Auton Agent Multi-Agent Syst
DOI 10.1007/s10458-017-9364-x

Establishing norms with metanorms over interaction topologies

Samhar Mahmoud[1] · **Nathan Griffiths**[2] · **Jeroen Keppens**[1] · **Michael Luck**[1]

Abstract Norms are a valuable means of establishing coherent cooperative behaviour in decentralised systems in which there is no central authority. Axelrod's seminal model of norm establishment in populations of self-interested individuals provides some insight into the mechanisms needed to support this through the use of metanorms, but considers only limited scenarios and domains. While further developments of Axelrod's model have addressed some of the limitations, there is still only limited consideration of such metanorm models with more realistic topological configurations. In response, this paper tries to address such limitation by considering its application to different topological structures. Our results suggest that norm establishment is achievable in lattices and small worlds, while such establishment is not achievable in scale-free networks, due to the problematic effects of hubs. The paper offers a solution, first by adjusting the model to more appropriately reflect the characteristics of the problem, and second by offering a new dynamic policy adaptation approach to learning the right behaviour. Experimental results demonstrate that this dynamic policy adaptation overcomes the difficulties posed by the asymmetric distribution of links in scale-free networks, leading to an absence of norm violation, and instead to norm emergence.

Keywords Norm emergence · Metanorm · Topologies

✉ Samhar Mahmoud
samhar.mahmoud@kcl.ac.uk

Nathan Griffiths
Nathan.Griffiths@warwick.ac.uk

Jeroen Keppens
jeroen.keppens@kcl.ac.uk

Michael Luck
michael.luck@kcl.ac.uk

1 King's College London, London, UK

2 University of Warwick, Coventry, UK

Published online: 08 April 2017

1 Introduction

In peer-to-peer (P2P) systems, agents share resources (hardware, software or information) with one another. But if there is no cost or limit on accessing the resources provided by others, there is no incentive for agents to contribute resources for the benefit of others. More generally, when self-interested autonomous agents must exchange information without any central control, non-compliance (due to selfish interests) can compromise the entire system. Legal norms have been shown to be a very powerful mechanism in regulating the behaviour of multi-agent systems. Indeed, the field of Artificial Intelligence (AI) and Law is largely concerned with intelligent systems that reason with or about such norms, and with the possibility of a central authority that may enforce them. However, some systems, such as P2P systems, consist of large populations of interacting entities and are not controlled by a central authority; the nature and volume of interactions in such systems make it infeasible to enforce legal norms. These kinds of system can benefit from *social norms* that emerge through interactions, and are maintained by the individuals that participate within them. Such social norms are the focus of the work in this paper; we will refer to them as norms in the rest of the paper for simplicity.

The use of *norms* to provide a means of ensuring cooperative behaviour has been proposed by many [7,10,11,32,36,38,39,41] but, as shown by Axelrod [2], norms alone may not lead to the desired outcomes. In consequence, *metanorms* have been proposed as a means of ensuring not only that norms are complied with, but that they are enforced. Axelrod's model is interesting and valuable in examining how norms can be established in a population of agents. However, in real-world systems, such as peer-to-peer and wireless sensor networks, each agent can only observe the behaviour of a relatively small number of others, Yet, Axelrod's model assumes that the actions of each agent can be observed by all the other agents. While experiments have shown that metanorms are effective in fully-connected environments as used by Axelrod, there has been limited consideration of metanorms with different but more realistic topological configurations [26], which fundamentally change the mechanisms required to establish cooperation. Using our simulation model [24], we are able to replicate Axelrod's results (and in fact improve on them for extended runs, even with an observability constraint), as a result of allowing agents to learn from the payoff they receive from conducting certain actions. Learning eliminated the problem that was caused by the mutation involved in the original evolutionary approach suggested by Axelrod, as well as eliminating the need to access other agent's private policies, which are considered private information in many contexts. Although this provides a valuable illustration of the value of metanorms in avoiding norm collapse in a system in which there is no central control, it still assumes a fully connected network. Some work has already been undertaken to examine the impact of different topologies on norm establishment. For example, Savarimuthu et al. [30] consider the *ultimatum game* in the context of providing advice to agents on whether to change their norms in order to enhance cooperation in random and scale-free networks. Delgado et al. [9] study norm emergence in coordination games in scale-free networks, and Sen et al. [34] examine rings and scale-free networks in a related context. Additionally, Villatoro et al. [38] explore norm emergence with memory-based agents in lattices and scale-free networks.

While these efforts provide valuable and useful results, the context of application has been limited, with only two agents involved in each encounter, rather than a larger population of agents. This simplifies the problem when compared with those in which the actions of multiple interacting agents can impact on norm establishment. In response, this paper builds on the enhancement of Axelrod's model that is presented in [24] to address the context of different

topological configurations[1]. First, the model so far assumes a fully connected network, and is influenced by that for certain aspects, such as how one agent observes another's actions. In a variably connected structure, this part of the model is thus not meaningful and requires modification, causing some difficulties in establishing norms. Second, in scale-free networks, which contain both heavily connected nodes (*hubs*) and lightly connected nodes (*outliers*), hubs obstruct norm emergence since they require observation of, and interaction with, many others in the network, causing asymmetric effects. Such effects require further adaptation of the model, which are introduced in this paper.

The rest of the paper is organised as follows. Section 2 introduces related work, Sect. 3 gives an overview of Axelrod's model, and Sect. 4 outlines the metanorms game, adjusted to suit the purposes of this paper, and augmented with a learning mechanism. Sections 5 and 6 describe in detail the impact of applying the model in lattices and small world networks, respectively. In Sect. 7, we consider the problems that arise from the use of scale-free networks, and present an adaptation of the model that copes with their characteristics. Section 8 describes the results obtained from applying the model on samples of real world networks and, finally, Sect. 9 concludes this paper.

2 Related work

There has been much work that examines norm emergence in societies of interacting agents. Much of this work concentrates on analysing norm emergence over fully-connected networks [5,33,36,39], and it was only relatively recently that attention shifted towards the effect of the structure of these societies. In this section, we review the literature that addresses these concerns.

The earliest work concerned with the impact of topologies on convention emergence belongs to Kittock [18], who studied the effect of regular lattices on the iterated cooperation game (ICG) [17] and the iterated prisoner's dilemma (IPD) [1]. Both are iterative games in which two agents choose one of two actions with their reward depending on the combination of their choices. Agents choose actions based on the highest current reward (HCR) learning algorithm. The results obtained suggest that different convergence rates are observed with different topologies and, in particular, that a larger network diameter (being the longest path between any two nodes) makes it more difficult for the convergence to arise.

Urbano et al. [37] use a classical convention emergence game to study the influence of interaction topologies (random, regular, small world and scale free topologies) on the External Majority strategy update rule (EM) originally proposed by Shoham and Tennenholtz [35]. In particular, they investigate the effect of different memory sizes on the rate of convergence. Their empirical results show that a memory size of 3 is the best option for all types of topologies. More recently, Franks et al. [13] have shown that inserting a small number of influencer agents—those with specific conventions and strategies—is enough to manipulate the convention adopted by large societies. However, their results depend on the underlying network, with scale-free networks easing the emergence of conventions in comparison to small worlds, over which convention emergence is slower and with higher cost.

Delgado et al. [8,9] study the emergence of coordination in scale-free networks. Their study involves an interaction model of a multi-agent system, by which they try to analyse how fast coordination can spread among agents. Coordination here is represented through agents being in the same state, and coordination is considered to have emerged when 90%

[1] The work presented in this paper extends the work presented in [22] and [23].

of agents are in the same state. The results demonstrate that coordination can indeed be achieved over scale-free networks, but in a rather restricted setting. Similarly, Sen et al. [34] use a game to investigate norm emergence over lattices and scale-free networks. In particular, they analyse the effect of increasing the number of actions available to agents, as well as the effect, on the speed of norm emergence, of increasing the number of agents in both scale-free networks and lattices. Their results suggest that both increasing the number of actions *and* increasing the number of agents causes a delay to norm emergence in the population over a scale-free network. Similarly, norm emergence in lattices is much slower when agents have a larger set of actions to choose from, or when the number of agents in the population is increased. Overall, their analysis shows that, for a small set of actions, it is faster for a norm to spread in a ring than in other topologies (followed by fully connected structures, and then scale-free networks). In contrast, for a large set of actions, it turns out that this is much faster in scale-free networks than in rings and fully connected structures.

The models used in these previous pieces of work are relatively unsophisticated, with only two agents involved in a single interaction, and reward values remaining fixed and not changing during the game. In response, Villatoro et al. [38] adopted the same concept of two-agent interactions, but introduced the notion of the reward of an action being determined through the use of the memory of agents, thus adding some dynamism to the model. Here, the reward of a certain action is determined by whether the action represents the majority action in both agents' memories, and the reward is proportional to the number of occurrences of this majority action in their memories. However, it is not clear from where these rewards derive nor who applies them, as agents only have access to their individual memory. With regard to the interaction network, their work illustrates that increasing the neighbourhood size of a lattice accelerates norm emergence. In contrast, in the case of scale-free networks, norms do not emerge using the basic model. This is because of the development of *sub-conventions* that are persistent and hard to break, and these prevent the whole population from converging towards a single convention. A solution to this problem was found by giving *hub* agents (those with the majority of connections to others) more influence on the reward function.

