

Original citation:

Franks, Henry P. W. and Griffiths, Nathan. (2015) Robust reputation in decentralized markets. *Computational Intelligence*, 31 (4). pp. 569-592.

Permanent WRAP url:

<http://wrap.warwick.ac.uk/59995>

Copyright and reuse:

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions. Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Publisher's statement:

This is the accepted version of the following article Franks, Henry P. W. and Griffiths, Nathan. (2015) Robust reputation in decentralized markets. *Computational Intelligence*, 31 (4). pp. 569-592. which has been published in final form at <http://dx.doi.org/10.1111/coin.12036>

A note on versions:

The version presented here may differ from the published version or, version of record, if you wish to cite this item you are advised to consult the publisher's version. Please see the 'permanent WRAP url' above for details on accessing the published version and note that access may require a subscription. For more information, please contact the WRAP Team at: publications@warwick.ac.uk

warwick**publications**wrap

highlight your research

<http://wrap.warwick.ac.uk>

Robust reputation in decentralised markets

HENRY FRANKS

Department of Computer Science, University of Warwick, Coventry, CV4 7AL, UK

hpwfranks@gmail.com

NATHAN GRIFFITHS

Department of Computer Science, University of Warwick, Coventry, CV4 7AL, UK

Nathan.Griffiths@warwick.ac.uk

Establishing cooperation and protecting individuals from selfish and malicious behaviour are key goals in open Multi-Agent Systems (MAS). Incomplete information regarding potential interaction partners can undermine typical cooperation mechanisms such as trust and reputation, particularly in lightweight systems designed for individuals with significant resource constraints. In this paper, we (i) propose extending a low cost reputation mechanism to use gossiping to mitigate against the effect of incomplete information, (ii) define four simple aggregation strategies for incorporating gossiped information, and (iii) evaluate our model on a variety of synthetic and real-world topologies and under a range of configurations. We show that (i) gossiping can significantly reduce the potentially detrimental influence of incomplete information and the underlying network structure on lightweight reputation mechanisms, (ii) basing decisions on the most recently received gossip results in up to a 25% reduction in selfishness, and (iii) gossiping is particularly effective at aiding agents with little or no interaction history, such as when first entering a system.

Key words: Trust, Reputation, Evolution of trust, Image scoring, Gossiping, Network topology.

¹ Address correspondence to Nathan.Griffiths@warwick.ac.uk

1. INTRODUCTION

Typical approaches to increasing levels of cooperative behaviour in decentralised open Multi-Agent Systems (MAS) have involved biasing interactions towards cooperative individuals. Such mechanisms serve two purposes: (i) protecting agents from other individuals that are likely to engage in selfish behaviour and (ii) increasing the aggregate welfare of the population. The structure of many MAS domains implicitly creates incentives for selfish behaviour, such as free-riding in BitTorrent and other P2P networks (Ruberry and Seuken, 2012), or energy conservation in wireless sensor networks (Galstyan et al., 2004).

Trust and reputation mechanisms, which incorporate observations and individual experience to aid decision making, introduce this bias into agent partner selection through direct and indirect *reciprocity* (Nowak and Sigmund, 2005). An agent who has cooperated in the past is more likely to receive reciprocal cooperation from others. Direct reciprocity can be thought of as a form of *trust* and indirect reciprocity as a form of *reputation*. In domains in which the identity of interaction partners is known, trust and reputation can facilitate significant increases in aggregate welfare, but its efficacy is directly related to the quality and quantity of information available about individuals in the population (Sommerfeld et al., 2008). Trust, which is typically assessed using direct observations of behaviour (Huynh et al., 2006; Pirzada and McDonald, 2006), can only be effective once historical interaction data is available, and reputation, which relies on observation or propagation of third-party agent behaviour, may be undermined by *incomplete information*, in which agents make decisions based on unrepresentative sets of observations. Direct and

indirect reciprocity involve feedback effects: a cooperative action can cause many subsequent cooperative actions, and vice-versa. Consequently, decisions made on incomplete information may be incorrect, in the sense that given full information the agent would have acted otherwise, and these mistakes will be amplified by the feedback of reciprocity.

Network topology also plays a significant role in the dynamics of trust and reputation mechanisms. By definition, agents are constrained to interact only with their direct neighbour set, and interaction behaviour can only be observed by those directly connected to the participants. Various network properties can consequently influence the efficacy of trust and reputation mechanisms. For example, in areas of high clustering individuals are more likely to observe the behaviour of agents that they subsequently interact with (and vice-versa, such that individuals in low clustered regions are less likely to have observed potential interaction partners). Network structures constrain information propagation, which in turn affects the efficacy of mechanisms that rely on the transmission of relevant information. The specific network topology may therefore be detrimental (e.g. if the constraints lead to information being incomplete or out-of-date) or beneficial (e.g. if the network supports reliable and efficient throughput) for trust and reputation mechanisms. Research has shown that networks can support isolated communities of cooperators (Nowak and Sigmund, 2005) and the role of network structure in facilitating information propagation is well studied (e.g. Glinton et al. (2010); Newman (2003); Wang (2003)).

Trust and reputation mechanisms are thus highly suited to open MAS domains such as Mobile or Vehicular Ad-hoc Networks (MANETs and VANETs respectively), and present a useful setting for investigating the impacts of incomplete in-

formation and network structure on levels of emergent cooperative behaviour. In this paper, we empirically analyse the conditions under which mechanism efficacy is reduced and demonstrate a possible mechanism, in the form of gossiping, to mitigate the effects of incomplete information and exploit the ease of information transmission in typical network structures. Specifically, we show that incomplete information can result in inaccurate reputation assessments that subsequently reduce cooperation, and that the underlying network structure significantly influences emergent behaviour, both positively and negatively depending on the configuration. We supplement trust and reputation with gossiping, which exhibits low space and time complexity, using one of four aggregation rules. We show that gossiping can reduce selfishness in the population by up to 25%, and is particularly effective on real-world network topologies.

2. BACKGROUND

Trust and reputation are highly successful mechanisms for supporting cooperative and coordinated behaviour (Josang et al., 2007; Nowak and Sigmund, 1998; Pirzada and McDonald, 2006; Ramchurn et al., 2005). Broadly speaking, trust is a subjectively held belief that another individual will reciprocate given an opportunity to defect and receive higher payoff, and supports cooperation through direct reciprocity, namely the notion of “you scratch my back and I’ll scratch yours” (Nowak and Sigmund, 2005). Reputation is typically further defined as a socially accepted trust assessment for a given individual (Josang et al., 2007), and supports cooperation through indirect reciprocity. Trust often requires significant historical interaction

data for accurate assessments, and instantiations often make use of multiple dimensions of information (Huynh et al., 2006; Sabater et al., 2006). Reputation is a key mechanism when individuals have insufficient direct interaction history with which to assess a partner's trust value, and indirect reciprocity has been shown to be a greater force in encouraging cooperative behaviour than direct reciprocity in domains with low probability of repeat interaction (Bravo and Tamburino, 2008).

Trust and reputation systems that have exhibited the most promising results also tend to be the most complex architecturally (Huynh et al., 2006; Sabater et al., 2006). Given the properties of typical open MAS domains (including potentially limited computational and communication resources), such mechanisms may not be as suitable as mechanisms with less complex requirements. However, while robust systems tend to incorporate more complex architectures, systems have been demonstrated that are remarkably simple and support cooperative behaviour in low-overhead environments, such as image scoring (Nowak and Sigmund, 1998), as used in this paper, or the ad-hoc network oriented model of Pirzada and McDonald (2006).

Despite reputation's simplicity and effectiveness in supporting cooperative behaviour in highly decentralised systems, it still requires agents to obtain a sufficient amount of information regarding individuals in order to make accurate assessments. While this is easier than in pure trust systems, since the information can be gathered through observation or propagation of information through the social network, we can identify two situations in which this information may be hard to completely acquire: (i) when agents have just joined a system and there is little information regarding their history, and (ii) when the rate of interactions is high enough that

agents cannot observe a complete set of interactions for potential partners. These situations are the focus of this paper.

Nowak and Sigmund (N&S) introduced and extensively investigated *image scoring*, a simple instantiation of reputation with low computational and bandwidth overheads based on the notion of indirect reciprocity, in which cooperation emerges without requiring subsequent interactions between the same individuals (Nowak and Sigmund, 1998, 2005). This property is key to its suitability in open decentralised systems. Each agent maintains an image score for every individual they encounter, either through (i) direct interactions or (ii) observing the interactions of neighbours. Cooperative actions increase the image score by one, and selfish actions decrease it by one. When deciding whether to cooperate or not, an agent compares its strategy, an integer, with the perceived image score of the potential partner (if no data is available, it is assumed to be zero). If the image score is greater than or equal to the strategy then the agent cooperates. We describe the image scoring model in more detail in Section 3.

N&S found that cooperation emerges, but is often cyclic as non-cooperative agents invade populations of unconditionally cooperative agents and gain higher payoffs, causing the population to be subsequently dominated by conditionally cooperative agents, who are then superseded by unconditionally cooperative agents. In the setup used by N&S, agents are randomly chosen and paired from the entire population for repeated rounds, where a round is a single generation of agents engaging in one or more interactions. The total number of interactions per round (m) is at most one order of magnitude larger than the number of agents in the population (n) in the configuration used by N&S.

