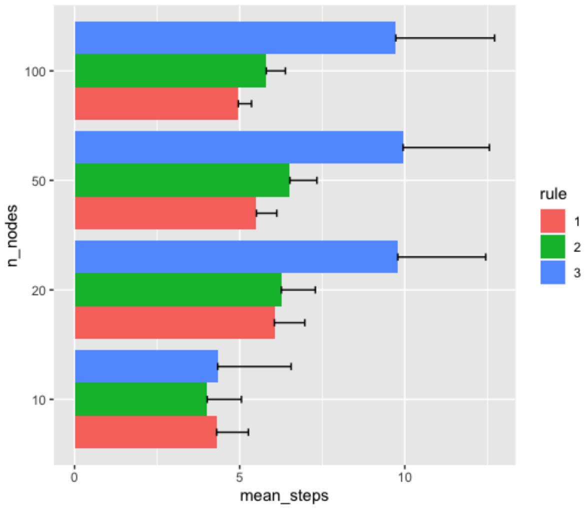

Figure 6.3: Average steps until stabilisation. Y-axis displays network size. X-axis displays average step values. Rules are categorised by colour; R1 (orange), R2 (green), R3 (blue).

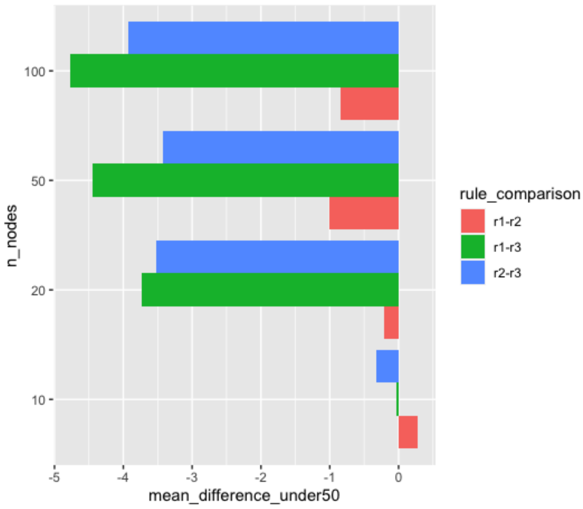

Figure 6.4: Deltas of stabilisation. Y-axis displays network size. X-axis shows delta values. Comparisons are categorised by colour; R1-R2 (orange), R1-R3 (green), R2-R3 (blue).

verse relationship between network size and steps resulted only from networks that never reached a stable state. The mean steps increases from 10 node networks to 20 for all rules. However this increase does not happen for all other network sizes. Rule 3 performs similarly to rules 1 and 2 in 10 node networks whereas before it was significantly slower. Otherwise, there is a significant difference between mean steps for each rule. We perform a two-way anova to verify both the significance of the rule and network size as predictors of the number of steps until stabilisation. The results show that both network size and rule are significant predictors ($p < 2e-16 < 0.05$). We run the same pairwise t-test with bonferroni correction for stable network data and find that there is a significant difference in steps between rules 1 and 2 ($p < 2e-16 < 0.05$). This is in contrast to the aggregate data wherein the difference in rules 1 and 2 was insignificant. This is visualised in figure 6.4. Except for 10 node networks, the difference between rules 1 and 2 is more pronounced. 10 node networks appear to be anomalous in that rules 1 and 2 perform similarly to rule 3. This supports our conjecture that link formation increases the speed of diffusion in networks only when the longest path between an adopted node and non-adopter is greater than 2. However, it is also possible that the decrease in steps results from behaviours being dropped from the network. There are so few nodes and edges that adoption is less likely to occur in these networks even at low thresholds. As such, these stable states stem from the abandonment of behaviours rather than the adoption. Additionally, only in 10 node networks we see rule 2 stabilises faster than rule 1. The removal of links may lead to 10 node networks becoming fully disconnected such that nodes cannot influence each other or form new links. Other than for 10 node networks, the relationship between rules is similar to those shown in figure 6.2.

6.3 Stable and Full Cascade

We restrict the data further to the networks that both stabilised and had a full cascade. Figure 6.6 shows the deltas in steps in networks that both stabilised and had a full cascade and figure 6.5 shows the mean steps until stabilisation in these networks. It is important to note that this restriction shrank the data for both 10 and 20 node networks significantly. This suggests that networks of those sizes rarely reached a full cascade. This likely results from the sparse distribution of initial adopters and edges in comparison to larger networks. In 100 node networks there are 20 initial adopters per behaviour, whereas in 10 node networks there are only 2. Considering the adoption threshold $P_{adopt} = 0.3$ the existence of 2 adopters has fewer networks where there exist node(s) with a sufficient proportion of neighbours to cascade a behaviour. Except for 10 node networks, the data is strikingly similar to our analysis of figure 6.3 and 6.4.