An interesting property that has been explored with regard to networks is that of community structure [28,29], which has been shown to exist in many real-world social networks [27]. If the nodes of a graph form a set of groups that are themselves densely connected, but loosely connected with other groups, then such a graph is said to have the community structure property. O'Riordan et al. have shown that given a strong community structure, robust cooperation emerges among a population of agents that are playing the N-player prisoner dilemma [6]. This has been shown to be effective over both lattice [28] and small world networks [29].

Savarimuthu et al. [30] analyse the effect of *advice* on norm emergence over random and scale-free networks. For this reason, they use the *ultimatum game* and their results show that norm emergence increases in speed over both random and scale-free networks with an increase in the average degree of connectivity. Furthermore, they have shown [31] that their model works over dynamic network topologies that are generated using Gonzalez's model [15]. More recently, Mungovan et al. [25] introduced the idea of weighted random interaction by which agents are able to interact with random members of the population based on distance, and so the closer an agent is to another, the more likely there will be an interaction between these two agents. Their results suggest that dynamic interaction helps in easing emergence especially in breaking local biases that are normally hard to break.

3 Axelrod's model

Inspired by Axelrod's model [2], our simulation focusses only on the essential features of norm emergence in a community of self-interested agents. In the simulation, the agents play the following game iteratively; in each iteration, they make a number of binary decisions. First, each agent decides whether to comply with *the norm* or to defect. Here, the norm refers to an abstract norm, by which an agent has to make a choice between two alternative actions. This could be deciding whether to drop litter in a park, or whether to share files after downloading them in a P2P system. In this context, defection brings a reward for the defecting agent, and a penalty to all other agents. For example, if an agent decides not to share files with others, then all those expecting to receive (at least part of) the file from this agent will suffer from a longer download time in order to receive it. Each defector risks being observed by the other agents and may be punished as a result. When agents observe another agent defecting, they decide whether to punish that agent, with a low penalty for the punisher and a high penalty for the punished agent. Agents that do not punish those observed defecting risk being observed themselves, and potentially incur metapunishment. Thus, finally, each agent decides whether to metapunish agents observed to spare defecting agents. Again, metapunishment comes at a high penalty for the punished agent and a low penalty for the punisher. Thus, in each round, each agent must make the decision of whether to defect or to comply, while making multiple other decisions of whether to punish, metapunish, or spare other agents that it observes defecting or sparing a defector.

The behaviour of agents in each round of the game is random, but is governed by three variables: the probability of being seen S, boldness B, and vengefulness V. Each round agents are given a fixed number of opportunities o to defect or comply, each of which has a randomly selected probability of a defection being seen. Boldness determines the probability that an agent defects, such that if an agent's boldness exceeds the probability of a defection being seen then the agent defects. Vengefulness is the probability that an agent punishes or metapunishes another agent. Thus, boldness and vengefulness of an agent are said to comprise that agent's strategy. After several rounds of the game, each agent's rewards and penalties are tallied, and successful and unsuccessful strategies are identified. By comparing themselves to other agents on this basis, the strategies of poorly performing agents are revised such that features of successful strategies are more likely to be retained than those of unsuccessful ones. We need not be concerned with the details of the learning algorithm in this paper, beyond the fact that boldness and vengefulness are simply revised upward or downward as appropriate, in line with a specified learning rate. If most agents employ a strategy of low boldness and high vengefulness, it can be argued that the norm has become *established* in that community, because strategies that lead to defection or to sparing defecting agents are unlikely (due to low boldness) and are penalised severely (due to high vengefulness).

3.1 Our simulation algorithm

Given Axelrod's model as a starting point, we have previously developed refinements of it that are better suited to real-world distributed systems, by not requiring agents to have information on the private strategies of others, and by allowing agents to improve performance, via a reinforcement learning technique. Since this is not the focus of this paper, we will not provide a full explanation; the full details of why and how are provided in a sister paper [24]. Nevertheless, since these refinements are the starting point for our work here, in this section we briefly review the setting presented in [24].

Algorithm 1 The Simulation Control Loop: *simulation*$(T, H, P, E, \gamma, \delta)$

1. **for** each round **do**
2. interact(T, H, P, E)
3. learn(γ, δ)

First, in order to determine the unique effect of each individual action on agent performance, each agent keeps track of three different cost/utility values: the *defection score* (DS), which is the total of temptation rewards received minus the total punishments incurred from defections; the *punishment score* (PS) incurred by an agent who punishes or metapunishes another (as a result of an enforcement cost); and the *punishment omission score* (POS) incurred by an agent who does not punish another when it should, and is consequently metapunished. These are summed to form a total score (TS).

In this context, we can consider the algorithms used in our simulation, in two phases, as represented in Algorithms 2 and 3, called from the main simulation loop in Algorithm 1. More precisely, in Algorithm 2, the scores of each agent are set to zero to isolate the effect of the current round from previous rounds (Lines 1–5). Each agent has defection opportunities (o), and defects if its boldness is greater than the probability of its defection being seen. If an agent defects (Line 8), its DS increases by a *temptation payoff*, T (Line 9), but it *hurts* all others in the population, whose scores decrease by H (Line 11), where H is a negative number that is thus added to the score. If an agent cooperates, no scores change. DS thus determines whether an agent should increase or decrease boldness in relation to its utility.

However, each hurt agent may in turn observe the defection and react to it with a probability that is proportional to its vengefulness. Punishment and metapunishment both have two-sided consequences: if an agent j sees agent i defect in one of its opportunities (o) to do so, with probability S_o (Line 12), and decides to punish it (which it does with probability V_j, Line 13), i incurs a punishment cost, P, to its DS (Line 14), while the punishing agent incurs an enforcement cost, E, to its PS (Line 15). Note that both P and E are negative values, so they are added to the total when determining an overall value. If j does not punish i, and another agent k sees this in the same way as previously (Line 18), and decides to metapunish (Line 19), then k incurs an enforcement cost, E, to its PS, and j incurs a punishment cost P to its POS.

In the learning phase, in Algorithm 3, and as mentioned above, each agent uses the various scores to determine how to improve its actions in the future. At the beginning of the learning procedure, the agent calculates its total score by combining all the other scores. In order to ensure a degree of exploration (similar to mutation in the original model's evolutionary approach, to provide comparability), we adopt an *exploration rate*, γ, which regulates adoption of random strategies from the available strategies universe (Line 8).

If the agent does not explore, and its defection score is negative (Line 12), this means that defecting brings negative consequences. In this case, the agent decreases its boldness. Conversely, if its defection score is positive, because the agent managed to avoid being punished by other agents, it increases its boldness, leading to defecting even more. Similarly, agents increase their vengefulness if they find that the effect of not punishing is worse than the effect of punishing (Line 22), and decrease vengefulness if the situation is reversed. The amount of increase or decrease is equal to predetermined learning step δ. In Axelrod's original model, vengefulness and boldness have eight possible values from $\frac{0}{7}$ to $\frac{7}{7}$. In order to have comparable results to Axelrod's model, we adopt the conservative approach of increasing or decreasing by one level at each point, corresponding to a learning rate of $\delta = \frac{1}{7}$. Thus, an agent with boldness of $\frac{5}{7}$ and vengefulness of $\frac{3}{7}$ that decides to defect less and punish

Algorithm 2 interact(T, H, P, E)

```
1. for each agent i do
2.    DS_i = 0
3.    PS_i = 0
4.    POS_i = 0
5.    TS_i = 0
6. for each agent i do
7.    for each opportunity to defect o do
8.       if B_i > S_o then
9.          DS_i = DS_i + T
10.         for each agent j : j ≠ i do
11.            TS_j = TS_j + H
12.            if see(j,i,S_o) then
13.               if punish (j, i, V_j) then
14.                  DS_i = DS_i + P
15.                  PS_j = PS_j + E
16.               else
17.                  for each agent k : k ≠ i ∧ k ≠ j do
18.                     if see(k,j,S_o) then
19.                        if punish (k, j, V_j) then
20.                           PS_k = PS_k + E
21.                           POS_j = POS_j + P
```

more will decrease its boldness to $\frac{4}{7}$ and increase its vengefulness to $\frac{4}{7}$. As both PS and POS represent the result of two mutually exclusive actions, their difference for a particular agent determines the change to be applied to vengefulness. For example, if $PS > POS$, then punishment has some value, and vengefulness should be increased. As indicated previously, this is covered in more detail in [24], and we provide no further details here.

4 Imposing topologies on metanorms

Axelrod's model assumes that each agent could potentially observe the behaviour of all other agents. In terms of network topology, it organises agents in a fully connected network. However, in real-world problems, agents would only interact with, and, therefore potentially observe a (small) subset of the agents in the community. In other words, agent interactions are governed by more constrained topologies. This constraint on connectivity between agents implies some adjustments to Axelrod's model, as follows.