While image scoring is effective at supporting cooperation, we can identify situations in which it might be undermined by agents having incomplete or insufficient information regarding potential interaction partners. Firstly, if there are a large number of interactions per round compared to the number of agents (i.e. a high ratio of m/n), agents may only have observed a proportion of the interaction history of a potential interaction partner. If the observed sub-set of interactions is unrepresentative, this may result in a decision that the agent would not have taken given complete information. Similarly, if there are relatively few interactions (i.e. a low ratio of m/n), or agents have only recently entered the system, then an agent may have insufficient information with which to make an accurate decision. In this paper, we evaluate the extent to which these hypotheses are correct, and propose gossiping as a mitigating solution.

Gossiping algorithms, initially introduced by Frieze and Grimmett (1985), perform data aggregation and spreading in distributed systems. Loosely modelled on the dynamics of human gossip, they are effective at spreading information through networks, and have low space and time complexity and minimal bandwidth requirements when compared to traditional spreading mechanisms (Fernandess and Malkhi, 2007; Kempe and Kleinberg, 2003). They have previously been applied to constrained trust and reputation problems (Bachrach et al., 2008; Ramchurn et al., 2004; Zhou and Hwang, 2007), and can efficiently aggregate trust values without the need for complex data structures.

Gossiping is an attractive solution to the problems inherent in local perception of information by agents. Sommerfeld et al. (2007) have extensively investigated gossiping in humans and show that gossiping of information is an effective substitute

for direct observation. Subsequent work by Sommerfeld et al. (2008) demonstrates that gossip is robust to propagation of inaccurate information, and they conclude that humans use a majority rule: if the majority of gossips are positive, then the individual forms a positive opinion of the subject. The low overheads, high robustness to inaccurate information, and ability to efficiently spread and aggregate information in decentralised domains make gossiping highly applicable to our model.

3. INCORPORATING GOSSIPING INTO IMAGE SCORING

We adopt the image scoring model introduced by N&S, using their original setup as follows: each agent i is associated with a strategy k_i , chosen uniformly at random in the range $[-5, 6]$. Each agent maintains an image score I_i for each agent i it has observed interacting. Image scores are initialised at 0 and constrained to the range $[-5, 5]$. Each round m pairs of agents are randomly chosen from a population, with one agent in each pair being designated as the donor and the other as the recipient. If the donor's strategy is less than or equal to its perception of the image score of the recipient, $k_{donor} \leq I_{recipient}$, then it confers a benefit b on the recipient at a cost c to itself. We adopt the values of $b = 1, c = 0.1$ used by N&S. An agent assumes an image score of 0 if it has no data on the recipient. If the donor donates (cooperates), then the observers of that interaction increase their perception of the donor's image score by one (the recipient's image score remains the same). If the donor does not cooperate, the perceived image score of the donor as held by the observers is reduced by one. If the image score is moved out of the range $[-5, 5]$ as a result of the update, then it is kept at the limit of -5 or 5 as appropriate. An agent's strategy k_i thus

represents the degree of selfishness of potential interaction partners that the agent is willing to tolerate and still act cooperatively.

N&S consider both *complete* and *partial* observability of interactions. In the partial observability settings, N&S allow a set of agents (10 in their formulation), chosen at random, to observe each interaction. We model partial observability using an observability parameter, o , in the range $[0, 1]$, as the probability of each neighbour being selected to observe. If N_i denotes the set of neighbours for a given agent i , then, on average, $o \times |N_{donor} \cup N_{recipient}|$ observers are selected at random to observe for each interaction. Observations are perfect, in that the interaction is observed fully without noise. Given $n = 100$, $o = 0.1$ is equivalent on a completely connected topology to the original setup of N&S. Observability, in the static connection topologies investigated in this paper, can be viewed as a simple abstract model of typical resource constraints, or hardware or communications failure.

After m interactions have been performed, offspring are generated in proportion to an agent's final payoff. If agent i has *fitness* f_i , where f_i is equal to its net benefit (the sum of the costs incurred and benefits received), then F is the net *population benefit* such that $F = \sum_{i=0}^n f_i$. An agent will produce $\frac{n \times f_i}{F}$ offspring. The strategy of the offspring is an exact copy of the parent strategy, with a small probability μ of mutation such that the strategy is set to a random value (we adopt the value of $\mu = 0.001$ used by N&S). N&S found that strategies do not converge to a single value except for when $o = 1$ and $\mu = 0$, but instead go through cycles as selfish agents become dominated by conditionally cooperative agents (called *discriminators* by N&S), who only help other cooperative individuals. These agents are then

superseded by unconditionally cooperative agents (also called *altruists* by N&S), who are subsequently invaded by selfish agents (called *defectors* by N&S).

N&S characterise the strategy space as: $k \leq 0$ denotes cooperation, since agents will interact with most other agents, and $k > 0$ denotes defection (also called *selfish* by N&S). We further divide the cooperative strategy space into *unconditionally cooperative* ($-5 \leq k \leq -2$) and *conditionally cooperative* ($-2 < k \leq 0$). We describe interaction choices as follows. Interactions in which an agent cooperated based on its *perceived* image score of the recipient, when it should have defected based on the *actual* image score, or vice-versa, are *misclassified interactions*. An interaction is called an *incorrect cooperation* if an agent cooperates (i.e. confers a benefit on an individual) when it should have defected (given complete information). An *incorrect defection* is an interaction in which an agent defects (i.e. does not donate to the recipient) when it should have cooperated (given complete information). The number of misclassified interactions is the sum of the incorrect cooperations and incorrect defections. Incorrect defections are undesirable since they reduce the donor’s image score, leading to fewer subsequent donations to the donor. Incorrect cooperations are undesirable since they allow selfish agents to gain higher payoff, and become more likely to be reproduced. The absolute value of an agent’s image score that is maintained (to allow calculation of misclassified interactions) includes any incorrect cooperations or defections that that agent has made — it is the result of an agent’s actual actions rather than how they should have acted given complete information.

Note that in this paper we do not investigate a corresponding “conditionally self-*ish*” delineation of strategies between $0 < k \leq 2$. With cooperation, it is useful to determine whether individuals are conditionally cooperative or not since these groups

are vulnerable to invasion by different strategies (i.e. unconditional cooperators can be exploited by selfishness, and conditional cooperators can be outperformed by unconditional cooperators). While there may be similar vulnerable configurations for conditionally selfish agents, in this paper we focus on reducing the ability of any selfishness to survive in the population. The strategy space might alternatively be separated into ‘selfish’, ‘cooperative’, and ‘conditional’ areas, where a ‘conditional’ strategy would imply $-2 < k < 2$. While it would be interesting to analyse results according to this schema and determine if it facilitates any additional insight into the dynamics of our model, it is beyond the scope of this paper and we leave it for future work.

Our simple gossip mechanism spreads perceived image scores as follows: each agent maintains a queue of received gossips, which are processed in a separate gossip phase. After an interaction, each observer starts a gossip with probability ogp (observer gossip probability) by sending a gossip packet to a randomly chosen neighbour. The probability of any given agent starting a gossip thus depends both on o , the probability it is chosen as an observer, and on ogp , the probability that an observer starts a gossip. Each gossip packet contains the image score of the donor, as perceived by the gossip starter, the unique ID of the donor, the unique ID of the gossip starter, and a time to live (TTL). In this paper our focus is on the effect of gossiping on situations of incomplete and insufficient information, and so we assume that gossips are accurate, however it is important to note that gossiping has been shown to be resilient to inaccurate information, such as untrustworthy or malicious agents might provide (Sommerfeld et al., 2008).

Every $gossipRate$ interactions, there is a gossip phase. Each agent in turn up-

dates their image score values for each agent that they have received gossips about using some update rule, and propagates the gossip with $TTL_{t+1} = TTL_t - 1$ to a single randomly chosen neighbour that does not yet have the gossip until $TTL = 0$. It is assumed that an agent can check if a neighbour has received a gossip already.

We propose four update rules that gossip receivers can use to incorporate received gossip information.

- (1) *Aggregate Average (AA)*: The agent replaces its perceived image score for agent i with the average of its previous perceived score for i and the values contained in all the received gossips concerning i . If $G_{i,a}$ denotes the set of gossips received by agent i about agent a and $G(j)_{i,a}$ denotes the j^{th} gossip received by i regarding a , then the new image score held by i after applying AA is,

$$I_{i,a,t+1} = \frac{I_{i,a,t} + \sum_{j=1}^{|G_{i,a}|} G(j)_{i,a}}{|G_{i,a}| + 1}$$

- (2) *Average Replace (AR)*: The agent replaces its perceived image score for agent i with the average of the values contained in all received gossips concerning i ,

$$I_{i,a,t+1} = \frac{\sum_{j=1}^{|G_{i,a}|} G(j)_{i,a}}{|G_{i,a}|}$$

- (3) *Majority Replace (MR)*: The agent replaces its perceived image score for i with the median value contained in all received gossips concerning i . As noted above, it is thought that this is approximately how humans process gossip (Sommerfeld et al., 2008),

$$I_{i,a,t+1} = \text{median}(G_{i,a})$$

- (4) *Most Recent (MRec)*: The agent replaces its perceived image score for i with the most recent value received concerning i ,

$$I_{i,a,t+1} = G(|G_{i,a}| - 1)_{i,a}$$

We summarise the notation used in our model in Table 1.