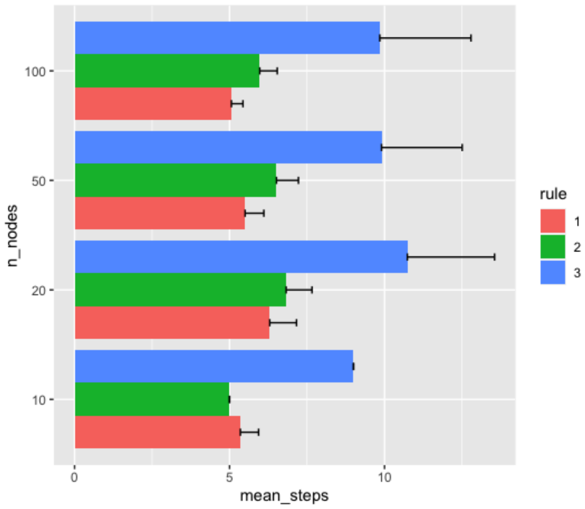

Figure 6.5: Average steps until stabilisation in full cascade networks. Y-axis shows the network size. X-axis shows mean step values. Rules are categorised by colour; R1 (orange), R2 (green), R3 (blue).

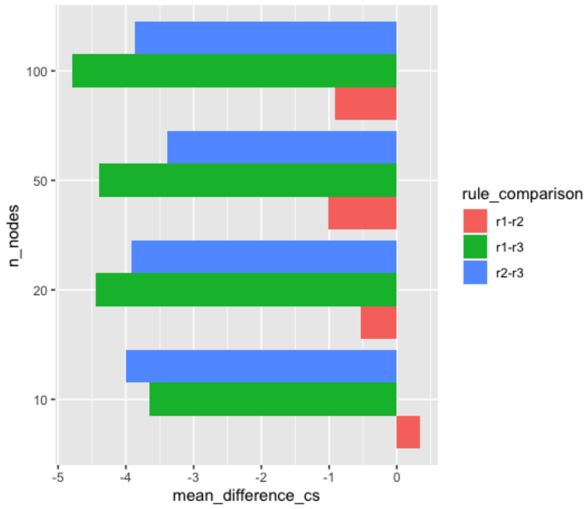

Figure 6.6: Deltas of stabilisation in full cascade networks. Y-axis shows the network size. X-axis shows delta values. Comparisons are categorised by colour; R1-R2 (orange), R1-R3 (green), R2-R3 (blue).

7 Conclusion

In this paper we explored different updates on threshold models. These updates revolved around social influence and selection and their relationship. First we defined social influence that allowed agents to learn new behaviours and social selection that connected nodes that were related to similar agents. It was found that this form of social selection can lead to faster cascades of information through a network. However we conjecture that this phenomenon is only the case under two conditions; first, only networks in which the social selection doesn't prevent information from spreading and second only if there is a significant distance between an agent that has adopted a behaviour and an agent that hasn't. Next it was shown that unique sequences of network updates involving social influence and selection can lead to the same final model.

Then we extended these definitions of influence and selection to allow nodes to abandon behaviours as well as abandon connections. Furthermore, the redefined social selection required that the nodes had enough similar behaviours to be connected. This allowed agents to evaluate their similarity to each other and conversely decide if they are similar enough to remain connected. These are more complex updates and were explored empirically by building a simulation in python. Data from the simulations lead to the following observations. In networks that only featured the redefined influence update networks did not readily reach stable states. Selection lead to more stable networks. There was a significant difference in network stabilisation when agents could and could not remove connections from their networks. Finally, network size is a significant predictor for network behaviour, however there is not a clear trend presented on how these are interrelated.

In the properties of the model outlined in section 5, variation in both θ_a and θ_E have significant impacts on the final network. Clearly, increasing θ_a increases the resistance towards a diffusing behaviour. Interestingly, there was a soft boundary between 0.3 and 0.4 for the adoption threshold with all things equal where behaviours would begin being collectively unlearnt. All adoptions would trend towards zero adopters.

There are many ways in which further research can adapt this study. We used random network generation, however there are many other types of graphs that can be explored with this framework. For example, small world networks (Watts & Strogatz, 1998) have been shown to be a more realistic distribution of populations. Nodes are grouped according to neighbour clusters, that most nodes are not neighbours but the neighbours of a node are highly likely to be each others neighbours.

Additionally we used a homogeneous adoption threshold, however there is emerging research on the idea of heterogeneous thresholds that assigns a threshold to each node based on some distribution of values e.g. (Ryan & Tucker, 2012), (Bauer & Hein, 2006), (Chatterjee & Eliashberg, 1990). This is to show that different entities in a network have different standards of adoption.