First, in Axelrod's model it is assumed that an agent's defection penalises all other agents in the population. The introduction of a topology enables us to restrict the penalty to only those agents with whom the defector interacts. Second, in Axelrod's model, agents are assumed to be able to observe the entire population. By introducing a topology, we capture an important characteristic of real word computational systems in which an agent can only observe those agents with whom it interacts. Third, punishment requires observation of misbehaviour. In Axelrod's model, this requirement is implicit as it makes no meaningful distinction. However, by introducing constraints on observation and rendering the model more realistic, a further refinement is required: an agent can only punish a defector if the agent can observe the defector. In addition, an agent can only metapunish an agent that fails to punish a defector if the topology allows it to observe both the defector *and* the agent that fails to punish the defector. Finally, in order to enhance an agent's individual performance, it compares itself to others in the population before deciding whether to modify its strategy. However, since

Algorithm 3 learn(γ, δ)

```
1. Temp = 0
2. for each agent i do
3.    TS_i = TS_i + DS_i + PS_i + POS_i
4.    Temp = Temp + TS_i
5. AvgS = Temp/no_agents
6. for each agent i do
7.    if TS_i < AvgS then
8.       if explore(γ) then
9.          B_i = random()
10.         V_i = random()
11.      else
12.         if DS_i < 0 then
13.            if B_i − δ < 0 then
14.               B_i = 0
15.            else
16.               B_i = B_i − δ
17.         else
18.            if B_i + δ > 1 then
19.               B_i = 1
20.            else
21.               B_i = B_i + δ
22.         if PS_i < POS_i then
23.            if V_i − δ < 0 then
24.               V_i = 0
25.            else
26.               V_i = V_i − δ
27.         else
28.            if V_i + δ > 1 then
29.               V_i = 1
30.            else
31.               V_i = V_i + δ
```

agents can only observe their neighbours, these are the only agents they are able to learn from.

In consequence, Algorithms 2 and 3 presented above are no longer adequate, and need to be replaced with Algorithms 4, and 5. Specifically, the changes are as follows. First, in Algorithm 4, Line 5 considers only agent i's neighbours NB_i rather than all the agents in the population, and Line 12 considers only agent j's neighbours NB_j. In Algorithm 5, the average score in Line 3, $AvsS_{NB_i}$ refers to the average score of agent i's neighbourhood NB_i (that is, those agents to which agent i is connected). In this way, and with these simple modifications, our algorithms now address the needs of different topological structures.

In what follows, we consider these modifications to the model in the context of different kinds of topology, in particular small world models and scale-free networks. However, to start, we introduce lattices, since they provide the foundation on which small world networks are based. Note that all experiments in this paper have been performed with parameters as specified in Table 1. It is worth mentioning here that Axelrod has assumed the values of P and P' to be equal. This is to emphasise that the violation of the original norm is equally significant to the violation of sparing a norm violator, and they both should be punished with equal amount. This is an assumption that we adopt in this paper as well.

Algorithm 4 interact(T, H, P, E)

```
1. for each agent i do
2.   for each opportunity to defect o do
3.     if B_i > S_o then
4.       DS_i = DS_i + T
5.       for each agent j ∈ NB_i do
6.         TS_j = TS_j + H
7.         if see(j,i,S_o) then
8.           if punish (j, i, V_j) then
9.             DS_i = DS_i + P
10.            PS_j = PS_j + E
11.          else
12.            for each agent k ∈ NB_j : k ≠ i do
13.              if see(k,j,S_o) then
14.                if punish (k, j, V_k) then
15.                  PS_k = PS_k + E
16.                  POS_j = POS_j + P
```

Algorithm 5 learn(γ, δ)

```
1. for each agent i do
2.   TS_i = TS_i + DS_i + PS_i + POS_i
3.   if TS_i < AvgS_{NB_i} then
4.     if explore(γ) then
5.       B_i = random()
6.       V_i = random()
7.     else
8.       if DS_i < 0 then
9.         if B_i − δ < 0 then
10.          B_i = 0
11.        else
12.          B_i = B_i − δ
13.      else
14.        if B_i + δ > 1 then
15.          B_i = 1
16.        else
17.          B_i = B_i + δ
18.      if PS_i < POS_i then
19.        if V_i − δ < 0 then
20.          V_i = 0
21.        else
22.          V_i = V_i − δ
23.      else
24.        if V_i + δ > 1 then
25.          V_i = 1
26.        else
27.          V_i = V_i + δ
```

5 Metanorms in lattices

A lattice (typically a simple ring structure) is the simplest network topology we consider, in particular, because it is used as a base for more interesting and realistic topologies. In a (one-dimensional) lattice with neighbourhood size n, agents are situated on a ring, with each agent connected to its neighbours n or fewer hops (lattice spacings) away, so that each agent is connected to exactly $2n$ other agents. Thus, in a lattice topology with $n = 1$, each agent

Table 1 Parameter initialisation

Term	Description	Value
i, j	Individuals	A number to identify individual agents
S	Probability of a defection being seen by any given individual	Uniform distribution from 0 to 1
B_i	Boldness of i	Uniform distribution from 0 to 1
V_i	Vengefulness of i	Uniform distribution from 0 to 1
T	Player's temptation to defect	$+3$
H	Hurt suffered by others as a result of an agent's defection	-1
P	Cost of being punished	-9
E	Enforcement cost, i.e. cost of applying punishment	-2
P'	Cost of being punished for not punishing a defection	-9
E'	Cost of punishing someone for not punishing a defection	-2
δ	*Learning step*	$\frac{1}{7}$
γ	*Exploration rate*	0.01

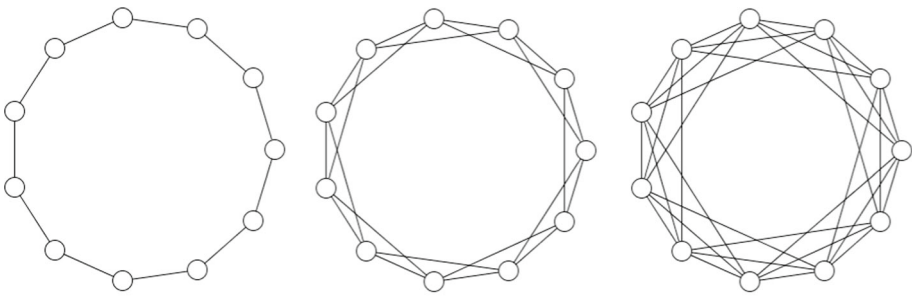

Fig. 1 Lattice topologies. **a** NB Size 1. **b** NB Size 2. **c** NB Size 3

has two neighbours and the network forms a ring as shown in Fig. 1a. In a lattice topology with $n = 3$, each agent is connected to 6 neighbours, as shown in Fig. 1c.

5.1 Neighbourhood size

It is clear that, depending on the neighbourhood size, lattices may be more or less connected. Those with larger neighbourhood sizes are more similar to Axelrod's fully connected model. Our hypothesis is that as the neighbourhood size increases, the greater connections between agents enable punishment and metapunishment to become more effective in reducing boldness and increasing vengefulness. In order to investigate this hypothesis, we ran several experiments.

In our first set of experiments, we used 51 agents (so we have an even number, plus one, to account for the $2n$ neighbours plus our original agent), and varied the neighbourhood

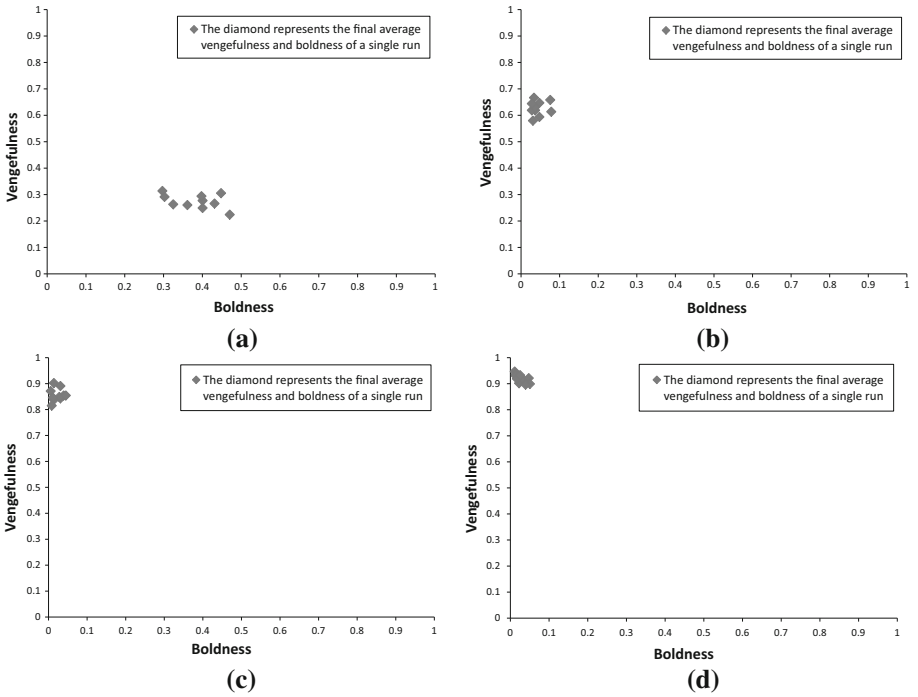

Fig. 2 Lattice Topologies: impact of neighbourhood size. **a** Lattice with neighbourhood size 1, and 1,000,000 timesteps. **b** Lattice with neighbourhood size 3, and 1,000,000 timesteps. **c** Lattice with neighbourhood size 13, and 1,000,000 timesteps. **d** Lattice with neighbourhood size 19, and 1,000,000 timesteps

size between the least connected lattice (the ring topology) and the most connected lattice ($n = 25$). Each experiment involved 10 separate runs, with each run comprising 1,000,000 timesteps for a particular neighbourhood size.

For the least connected lattice (n of 1), no norm is established, as runs ended in both relatively low boldness and relatively low vengefulness (see Fig. 2a). In this case, though agents rarely defect, they also rarely punish a defection. This constitutes an unstable situation in which defecting could be a rewarding behaviour for agents as it is relatively unlikely to be penalised. However, increasing the neighbourhood size slightly to 3 (Fig. 2b) has a noticeable impact on the results, as the boldness of the population drops almost to 0, which means that agents do not defect. While the level of vengefulness increases, it is still not at a level that can be considered to correspond to norm emergence, since agents might still not punish a defection without being metapunished for not doing so.