[Table 1 about here.]

4. EXPERIMENTAL SETUP

We model two primary situations in which incomplete information may undermine the efficacy of reputation: (i) when there is a very low probability of having observed any interactions, such as when first entering a system, and (ii) when there is a very low probability of observing a complete set of interactions. We model situation (i) by using a low ratio of interaction rate to population size, and situation (ii) by using a very high ratio of interaction rate to population size. N&S used parameters of $n = \{20, 50, 100\}$ and $m = \{125, 200, 300, 500, 1000\}$, which is sufficient for modelling situation (i) but limited for situation (ii). To investigate (ii), we simulated $m = \{1000, 5000, 10000, 20000, 50000\}$ for $n = 100$, i.e. we consider a maximum ratio of $m/n = 500$. We use $o = 0.1, \mu = 0.001, b = 1, c = 0.1$, and, unless otherwise stated, we use an observer gossip probability $ogp = 1.0$ and $gossipRate = 1$. Since the diameter of the synthetic networks we generate in our simulations is typically less than 5 we use a TTL of 5 for both the synthetic networks and real-world network samples. We performed a number of simulations scaling the population to $n = 1000$ to allow us to test the effects of group size.

We situate agents on a variety of network structures. We replicate the com-

pletely connected topology used by N&S, and implement random (such that each pair of nodes is connected with probability p), scale-free and small-world synthetic networks¹. Scale-free networks are generated using the Eppstein and Wang (2002) algorithm and small-world networks using the generation algorithm proposed by Kleinberg (2000). Additionally, we use 8 networks sampled using Breadth-First Search (BFS) from the Enron email dataset and the arXiv general relativity section collaboration network² to corroborate our results on networks that are structurally closer to those found in the real world. BFS is used as a sampling method as, although it is known to be biased towards high degree nodes, it accurately retains the local network structures within the sample (Gjoka et al., 2010).

Our investigation focused on two main metrics: the strategy distribution for the population and the numbers of misclassified interactions. As discussed in Section 2 cooperation is cyclic with non-cooperative agents invading populations of unconditionally cooperative agents leading to a temporary reduction of cooperation before the population recovers. The state of the population at an arbitrarily chosen generation (e.g. the final generation of $t = 10000$) is unlikely, therefore, to provide a representative view of the simulation, and so we adopt the methodology of N&S of taking a sample over many generations (Nowak and Sigmund, 1998, 2005). Specifically, the results given are averaged over the course of the simulation for 20 runs with each parameter configuration, giving a standard deviation ranging from 1–14%. We used $t = 10000$ generations of evolution.

There are a number of significant differences between the experiments performed

¹Generated using the Java Universal Network/Graph Framework <http://jung.sourceforge.net/>

²Both datasets are taken from <http://snap.stanford.edu/data/>

by N&S and our configuration. N&S tested only a completely-connected network topology, whereas we use a number of synthetic and real-world networks with which to evaluate the model. Secondly, we extend the number of interactions per round significantly, from N&S' maximum of 1000 to 50000. N&S experimented with various population sizes from 20 to 100. The majority of our experiments are with $n = 100$ agents, but we also perform experiments with $n = 1000$ to test the effects of population size.

5. RESULTS AND DISCUSSION

In this section we present our results, beginning with the effects of incomplete and insufficient information, followed by the effects of population size, and finally the effects of gossiping and the parameters that define how it operates.

5.1. Incomplete information due to high interaction rate

[Figure 1 about here.]

Initially, we evaluate the effect of high ratios of m/n (i.e. the number of interactions per round / population size). As discussed above, a high interaction rate relative to the number of agents increases the probability that agents will have incomplete information regarding potential interaction partners. Figure 1(a) shows the effect on the population strategy distribution of varying m , the number of interactions each round, using $n = 100$ agents, $o = 0.1$ (i.e. an average of 10 agents observing each interaction), $b = 1$, $c = 0.1$, $\mu = 0.001$, and a fully-connected network. This is equivalent to N&S' original setup. The society is highly cooperative at $m =$

1000 (i.e. $m/n = 10$), with less than 10% of agents adopting selfish strategies (i.e. $k > 0$). Figure 2 shows the average strategy over time under this configuration, with the horizontal lines delineating selfishness (above the top line), conditionally cooperative (between the two lines), and unconditionally cooperative (below the bottom line) strategies. We can clearly see the cyclic behaviour noted by N&S. At $m = 1000$, 32 interactions were misclassified per round on average, or 3.2%. As m increases, there are increases in both the levels of selfishness and the proportion of misclassified interactions. Figure 1(b) plots the percentage of interactions that were misclassified over the entire simulation. When $m = 50000$ (i.e. $m/n = 500$), 40% of the population has adopted a selfish strategy and 5.7% of interactions are misclassified (i.e. an average of 2850 per round).

[Figure 2 about here.]

The proportion of misclassified interactions falls slightly between $m = 20000$ and $m = 50000$, despite an increase in selfishness. We believe that the strategy distribution of the population is an important determinant of levels of incomplete information. Recall that image scoring, through indirect reciprocity, induces a feedback effect in which cooperative actions cause subsequent cooperative actions, and vice-versa for defection. In a highly cooperative society, a choice by a donor to cooperate is likely to be correct (due to the high numbers of cooperators) even if made on the basis of highly incomplete information. The same is true, vice-versa, for defecting societies. However, when the strategy distribution is mixed, uncertainty regarding a recipient's strategy is higher, and consequently choices made on the basis of incomplete information are more likely to be incorrect. This is evident in the data

for $m = \{20000, 50000\}$, in which selfishness rises by 9%, while the proportion of misclassified interactions drops by 1%.

[Figure 3 about here.]

Figure 3 separates the misclassification of cooperative and defective actions for the data in Figure 1(b). At $m = 1000$, 0.24% of cooperative actions are incorrect — in all interactions in which a donor donated, it was the same decision that an individual would have taken given complete information 99.76% of the time. Conversely, of all the interactions in which the donor defected, 26.4% would have been cooperative had the donor had complete information (i.e. the agent made the same decision as they would have given complete information only 73.6% of the time). This corroborates the discussion above, since the society is more than 90% cooperative. As m rises, the proportion of defections that are misclassified rises (to a peak of 58%) and then falls, as the rising proportion of selfish agents reduces the probability that an interaction partner is cooperative, and subsequently that the decision to defect is incorrect.

These results demonstrate two relationships: where there is a high interaction rate the proportion of misclassified interactions and selfish strategies both increase as m/n increases. We believe that incomplete information is a key component of the mechanism by which the increasing interaction rate results in reduced support for cooperative behaviour.

5.2. Network structure

It is also important to investigate the relationship between network structure and the impact of incomplete information. We initially investigate random, scale-free and small-world networks. Random networks are a useful middle ground between completely connected networks and scale-free or small-world networks. Scale-free and small-world networks are known to model features of networks found in the real world (Albert and Barabási, 2002). To explore the effect of network topology we modify the interaction model so that instead of pairs of agents being chosen randomly, the donor is now selected at random from the population and the recipient is chosen at random from the donor's neighbour set.

[Figure 4 about here.]

[Figure 5 about here.]

In random networks, each pair of agents is connected with probability p . Assuming an undirected graph and $p = 0.1, n = 100$, each agent is, on average, connected to 4.95 neighbours. Given that sparse connectivity is a feature of many open MAS, especially VANETs (Ott and Kutscher, 2004), it is useful to evaluate the effect of varying p . Figure 4(a) plots the strategy distribution for the configuration in Figure 1(a), in random networks for various connection probabilities. Selfishness dominates at $p = 0.01$, since there are so few agents observing interactions that indirect reciprocity cannot take hold. Cooperation takes hold as p rises to 0.1, but selfishness again rises as p increases up to 0.5. For these values of p , there are sufficient neighbours that incomplete information, due to any given neighbour only observing a small sub-set of a potential recipient's history, becomes significant.

Figure 4(b), which plots the levels of misclassified interactions, demonstrates this: at $p = 0.5$, 2.09% of interactions are misclassified, whereas there are negligible misclassified interactions at $p = 0.01$. Random networks show statistically significant differences in the levels of selfishness compared to completely connected networks. The p-values for a two-tailed t-test range from 0.031 at $m = 1000$ to $6.37 \times e^{-20}$ at $m = 50000$, demonstrating that the introduction of network topology has significantly impacted the operation of image scoring.

Scale-free networks show strong support for cooperative behaviour, as shown in Figure 5(a). Although a certain number of interactions are required to initially support cooperative behaviour (i.e. at $m = 1000$), as m rises we no longer see the characteristic rise in selfishness observed in completely-connected or random networks. Unconditional cooperators in particular are dominant, suggesting that the structure of scale-free networks allows groups of agents to cooperate with reduced vulnerability to selfish invaders. Figure 5(b), which shows the misclassified interactions for data in Figure 5(a), corroborates this: with a maximum total of 1.13% at $m = 1000$.

