Cambridge University
Statistical Laboratory Research Report 2005-11

The Impact of Liars on Reputation in Social Networks

Jochen Mundinger[*] Jean-Yves Le Boudec[†]

June 15, 2005

Abstract

In this paper we consider a closed social network with a certain proportion of liars who are trying to influence their peers' reputation about some subject. Each person's reputation about this subject is based on both own direct experience and second hand information from their peers which cannot be verified. Given certain assumptions on when people believe or do not believe second hand information, we investigate the liars' impact on their peers' reputation about the subject. We present a mathematical model for this situation and show that there is a threshold proportion of liars below which they have no impact. Above it, liars do have an impact. We quantify this impact and give the threshold proportions. We compare our results in two fundamentally different scenarios: In the first one, reputation is passed on as second hand information. In the second one, direct experience only is passed on as second hand information. We find that in the latter scenario liars have less impact.

Keywords: Social networks, reputation, liars, model, phase transition

1 Introduction

In this paper we consider a dense, closed social network. By this we mean that everyone in it is connected to everyone else by similarly strong relationships. People in the network are assumed to take an interest in the behaviour of some subject which can be either positive or negative. They interact with this subject directly. They also interact with each other, e.g. in conversations, and thereby pass on their own experiences with the subject to their peers. Based on both direct experience and indirect information they form their reputation about the subject. An example is the social network of truck drivers interested in the quality of food of a highway restaurant. Alternatively, the subject might be part of the social network itself and there might be more than one subject. This is the case when people in the network gossip about each other.

Sharing experiences with one's peers serves the purpose of using information more efficiently: By also considering other people's experiences, one is able to get a more accurate idea about the actual subject behaviour faster. However, it might be the case that not every peer in the social network passes on their experiences with the subject truthfully. There might be liars. In the absence of trust, these might have have a detrimental effect on the overall reputation of the subject in the network.

[*] Statistical Laboratory, Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WB, UK
[†] EPFL-IC-LCA, CH-1015 Lausanne, Switzerland

So the question arises whether second hand information is or should always be believed. We assume that this is not so. Rather, if a person is confronted with information that is not verifiable, they will probably believe it only if, to them, it seems likely. However, they will ignore it if, to them, it seems unlikely. Moreover, they will not necessarily attach the same weight to an experience reported by a peer compared to their own direct experiences. We also assume that people gradually forget experiences they have made a long time ago and that in the current reputation about the subject recent experiences are given greater impact as a result.

We present a mathematical model for this situation and analyze the impact that liars can have when they consistently report either negative or positive behaviour about the subject in an attempt to influence the overall reputation of the subject. We show that there is a threshold proportion of liars below which they have no impact. Above it, liars do have an impact. We quantify this impact and give the threshold proportions.

We compare the liars' impact in two fundamentally different scenarios: In the first one, reputation, which is based on all their experiences including indirect ones, is passed on as second hand information. In the second one, only direct experience is passed on as second hand information. We find that in the latter scenario liars have less impact.

The idea of reputation also plays an important role in Peer-to-Peer communication systems where one often encounters free-riding, a well-known problem in economics [Sam54]. That is, users consume without contributing which leads to a loss in performance. Reputation systems are one of the approaches that have been introduced to solve the free-rider problem in Peer-to-Peer communication systems. A popular example of a reputation mechanism is the rating used in EBAY [RZ02]. However, this is based on a centralized mechanism, which is not appropriate for Peer-to-Peer communication systems.

In a decentralized reputation system users keep track of their peers' behaviour and exchange this information with others. Each user merges their own first hand information with the second hand information they receive in order to compute a reputation value about each of their peers. This might be an automated procedure. Users with a good reputation are then favoured, thus providing an incentive to behave well. However, reputation system must be robust against liars. A simple idea to address this problem was suggested originally in the context of Mobile Ad-Hoc Communication Networks by Buchegger and Le Boudec [BLB04]. Here, a user believes second hand information only if it does not differ too much from the user's reputation value. This is called the deviation test. In fact, the authors considered a more complex system that also allows the use of second hand information from trusted peers, but in simulations the deviation test on its own was found to perform surprisingly well.

It has since been analyzed in detail in the case of 2 users, one honest the other a liar. In [MLB05b] we considered a model that simplifies the reputation counters in order to obtain a one-dimensional system which is easier to analyze. In [MLB05a] we then looked at the original two-dimensional system. In the context of social networks, the interpretation is now different, but the analytical results of the present paper can be viewed as a generalization to a network of N people of our results in [MLB05a].

Similar questions have also been considered in the context of social networks. Moerbeek and Need [MN03], for example, examine to what extent foes deteriorate a person's labor market position. In comparison to their work, we address a more general question, not specific to the labor market context. Another difference is in our approach. We do not collect and analyze a data set. Instead, we specify a set of assumptions that we consider reasonable.

On the basis of this, we build and analyze a mathematical model. Other authors have also used mathematical techniques in social networks. Barnes et al. [BCL98], for example, model interpersonal relationships using algebraic semigroups.

Burt [Bur01] investigates how the competitive advantage known as social capital depends on the structure of the social network. He also focuses on closed networks rather than networks of interdependent groups (brokerage, cf. also [Bur99]) and evaluates two competing hypotheses. Firstly, the so-called bandwidth hypothesis that network closure enhances information flow. This is found in closure models of social capital and as well as in reputation models in economics. Secondly, the so-called echo hypothesis that closure models merely create an echo that reinforces predispositions and leads to ignorant certainty. This is found in the social psychology of selective disclosure. Evidence considered in the literature as well as in [Bur01] supports echo over bandwidth. Bandwidth and echo models represent a fundamental choice for theoretical models of trust. The lying in our model can be interpreted as selective disclosure in this context. We shall see that passing on reputation as second hand information more likely creates an echo scenario. Passing on only direct experience more likely creates a bandwidth scenario.

The rest of this paper is structured as follows. The precise modeling assumptions are listed in Section 2. In Section 3 we introduce our model and investigate the liars' impact in the scenario where reputation is passed on as second hand information. In Section 4 we repeat this analysis in the scenario where direct experience only is passed on. Further details and the simulation results confirming our analysis for both these scenarios can be found in the appendix. An interpretation of our results and directions for further work are given in Section 5.

2 Modeling assumptions

2.1 Subject behaviour

We study the case when there is a single subject whose reputation is considered. Its actual behaviour is assumed to be either positive or negative with probabilities θ and $1-\theta$ respectively. Thus, when a person interacts with the subject itself it experiences positive behaviour with probability θ and negative behaviour otherwise. This is assumed to be independent of all other experiences. Hence the actual behaviour is represented by the parameter θ, a real number in $[0, 1]$.

This subject might or might not be one of the N people in the network. People in the network might also be interested in the behaviour of some external subject such as the quality of service of a restaurant.

The case when when there are M subjects of interest can be decomposed into M separate instances of our model. The M sets of reputation values do not interfere with each other and can be considered independently. In particular, we might take N subjects, one for each person in the network. This is the case when people in the network gossip about each other.

2.2 Reputation

The total of N people splits into N_h honest people $1, 2, \ldots, N_h$ and N_l liars $N_h + 1, \ldots, N$. They have corresponding reputation values R^i about the subject. Notice that liars can change their own opinion anyway they want. The question is whether they can influence the reputation values of the honest people, so these are the ones that we will consider.

The reputation values are also real numbers in $[0, 1]$ and reflect the belief that person i has about the actual subject behaviour θ at a given time. This opinion might change with new information obtained from interactions with the subject itself or with a peer in the network.

Each person i in the network remembers both positive and negative information. (x^i_n) and (y^i_n) are the components of the reputation counter that keep track of positive experiences and negative experiences respectively. The actual reputation value R^i_n is then computed as the proportion $x^i_n/(x^i_n + y^i_n)$.

A direct (first hand) experience is an experience of the subject's behaviour. People always believe their own direct experiences so that reputation values are updated accordingly. An indirect (second hand) experience arises from interactions with peers. A person believes an indirect experience only if the reported information does not deviate too far from their current reputation value R^i_n. This behaviour is controlled by the threshold parameter $\Delta \in (0, 1)$. A large Δ means that this person naively believes most second hand information. A small Δ means a more critical attitude. Even if believed, the indirect experience does not necessarily have the same impact as a direct experience. Thus, it is scaled by a weighting parameter $\omega_{weight} > 0$. See Table 1 for a summary of the notation.

There are two fundamentally different possible assumptions as regards to reports of second hand information by honest people. They might report their current reputation values and thus their actual opinion based on all their experiences including indirect ones. Or they might report their own direct experiences only. Both are acceptable assumptions and it will be interesting to see how exactly they compare (cf. Sections 3 and 4).