In addition, increasing the neighbourhood size to 13 has the same effect on boldness and a stronger effect on vengefulness (see Fig. 2c), as vengefulness increases further, and almost to its maximum of 1, when the neighbourhood size of 19 is used (see Fig. 2d). These results suggest that increasing neighbourhood size strengthens norm emergence, by virtue of agents being more willing to punish norm violators.

In seeking to provide more detail for analysis, the results of all runs were averaged, and shown on the graph in Fig. 3, with neighbourhood size plotted against boldness and vengefulness. This shows that a neighbourhood size as small as 2 is enough to maintain boldness near 0, indicating that agents do not defect except when they *explore* as a result of

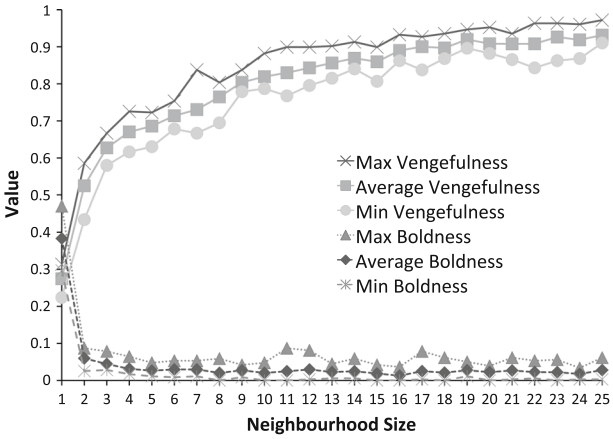

Fig. 3 Lattice: impact of neighbourhood size on final B and V

sometimes adopting random strategies (introduced for comparability with Axelrod's model). Conversely, increasing the neighbourhood size has a major impact on vengefulness, until the neighbourhood size reaches around 15 (at which point an agent is connected to half the population) when it brings only very minor change. This is because, in a poorly connected environment, agents that do not punish defection can more easily escape metapunishment than in a more connected environment.

As we hypothesised, increasing neighbourhood size brings a corresponding effect on the strategy of agents (in terms of boldness and vengefulness). Only the most poorly connected lattices have moderate levels of boldness, with vengefulness increasing monotonically over a longer period before it stabilises at a level consistent with norm establishment. The connections between agents give rise to this behaviour, with an increase in connections providing more opportunities for agents to respond to defectors appropriately.

5.2 Population size

Now, if we increase the population size while keeping the neighbourhood size static, we decrease the relative number of connections among the overall population. This suggests that convergence to norm establishment should decrease, in line with the results obtained above. In the second set of experiments, therefore, the neighbourhood size was fixed and the population size varied between 51 and 1001 agents. However, the results obtained, shown in Fig. 4 for a neighbourhood size of 3 (though other values gave similar results), are not as expected, and suggest that increasing the population size has no effect on the rate of norm emergence, as all runs for all sizes of population end almost with the same level of boldness and vengefulness.

These results suggest that norm emergence in a community of agents that interact in a lattice is not affected by total population size but by neighbourhood size. By increasing the number of neighbours, norm establishment becomes more likely, irrespective of the size of the population. In other words, the likelihood of norm establishment is governed by the total amount of punishment that could potentially be brought upon a defector or an agent failing to punish a defector, which may be termed the *potential peer pressure* of a lattice. This is because such lattices essentially comprise multiple overlapping localities in which agents are highly connected: via punishments, the agents in these localities impose a strong influence

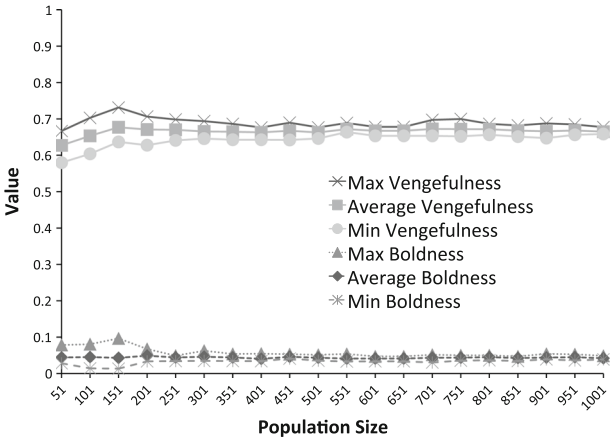

Fig. 4 Lattice: impact of population size on final B and V (where neighbourhood size, $n = 3$)

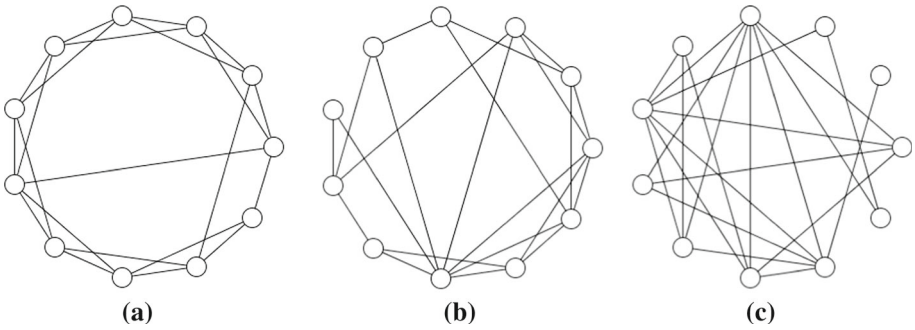

Fig. 5 Small world topologies. **a** RP = 0.1, **b** RP = 0.2, **c** RP = 0.8

on their neighbours. Increasing the population size simply increases the number of such overlapping regions. It is worth noting here that by overlapping localities we are referring to what is known as the *clustering coefficient* in graph theory, which refers to the likelihood of nodes clustering with each other.

6 Metanorms in small worlds

While lattices are regular structures, as opposed to random structures, Watts and Strogatz [40] noted that many biological, technological and social networks lie somewhere between the two: neither completely regular nor completely random. They instead proposed *small world networks* as a variation of lattices in which agents are connected to others k or fewer hops (on the ring) away, but with some of the connections replaced by connections to other randomly selected nodes in the network, in line with some specific rewiring probability (RP). Examples of such networks with different rewiring probabilities are shown in Fig. 5.

Thus, while lattices essentially create overlapping localities of well connected agents (since agents are connected to $2n$ agents immediately surrounding them), the effect of small world networks is to break some of these connections, replacing them by others. Though the

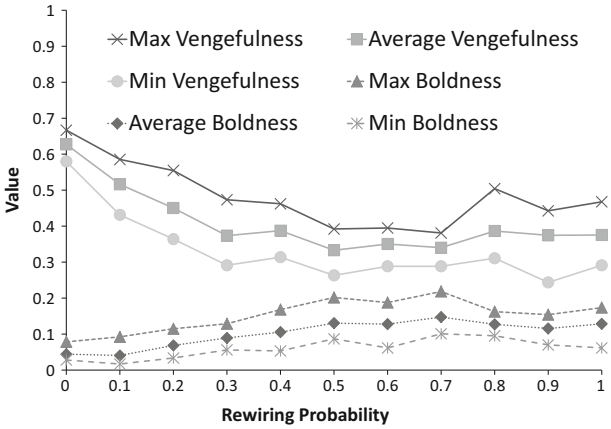

Fig. 6 Small world: impact of rewiring on final B and V (where neighbourhood size, $n = 3$)

number of connections does not change, the locality effect does, since there may no longer be localities of well connected agents, but instead agents with some connections to their local neighbours, and some connections to others elsewhere in the network. A small world network with a rewiring probability of 0 is a regular lattice, while a small world network with a rewiring probability of 1 is a completely random network. As these local regions break down, the strong influence of an agent's local neighbours, causing compliance with norms, should also break down because of the more sparse connections.

To verify this hypothesis, we investigated the impact of the rewiring probability by running the model with different values, in populations of 51 agents, for different neighbourhood sizes. The results of the experiment with a neighbourhood size of 3 are shown in Fig. 6, which indicates that increasing the RP decreases the final average vengefulness in the population, with other neighbourhood sizes giving similar results.

The results obtained are due to the fact that, as a result of rewiring, agents no longer affect just their locality, but now affect agents that are much further away, consequently requiring establishment of the norm in multiple localities. For example, in the case of neighbourhood size of 3, it is clear that not only is the norm not established, but as the RP rises above small values, the trend moves further away from establishment, since the connections of agents are increasingly rewired, giving a locality effect similar to lattices with a neighbourhood size of 2 (discussed in Sect. 5.1). In addition, rewiring to other agents further away brings the need to establish the norm in all those localities to which an agent is connected, making it much more difficult.

In terms of boldness, it is clear that the RP of small world networks has very little impact on the level of defection in the population since, independently, boldness remains very low, indicating that agents are very unlikely to defect.