Scale-free topologies are known to have beneficial properties regarding information propagation and robustness to untargeted malicious action. For example, Delgado (2002) used a model of social convention emergence to show that complex (i.e. scale-free and/or small-world) networks are more efficient than regular graphs with the same average node degrees, and that scale-free networks are as efficient at spreading information as fully-connected graphs. Albert and Barabási (2002) also noted the remarkable fault-tolerance of scale-free networks. The robustness of scale-free networks is partially derived from their clustering: there are highly-internally

connected groups with relatively few links to the rest of the population. In the context of our investigation, we hypothesise that this grouping effect allows image scoring to act with a much smaller average connectivity, since there will be many such groups in which agents are highly visible to other agents within that group. As discussed previously, visibility of agents is important for the efficacy of image scoring. We use *visibility* to denote the combination of observability and topological connectivity, since both influence how many agents might observe an interaction. Sen (2008) demonstrated the existence of scale-free topological structure in mobile ad-hoc networks, and many other real-world networks are known to be scale-free (Albert and Barabási, 2002). The robustness of image scoring on scale-free networks is thus highly important, as it demonstrates the broad applicability of the technique.

We draw three main conclusions from these experimental results:

- (1) **The level of incorrect interaction choices is dependent on the probability of having witnessed a recipient's interactions.**

This probability is based on a number of factors, including the degree of a node, the observability in the population, and the number of interactions.

- (2) **Incomplete information has an observable effect on levels of emergent cooperation in the image scoring model.**

N&S note in their original paper that when moving from their initial model, equivalent to observability = 1, to observability of 0.1 a larger number of interactions are needed to establish cooperation. Our results corroborate this and establish that higher levels of incomplete information (whether caused by low node degree or high numbers of interactions) lead to more selfish societies.

(3) The levels of cooperation in the image scoring model are highly dependent on the underlying topological structure of the social network.

Random graphs, and to a lesser extent, scale-free networks, significantly reduce the detrimental effects of incomplete information and aid the emergence of high levels of cooperation.

5.3. Insufficient information due to low interaction rate

Scaling the ratio of m/n to high values means that agents are likely to have incomplete information regarding the set of interactions a potential recipient has participated in. This represents a vulnerable configuration, as our previous results show, but agents are likely to also be vulnerable when this ratio is very low. Under this configuration, while agents may have complete information regarding a potential recipient's history, there may be insufficient historical data to make accurate decisions. This is particularly true when agents first enter a system.

[Figure 6 about here.]

Figure 6 shows the strategy distribution in the population with $n = 100$, $m = \{125, 300\}$ (i.e. a ratio of $m/n = \{1.25, 3\}$), across a variety of synthetic networks. For the majority of network classes, selfishness dominates within the population. The proportion of selfish agents is particularly high for $m = 125$, and tends to decrease as m increases to 300. This supports our hypothesis that the interaction history is insufficient at low levels of m , reducing the ability of indirect reciprocity to support cooperation. Small-world networks appear to give strong support for cooperative behaviour, and we believe this is because these networks make agent interactions

highly visible to potential future donors. Small-world networks are characterised by high levels of clustering, which indicates that the neighbours of any given agent are more likely to know each other than in other network classes (which don't correspondingly show high levels of clustering). Our hypothesis is that this clustering increases the probability of cooperative actions being subsequently rewarded via indirect reciprocity, leading to the disparity between network classes shown in Figure 6. These results suggest that both configurations that we examine, namely (i) a very high rate of interactions, and (ii) a very low rate of interactions, are vulnerable settings in which indirect reciprocity is less effective due to either incomplete information (as in (i)) or insufficient information (as in (ii)).

5.4. Larger population sizes

[Table 2 about here.]

[Table 3 about here.]

Our results thus far are limited in the sense that we have only simulated 100 agents (as per N&S). In this section, we describe results from simulations using a population size of 1000.

As Table 2 shows, scaling up the population to 1,000 agents introduces a smoothing effect. Overall, the model behaviour is very similar to where we have a smaller population ($n = 100$). The influence of incomplete information appears slightly reduced, but the populations are more evenly distributed with selfishness remaining significant. Interestingly, the support that small-world networks displayed for cooperative behaviour is no longer present, and selfishness levels are similar to

scale-free networks. Increasing the number of edges in scale-free networks slightly reduces the level of selfishness, corroborating our hypothesis regarding visibility of agent interactions. This has important implications. Consider mechanisms to aid the emergence of social norms, which reduce the strategy choices available to agents. In systems with high levels of normative control, incomplete information is reduced due to lower uncertainty about agent strategies, and reputation mechanisms may subsequently become more effective.

Our results show significant levels of selfishness across a variety of configurations. We conclude that there are three primary influences on levels of selfishness:

(1) **Underlying network topology**

Selfishness is more evident in scale-free networks (structurally closest to the real-world) than small-world networks, but small-world networks are particularly supportive of cooperative behaviour. Small-world networks have low geodesic path lengths and high clustering, implying a higher probability of connection between observers of an interaction and potential interaction pairs³.

(2) **Interaction rate**

At very low rates (i.e. $m = 125$), there is not time for indirect reciprocity to take hold and selfishness increases (i.e. image scoring suffers from a *cold start* problem). As m increases selfishness is slightly reduced (down to 1.07% at $m = 1000$), but again rises as we approach $m = 50000$ (up to 31.4%). At low and high values of m , there is a higher probability of an agent having insufficient information with which to make an accurate assessment of a potential partner.

³Recall that although observers may be connected with the recipient of an interaction, they only update the score of the donor. For the observation to be of use, the observer must then also interact with the donor.

As a result, the efficacy of image scoring is drastically reduced, and selfishness rises. These represent vulnerable configurations for reputation mechanisms.

(3) **Population strategy distribution**

A population with an equal strategy distribution increases the effect of incomplete information, by increasing the uncertainty about a potential partner's strategy and making a decision based on incomplete information more likely to be incorrect.

5.5. Gossiping

[Table 4 about here.]

[Figure 7 about here.]

[Table 5 about here.]

In this section, we present results from implementing gossiping and the effect of which aggregation rule is adopted. Table 4 compares levels of selfishness in the population for the same configuration as Figure 1, except that agents gossip and use our Average Replace update rule. Figure 7 shows the strategy distribution using each of the update rules and a control with no gossiping, on a scale-free topology with $m = 1000$. Finally, Table 5 shows the results from using gossiping on the real-world network samples, with $n = 1000$. From these results, we can conclude the following:

- (1) **Gossiping significantly reduces levels of selfishness in populations under the image scoring model**

On average, the introduction of gossiping reduces levels of selfishness by around 10% in the synthetic networks and around 18% in the real-world networks.

- (2) **There does not appear to be a relationship between the number of interactions and the reduction in selfishness in the image scoring model, whereas there is between network structure and gossip efficacy**

The real-world networks and scale-free synthetic networks in particular show significant reductions in selfish behaviour. Given the ubiquity of scale-free degree distributions in real-world open MAS domains, these results suggest that gossiping can be practically applied. Random networks are not as conducive to effective gossiping as other network classes, which may be a consequence of their reduced clustering. As argued above, clustering increases the probability of observations being of use, and since gossips are a substitute for direct observation this property translates across.

- (3) **All update rules except Aggregate Average show a statistically significant decrease in selfishness ($\alpha = 0.05$)**

In the synthetic networks, Aggregate Average performs worse than the other update rules, whereas the other update rules perform fairly equally. In the real-world networks, Most Recent performs consistently and with the greatest reduction in selfish behaviour, but the other update rules occasionally result in an increase in selfish behaviour.

On average over the four update rules, 331.7 million gossips were started, with 1.436 billion gossip packets sent over 10 million interactions, or 143 packets per interaction. Agents adopted a new image score for a given individual 496.4 million

times. On average, a single gossip causes 1.50 image score changes. Aggregate Average is the only rule to incorporate the agent's current image score perception of the gossip subject, whereas the other three rules only take into account the gossips received about a given subject. A gossip using the Aggregate Average rule causes, on average, 1.09 image score changes, whereas using the other rules causes 1.67 (Average Replace), 1.63 (Most Recent), or 1.60 (Majority Replace) changes in image score. Clearly, update rules that do not incorporate the *current* perception of the subject's image score perform better. That Most Recent performs as well as the other rules suggests that many of the updates are for when agents have no information (i.e. they assume an image score of 0), and the gossip provides initial data to make a choice with. Aggregate Average incorporates the assumption of an image score of 0, biasing the resultant updated value. These results suggest that gossiping is a useful mechanism by which new entrants to a system can start interacting quickly without having to observe the population to gain sufficient information.

In the real-world networks, Aggregate Average still performs poorly, while Most Recent gives the most consistently beneficial effects. These results are given for $m = 1000$, and as such agents are likely to have very little or no information on potential interaction partners. The Most Recent rule is equivalent to allowing each of the gossip recipients to act as an observer of the interaction being gossiped about, and thus reduces the number of interactions necessary for indirect reciprocity to take hold. This corroborates our conclusion that gossiping is a particularly useful supplement to reputation for new entrants to a system, or in systems characterised by high levels of population churn. While the difference gossiping makes in the real-world networks is generally larger than in the synthetic networks, sometimes

the introduction of gossiping results in an increase in selfishness (particularly with Aggregate Average, but never with Most Recent). Why this is the case requires further investigation, but these results imply careful consideration must be given to how agents incorporate information attained through gossiping.

5.6. Gossiping without observation of interactions

[Figure 8 about here.]