Moreover, we assume a discount factor $0 < \rho < 1$, typically very close to 1. Discounting takes place whenever there is a new experience. This it to account for people gradually forgetting experiences they have made a long time ago. As a result, recent experiences are given greater impact. Discounting also allows for people to adapt to changes in the subject's behaviour.

2.3 Interaction

The interaction model describes how people interact with the subject and their peers in the network. The idea here is that person i meets or has a conversation with the subject or a peer j in the network and thus receives new information. We shall assume that each person i makes direct experiences at the points of a Poisson process in time, at rate $\lambda_d > 0$. Interactions of person i with j (such that i receives second hand information from j) occur according to a Poisson process with rate $\lambda_i > 0$. All processes are assumed to be independent. Without loss of generality, we might assume $\lambda_d = 1$ by rescaling time.

Thus, each person makes experiences at the points of a Poisson process with rate $\lambda = \lambda_d + \lambda_i$. The subscripts n in x^i_n, y^i_n and R^i_n are thus to be interpreted as the jump times T^i_n of a Poisson process at rate λ. We shall often write n instead of T_n for ease of notation.

A given event is a direct experience with probability $p = \lambda_d/(\lambda_d + \lambda_i)$. It is an indirect experience from a liar with probability $q = \lambda_i N_l/(\lambda_d + \lambda_i)(N_h + N_l - 1)$ and from an honest peer with probability $r = \lambda_i(N_h - 1)/(\lambda_d + \lambda_i)(N_h + N_l - 1)$, all independent of other events.

Although the interaction pattern might differ between scenarios, the model above is a natural one to examine. It assumes a symmetric closed network, that is, people interact with one another equally frequently. Social Networks often have a much more asymmetric,

symbol	meaning
θ	probability of positive subject behaviour
N_h	number of honest people
N_l	number of liars
λ_d	direct experience rate of person i for all $i = 1, 2, \ldots, N_h$
λ_i	indirect experience rate of person i from j for all $j = 1, 2, \ldots, N$
p	probability of an experience being direct
q	probability of an experience being indirect from a liar
r	probability of an experience being indirect from an honest peer
$x_n = x_{T_n}$	positive component of reputation counter
$y_n = y_{T_n}$	negative component of reputation counter
$R_n = R_{T_n}$	reputation value of honest person after n^{th} interaction
x_0, y_0, R_0	initial values of reputation counters and values
Δ	threshold parameter for indirect experiences
ω_{weight}	weighting factor attached to indirect experiences
ω_{max}	maximal impact of a second hand report
ω	$\omega_{weight}\omega_{max}$
ρ	discount factor

Table 1: Summary of notation.

complex structure. However, one has to start with a model that is tractable. Moreover, this model is quite appropriate for certain scenarios, e.g. online communities.

2.4 Liars

We have not yet specified what exactly to expect of the liars. We shall assume that liars always lie maximally, i.e. they will always report either extremely negative or extremely positive behaviour about the subject when interacting with their peers. They do so in attempt to achieve maximal impact. In fact, we focus on the extremely negative part, as the other one is similar by symmetry.

3 The liars' impact when reputation is passed on

In this section we analyze the liars' impact in the scenario where reputation values are passed on as second hand information, that is the actual opinion based on all previous experiences including indirect ones. The other scenario will be investigated in Section 4.

3.1 Model formulation

For each person $i = 1, 2, \ldots, N_h$ we consider the two-dimensional process formed by the positive and negative component of the reputation counters $(x^i_{T^i_n}, y^i_{T^i_n})$ as given in (1) for $n \geq 0$. The T^i_n are the jump times of the Poisson process associated with person i. The four possible cases correspond to a positive direct experience, a negative direct experience and an indirect experience from a liar or from an honest peer respectively (cf. Sections 2.1 and 2.3).

$$\begin{aligned}(x^i_{T^i_{n+1}}, y^i_{T^i_{n+1}}) &= \rho(x^i_{T^i_n}, y^i_{T^i_n}) \\ &+ \begin{cases}(1,0) & \text{w.p. } p\theta \\ (0,1) & \text{w.p. } p(1-\theta) \\ \omega_{weight}(0,\omega_{max})\mathbf{1}\left\{\frac{x^i_{T^i_n}}{x^i_{T^i_n}+y^i_{T^i_n}} \leq \Delta\right\} & \text{w.p. } q \\ \omega_{weight}(.,.)\mathbf{1}\left\{\left|\frac{x^i_{T^i_n}}{x^i_{T^i_n}+y^i_{T^i_n}} - \frac{a}{a+b}\right| \leq \Delta\right\} & \text{w.p. } r\end{cases}\end{aligned} \tag{1}$$

Upon each interaction, both components are discounted individually. Whereas direct experiences are always believed and counted with 1, indirect experiences are only believed if they do not deviate from the person's reputation value by more than Δ. This is modeled by the indicator function. They are also weighted by ω_{weight} as described in Section 2.2. We assume that a second hand report is bounded by some maximal $\omega_{max} > 0$ on the sum of the components. This is to ensure that a liar cannot simply report a very large number of interactions to in order to increase impact. The initial reputation counter is (x_0, y_0). The quantity we are interested in is $R_n = x_n/(x_n + y_n)$, in some sense the proportion of positive experiences. We examine how well this compares to the true θ, that is the actual proportion of positive behaviour of the subject in question.

We still need to specify the increment $(.,.)$ in the fourth case of a second hand report (a, b) from an honest peer. Essentially, there are three ways. Namely,

$$(a,b),\ \frac{1}{a+b}(a,b) \text{ or } \frac{1}{1+a+b}(a,b) \tag{2}$$

all with the same reputation value $a/(a + b)$. The latter two will be relevant in the next section with direct experience, because in that case scaling is necessary to ensure finiteness. For our current scenario with reputation, however, the first alternative is the most suitable one. Scaling is not necessary, because reputation counters are known to lie in a bounded region if ω_{weight} is not too big.

Let us assume that the componentwise sum of the reported reputation counters does not exceed $1/(1 - \rho)$. Reputation counters outside this region are considered invalid. Thus, $\omega_{max} = 1/(1 - \rho)$. Choosing $\omega_{weight} \leq 1 - \rho$ now ensures that reputation counters do not blow up. In the formulation below we have written the process in terms of $\omega = \omega_{weight}\omega_{max}$, ranging from 0 to 1. This is done in a way compatible with our assumptions in the previous section.

$$\begin{aligned}(x^i_{T^i_{n+1}}, y^i_{T^i_{n+1}}) &= \rho(x^i_{T^i_n}, y^i_{T^i_n}) \\ &+ \begin{cases}(1,0) & \text{w.p. } p\theta \\ (0,1) & \text{w.p. } p(1-\theta) \\ (0,\omega)\mathbf{1}\left\{\frac{x^i_{T^i_n}}{x^i_{T^i_n}+y^i_{T^i_n}} \leq \Delta\right\} & \text{w.p. } q \\ \omega(1-\rho)(x^j_{T^j_n}, y^j_{T^j_n})\mathbf{1}\left\{\left|\frac{x^i_{T^i_n}}{x^i_{T^i_n}+y^i_{T^i_n}} - \frac{x^j_{T^j_n}}{x^j_{T^j_n}+y^j_{T^j_n}}\right| \leq \Delta\right\} & \text{w.p. } r\end{cases}\end{aligned} \tag{3}$$

From our interaction model it follows that at each time T^i_n where there is an indirect experience from an honest peer j, that person j is chosen uniformly at random from the honest people other than i. Reputation counters only change with interactions, i.e. $(x_t^j, y_t^j) = (x_{T_n^j}^j, y_{T_n^j}^j)$ for $T_n^j \leq t < T_{n+1}^j$.

3.2 Mean-field approach

Instead of considering the honest people's reputation counters individually, we will consider their average. Let T_n denote the jump times of the Poisson process associated with the average reputation. That is, the set of these jump times is the union of the sets of jump times of the individual people's processes. The average reputation counter is then given by

$$(x_{T_n}, y_{T_n}) = (1/N_h) \sum_{i=1}^{N_h} (x_{T_n}^i, y_{T_n}^i). \tag{4}$$

This can be interpreted as the overall reputation of the subject in the system. Moreover, we make the (strong) assumption that all honest people's reputation counters are equal and hence the average. We will judge this assumption by how well the theoretical predictions will match the simulation results.