6.1 Neighbourhood size and rewiring probabilities

As discussed in Sect. 5.1, increasing neighbourhood size causes an increase in vengefulness in lattices. In seeking to understand the impact in small world networks, we repeated the lattice experiments in this new context, for different values of the RP. Results for a rewiring probability of 0.4 are shown in Fig. 7 (with results for other values of the RP being similar in trend), again showing that neighbourhood size increases vengefulness. However, note

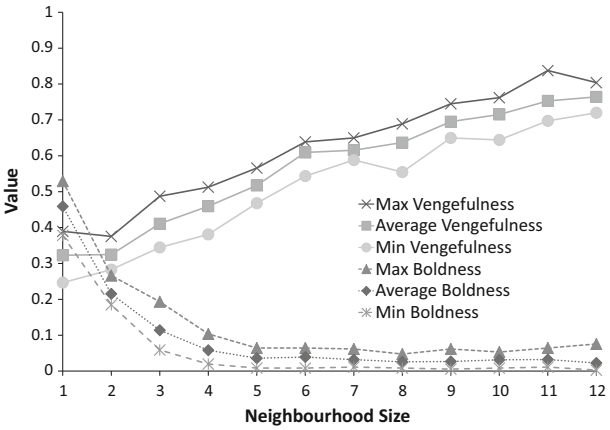

Fig. 7 Small world: impact of neighbourhood size on final B and V (RP = 0.4)

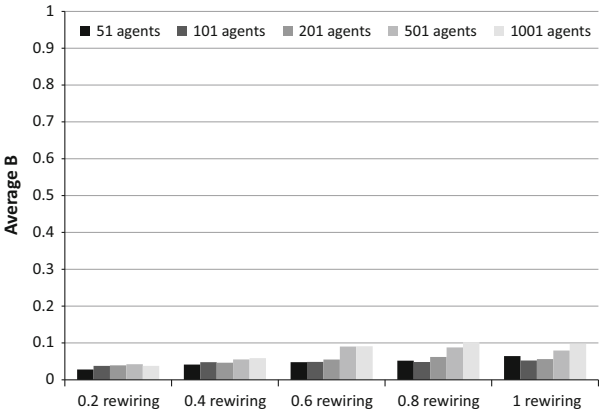

Fig. 8 Small world: impact of rewiring and population size on final boldness (where neighbourhood size $n = 5$)

that, in comparison to lattices, vengefulness in a small world network is lower for the same neighbourhood size. This is because the agents must now respond to defections in different regions of the network, where there is less influence on behaviour, and thus potentially incur greater enforcement costs.

6.2 Population size and rewiring probabilities

Population size has been shown to have no effect on norm establishment in lattices due to the *potential peer pressure* arising from the size of each agent's neighbourhood rather than the total population size. However, since these concentrated local regions of connected agents are weakened in small world networks, we repeated the previous experiments to determine the effect of population size with RPs of 0.2, 0.4, 0.6, 0.8 and 1.0, and n of 5. The results indicate that boldness is not affected by the changes of the population size as boldness is always close to zero, as shown in Fig. 8. But, vengefulness decreases as the RP increases for large population size. More specifically, when the RP is 0.2, increasing the population

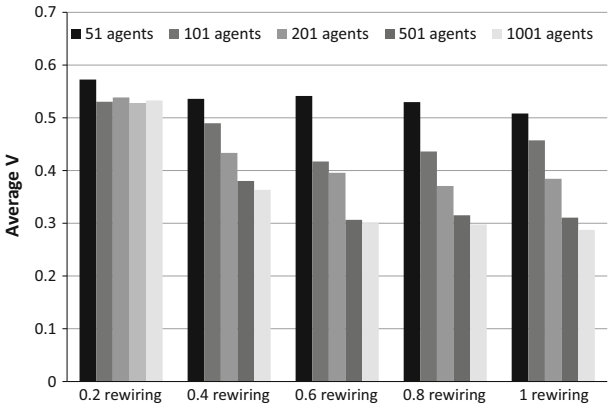

Fig. 9 Small world: impact of rewiring and population size on final vengefulness (where neighbourhood size $n = 5$)

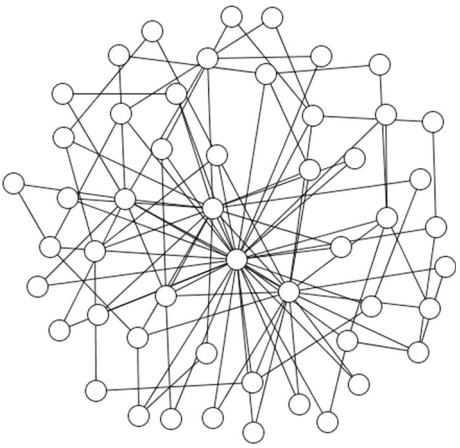

Fig. 10 Scale-free network

size has little effect, as shown in Fig. 9. However, for the other RP values, increasing the population size decreases vengefulness. Again, this is due to rewiring breaking down the strong locality effect, and this is magnified with increasing population sizes, since there is a greater opportunity for connections to other localities, causing a greater cost for agents seeking to bring about norm establishment in all these localities at once.

7 Metanorms in scale-free networks

The topologies considered above are similar in that each agent has exactly the same number of connections, in contrast to scale-free networks [4], in which connectivity of nodes follow the power law distribution. Thus, some nodes have a vast number of connections, but the majority have very few connections, as illustrated in Fig. 10. These properties of scale-free networks suggest an imbalance in connections. In turn, this has an impact on the results that are obtained, due both to punishment and to enforcement costs, which dramatically modify the dynamics of the system. To investigate this, we ran 1000 experiments on 1000 different

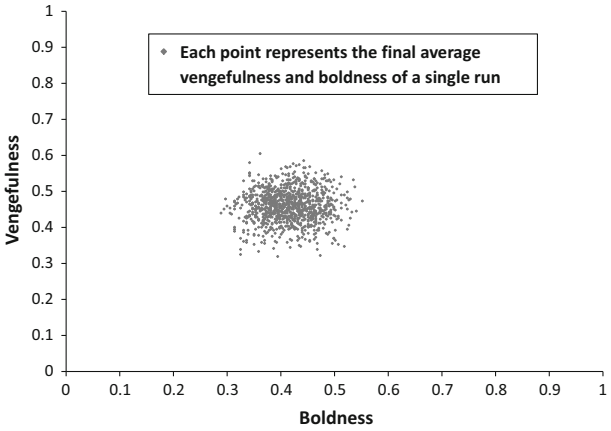

Fig. 11 Scale-free network, 1,000,000 timesteps

scale-free networks with 1000 agents each, five of which were *hubs* (having a large number of connections) and the others (which we call *outliers*) with at least two connections to other agents in the population, and typically no more than four connections (according to Barabási's algorithm [4]). Each experiment was run for 1,000,000 timesteps, and parameters for the experiments were as follows (and are the same for all subsequent experiments reported in this paper): Temptation Value ($T = 3$), Enforcement Cost ($E = -2$), Punishment Cost ($P = -9$) and Hurt Value ($H = -1$). The results, shown in Fig. 11, indicate that all runs end with both average boldness and average vengefulness of midrange values, so that no norm is established. A detailed analysis of individual runs reveals that, overall, there is no significant change to the average vengefulness and boldness, with both fluctuating around a midrange value from the start of the run until the end.

However, certain patterns emerge when examining hubs and outliers separately. Specifically, the model succeeds in lowering the boldness of hubs, but their vengefulness remains near the midrange. Because hubs are connected to many other agents and are thus punished many times for a defection, boldness decreases. Conversely, they also punish many of these other agents for defecting, and consequently pay a very high cumulative enforcement cost that causes them to lower their vengefulness. In turn, this lower vengefulness causes them subsequently not to punish others and as a result to receive metapunishment from other agents, leading to an increase in vengefulness again. Over time, this repeats, with vengefulness decreasing and then increasing back to the midrange, as shown in Fig. 12. For the remaining *outlier* agents, changes to boldness and vengefulness are indicative of the overall boldness and vengefulness because they comprise the majority of the population. They are typically connected to one or more of the hubs, and while they too defect and punish, they do so much less frequently than the hubs to which they are connected. Thus, their scores are generally higher than the scores of the hubs. Since those agents with higher scores do not learn from others (since there are no higher scoring others to learn from), they do not change their strategies, and their boldness and vengefulness remain close to the midrange value, as shown in Fig. 13. These results demonstrate that our algorithm is not effective in scale-free networks. Importantly, as the burden of punishment falls largely on hubs rather than outliers, hubs perform worst in the population. To address this, we modify the learning technique so that it can cope with the nature of scale-free networks. The updated algorithm is discussed next.