A key feature of our model is the observation of interaction results by the neighbours of the participants. Observers underpin indirect reciprocity through two mechanisms: (i) updating their own perception of the donor's image score, for use in subsequent interactions with the donor, and (ii) initiating gossip information regarding the donor to other individuals who may also subsequently interact with the donor. In some domains, we may not be able to assume that interactions are observable: in such cases, we would like to know whether image scoring still effective at promoting indirect reciprocity, and if not, whether gossiping can recover that efficacy. Recall that N&S note that reduced observability means that a larger number of interactions are required to sustain cooperation.

Accordingly, we performed simulations in which interactions were not observable. Only the recipient, therefore, updates their perceived image score for each interaction, and only recipients or donors start gossips. Figures 8(a) and 8(b) show the strategy distribution, averaged over all network topologies, for a no-observation configuration, with $m = 1000$, and $n = \{100, 1000\}$ with and without gossiping. Selfishness dominates, and we witness a sharp increase in misclassified interactions:

with 100 agents, 3% of interactions are misclassified, compared with 0.6% for the identical configuration with observation.

From Figure 8, we can see that gossiping retains some efficacy but only produces around a 10% decrease in selfishness for $n = 100$, falling to less than 5% for $n = 1000$. Since gossips are only started by interaction participants in this configuration, fewer gossips regarding each agent are circulated. This reduces the efficacy of the aggregation rules, and makes errors in perception more likely to be propagated (if agents receive any gossips at all). We see this in the misclassified interaction data: there is no statistically significant drop in misclassified interactions when introducing gossiping, despite the fall in selfishness. The decrease in selfishness for $n = 1000$ is lower than for $n = 100$. We believe this to be a result of each agent having an increased neighbourhood size (due to the increase in population size), which, in combination with the sparser rate of gossips, means that agents frequently do not benefit from gossiping. Interestingly, the number of unconditional cooperators falls with the introduction of gossiping, while the number of conditional cooperators rises. We believe that the presence of gossiping allows conditional cooperators to make more accurate choices, while unconditional cooperators are consistently exploited by the large proportion of selfish agents.

5.7. Gossiping with high cost of cooperation

Throughout our results, the cost of cooperation is kept at 0.1 of the benefit. The cost of cooperation has been the subject of much research, e.g. by Ohtsuki et al. (2009), and it is important to investigate whether gossiping can help support cooperation in the face of a low benefit/cost ratio. Figure 9 shows the strategy distribution,

averaged across all network topologies, for $c = \{0.1, 0.5, 0.9\}$, $n = 1000$, with and without gossiping.

[Figure 9 about here.]

Image scoring appears to be ineffective at supporting cooperative behaviour as the cost of cooperation rises past 0.5 of the benefit conferred. The high cost of cooperation means that it is evolutionarily advantageous for agents to defect consistently. The introduction of gossiping mitigates this effect slightly at $c = 0.5$, but appears entirely ineffective at $c = 0.9$. At $c = 0.5$, the introduction of gossiping allows those agents that are cooperative to make accurate assessments regarding which (small) subset of agents they may cooperate with. However, at $c = 0.9$, this benefit is overcome by the cost of cooperation, and there is very little incentive for agents to cooperate at all. As such, gossips will in this case simply confirm the selfishness of potential recipients, rather than support cooperation in the face of selfishness.

6. CONCLUSIONS

Highly decentralised open MAS require robust, low-cost mechanisms for supporting cooperation and protecting individuals from selfish or malicious behaviour, particularly for new entrants to a system. We have investigated two issues that can influence the efficacy of image scoring, a simplified model of reputation, namely (i) incomplete and insufficient information and (ii) underlying network structure, in settings that model those found in practical domains. Our results show that incomplete or insufficient information can undermine the feedback effects that indirect reciprocity introduces into a system and subsequently increase the levels of selfishness

in the population, and that network structures found in the real-world act to mitigate selfishness.

We have shown that (i) incomplete information can significantly undermine light-weight reputation mechanisms, with up to 62% of defection actions (in the completely connected topology scenario) taken incorrectly, and (ii) that the underlying network topology has a significant influence on levels of selfishness in the population. We applied gossiping algorithms and showed that (iii) they reduce levels of selfishness by up to 25%, with the biggest gains found on real-world topologies. We found that (iv) simply using the most recently gossiped information about a potential partner results in the most consistent benefits, suggesting that gossiping may be particularly useful for agents first entering a system. However, we have also shown that gossiping is not effective in all situations, and that when the cost of cooperation is high, or there is no external observability of agent interactions, the efficacy of gossiping is drastically reduced.

Our results are based on incorporating gossiping into the image scoring model proposed by N&S (Nowak and Sigmund, 1998, 2005), which is a simple reputation mechanism that is representative of the key features of more complex approaches (Huynh et al., 2006; Sabater et al., 2006). As such, it is important to note that our conclusions are only demonstrated in this representative reputation mechanism. However, given that more complex approaches share the core aspects of a reliance on information from direct experience and indirect recommendations and opinions, we hypothesise that issues of incomplete and incorrect information are likely to have similar impacts on more complex reputation mechanisms. An analysis of the extent of the impact in more complex reputation mechanisms is outside the

scope of this paper, but we argue that the conclusions identified above represent a starting point for such analysis. It is also important to note that although there is a wide variety of work suggesting that reputation spread by gossiping is robust to malicious gossip strategies (Sommerfeld et al., 2008), we have not investigated the impact of malicious gossips on the levels of selfishness demonstrated by our model. Investigating our model's robustness to malicious gossip is a key part of future work.

We have shown that the ratio of m/n significantly influences the level of incomplete information in the society. The two extremes of this ratio both represent vulnerable situations for simple reputation mechanisms. Our results are corroborated on real-world networks, but we note that our investigation in this area is limited. In future work, we plan to repeat the analysis on real-world networks sampled using an unbiased sampling algorithm (but we note that such algorithms typically do not retain the local network structure), and determine the risks of incomplete information in such settings.

In this paper we have assumed that agents have a neutral trust disposition, rather than being optimistic or pessimistic, through the initialisation of the image scores to 0 and using the same size of increment and decrement after cooperation or non-cooperation. Future work will investigate the effect of trust disposition on the dynamics of cooperation, with inclusion of optimistic agents whose default image score is > 0 and the increment for cooperation outweighs the decrement for non-cooperation, and pessimistic agents who have the opposite configuration.

A key effect to note from our results is that of the visibility of interactions on the levels of cooperation observed. Many open MAS domains are characterised by sparse topologies and our results appear to show the efficacy of image scoring

is reduced in such settings. Implementing gossiping with the Most Recent update rule shows significant reductions in selfishness, and is equivalent to increasing the visibility of interactions (for example by increasing network connectivity). Gossiping has been applied successfully within the specific topological challenges of VANETs (Bako et al., 2007; Costa et al., 2008) and MANETs (Buechegger, 2005), and also within the domain of reputation mechanisms (Bachrach et al., 2008; Munding and Le Boudec, 2006). Our results with no observability demonstrate the challenges of supporting cooperation in systems with low interaction visibility. In future work, we aim to extend our gossiping mechanism to better deal with systems with low observability of interactions.

REFERENCES

- ALBERT, R., and AL. BARABÁSI. 2002. Statistical mechanics of complex networks. *Reviews of Modern Physics*, **74**(1):47–98.
- BACHRACH, Y., A. PARNES, A. D. PROCACCIA, and J. S. ROSENSCHEIN. 2008. Gossip-based aggregation of trust in decentralized reputation systems. *Autonomous Agents and Multi-Agent Systems*, **19**(2):153–72.
- BAKO, B., I. RIKANOVIC, F. KARGL, and E. SCHOCH. 2007. Adaptive topology based gossiping in VANETs using position information. *In Proceedings of the 3rd International Conference on Mobile Ad-hoc and Sensor Networks*, Springer, pp. 66–78.
- BRAVO, G., and L. TAMBURINO. 2008. The evolution of trust in non-simultaneous exchange situations. *Rationality and Society*, **20**(1):85–113.
- BUHEGGER, S. 2005. Coping with misbehavior in mobile ad-hoc networks. Ph.d dissertation, Ecole Polytechnique Federale De Lausanne.
- COSTA, P., D. GAVIDIA, B. KOLDEHOFE, H. MIRANDA, M. MUSOLESI, and O. RIVA. 2008. When cars start gossiping. *In Proceedings of the 6th Workshop on Middleware for Network Eccentric and Mobile Applications*, pp. 1–4.
- DELGADO, J. 2002. Emergence of social conventions in complex networks. *Artificial Intelligence*, **141**(1-

2):171–85.

EPPSTEIN, D., and J. WANG. 2002. A steady state model for graph power laws. *In* 2nd International Workshop on Web Dynamics.

FERNANDESS, Y., and D. MALKHI. 2007. On collaborative content distribution using multi-message gossip. *Journal of Parallel and Distributed Computing*, **67**:1232–39. ISSN 07437315. .

FRIEZE, A.M., and G.R. GRIMMETT. 1985. Shortest-path problem for graphs with random arc-lengths. *Discrete Applied Mathematics*, **10**(1):57–77.

GALSTYAN, A., B. KRISHNAMACHARI, and K. LERMAN. 2004. Resource allocation and emergent coordination in wireless sensor networks. *In* AAAI Workshop on Sensor Networks.