We now consider the following mean deterministic differential equation for the average reputation counter.

$$\begin{aligned} \dot{x}(t) &= -\lambda(1 - r\omega)(1 - \rho)\, x(t) + \lambda p\theta \\ \dot{y}(t) &= -\lambda(1 - r\omega)(1 - \rho)\, y(t) + \lambda p(1 - \theta) + \lambda q\omega \mathbf{1}\left\{\tfrac{x(t)}{x(t)+y(t)} \leq \Delta\right\} \end{aligned} \tag{5}$$

This can be obtained from (3) via a fast-time scaling and by means of averaging the dynamics as shown in Appendix A.1. The system is discontinuous, but linear above and below the line of discontinuity $x(t)/(x(t) + y(t)) = \Delta$. We can solve it separately on each region. The exponential decay term is $e^{-\lambda(1-\rho)(1-r\omega)}$, thus the speed of convergence decreases in r.

$$(x, y) = \frac{1}{(1 - r\omega)(1 - \rho)} (p\theta, p(1 - \theta)) \tag{6}$$

is a fixed point if $\Delta < \Delta_{c_4} = \theta$. If it exists, it is asymptotically stable and trajectories from $x(t)/(x(t) + y(t)) > \Delta$ are attracted to it. The corresponding reputation value is $R_1^* = \theta$.

$$(x, y) = \frac{1}{(1 - r\omega)(1 - \rho)} (p\theta, p(1 - \theta) + q\omega) \tag{7}$$

is a fixed point if $\Delta \geq \Delta_{c_1} = (p\theta)/(p + q\omega)$. If it exists, it is asymptotically stable and trajectories from $x(t)/(x(t) + y(t)) \leq \Delta$ are attracted to it. The corresponding reputation value is $R_3^* = \theta p/(p+\omega q)$. If only one of the two fixed points exists then the trajectories from the other region lead into its region and thus are also attracted to it. That is all trajectories are attracted to it. Otherwise, both are asymptotically stable on their respective region. Thus, we have the following result.

Theorem 1 *If $\Delta < \Delta_{c_1} = p\theta/(p + q\omega)$,* (6) *is the unique fixed point of the mean differential equation (5). It is asymptotically stable and all trajectories are attracted to it. If $\Delta_{c_1} \leq \Delta < \Delta_{c_4} = \theta$ there is a second, false fixed point* (7) *and both are asymptotically stable, attracting trajectories from $x(t)/(x(t) + y(t)) > \Delta$ and $x(t)/(x(t) + y(t)) \leq \Delta$ respectively. Finally, if $\Delta_{c_4} \leq \Delta$, then only the latter, false one is asymptotically stable and all trajectories are attracted to it.*

Note that it is the ratio of p and ωq and only this ratio that is important for the false fixed point.

As a result, the differential equation for the reputation system exhibits a phase transition behaviour. We have phrased this in terms of the threshold Δ. This is also visualized in blue in Figure 4.2: For sufficiently small Δ, liars do not have any impact. Only for intermediate values, liars have some impact. For large Δ liars have maximal impact. They brainwash everyone in the network.

These results can also be stated in terms of the lying probability q. Liars do not have any impact if

$$q < \frac{1}{\omega}\frac{\theta - \Delta}{\Delta} p \tag{8}$$

Otherwise, they have some impact and if moreover $\Delta < \theta$ liars have maximal impact and brainwash everyone in the network.

Under the assumption of a symmetric social network we can rephrase these results once more in terms of the proportion of liars in the network. Liars do not have any impact if their proportion is below some threshold. Otherwise, they have some impact. If moreover $\Delta < \theta$ liars have maximal impact and brainwash everyone in the network.

In Appendix B.1 we provide simulation results confirming the analytical results of this Section.

4 The liars' impact when direct experience only is passed on

In the previous section we analyzed the liars' impact in the scenario where reputation values are passed on as second hand information. We now consider the scenario where only direct experience is passed on.

4.1 Model formulation

For each $i = 1, 2, \ldots, N_h$ we consider the two-dimensional process $(x^i_{T^i_n}, y^i_{T^i_n})$ given in (9) below for $n \geq 0$ where, as before, the T^i_n are the jump times of the Poisson process associated with person i.

$$\begin{aligned}
(x^i_{T^i_{n+1}}, y^i_{T^i_{n+1}}) &= \rho(x^i_{T^i_n}, y^i_{T^i_n}) \\
&+ \begin{cases}
(1,0) & \text{w.p. } p\theta \\
(0,1) & \text{w.p. } p(1-\theta) \\
\omega_{weight}(0, \omega_{max})\mathbf{1}\left\{\frac{x^i_{T^i_n}}{x^i_{T^i_n}+y^i_{T^i_n}} \leq \Delta\right\} & \text{w.p. } q \\
\frac{\omega_{weight}}{x^{ij}_{T^i_n}+y^{ij}_{T^i_n}+1}(x^{ij}_{T^i_n}, y^{ij}_{T^i_n})\mathbf{1}\left\{\left|\frac{x^i_{T^i_n}}{x^i_{T^i_n}+y^i_{T^i_n}} - \frac{x^{ij}_{T^i_n}}{x^{ij}_{T^i_n}+y^{ij}_{T^i_n}}\right| \leq \Delta\right\} & \text{w.p. } r
\end{cases}
\end{aligned} \tag{9}$$

for suitably defined $(x^{ij}_{T^i_n}, y^{ij}_{T^i_n})$, that is $(x^{ij}_{T^i_n}, y^{ij}_{T^i_n})$ is a counter of the direct experiences made by person j that they have not yet reported to person i by time n. From our interaction model it follows that at each time T^i_n where there is an indirect experience from an honest peer, person j is chosen uniformly at random from the honest people other than i.

Recall the three ways of choosing the increment $(.,.)$ given a second hand report (a,b) in (2). We now also require scaling to ensure finiteness, otherwise a second hand information from a single interaction could have arbitrary impact. This does not seem a realistic assumption. Thus we only consider the latter two

$$\frac{1}{a+b}(a,b) \text{ or } \frac{1}{1+a+b}(a,b). \tag{10}$$

The latter is increasing in $a+b$ whereas the second is not. This is another desirable property, because then 10 positive out of 20 experiences will be given more impact than 1 positive out of 2. Thus, we shall focus on the latter increment. Independently of this choice, in fact, we obtain $\omega_{max} = 1$, that is $\omega = \omega_{weight}$. Also, independently of this choice, note that we require a large enough expected number of experiences for this not to introduce a skew. For, suppose a person reports 100 experiences, 80 of which are positive. It might be the case that if this is done in 100 separate reports of one experience each, the 20 negative ones are not believed, whereas if the aggregate is reported in a single report they are. In the first case the reputation value for the report would be 1, in the latter case it would be 0.8. We shall return to the effects of this later.

$$\begin{aligned} (x^i_{T^i_{n+1}}, y^i_{T^i_{n+1}}) &= \rho(x^i_{T^i_n}, y^i_{T^i_n}) \\ &+ \begin{cases} (1,0) & \text{w.p. } p\theta \\ (0,1) & \text{w.p. } p(1-\theta) \\ (0,\omega)\mathbf{1}\left\{\frac{x^i_{T^i_n}}{x^i_{T^i_n}+y^i_{T^i_n}} \leq \Delta\right\} & \text{w.p. } q \\ \frac{\omega}{x^{ij}_{T^i_n}+y^{ij}_{T^i_n}+1}(x^{ij}_{T^i_n}, y^{ij}_{T^i_n})\mathbf{1}\left\{\left|\frac{x^i_{T^i_n}}{x^i_{T^i_n}+y^i_{T^i_n}} - \frac{x^{ij}_{T^i_n}}{x^{ij}_{T^i_n}+y^{ij}_{T^i_n}}\right| \leq \Delta\right\} & \text{w.p. } r \end{cases} \end{aligned} \tag{11}$$

4.2 Mean-field approach

As in the previous section, we will consider the average reputation counter $(x_{T_n}, y_{T_n}) = (1/N_h)\sum_{i=1}^{N_h}(x^i_{T_n}, y^i_{T_n})$ and make the (strong) assumption that all honest people's counters are equal and hence the average. Moreover, we now consider the average of the second hand information and assume all are equal to that average. We obtain this average as the expected value and substitute in. The probability of a person passing on information to a peer before obtaining another direct experience is

$$\gamma = \frac{\lambda_i/(N_h+N_l-1)}{\lambda_d + \lambda_i/(N_h+N_l-1)} = \frac{r}{p(N_h-1)+r}. \tag{12}$$

Thus the number of experiences passed on as second hand information is distributed Geo(γ). The expected number is $(1-\gamma)/\gamma = (N_h-1)p/r$, a fraction θ of which positive. This assumption is not appropriate for a very small number of expected reports $(N_h-1)p/r$, i.e. for a very small number of people. Recall, however, from the previous section that the system itself introduces a skew in this case. We demonstrate this effect in AppendixB.2.