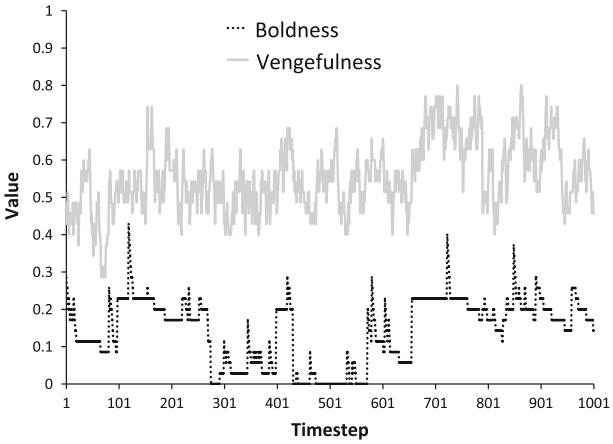

Fig. 12 Hubs in scale-free networks

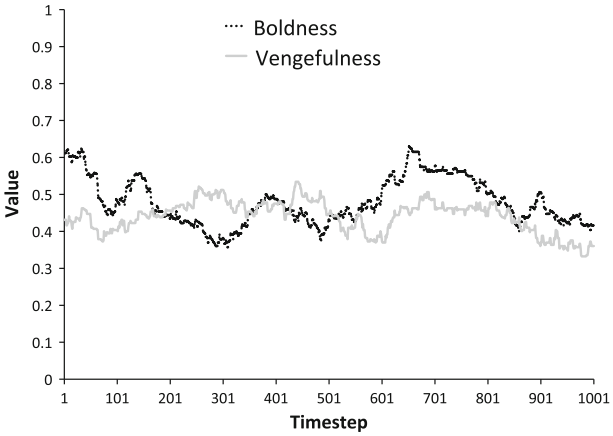

Fig. 13 Outliers in scale-free networks

7.1 Universal learning

The algorithm proposed earlier suffers from the limitation that it requires knowledge of the average score in the population in order for an agent to determine whether to modify its policies. In some domains, for example with online games, such information is public knowledge. However, in domains like P2P file sharing, where the actual underlying structure is of a scale free nature [21], such information may not readily be available. It therefore makes little sense to assume that agents have access to an average population score against which to compare themselves before deciding whether to modify their policies. For this reason, we consider here an alternative approach, in which agents always modify their policies to improve performance, regardless of the behaviour of others, and only in relation to their own score. This modification is simple, and involves removing line 3 of Algorithm 5. Following this, agents ignore the performance of others in the system, and change their policies based only on their own performance, which can be inferred from the different scores they have accumulated in a specific round.

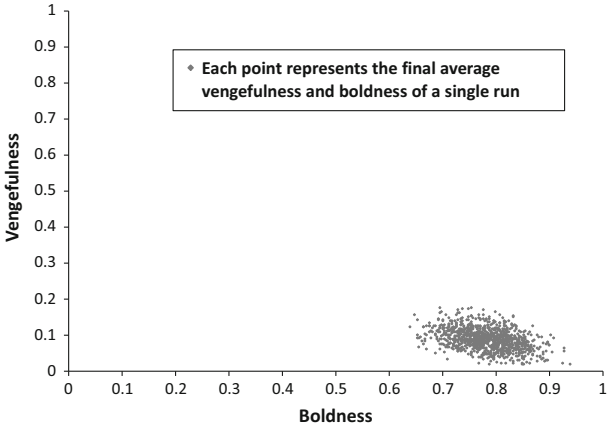

Fig. 14 Universal learning, 1,000,000 timesteps

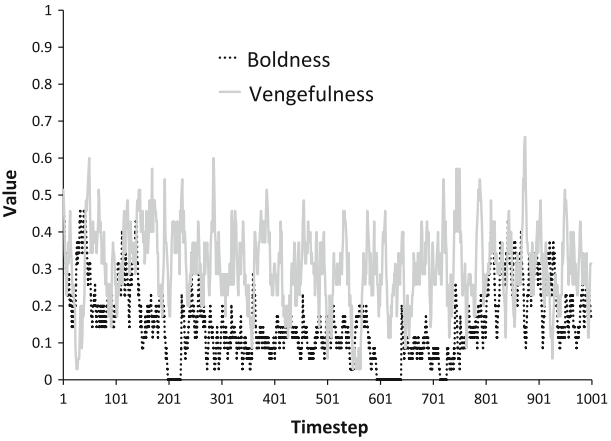

Fig. 15 Universal learning: Hubs

Experiments with this new approach give the results shown in Fig. 14. Counterintuitively, the results indicate norm collapse, as all runs end with high boldness and low vengefulness. By analysing the performance of the different types of agents, we are able to explain this behaviour. We illustrate by reference to a sample run for a hub in Fig. 15, and a sample run for an outlier agent shown in Fig. 16.

Outliers have few connections, but are connected to one or more hubs. When agents punish others, they pay an enforcement cost but risk metapunishment when they do not. However, since these outliers have very low connectivity, the risk of metapunishment is also very low, so they avoid punishing others and vengefulness consequently decreases. Metanorms are not effective here because of the lack of connectivity between agents. As a result, outliers always have high boldness and low vengefulness levels. In addition, and as we will see, the vengefulness of hubs also drops and is never higher than the midrange level, so agents can defect and gain the benefit of doing so, without being punished by hubs. Outliers thus increase their boldness, causing norm collapse in the whole population.

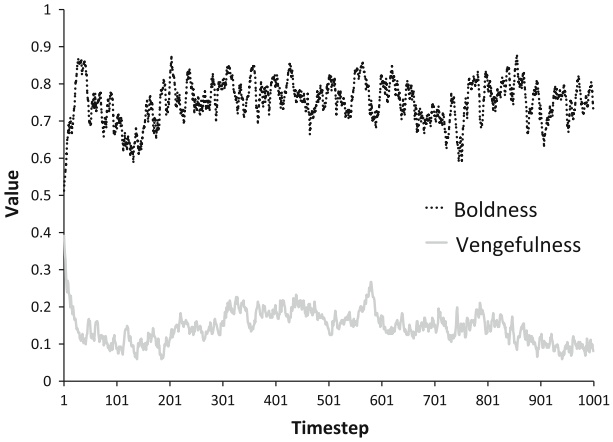

Fig. 16 Universal learning: outliers

In contrast to outliers, hubs are highly connected. Therefore, vengeful hubs apply punishments to many others, incurring high enforcement costs. To address this, they decrease their vengefulness, resulting in metapunishment from the many nodes to which they are connected, in turn causing hubs to increase their vengefulness (but only to a mid-range level). In addition, because of the high boldness of outliers, there is a high rate of defection in the population, causing oscillation between mid-range and low vengefulness. Boldness of hubs is kept at a low level, however, due to the amount of punishment that the hubs are exposed to.

7.2 Connection-based observation

As in Axelrod's original model, our experiments have assumed that deviant behaviour has a small risk (probability) of being observed. In the context of a fully connected network, this is a reasonable assumption to incorporate as agents are unlikely to continually observe all others in the community. However, in the kinds of topologies we are concerned with, such as those that reflect the situation in peer-to-peer (P2P) networks or wireless sensor networks, for example, observation of the behaviour of others arises from a direct connection between agents. Thus, if a peer x is connected to another peer y, then x is able to observe all communication from y. As a result, if y defects by, for example, not sharing files in the case of a file-sharing P2P network, this is observed by x. To reflect this property in our model, Axelrod's probability of being seen requires replacing with the notion that each agent observes all actions of its direct neighbours. This modification to the model gives rise to rather different results.

In particular, the results of running the model on a scale-free network, in Fig. 17, show that all runs end in low boldness and low vengefulness, indicating that defection is very rare in the population because of the low boldness. In addition, punishment is not common since agents rarely punish defectors, due to their low vengefulness. To understand this better, the results of the first 1,000 timesteps of a sample run, for outliers and hubs, are shown in Figs. 18 and 19, respectively.

More specifically, Fig. 18 shows that outliers start the run by decreasing both vengefulness and boldness to a low level where they remain, with some small degree of fluctation. Figure 19

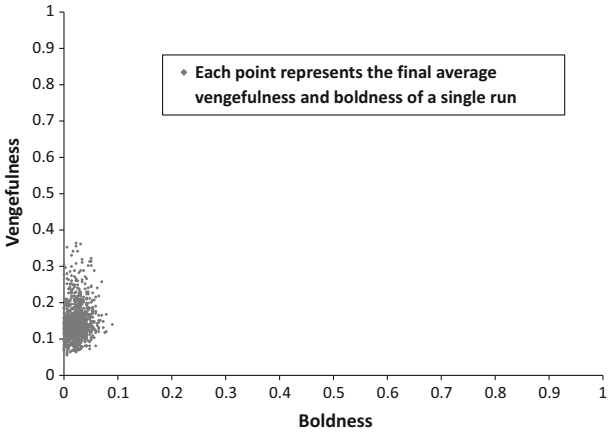

Fig. 17 Connection-based observation, 1,000,000 timesteps

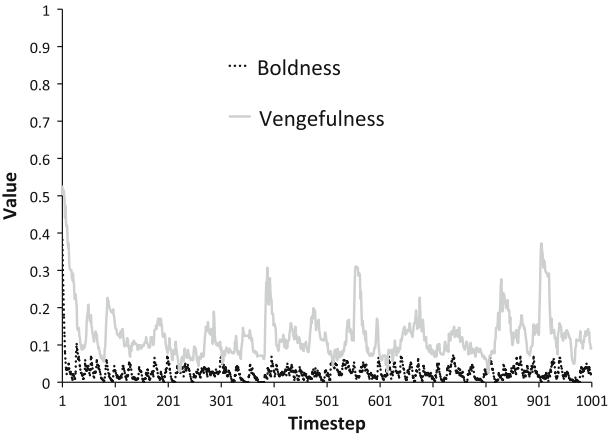

Fig. 18 Connection-based observation: outliers

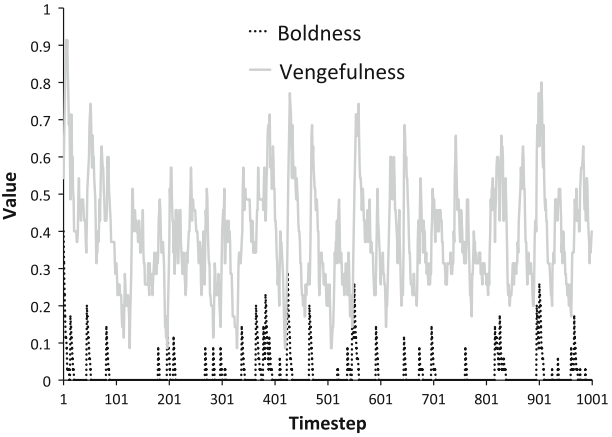

Fig. 19 Connection-based observation: Hubs

suggests that hubs start the run by increasing their vengefulness to a high level and decreasing their boldness to a very low level. After a few timesteps, vengefulness decreases to a mid-range level, from which it decreases further to a low level. However, it does not stabilise there, since it moves up again, and this pattern is repeated throughout the run. Similarly, boldness initially decreases to zero and then jumps to a low level, before decreasing back to zero. Hubs thus have a fluctuating mid-range level of vengefulness, and a very low level of boldness.