GJOKA, M., M. KURANT, C.T. BUTTS, and A. MARKOPOULOU. 2010. Practical recommendations on crawling online social networks. *In* Proceedings of the 29th Conference on Information Communications, Number December, pp. 2498–506.

GLINTON, R., P. SCERRI, and K. SYCARA. 2010. Exploiting scale invariant dynamics for efficient information propagation in large teams. *In* Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems, International Foundation for Autonomous Agents and Multiagent Systems, pp. 21–30.

HUYNH, T. D., N. R. JENNINGS, and N. R. SHADBOLT. 2006. An integrated trust and reputation model for open multi-agent systems. *Autonomous Agents and Multi-Agent Systems*, **13**(2):119–54.

JOSANG, A., R. ISMAIL, and C. BOYD. 2007. A survey of trust and reputation systems for online service provision. *Decision Support Systems*, **43**(2):618–44.

KEMPE, D., and J. KLEINBERG. 2003. Maximizing the spread of influence through a social network. *In* Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 137–146.

KLEINBERG, J. 2000. Navigation in a small world. *Nature*, **406**(3):845.

MUNDINGER, J., and J. Y. LE BOUDEC. 2006. Reputation in self-organized communication systems and beyond. *In* Proceedings of Workshop on Interdisciplinary Systems Approach in Performance Evaluation and Design of Computer & Communications Systems.

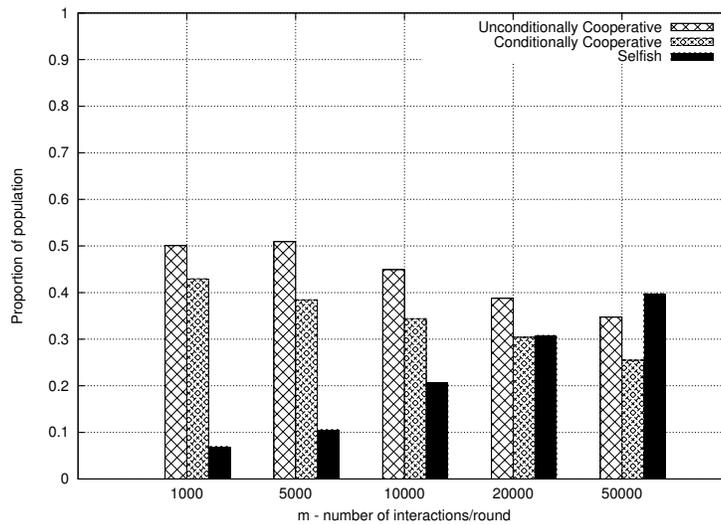
NEWMAN, M. 2003. The structure and function of complex networks. *SIAM review*, **45**(2):167–256.

NOWAK, M., and K. SIGMUND. 1998. The dynamics of indirect reciprocity. *Journal of theoretical biology*, **194**(4):561–74.

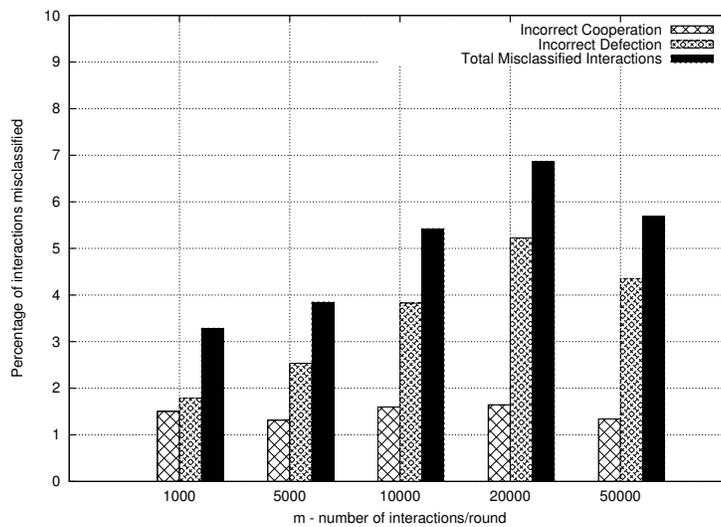
NOWAK, M., and K. SIGMUND. 2005. Evolution of indirect reciprocity. *Nature*, **437**(7063):1291–98.

OHTSUKI, H., Y. IWASA, and M. NOWAK. 2009. Indirect reciprocity provides only a narrow margin of

- efficiency for costly punishment. *Nature*, **457**(7225):79–82.
- OTT, J., and D. KUTSCHER. 2004. Drive-thru internet: IEEE 802.11b for “automobile” users. *In* 23rd Annual Joint Conference of the IEEE Computer and Communications Societies, Volume 1, pp. 362–73.
- PIRZADA, A. A., and C. MCDONALD. 2006. Trust establishment in pure ad-hoc networks. *Wireless Personal Communications*, **37**(1-2):139–168. ISSN 0929-6212. .
- RAMCHURN, S., N. JENNINGS, C. SIERRA, and L. GODO. 2004. Devising a trust model for multi-agent interactions using confidence and reputation. *Applied Artificial Intelligence*, **18**(9-10):833–52.
- RAMCHURN, S. D., D. HUYNH, and N. R. JENNINGS. 2005. Trust in multi-agent systems. *The Knowledge Engineering Review*, **19**(01):1–25.
- RUBERRY, M., and S. SEUKEN. 2012. Sharing in bittorrent can be rational. *In* *Auctions, Market Mechanisms, and Their Applications*. Harvard University Press, pp. 34–35.
- SABATER, J., M. PAOLUCCI, and R. CONTE. 2006. Repage: reputation and image among limited autonomous partners. *Journal of Artificial Societies and Social Simulation*, **9**(2):3–21.
- SEN, Q. 2008. Scale-free topology structure in ad hoc networks. *In* *Proceedings of the 11th IEEE International Conference on Communication Technology*, pp. 21–24.
- SOMMERFELD, R. D., H. J. KRAMBECK, and M. MILINSKI. 2008. Multiple gossip statements and their effect on reputation and trustworthiness. *Biological sciences*, **275**(1650):2529–36.
- SOMMERFELD, R. D., H. J. KRAMBECK, D. SEMMANN, and M. MILINSKI. 2007. Gossip as an alternative for direct observation in games of indirect reciprocity. *Proceedings of the National Academy of Sciences of the United States of America*, **104**(44):17435–40.
- WANG, X. F. 2003. Complex networks: small-world, scale-free and beyond. *IEEE Circuits and Systems Magazine*, **3**(1):6–20.
- ZHOU, R., and K. HWANG. 2007. Gossip-based reputation aggregation for unstructured peer-to-peer networks. *In* *IEEE International Parallel and Distributed Processing Symposium*, pp. 1–10.



(a)



(b)

FIGURE 1: (a) Strategy distributions and (b) levels of misclassified interactions, using a completely connected topology, $n = 100, o = 0.1, \mu = 0.001$, varying m .

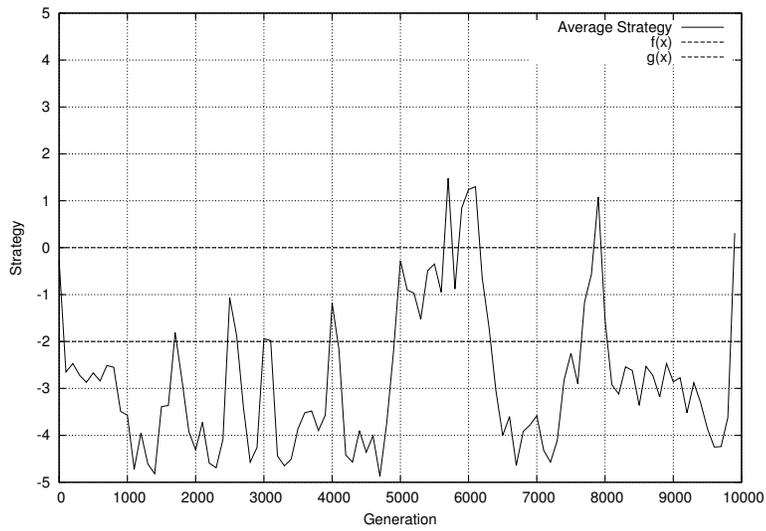


FIGURE 2: Average strategy over time, sampled every 100 generations, for $m = 1000$, for a representative run from Figure 1.

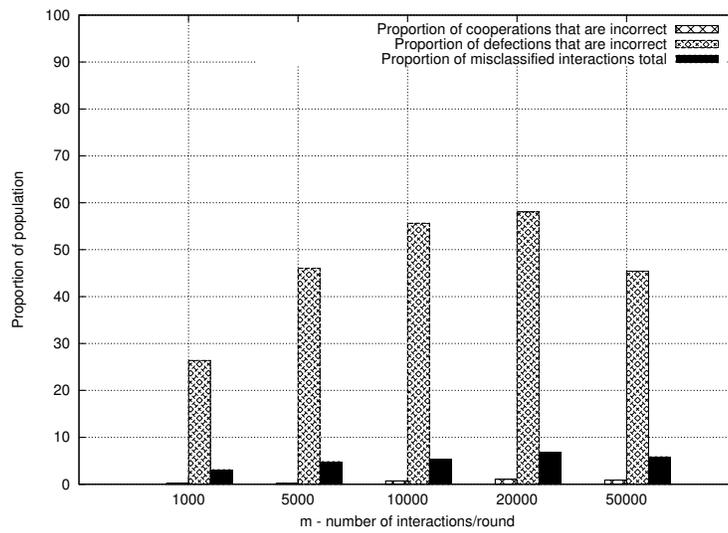


FIGURE 3: Interactions misclassified as proportion of interaction type for simulation runs in Figure 1, that is, the proportion of cooperations that are incorrect is the number of interactions in which the donor donated incorrectly divided by the total number in which the donor donated (as opposed to the total number of interactions as denominator).