We now consider the following mean deterministic differential equation for the average reputation counter.

$$\begin{aligned}\dot{x}(t) &= -\lambda(1-\rho)\,x(t) + \lambda p\theta + \lambda r\omega\theta(1-\gamma)\mathbf{1}\left\{\left|\tfrac{x(t)}{x(t)+y(t)} - \theta\right| \le \Delta\right\}\\ \dot{y}(t) &= -\lambda(1-\rho)\,y(t) + \lambda p(1-\theta) + \lambda r\omega(1-\theta)(1-\gamma)\mathbf{1}\left\{\left|\tfrac{x(t)}{x(t)+y(t)} - \theta\right| \le \Delta\right\}\\ &\qquad +\lambda q\omega\mathbf{1}\left\{\tfrac{x(t)}{x(t)+y(t)} \le \Delta\right\}\end{aligned} \tag{13}$$

This can be obtained from (11) via a fast-time scaling and by means of averaging the dynamics as shown in Appendix A.2. This system is discontinuous along the line $x(t)/(x(t)+y(t)) = \Delta$. Moreover, it might be discontinuous along $x(t)/(x(t)+y(t)) = \theta-\Delta$, $x(t)/(x(t)+y(t)) = \theta+\Delta$ or both. So the system might have one, two or three lines of discontinuity, but is linear in between. We can solve it separately on each of the possible regions, similarly to the solution of (5). The exponential decay term is $e^{\lambda(1-\rho)}$.

$$(x,y) = \frac{1}{1-\rho}\left(p\theta, p(1-\theta)\right) \tag{14}$$

is a fixed point if $\Delta < \theta$ and $\Delta < 0$. That is, never.

$$(x,y) = \frac{1}{1-\rho}\left(p\theta + r\omega(1-\gamma)\theta, p(1-\theta) + r\omega(1-\gamma)(1-\theta)\right) \tag{15}$$

is a fixed point if $\Delta < \theta$ and $\Delta \ge 0$. That is if $\Delta < \theta$. If it exists, it is asymptotically stable and trajectories from $x(t)/(x(t)+y(t)) > \Delta$, $|x(t)/(x(t)+y(t)) - \theta| \le \Delta$ are attracted to it. Moreover, trajectories from $x(t)/(x(t)+y(t)) > \Delta$, $|x(t)/(x(t)+y(t)) - \theta| > \Delta$ (if this is non-empty) lead into the same region and thus are also attracted to it. That is, trajectories from $x(t)/(x(t)+y(t)) > \Delta$ are attracted to it.

$$(x,y) = \frac{1}{1-\rho}\left(p\theta + r\omega(1-\gamma)\theta, p(1-\theta) + r\omega(1-\gamma)(1-\theta) + q\omega\right) \tag{16}$$

is a fixed point if $\Delta \ge \theta\frac{p+r\omega(1-\gamma)}{p+r\omega(1-\gamma)+q\omega}$ and $\Delta \ge \theta\left(1 - \frac{p+r\omega(1-\gamma)}{p+r\omega(1-\gamma)+q\omega}\right)$. If it exists, it is asymptotically stable and trajectories from $x(t)/(x(t)+y(t)) \le \Delta$, $|x(t)/(x(t)+y(t)) - \theta| \le \Delta$ are attracted to it.

$$(x,y) = \frac{1}{1-\rho}\left(p\theta, p(1-\theta) + q\omega\right) \tag{17}$$

is a fixed point if $\Delta \ge \theta\frac{p}{p+q\omega}$ and $\Delta < \theta\left(1 - \frac{p}{p+q\omega}\right)$. If it exists, it is asymptotically stable and trajectories from $x(t)/(x(t)+y(t)) \le \Delta$, $|x(t)/(x(t)+y(t)) - \theta| > \Delta$ are attracted to it.

(14) and (15) correspond to the true subject behaviour θ. At this value, honest reports are believed and accepted, thus the conditions for (14) cannot be satisfied. At (16), lies are accepted, but the honest peers reports, too. Finally, at 17, lies are accepted whereas honest peers' reports are not. The reputation values corresponding to the three possible fixed points and the simplified conditions are summarized in Table 4.2. The bifurcation plot is given in Figure 4.2.

Thus, there are four critical points.

$$\Delta_{c_1} = \theta\frac{p}{p+q\omega} \tag{18}$$

Fixed point	Reputation value	Conditions
True (15)	$R_1^* = \theta$	$\Delta < \theta$
Intermediate (16)	$R_2^* = \theta\frac{p+r\omega(1-\gamma)}{p+r\omega(1-\gamma)+q\omega}$	$\Delta \geq \theta\frac{\max(p+r\omega(1-\gamma),\, q\omega)}{p+r\omega(1-\gamma)+q\omega}$
False (17)	$R_3^* = \theta\frac{p}{p+q\omega}$	$\Delta \in \left[\theta\frac{p}{p+q\omega},\, \theta\frac{q\omega}{p+q\omega}\right)$

Table 2: Summary of fixed point reputation values and their conditions.

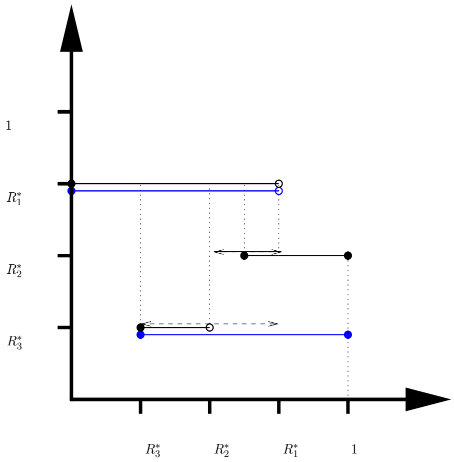

Figure 1: Bifurcation plot showing the existence of fixed points as a function of Δ: As Δ increases from 0 to 1 the number of fixed points changes. Black: In the scenario where direct experience only is passed on. Depending on the parameters there might be one, two, or three fixed points. Blue: In the scenario where reputation is passed on. As Δ increases from 0 to 1 the number of fixed points changes from one to two and back to one.

$$\Delta_{c_2} = \theta\frac{\max\left(p, q\omega\right)}{p + q\omega} \tag{19}$$

$$\Delta_{c_3} = \theta\frac{\max\left(p + r\omega(1-\gamma), q\omega\right)}{p + r\omega(1-\gamma) + q\omega} \tag{20}$$

$$\Delta_{c_4} = \theta. \tag{21}$$

Note that $\Delta_{c_1} \leq \Delta_{c_2}$ and $\Delta_{c_3} < \Delta_{c_4}$, but Δ_{c_2} might be less, more or equal to Δ_{c_3}. Hence there might be two, three or four distinct critical points.

It can be checked that if only the true fixed point (15) exists, then the trajectories from the other regions lead into its region and thus are also attracted to it. Similarly, if only the intermediate fixed point (16) exists, then all trajectories are attracted to it. If both exist but not the false one (17), then trajectories from the potential other region are also attracted to the intermediate fixed point (16). If the true and the false fixed points (6) and (17) exist, then trajectories from the potential intermediate region might be attracted to either one, depending on the parameter values. Finally, if all three exist, they are asymptotically stable and attract trajectories from their respective region only. Thus, we have the following result.

Theorem 2 *(i) Suppose $q\omega \leq p$. If $\Delta < \Delta_{c_3}$,* (15) *is the unique fixed point of the mean differential equation (13). It is asymptotically stable and all trajectories are attracted to it. If*

$\Delta_{c_3} \leq \Delta < \Delta_{c_4}$ *there is a second, intermediate fixed point* (16) *and both are asymptotically stable, attracting trajectories from* $x(t)/(x(t) + y(t)) > \Delta$ *and* $x(t)/(x(t) + y(t)) \leq \Delta$ *respectively. Finally, if* $\Delta_{c_4} \leq \Delta$*, then only the latter, intermediate one is asymptotically stable and all trajectories are attracted to it.*

(ii) Otherwise, $q\omega > p$*. Suppose* $\Delta_{c_2} < \Delta_{c_3}$*. If* $\Delta < \Delta_{c_1}$*,* (15) *is the unique fixed point. It is asymptotically stable and all trajectories are attracted to it. If* $\Delta_{c_1} \leq \Delta < \Delta_{c_2}$ *there is a second, false fixed point* (17) *and both are asymptotically stable, attracting trajectories from their respective region and one of them from the potential intermediate region, too. If* $\Delta_{c_2} \leq \Delta < \Delta_{c_3}$*,* (15) *is again the unique fixed point of the mean differential equation (13). It is asymptotically stable and all trajectories are attracted to it. If* $\Delta_{c_3} \leq \Delta < \Delta_{c_4}$ *there is a second, intermediate fixed point* (16) *and both are asymptotically stable, attracting trajectories from* $x(t)/(x(t)+y(t)) > \Delta$ *and* $x(t)/(x(t)+y(t)) \leq \Delta$ *respectively. Finally, if* $\Delta_{c_4} \leq \Delta$*, then only the latter, intermediate one is asymptotically stable and all trajectories are attracted to it.*

(iii) The case $\Delta_{c_2} > \Delta_{c_3}$ *is essentially the same except that now if* $\Delta_{c_3} \leq \Delta < \Delta_{c_2}$ *all three fixed points are asymptotically stable and attract trajectories from their respective region.*

(iv) Finally, the case $\Delta_{c_2} = \Delta_{c_3}$ *is the same except that now there is no such intermediate regime.*

As a result, the differential equation for this scenario, too, exhibits a phase transition behaviour. For sufficiently small Δ, liars do not have any impact. Only for intermediate values, liars have some impact. For large Δ liars have maximal impact. They brainwash everyone in the network.