There are two distinctive features that can be observed here, in contrast to the results obtained by the universal learning approach. First, hubs reach a high level of vengefulness, which is limited to mid-range vengefulness in the previous approach. This is mainly because the new technique raises the action observation probability to 100%, which allows a high possibility for metapunishment to occur and, as a result, forces hubs to increase their vengefulness to a high level. However, as before, this does not persist because of the high enforcement cost observed with such a high level of vengefulness. Second, the boldness of outliers is low here, mainly due to the combination of the high vengefulness among hubs and the 100% defection observation, which together produce sufficient punishments to force outliers to decrease their boldness.

7.3 Dynamic policy adaptation

As we have seen, universal learning has a negative impact on the results, causing boldness to increase and vengefulness to decrease. However, there is a more important weakness of the model in that the learning rate is uniform in the face of differing punishment levels. More specifically, all agents use the same learning rate, regardless of how much utility gain or loss they suffer. Thus, for example, an agent that incurs a very small punishment score modifies its vengefulness to exactly the same degree as one whose punishment score is very large. While the direction of change is appropriate, the degree of change does not reflect the severity of the sanction; a more appropriate approach would be to change policy in line with performance. In this view, a very badly performing agent should modify its policy much more significantly than one that does not perform as poorly. In this section, we consider dynamic policy adaptation to address this weakness, and to bring about changes to vengefulness and boldness that reflect performance.

The key notion underlying our technique is to measure the *level* of performance rather than just the *direction*. This can be achieved through comparison of an agent's actual utility or *score* in our terms, and the maximum or minimum that could be obtained. We apply this principle to boldness and vengefulness in turn. Before proceeding, we introduce some notation. Let NDD be the number of available defection decisions, where each agent can have more than one chance to defect in a single round (as specified earlier), $|NB_i|$ be the number of i's neighbours, T be the utility that can be gained from a single defection, and P be the punishment cost that represents the utility lost from being punished.

7.3.1 Boldness

To learn the optimal boldness level, the relevant part of the total score is the *defection score*, which can be either positive or negative, requiring consideration of both maximum and minimum possible values. The maximum possible defection score $MaxDS_i$ arises when an agent i always defects but is never punished, as follows.

$$MaxDS_i = NDD \times T \tag{1}$$

In contrast, the minimum defection score that can be obtained by an agent arises when the agent always defects and is always punished by all of its neighbours, as follows.

$$MinDS_i = NDD \times (T + (|NB_i| \times P)) \tag{2}$$

Then, in order to determine the degree of change to an agent i's boldness ($FactorB_i$, see Eq. 3), we must consider three different situations. First, when the defection score is positive (so that boldness should increase), the degree of change is determined by dividing the obtained defection score by the maximum possible defection score. Second, when it is negative, (so that boldness should decrease), the obtained defection score is divided by the minimum possible defection score. Finally, if the defection score is zero, no change is required.

$$FactorB_i = \begin{cases} \frac{DS_i}{MaxDS_i} & \text{if } DS_i > 0 \\ \frac{DS_i}{MinDS_i} & \text{if } DS_i < 0 \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

Given this, we now need to determine how $FactorB_i$ can be used to change agent i's policy. In order to avoid dramatic policy movements that could lead to violent fluctuations, we limit the change that can be applied to a maximum value. In this case, the maximum value is the difference between two levels as in Axelrod's original model, of $\frac{1}{7}$. Thus, in terms of boldness, agent i modifies its boldness in line with its DS_i, as follows.

$$B_i = B_i + \begin{cases} \frac{1}{7} \times FactorB_i & \text{if } DS_i > 0 \\ -\frac{1}{7} \times FactorB_i & \text{if } DS_i < 0 \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

This means that an agent can maximally change its boldness by one level (or by $\frac{1}{7}$) when $FactorB$ is 1.

7.3.2 Vengefulness

An agent modifies its vengefulness depending on whether it is valuable to punish others, determined by comparing the utility lost from punishing others (the punishment score, PS) against the utility lost from not punishing them (the punishment omission score, POS). If PS is worse than POS, then an agent decreases vengefulness and increases it otherwise. Clearly, the magnitude of the difference between these two values gives an indication of the degree of change that should be applied to vengefulness. For example, if PS is -24 and POS is -20, then the amount of decrease to V should be significantly lower than when PS is -600 and POS is -20. We call this difference $DiffV$:

$$DiffV_i = |PS_i - POS_i| \tag{5}$$

Since the values of PS and POS are integers, the absolute value of their difference, $DiffV$, is 1 or more (when the values are not equal). This cannot be used directly to update an agent's V value, because V must always lie between 0 and 1. It must thus be *normalised* so that it can be applied to V, for which we use a scaled value, $FactorV$. This is determined by dividing $DiffV$ by the minimum of PS and POS. Since both PS and POS are negative, the absolute value of the minimum of these two scores is used for the scaling:

$$FactorV_i = \frac{DiffV_i}{|\min\{PS_i, POS_i\}|} \tag{6}$$

While this always produces a value between 0 and 1, it does not provide the same value for the same magnitude of difference. For example, if PS is -14 and POS is -20, we want $FactorV_i$ to be the same as when PS is 0 and POS is -6. We can achieve this by replacing $|\min\{PS_i, POS_i\}|$ with the maximum possible difference between PS and POS. This maximum difference is the difference from 0 (obtained when there is no cost at all from punishing or from not punishing) and the greatest possible magnitude of PS or POS. The highest punishment score HPS (the maximum in magnitude, and lowest in numerical terms — we use HPS to indicate the *highest* score to avoid ambiguity of minimum and maximum) is received by an agent punishing all of its neighbours for defection, and metapunishing all of its neighbours for not punishing all of their neighbours for defection.

To determine the value of HPS, we need to consider both the punishment enforcement cost and the metapunishment enforcement cost. First, the highest (maximum in magnitude, but minimum numerically) *punishment* enforcement cost ($HPEC$) arises when all of an agent's neighbours defect and the agent punishes all of them:

$$HPEC_i = NDD \times |NB_i| \times E \tag{7}$$

where E is the enforcement cost of a single punishment. Similarly, the highest *metapunishment* enforcement cost ($HMPEC$) arises when all of an agent's neighbours do not punish all of their neighbours for defecting, and the agent metapunishes all of them:

$$HMPEC_i = NDD \times |NBB_i| \times E \tag{8}$$

where $|NBB_i|$ is the total number of neighbours of all of agent i's neighbours.

Based on this, the highest punishment score of agent i is defined as follows:

$$HPS_i = HPEC_i + HMPEC_i \tag{9}$$

In the same way, the highest punishment omission score $HPOS$ (greatest in magnitude, lowest numerically) can be obtained when an agent does not punish any defectors, but is metapunished by all of its neighbours, as follows:

$$HPOS_i = NDD \times |NB_i| \times (|NB_i| - 1) \times P \tag{10}$$

where the maximum number of defectors is all of an agent's neighbours (NB), the maximum number of metapunishers is the same but excluding the defecting agent, and P is the punishment cost obtained from being metapunished (which is the same as for simply being punished).

Given this, $FactorV_i$ of agent i can be calculated (see Eq. 11) by dividing $DiffV$ by one of these values, as follows. If punishing brings a greater utility reduction than not punishing ($PS < POS$), then we use the highest punishment score HPS. Conversely, if $PS > POS$, then we use the highest punishment omission score $HPOS$. If there is no difference, then there is no change and $FactorV$ is equal to 0.

$$FactorV_i = \begin{cases} \frac{DiffV_i}{|HPS_i|} & \text{if } POS_i > PS_i \\ \frac{DiffV_i}{|HPOS_i|} & \text{if } POS_i < PS_i \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

This guarantees that the change made to V is always the same given the same difference in scores, since both HPS and $HPOS$ are fixed for each agent. Moreover, this approach allows hubs to change much less quickly than outliers, because the highest (maximum in magnitude) scores for hubs are much higher than for outliers, so that the results achieved by

using $FactorV$, and dividing by the difference in scores obtained for hubs, is much less than for outliers.