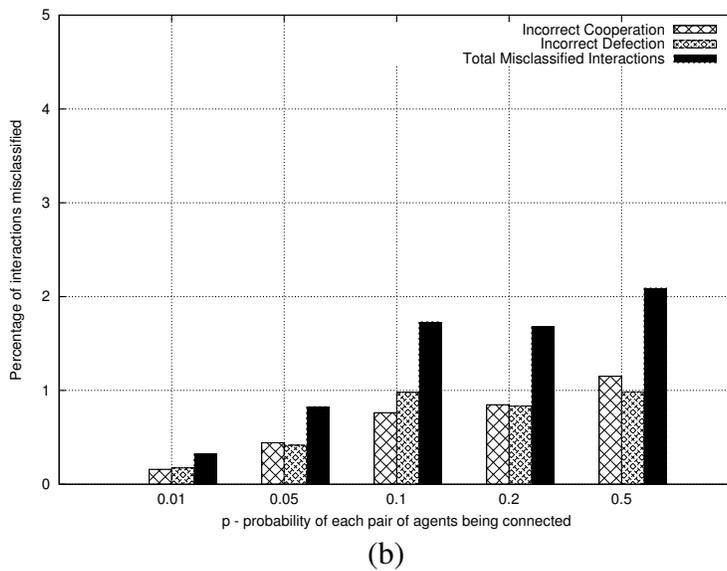
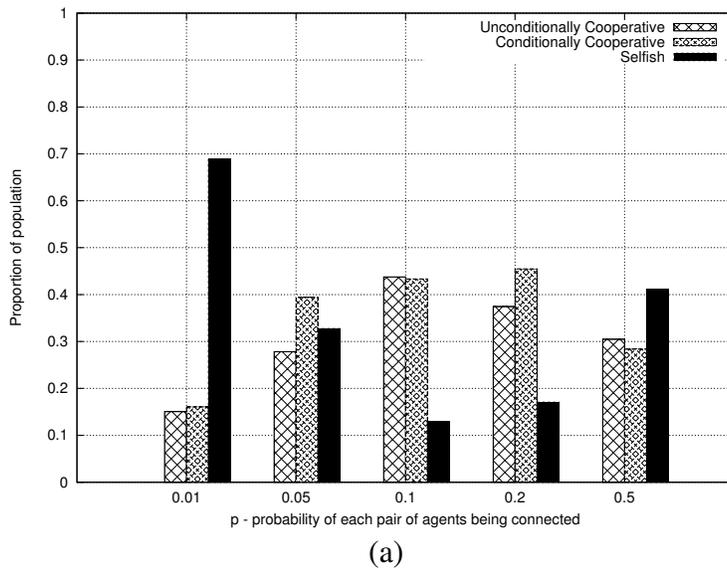


FIGURE 4: (a) Strategy classifications and (b) levels of misclassified interactions, using a random topology, varying p , with $m = 1000$ and all other settings as Figure 1.

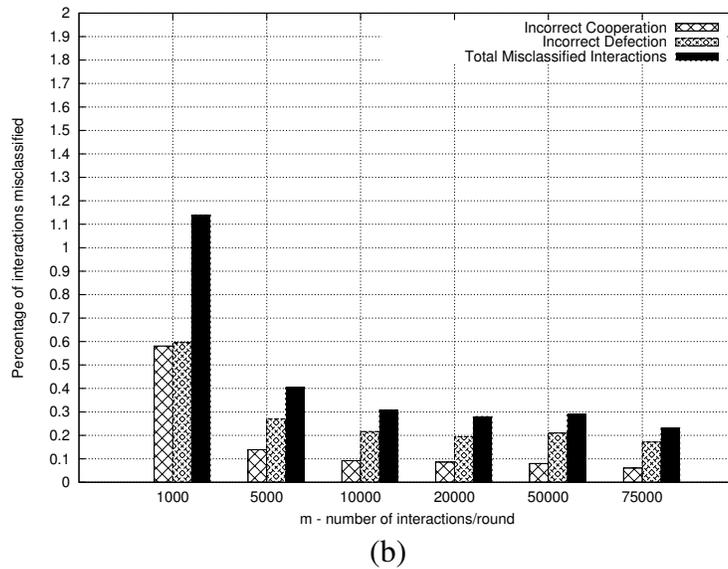
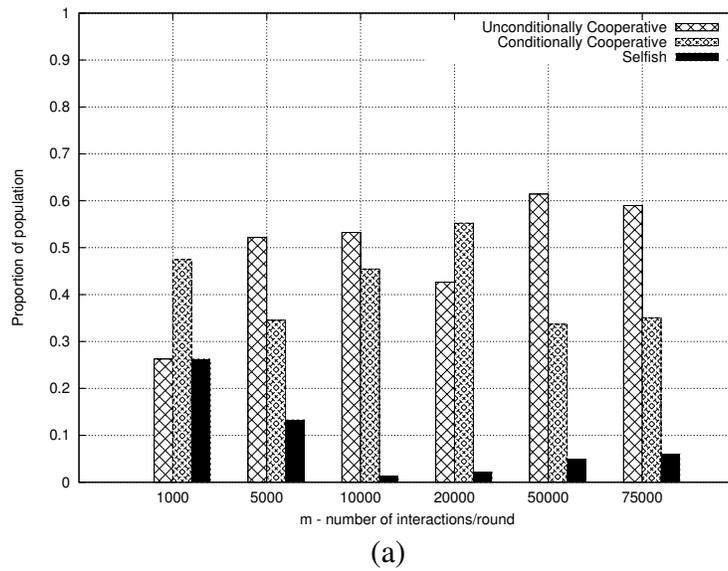


FIGURE 5: (a) Strategy classifications and (b) levels of misclassified interactions, for scale-free topology, varying m , using 1000 edges in total, and with all other settings as Figure 1.

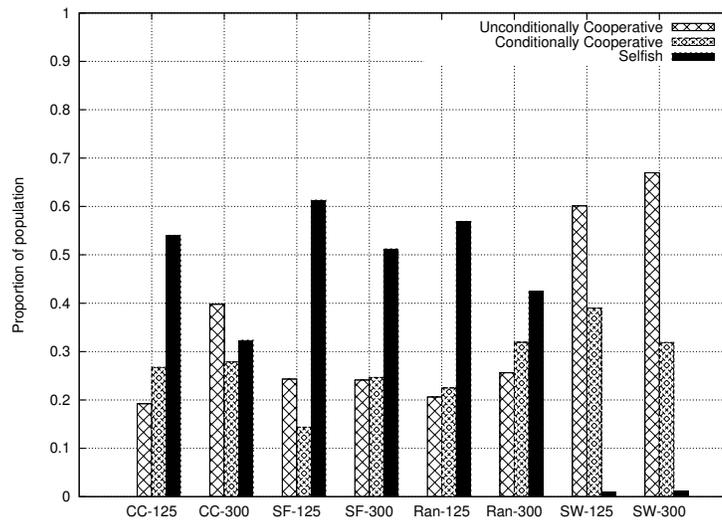


FIGURE 6: Comparison of population strategy distribution over 4 classes of topology: Completely Connected (CC), Scale-Free (SF), Random (Ran), and Small World (SW), using $m = \{125, 300\}$, $n = 100$.

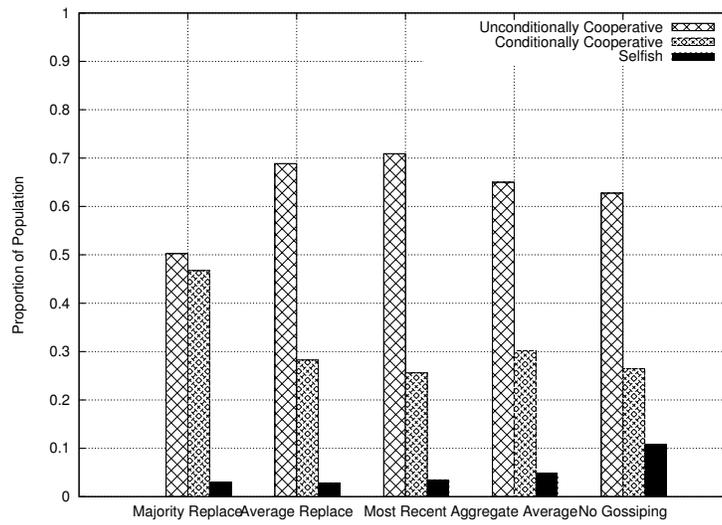


FIGURE 7: Strategy distribution using gossiping with $m = 1000$, on a scale-free network with 1000 edges, for each of the update rules, all other parameters as in Figure 1.