These results can also be stated in terms of the lying probability q. Liars do not have any impact if

$$q < \frac{1}{\omega}\frac{\theta - \Delta}{\Delta}p \tag{22}$$

when $\theta > 2\Delta$ and if

$$q \leq \frac{1}{\omega}p \text{ and } q < \frac{1}{\omega}\frac{\theta - \Delta}{\Delta}(p + r\omega(1-\gamma)) \tag{23}$$

when $\theta \leq 2\Delta$. Otherwise, they have some impact and if moreover $\Delta < \theta$ liars have maximal impact and brainwash everyone in the network.

Again, under the assumption of a symmetric social network we can rephrase these results in terms of the proportion of liars in the network. Liars do not have any impact if their proportion is below some threshold. Otherwise, they have some impact. If moreover $\Delta < \theta$ liars have maximal impact and brainwash everyone in the network.

In Appendix B.1 we provide simulation results confirming the analytical results of this Section.

5 Conclusions and Further Work

In this paper, we have introduced a mathematical model for the formation of reputation in a social network and investigated the impact of liars. We have observed a phase transition behaviour and investigated it in detail via a mean-field approach. Thus, we can give precise conditions under which the liars will have an impact and we can specify what this impact will be.

We first investigated the scenario where reputation is passed on as second hand information. We found that second hand information from honest peers does not help with getting a more accurate opinion about the actual subject behaviour. This is as one might expect, because their reputation values are subjected to the liars also and experiences are echoed back and forth. We then looked at passing on direct experiences only to account for the situation where people pass on information that they have seen with their own eyes exclusively. Here, the liars' impact is more complex to describe. Second hand information from honest peers now does have an impact on accuracy of opinion.

We now compare the results in more detail, first phrased in terms of the threshold Δ. A small threshold means that people do not naively believe all second hand information. There are two critical values R_1^* and R_3^* such that: For small $\Delta < R_3^*$, there is no difference, the true fixed point being unique. For $R_3^* \leq \Delta < R_1^*$ there might or might not be a difference. If there is, liars have less impact in the scenario where direct experience only in passed on. For $R_1^* \leq \Delta$, liars have less impact in the direct experience scenario with the intermediate fixed point being unique rather than the false one.

We now compare the results phrased in terms of the lying probability q. When $\theta > 2\Delta$ there is no difference. When $\theta \leq 2\Delta$, the results are the same qualitatively. However, the condition on the lying probability q for the liars not to have an impact is less strict in the scenario with direct experience only. That is, liars do not have an impact even when the lying probability q is larger (cf. (8) and (23)). Thus, liars have less impact in the scenario where direct experience only is passed on.

We have assumed independent subject behaviour. It would be interesting to consider the case when direct experiences are correlated.

Another extension is to consider strategic lying, that is liars attempting something more subtle than simply telling extreme lies. For example, they could lie in some proportion of reports only or they could always report intermediate behaviour in an attempt to conceal their lies.

Finally, it would be interesting to extend our results from the symmetric situation we have considered thus far to an asymmetric situation. In many social networks, people are not symmetric in terms of their interactions. Lai and Wong [LW02], for example, have examined the tie effect on information dissemination in the context of rumour spreading. They find that information transmitted via kin ties tends to arrive at the respondent faster than via non-kin ties or other communication channels.

One might even want to account for a proportion of people that never interacts with the subject directly. People might also differ in terms of their thresholds, some believing even rather unlikely reports, others hardly believing anything that they have not witnessed or verified themselves.

A Derivation of the Differential Equation

In this section, we show how the deterministic mean differential equation can be obtained from the stochastic process as used in Sections 3.2 and 4.2 via a fast-time scaling and by means of averaging the dynamics.

A.1 Reputation is passed on

First, for the scenario where reputation is passed on (Section 3.2). We consider a family of processes indexed by N and then a continuous-time rescaled version where the number of jumps of the average reputation in the interval $[t, t+\epsilon)$ is Poisson$(N\epsilon\lambda N_h)$ and the average jump is of size

$$\begin{array}{l} \frac{1}{NN_h}\left[-(1-\rho)\,x_t^N + p\theta + r\omega(1-\rho)x_t^N\right] \\ \frac{1}{NN_h}\left[-(1-\rho)\,y_t^N + p(1-\theta) + q\omega\mathbf{1}\left\{\frac{x_t^N}{x_n^N+y_t^N} \leq \Delta\right\} + r\omega(1-\rho)y_t^N\right] \end{array} \tag{24}$$

Note that the indicator function in the fourth possible case is trivially 1 under the assumption that all reputation counters are the same. We obtain

$$\begin{array}{l} x_{t+\epsilon}^N - x_t^N = \frac{N\epsilon\lambda N_h}{NN_h}\left[-(1-\rho)\,x_t^N + p\theta + r\omega(1-\rho)x_t^N\right] \\ y_{t+\epsilon}^N - y_t^N = \frac{N\epsilon\lambda N_h}{NN_h}\left[-(1-\rho)\,y_t^N + p(1-\theta) + q\omega\mathbf{1}\left\{\frac{x_t^N}{x_n^N+y_t^N} \leq \Delta\right\} + r\omega(1-\rho)y_t^N\right] \end{array} \tag{25}$$

Dividing by ϵ and taking the limit we are thus led to consider the deterministic mean differential equation 5.

A.2 Direct experience only is passed on

Next, for the scenario where direct experience only is passed on (Section 4.2). As before, we consider a family of processes indexed by N.

$$\begin{aligned} (x_{T_{n+1}/N}^N, y_{T_{n+1}/N}^N) \quad &= \quad (1 - \frac{1-\rho}{NN_h})(x_{T_n/N}^N, y_{T_n/N}^N) \\ + \quad & \quad \frac{1}{NN_h}\begin{cases} (1,0) & \text{w.p. } p\theta \\ (0,1) & \text{w.p. } p(1-\theta) \\ (0,\omega)\mathbf{1}\left\{\frac{x_{T_n/N}^N}{x_{T_n/N}^N+y_{T_n/N}^N} \leq \Delta\right\} & \text{w.p. } q \\ \omega(1-\gamma)(\theta, 1-\theta)\mathbf{1}\left\{\left|\frac{x_{T_n/N}^N}{x_{T_n/N}^N+y_{T_n/N}^N} - \theta\right| \leq \Delta\right\} & \text{w.p. } r \end{cases} \end{aligned} \tag{26}$$

We then consider a continuous-time rescaled version. The number of jumps in $[t, t+\epsilon)$ is Poisson$(N\epsilon\lambda N_h)$. The average jump is of size $\frac{1}{NN_h}$ times

$$\begin{array}{l} \left[-(1-\rho)x_t^N + p\theta + r\omega\theta(1-\gamma)\mathbf{1}\left\{\left|\frac{x_t^N}{x_t^N+y_t^N} - \theta\right| \leq \Delta\right\}\right] \\ \left[-(1-\rho)y_t^N + p(1-\theta) + r\omega(1-\theta)(1-\gamma)\mathbf{1}\left\{\left|\frac{x_t^N}{x_t^N+y_t^N} - \theta\right| \leq \Delta\right\} + q\omega\mathbf{1}\left\{\frac{x_t^N}{x_t^N+y_t^N} \leq \Delta\right\}\right] \end{array} \tag{27}$$

Thus, we obtain $x_{t+\epsilon}^N - x_t^N$ and $y_{t+\epsilon}^N - y_t^N$ as $\frac{N\epsilon\lambda N_h}{NN_h}$ times

$$\begin{aligned}&\left[-(1-\rho)x_t^N + p\theta + r\omega\theta(1-\gamma)\mathbf{1}\left\{\left|\tfrac{x_t^N}{x_t^N+y_t^N} - \theta\right| \leq \Delta\right\}\right]\\ &\left[-(1-\rho)y_t^N + p(1-\theta) + r\omega(1-\theta)(1-\gamma)\mathbf{1}\left\{\left|\tfrac{x_t^N}{x_t^N+y_t^N} - \theta\right| \leq \Delta\right\} + q\omega\mathbf{1}\left\{\tfrac{x_t^N}{x_t^N+y_t^N} \leq \Delta\right\}\right]\end{aligned} \tag{28}$$

Dividing by ϵ and taking the limit we are led to consider the the deterministic mean differential equation 13.

B Simulations

In this section, report our simulation results. Firstly, in the scenario where reputation is passed on (cf. Section 3) and then in the scenario where direct experiences only are passed on (cf. Section 4).