Variation in the adoption threshold used to simulate these results can lead to wildly different findings. Our goal with the simulation was to evaluate in practise, the environment required for information to fully cascade. It was important for our experiment that a behaviour be allowed to cascade through a network with adoption taking place more often than abandonment. This motivated these selections of network factors, however it is possible that there exists a different combinations of values that leads to only full cascades. This is especially important for 10 and 20 sized networks where the scarcity of edges and nodes lead to more cases of cyclic or empty networks where the number of adopters was reduced to zero.

The nature of friendship selection in this experiment is a combination of trait and network relationships. However, network distance became increasingly less relevant as network size grew. Longest paths of networks began converging. As network size grows, edges grow exponentially. The influence of a single node in the network increases exponentially with edges as they can influence more nodes. This helps explain why R1 and R2 did not find difficulty in converging to stable states. Many cyclic relationships are closed by triangular closure.

There are edges that connect nodes to each other symmetrically. With social media there has been more emphasis on exploring the idea of asymmetrical edges where one individual influences the other but not the other way around (Hangal et al., 2010). This paper will focused solely on symmetric edges as these edges exist more commonly in close interpersonal relationships that require both parties to participate for the connection to exist. However, further research can explore these updates with asymmetry by including a weighting factor to each edge. These would influence the strength of influence from one node to another.

Further research should assess the updates proposed in this paper with different threshold values to study how these rules lead to behaviours fully disappearing from a network i.e. instead of full cascade, all nodes in the network abandon some behaviour. It is not only important to understand how information spreads but also under what conditions it won't.

References

Aiello, L. M., Barrat, A., Schifanella, R., Cattuto, C., Markines, B., & Menczer, F. (2012). Friendship prediction and homophily in social media. *ACM Transactions on the Web (TWEB)*, *6*(2), 1–33.

Baltag, A., Christoff, Z., Rendsvig, R. K., & Smets, S. (2019). Dynamic epistemic logics of diffusion and prediction in social networks. *Studia Logica*, *107*(3), 489–531.

Bauer, K., & Hein, S. E. (2006). The effect of heterogeneous risk on the early adoption of internet banking technologies. *Journal of Banking & Finance*, *30*(6), 1713–1725.

Chatterjee, R. A., & Eliashberg, J. (1990). The innovation diffusion process in a heterogeneous population: A micromodeling approach. *Management science*, *36*(9), 1057–1079.

Christoff, Z., Baccini, E., & Verbrugge, R. (2022). Opinion diffusion in similarity-driven networks.

In *Logic and the foundations of game and decision theory* (Vol. 14).

Das, S., & Biswas, A. (2023). The ties that matter: From the perspective of similarity measure in online social networks. In *Machine learning, image processing, network security and data sciences: Select proceedings of 3rd international conference on mind 2021* (pp. 647–658).

Dodds, P., & Watts, D. J. (2011). Threshold models of social influence.

Easley, D., & Kleinberg, J. (2010). *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge university press.

Gilbert, E. N. (1959). Random graphs. *The Annals of Mathematical Statistics*, *30*(4), 1141–1144.

Granovetter, M., & Soong, R. (1983). Threshold models of diffusion and collective behavior. *Journal of Mathematical sociology*, *9*(3), 165–179.

Hangal, S., MacLean, D., Lam, M. S., & Heer, J. (2010). All friends are not equal: Using weights in social graphs to improve search. In *Workshop on social network mining & analysis, acm kdd* (Vol. 130).

Jeh, G., & Widom, J. (2002). Simrank: a measure of structural-context similarity. In *Proceedings of the eighth acm sigkdd international conference on knowledge discovery and data mining* (pp. 538–543).

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual review of sociology*, 415–444.

Ryan, S. P., & Tucker, C. (2012). Heterogeneity and the dynamics of technology adoption. *Quantitative Marketing and Economics*, *10*, 63–109.

Sarwar, B. M., Karypis, G., Konstan, J., & Riedl, J. (2002). Recommender systems for large-scale e-commerce: Scalable neighborhood formation using clustering. In *Proceedings of the fifth international conference on computer and information technology* (Vol. 1, pp. 291–324).

Schelling, T. C. (1969). Models of segregation. *The American economic review*, *59*(2), 488–493.

Smets, S., & Velázquez-Quesada, F. R. (2019). A logical analysis of the interplay between social influence and friendship selection. In *International workshop on dynamic logic* (pp. 71–87).

Smets, S., & Velázquez-Quesada, F. R. (2020). A closeness-and priority-based logical study of social network creation. *Journal of Logic, Language and Information*, *29*(1), 21–51.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world'networks. *nature*, *393*(6684), 440–442.