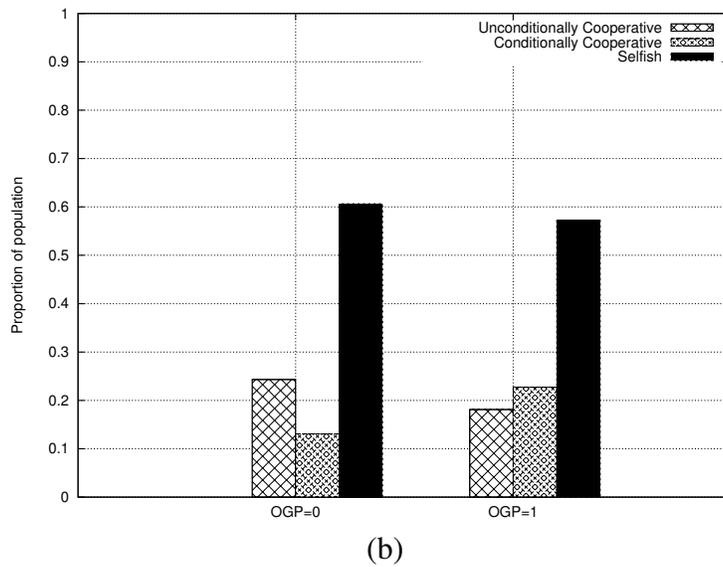
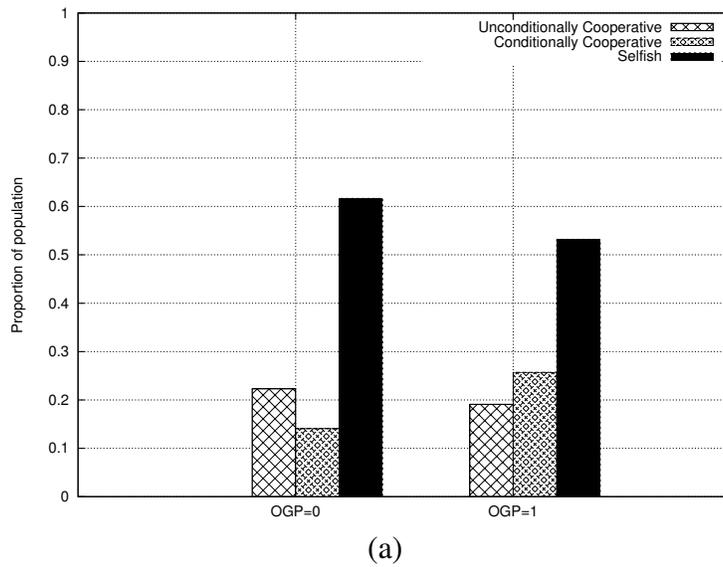


FIGURE 8: The population strategy distribution with gossiping ($ogp = 1$) and no gossiping ($ogp = 0$) when neighbours cannot observe interactions, for (a) $n = 100$ and (b) $n = 1000$. All other settings as in Figure 1.

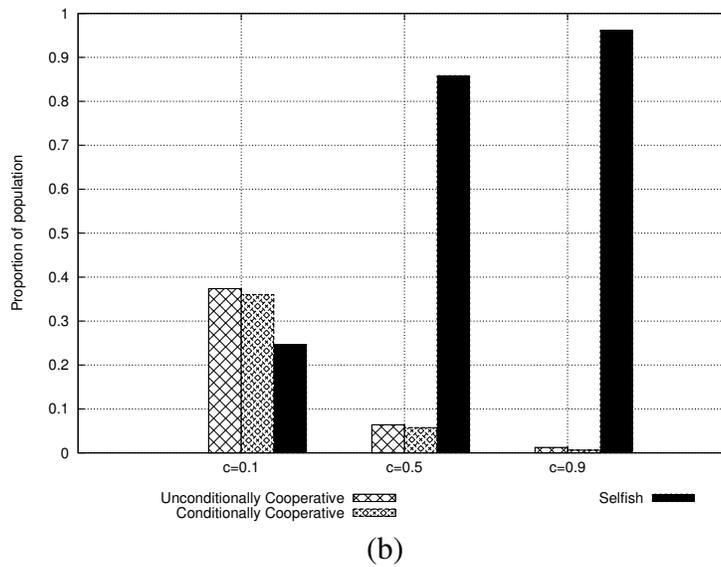
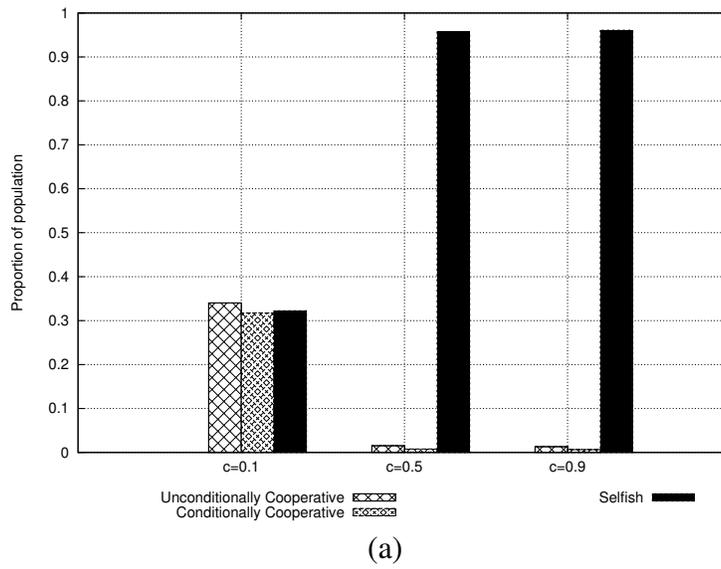


FIGURE 9: Population strategy distribution while increasing the cost of cooperation for (a) populations with no gossiping and (b) populations which gossip, averaged across all update rules and network topologies. All other settings as for Figure 1.

Table 1: Notation used in describing our model.

Symbol	Meaning
i, j	Lower case i and j are used to represent individual agents
I_i	Image score for agent i
b	Benefit received from cooperation (we adopt N&S' value of $b = 1$)
c	Cost of cooperation (we adopt N&S' value of $c = 0.1$)
N_i	The set of agents that are neighbours of i
o	Observability parameter (in the range $[0, 1]$)
n	The size of the population
m	The number of interactions per round
f_i	Fitness of agent i
F	Net fitness of the population
μ	Mutation probability (we adopt N&S' value of $\mu = 0.001$)
ogp	Observer gossip probability

Table 2: Strategy distribution for selected scale-free and small-world networks. UC is Unconditionally Cooperative, CC is Conditionally Cooperative, and S is Selfish. CE is Clustering Exponent, $n = 1000$, $m = 1000$. All other parameters as Figure 1.

Network	Parameter	Proportion of population		
		UC	CC	S
Scale-free (Eppstein)	1000 edges	0.33	0.20	0.46
Scale-free (Eppstein)	5000 edges	0.36	0.23	0.39
Scale-free (Eppstein)	10000 edges	0.41	0.25	0.32
Small-world (Kleinberg)	CE1	0.42	0.26	0.31
Small-world (Kleinberg)	CE5	0.43	0.20	0.36
Small-world (Kleinberg)	CE10	0.34	0.30	0.35

Table 3: Selfish proportion of population (S) and Percentage of Incorrect interactions (IP) for Fully Connected (FC), Scale Free (SF), Small World (SW) and Random (Ran) networks while varying m , $n = 100$. All other parameters as Figure 1.

m	FC		SF		SW		Ran	
	S	IP	S	IP	S	IP	S	IP
1000	0.13	3.71	0.91	1.2	0.01	0.1	0.03	2.0
5000	0.10	3.76	0.005	1.8	0.02	0.3	0.01	0.4
10000	0.21	5.71	0.004	1.1	0.03	0.4	0.02	0.3
20000	0.30	6.94	0.006	1.1	0.06	0.6	0.02	0.2
50000	0.39	5.93	0.01	2.2	0.11	0.6	0.04	0.2

Table 4: The average proportion of selfish agents in the population for the runs in Figure 6 ($ogp = 0$) compared with runs using the same configuration except that agents gossip using the Average Replace update rule ($ogp = 1$).

Topology	m	$ogp = 0$	$ogp = 1$	Diff.
Completely-Connected	125	0.540	0.418	0.122
Completely-Connected	300	0.324	0.221	0.103
Scale-free	125	0.613	0.479	0.134
Scale-free	300	0.512	0.330	0.182
Random	125	0.569	0.527	0.042
Random	300	0.424	0.256	0.168

Table 5: The effect of gossiping on selfishness in real-world networks with each update rule (Aggregate Average (AA), Average Replace (AR), Majority Replace (MR), Most Recent (Mrec)), with $ogp = 1$, $n = 1000$, $m = 1000$, and all other parameters as in Figure 1.

Topology	$ogp = 0$	$ogp = 1$				max. Diff.
		AA	AR	MR	MRec	
Enron-1	0.29	0.33	0.20	0.33	0.04	0.25
Enron-2	0.31	0.35	0.22	0.33	0.19	0.12
Enron-3	0.33	0.38	0.25	0.18	0.21	0.15
Enron-4	0.39	0.20	0.25	0.21	0.16	0.23
arXiv-1	0.33	0.39	0.21	0.19	0.15	0.18
arXiv-2	0.34	0.50	0.10	0.30	0.15	0.24
arXiv-3	0.34	0.40	0.19	0.35	0.22	0.15
arXiv-4	0.27	0.21	0.16	0.33	0.19	0.11