B.1 Reputation is passed on

In this section, we look at the simulations results in the scenario where reputation is passed on (cf. Section 3). We used formulation (3) to compute $40000N_h$ steps, keeping track of the average reputation values as well as two people's individual reputation values and then plotting them against n. As before, the lower and upper boundaries in the plots correspond to reputation values 0 and 1 respectively; the upper and lower intermediate lines correspond to R_1^* and R_3^* respectively. The average reputation values as used in the mean-field approach (4) are plotted in black. They are obtained by averaging over the two reputation counters first and then computing the resulting reputation value. The more intuitive average reputation which averages the individual reputation values is plotted in grey. Note that the two agree if the individual reputation counters are the same. We also plot the individual reputation values of two people in blue and in yellow. 25 independent runs were carried out for each set of parameters.

Parameter set 1 *$\theta = 0.8$, $p = 0.1$, $q = 0.3$ and $r = 0.6$ with various values of Δ. The number of honest people N_h is initially taken to be 5. This corresponds to having $N_l = (N_h - 1)q/r = 2$ liars. We also initially take $\omega = 1$ and $\rho = 0.995$. We used both the extreme initial values $R_0 = 0$ and $R_0 = 1$. Thus, from the previous mean-field results, the predicted fixed points are $R_1^* = 0.8$ and $R_3^* = 0.2$. The critical values are $\Delta_{c_1} = 0.2$ and $\Delta_{c_4} = 0.8$.*

In Figure 2 we show a typical sample path for $\Delta = 0.15 < \Delta_{c_1}$. The average reputation values increase from the extremely negative initial value past R_3^* to R_1^* and remain within its neighbourhood until the end of the simulation. This is as expected from the mean-field results. Moreover, the individual reputations and hence the intuitive average behave the same way.

In Figure 3 we show a typical sample path for $\Delta = 0.85 > \Delta_{c_4}$. Here, starting from the extremely positive initial value, the average and individual reputation values decrease past R_1^* to R_3^* and remain within its neighbourhood until the end of the simulation. This is again as expected from the mean-field results.

In both cases, we have chosen the discount factor large, $\rho = 0.995$ to illustrate the degree of convergence. For even larger values, convergence is slower but variability is even smaller. Thus, in particular the two graphs confirm the fixed points.

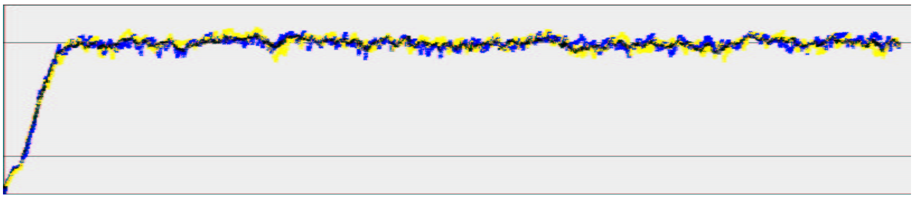

Figure 2: Typical sample path for parameter set 1. Here, $\Delta = 0.15 < \Delta_{c_1}$. The average (black and grey) and individual (blue and yellow) reputation values increase from the extremely negative initial value and then remain close to R_1^*.

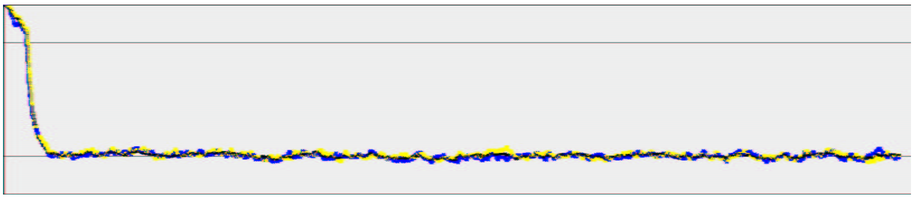

Figure 3: Typical sample path for parameter set 1. Here, $\Delta = 0.85 > \Delta_{c_4}$. The average and individual reputation values decrease from the extremely positive initial value and then remain close to R_3^*.

Figures 4 and 5 each show a typical sample path for $\Delta_{c_1} < \Delta = 0.4 < \Delta_{c_4}$ with extremely negative and positive initial values respectively. As expected, they show the same behaviour, namely that R_1^* and R_3^* are both fixed points and the average reputation settles down for some time in their neighbourhoods in an alternating fashion. Moreover, the individual reputations and hence the intuitive average behave the same way. In particular, they do so simultaneously, although this is not necessarily the case, as we will see below. Here, we have chosen the discount factor small, $\rho = 0.9$, for better illustration. This explains the higher variability compared to the earlier Figures.

For a larger number of people, both honest and a proportional number of liars, we observe the same behaviour, only the variability of the average reputation values is smaller. In particular, we have carried out simulations for $N_h = 101$ and $N_l = (N_h - 1)q/r = 50$. The sample paths for $\Delta = 0.15 < \Delta_{c_1}$ and $\Delta = 0.85 > \Delta_{c_4}$ are essentially the same as in Figures 2 and 3 respectively. We include the plots for $\Delta = 0.25$ and $\Delta = 0.75$ below (Figures 6 and 7). In particular, they confirm that $\Delta = 0.25 > \Delta_{c_1}$ and $\Delta = 0.75 < \Delta_{c_4}$, because in both

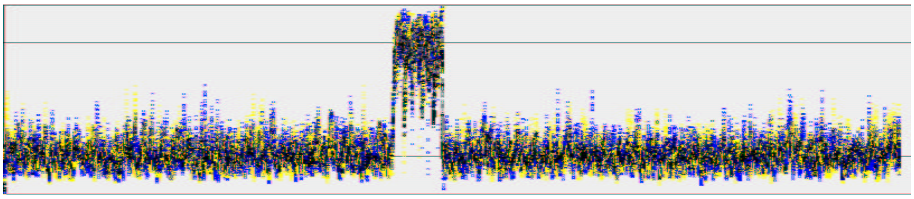

Figure 4: Typical sample path for parameter set 1 except $\rho = 0.9$. Here, $\Delta_{c_1} < \Delta = 0.4 < \Delta_{c_4}$. The average and individual reputation values increase from the extremely negative initial value and then settle down for some time in a neighbourhood of R_3^* and R_1^* in an alternating fashion.

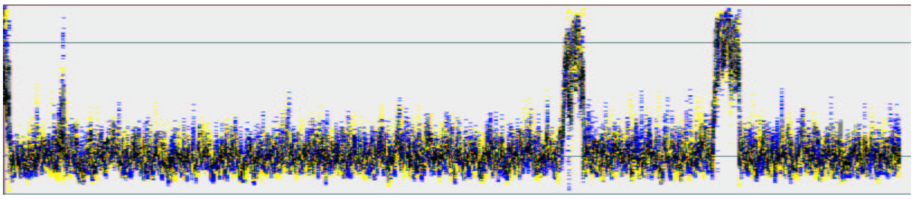

Figure 5: Typical sample path for the same parameters as in Figure 4. The average and individual reputation values decrease from the extremely positive initial value and then settle down for some time in a neighbourhood of the R_1^* and R_3^* in an alternating fashion.

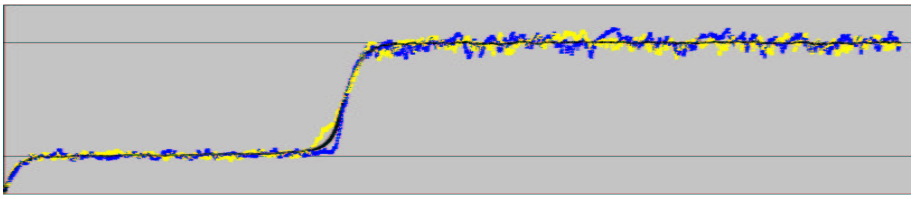

Figure 6: Typical sample path for parameter set 1 except that now $N_h = 101$. Here, $\Delta_{c_1} < \Delta = 0.25 < \Delta_{c_4}$. The average and individual reputation values increase from the extremely negative initial value and then settle down for some time in a neighbourhood of R_3^* and R_1^*.

cases both R_1^* and R_3^* appear to be fixed points. The corresponding plots for the earlier case $N_h = 5$ are again essentially the same, only the variability of the average reputation is larger.

Moreover, corresponding simulations have been carried out for parameter sets with $\omega < 1$, in particular for the same parameter set as above but now with $\omega = 0.1$. The theoretical results were confirmed in this case, too. Below we show a typical sample path for $\Delta_{c_1} = 8/13 \leq \Delta = 0.65 < \Delta_{c_4}$ with $\rho = 0.999$. We use this to illustrate that the individual reputation values can be at different fixed points, in which case the average reputation values take different intermediate values.

As a result, the simulations have confirmed the analytical results. They have shown that the individual reputation values are determined by the same fixed points and critical points, although variability is higher than the variability of the average reputation values. In the case of both fixed points, the individual reputations can behave as predicted for the average reputation, whereas the average reputation values take intermediate values. This is to be expected, because in this case the assumption of equal reputation values is no longer

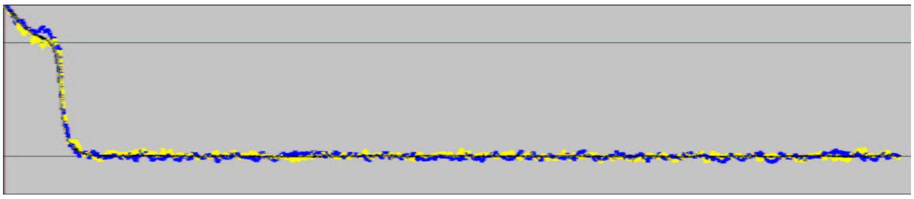

Figure 7: Typical sample path for the same parameters as in Figure 6. The average and individual reputation values decrease from the extremely positive initial value and then settle down for some time in a neighbourhood of R_1^* and R_3^*.

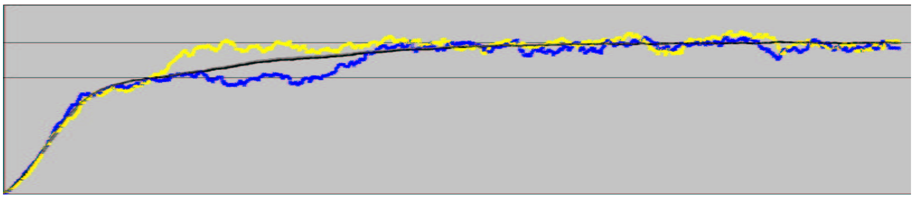

Figure 8: Typical sample path for parameter set 1 except that now $\omega = 0.1$ and $\rho = 0.999$. Here, $\Delta_{c_1} = 8/13 < \Delta = 0.65 < \Delta_{c_4}$. The average and individual reputation values increase from the extremely negative initial value. The individual reputation values settle down for some time in a neighbourhood of R_3^* and then one after the other increases further to R_1^*, the average reputation values taking different intermediate values.

satisfied. As such, our approach based on the assumption that all people have equal reputation is justified, although in the case of two fixed points the assumption itself might not hold.

B.2 Direct experience only is passed on

In this section, we look at the simulations results when direct experience only is passed on (cf. 4). From Theorem 2, we know that there are essentially four possible cases for the structure of the bifurcation plot 4.2 in terms of the overlap between the ranges of the two false fixed points, or rather three: (a) if $p \geq q\omega$ then R_3^* is not a fixed point; (b) if $p < q\omega$ and $\Delta_{c_3} \geq \Delta_{c_2}$ then R_3^* is a fixed point on a certain range and this does not overlap with the range on which R_2^* is a fixed point; (c) if $p < q\omega$ and $\Delta_{c_3} < \Delta_{c_2}$ then R_3^* is a fixed point on a certain range and this does overlap with the range on which R_2^* is a fixed point. For each case, simulations have been carried out to confirm the mean-field results. In this paper, for suitable length of presentation, we shall focus on one of the cases, however. Perhaps the most interesting is (c) where there can be three fixed points simultaneously, so this is the one we look at.

We used formulation (11) to compute $40000N_h$ steps, keeping track of the average reputation values as well as two people's individual reputation values and then plotting them against n. The colour code in the following figures is as before (cf. Section B.1). The three intermediate lines in the plots correspond to R_1^*, R_2^* and R_3^* in decreasing order. As before, 25 independent runs were carried out for each set of parameters.

Parameter set 2 $\theta = 0.8$, $p = 0.2$, $q = 0.6$, $r = 0.2$, $\omega = 0.75$ *with various values of* Δ. *The number of honest people* N_h *is initially taken to be* 100. *This corresponds to having* $N_l = (N_h - 1)q/r = 297$ *liars. We also initially take* $\rho = 0.999$. *We used both the extreme initial values* $R_0 = 0$ *and* $R_0 = 1$. *Thus, from the mean-field results, the predicted fixed points are* $R_1^* = 0.8$, $R_2^* = 2788/7985 = 0.34915466$ *and* $R_3^* = 0.2$ *respectively. From the values above we compute the critical values as* $\Delta_{c_1} = 0.2$, $\Delta_{c_3} = 3600/7985 = 0.45084534$, $\Delta_{c_2} = 0.6$ *and* $\Delta_{c_4} = 0.8$. *Recall that the values of* R_2^* *and* Δ_{c_3} *depend on* $\gamma = 1/N_h$.

In Figure 9 we show a typical sample path for $\Delta = 0.15 < \Delta_{c_1}$. Indeed, the average and individual reputation values increase from the extremely negative initial value past R_3^* and R_2^* to R_1^* and remain within its neighbourhood. A similar behaviour is observed when starting from the extremely positive initial value.

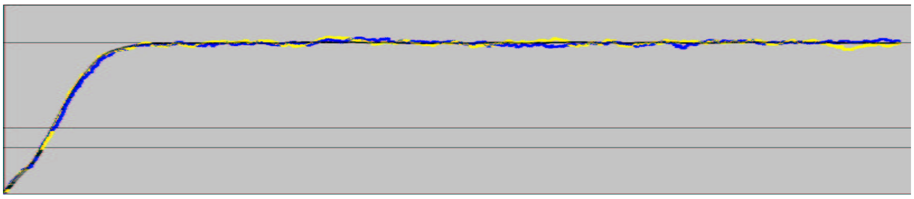

Figure 9: Typical sample path for parameter set 2 with $\Delta = 0.15 < \Delta_{c_1}$. The reputation values increase past R_3^* and R_3^*, but settle down at R_1^*.

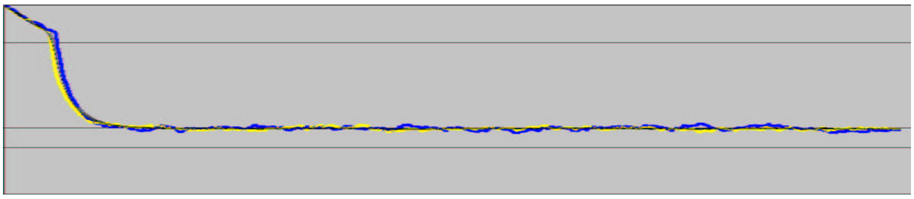

Figure 10: Typical sample path for parameter set 2 with $\Delta = 0.85 > \Delta_{c_4}$. The reputation values decrease past R_1^*, but settle down at R_2^*.

In Figure 10 we show a typical sample path for $\Delta = 0.85 > \Delta_{c_4}$. Here, the average and individual reputation values decrease from the extremely positive initial value past R_1^* to R_2^*. Similarly, starting from an extremely negative initial value, the reputation values are found to increase past R_3^* to R_2^*.

The two graphs confirm the values of R_1^* and R_2^* their being the unique fixed point on $\Delta \leq \Delta_{c_1}$ and $\Delta > \Delta_{c_4}$ respectively.

In Figure 11 we show a typical sample path for the same parameters as in Figure 10, only $N_h = 3$. The reputation values settle down lower than the intermediate false fixed point. The expected number of experiences in a second hand report is now 2, in the previous example it was 99. Thus the effect is as expected from the comments in the Section 4.2. The Figure also illustrates that some individual reputation values converge slower, but once approaching the true fixed point, they all decrease quickly.

In Figure 12 we show a typical sample path for parameter set 2 except $\rho = 0.995$ and we carry out $80000N_h$ steps with $\Delta_{c_2} < \Delta = 0.32 < \Delta_{c_3}$. The individual reputation values increase and settle down to R_3^* before increasing further, one by one, past R_2^* and settling down at R_1^*. The averaged reputation values take different intermediate values but also settle down at R_1^*. Starting from the extremely positive initial value, the reputation values are

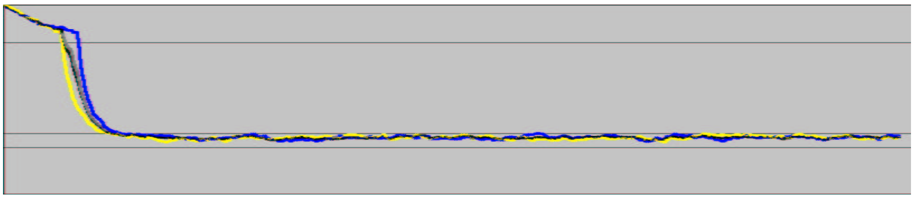

Figure 11: Typical sample path for the same parameters as in Figure 10, only now $N_h = 3$. The reputation values settle down below R_2^*; the small number of experiences reported per interaction has a noticeable effect.

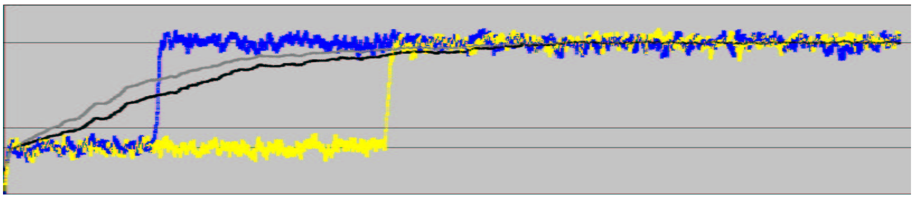

Figure 12: Typical sample path for parameter set 2 except $\rho = 0.995$ and we carry out $80000 N_h$ with $\Delta_{c_2} < \Delta = 0.32 < \Delta_{c_3}$. The individual reputation values increase and settle down to R_3^* before increasing further, one by one, past R_2^* and settling down at R_1^*.

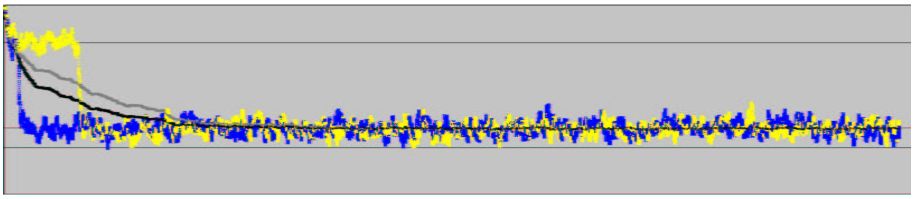

Figure 13: Typical sample path for parameter set 2 except $\rho = 0.99$ with $\Delta_{c_2} < \Delta = 0.7 < \Delta_{c_4}$. The individual reputation values decrease and settle down to R_1^* before decreasing further, one by one, and settling down at R_2^*.

found to decrease to R_1^*.

This confirms the value of R_3^* and also R_1^* and R_3^* being the stable fixed points on $\Delta_{c_1} \leq \Delta = \Delta_{c_3}$. It also suggests that the true fixed point is a stronger attractor than the false fixed point for $\Delta = 0.32$.

In Figure 13 we show a typical sample path for parameter set 2 except $\rho = 0.99$ with $\Delta_{c_2} < \Delta = 0.7 < \Delta_{c_4}$. The individual reputation values decrease and settle down to R_1^* before decreasing further, one by one, and settling down at R_2^*. The averaged reputation values take different intermediate values but also settle down at R_2^*.

In Figure 14 we show a typical sample path for the parameter set except $\rho = 0.995$ with the same $\Delta = 0.7$. The reputation values increase past R_3^* and settle down to R_2^*.

This confirms R_1^* and R_2^* being the stable fixed points on $\Delta_{c_2} \leq \Delta = \Delta_{c_4}$. It also suggests that the intermediate fixed point is a stronger attractor than the true fixed point for $\Delta = 0.7$.

Finally, in Figure 15 we show a typical sample path for parameter set 2 except $\rho = 0.96$ with $\Delta_{c_3} < \Delta = 0.5 < \Delta_{c_2}$. The individual reputation values settle down near R_3^* and R_2^* and R_1^* in an alternating fashion. The averaged reputation values take different intermediate

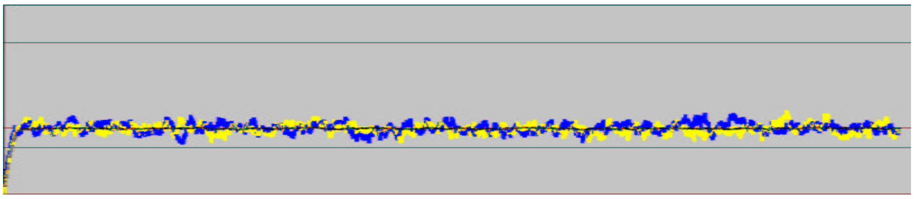

Figure 14: Typical sample path for the same parameters as in Figure 13 except $\rho = 0.995$. The individual reputation values increase past R_3^* and settle down at R_2^*.

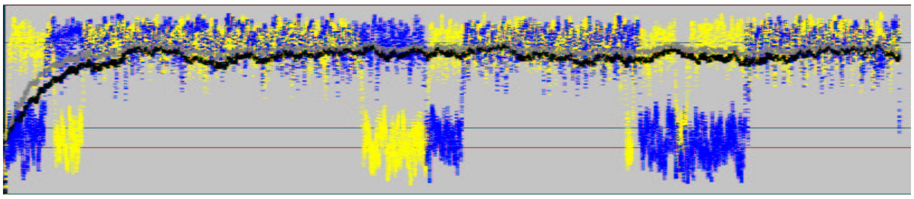

Figure 15: Typical sample path the parameter set 2 except $\rho = 0.96$ with $\Delta_{c_3} < \Delta = 0.5 < \Delta_{c_2}$. The individual reputation values settle down near R_3^* and R_2^* and R_1^* in an alternating fashion.

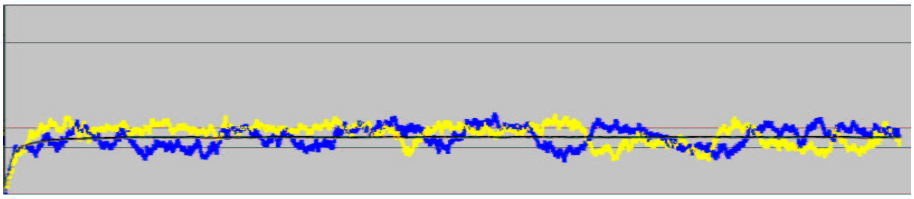

Figure 16: Typical sample path for the same parameters as in Figure 15, only $\rho = 0.995$ and $N_h = 1000$. The individual reputation values settle down at R_3^* and R_2^* in an alternating fashion.

values just below R_1^*.

It is hard to see what happens near R_3^* and R_2^*. We thus repeat the same simulations with $\rho = 0.995$. Figure 16 shows a typical sample path. We also chose $N_h = 1000$ here. It can now be seen that the individual reputation values settle down at R_3^* and R_2^* in an alternating fashion. The averaged reputation values take slightly different and intermediate values. For $N_h = 100$ the individual reputation values do not do so and also the averaged reputation values are lower. This appears to be be due to too small a number of honest people again.

This confirms R_1^*, R_2^* and R_3^* being stable fixed points on $\Delta_{c_3} \leq \Delta = \Delta_{c_2}$. It also suggests that the true fixed point is a stronger attractor than the others for $\Delta = 0.5$.

As a result, the simulations have confirmed the analytical results. They have shown that the individual reputation values are determined by the same fixed points and critical points, although variability is higher than the variability of the average reputation values. In the case of several fixed points, the individual reputation values can behave as predicted for the average reputation, whereas the average reputation values take intermediate values. Again, this is to be expected, because in this case the assumption of equal reputation values is no longer satisfied. As such, our approach based on the assumption that all people have equal reputation is justified, although in the case of two or three fixed points the assumption itself might not hold.

References

[BCL98] G. R. Barnes, P. B. Cerrito, and I. Levi. A mathematical model for interpersonal relationships in social networks. *Social Networks*, 20:179–196, 1998.

[BLB04] S. Buchegger and J.-Y. Le Boudec. A robust reputation system for peer-to-peer and mobile ad-hoc networks. In *Proceedings of P2PEcon 2004*, 2004.

[Bur99] R. S. Burt. The social capital of opinion leaders. *Annals of the American Academy of Political and Social Science*, 1999.

[Bur01] R. S. Burt. *Bandwidth and Echo: Trust, Information, and Gossip in Social Networks*, chapter in Networks and Markets: Contributions from Economics and Sociology (Casella, A. and J. E. Rauch (Edts.). Russell Sage Foundation, 2001.

[LW02] G. Lai and O. Wong. The tie effect on information dissemination: the spread of a commercial rumor in hong kong. *Social Networks*, 24:40–75, 2002.

[MLB05a] J. Mundinger and J.-Y. Le Boudec. Analysis of a reputation system for mobile ad-hoc networks with liars. In *Proceedings of WiOpt '05: Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks*, 2005.

[MLB05b] J. Mundinger and J.-Y. Le Boudec. Analysis of a robust reputation system for self-organised networks. *European Transactions on Telecommunications, Special Issue on Self-Organisation in Mobile Networking*, 16(5), October 2005.

[MN03] H. H. S. Moerbeek and A. Need. Enemies at work: can they hinder your career. *Social Networks*, 25:67–82, 2003.

[RZ02] P. Resnick and R. Zeckhauser. Trust among strangers in internet transactions: Empirical analysis of ebay's reputation system. *Advances in Applied Microeconomics: The Economics of the Internet and E-Commerce*, 2002.

[Sam54] P. A. Samuelson. The pure theory of public expenditure. *Review of Economics and Statistics*, 36(4):387–389, 1954.