According to $FactorV_i$, agent i thus increases vengefulness when it finds that not punishing is worse than punishing, and it decreases vengefulness when the converse is true, as follows:

$$V_i = V_i + \begin{cases} \frac{1}{7} \times FactorV_i & \text{if } PS_i > POS_i \\ -\frac{1}{7} \times FactorV_i & \text{if } PS_i < POS_i \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

7.3.3 Example

To illustrate, assume that a hub x is connected to 20 other agents, and that an outlier y is connected to only 2 other agents (one being a hub). As in Axelrod's seminal experiments and without loss of generality, let $NDD = 4$ for all agents, since every agent has 4 chances to defect in each round. $E = -2$ and is the same for all agents. Similarly, $P = -9$ and again is the same for all agents. The temptation value for all agents, received when they defect, is $T = 3$. Finally, suppose that x's neighbours have 50 other distinct neighbours in total (summed over all neighbours), while y's neighbours have 20 other distinct neighbours (again, over all). This is summarised in Table 2. Given these values, we can determine the relevant values needed as follows. Starting with defection scores and from Eqs. 1 and 2 respectively, we obtain the following:

$$\begin{aligned} MaxDS_x = MaxDS_y &= 4 \times 3 = 12 \\ MinDS_x &= 4 \times (3 + (20 \times -9)) = -708 \\ MinDS_y &= 4 \times (3 + (2 \times -9)) = -60 \end{aligned}$$

In terms of punishment scores, from Eqs. 7, 8 and 9, we obtain the following:

$$\begin{aligned} HPEC_x &= 4 \times 20 \times -2 = -160 \\ HMPEC_x &= 4 \times 50 \times -2 = -400 \\ HPS_x &= -160 - 400 = -560 \\ HPEC_y &= 4 \times 2 \times -2 = -16 \\ HMPEC_y &= 4 \times 20 \times -2 = -160 \\ HPS_y &= -16 - 160 = -176 \end{aligned}$$

Punishment omission scores using Eq. 10 are as follows:

$$\begin{aligned} HPOS_x &= 4 \times 20 \times 19 \times -9 = -13680 \\ HPOS_y &= 4 \times 2 \times 1 \times -9 = -72 \end{aligned}$$

Using this information (Table 2), we can determine the decisions for specific situations. For example, at the start of each run, the population has midrange average boldness and vengefulness (because of the uniform distribution function to generate initial policies). Now, suppose that both x and y also have mid-range boldness and vengefulness. If, after one round, both x and y defected twice (out of their four opportunities to defect), they each gain twice the temptation value T. However, since x is a hub, suppose it is punished 22 times, much more than y, which is only punished twice. This is because the defection score of a hub with

Table 2 Example values for Agents x and y

Agent	Pos	$\|NB\|$	NBB	$MinDS$	$MaxDS$	$LevB$	HPS	$HPOS$	$LevV$
x	Hub	20	50	−708	12	1/7	−560	−13680	1/7
y	outlier	2	20	−60	12	1/7	−176	−72	1/7

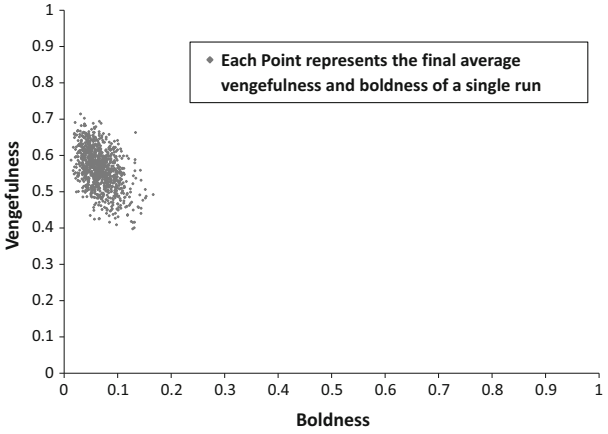

Fig. 20 Dynamic policy adaptation, 1,000,000 timesteps

mid-range boldness is typically much worse than that of a similar outlier, mainly due to the difference in their number of neighbours, and the midrange vengefulness in the population. Thus, x has a defection score of 2×3 from defecting, plus $22 \times -9 = -198$ from being punished, giving $DS_x = -192$. Similarly, $DS_y = ((2 \times 3) + (2 \times -9)) = -12$.

Given these defection scores, the degree of change that each agent applies to its boldness can be calculated as follows. First, from Eq. 3, $FactorB_x = \frac{-192}{-708} = 0.3$ and $FactorB_y = \frac{-12}{-60} = 0.2$. Now, using Eq. 4, and since both DS_x and DS_y are negative, B_x is decreased by $0.3 \times \frac{1}{7} = 0.04$, and B_y by $0.2 \times \frac{1}{7} = 0.03$.

In addition, if x punishes 20 other agents and metapunishes 10 more, and y punishes 2 other agents and metapunishes 1 more, their punishment scores are determined by multiplying the number of punishments issued by the enforcement cost E: $PS_x = ((20+10) \times -2) = -60$ and $PS_y = ((2 + 1) \times -2) = -6$. Then, if x has spared 10 defectors and has been metapunished 2 times for each instance of omitting punishment, and y has spared only one defector and been metapunished just once, the punishment omission scores are calculated by multiplying the number of metapunishments by the punishment cost P, as follows: $POS_x = (10 \times 2 \times -9) = -180$ and $POS_y = (1 \times 1 \times -9) = -9$. Thus, by Eq. 11, $FactorV_x = \frac{|-60-(-180)|}{13680} = 0.01$ and $FactorV_y = \frac{|-6-(-9)|}{72} = 0.04$. Then, since $PS_x > POS_x$, x increases its vengefulness V_x by $0.1 \times \frac{1}{7} = 0.001$ according to Eq. 12. Similarly, since $PS_y < POS_y$, y decreases its vengefulness by $0.04 \times \frac{1}{7} = 0.006$.

7.3.4 Experimental results

To determine the effect of introducing dynamic policy adaptation, we ran experiments, similar to the previous ones, on the new model. The results are visualised in Fig. 20. As can be

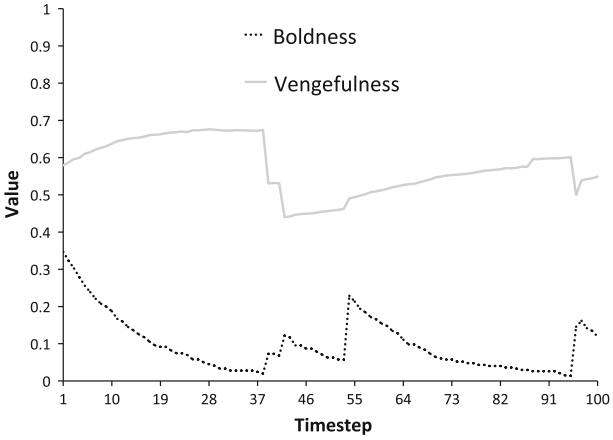

Fig. 21 Dynamic policy adaptation for Hubs

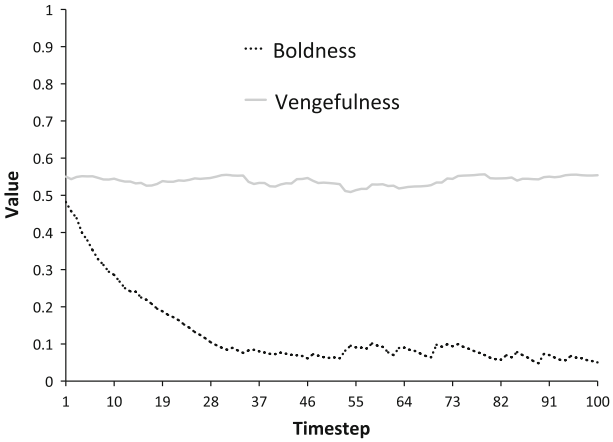

Fig. 22 Dynamic policy adaptation for outliers

seen from the figure, all runs result in populations with low average boldness and moderate vengefulness. As before, more detail on the evolution of average boldness and vengefulness for hubs and outliers was provided by examining runs of individual agents, as shown in Figs. 21 and 22, which confirm that outliers converge to a state of low boldness and moderate vengefulness consistently, while hubs do so with intermittent deviations.

The main difference from previous results is that the average vengefulness of outliers remains in the midrange level. This is because the change in agent's policies (boldness and vengefulness) is happening at a balanced rate depending on the amount of utility lost or gained. Outliers are decreasing their vengefulness at much slower rate, because the utility they are losing as a result of enforcing the norm is not considered as significant as before. This is allowing metapunishment to occur frequently enough to maintain this level of vengefulness among all agents. On the other hand, hubs are much more exposed to this metapunishment due to the number of connections they have, and so they maintain a high level of vengefulness. However, since there are few hubs in the population, an exploration (an agent adopting a very low level of vengefulness due to exploration) to a single hub can have a large effect on the

average vengefulness level among hubs. This is the reason for the sudden reductions shown in the figure, which are quickly restored to high level until the next significant exploration.

8 Metanorms in real world network

The results reported in this paper so far have focused on applying different variations of the metanorm model on artificially established topologies. A wide variety of synthetic network generators have been proposed, but tend to be poor models of real-world networks [20,26]. Thus, in this section, we show the outcome of applying the metanorm model on samples of real world networks. We use three networks: (i) a peer connection network from Gnutella (a P2P file-sharing platform), (ii) the Epinions social network, and (iii) the EuAll emails exchange dataset. Examples of these networks are shown in Fig. 23. The Epinions and EuAll networks are both based on human interactions, but are generated by different processes: the EuAll dataset is based on email exchange of members of a large European institute, while the Epinions dataset represents a trust relationship between members of general consumer review site, Epinions.com. Alternatively, Gnutella is a computational network of links in a P2P system. Since these networks are generated by different processes they display varied structural properties, allowing us to evaluate our methodology on a range of structures. Use of real-world networks is typically constrained by (i) impractically large node counts, and (ii) limited knowledge of the global structure. Consequently, sampling part of the network is often necessary. Ideally, the sampled structure should display similar properties to the full network. A wide variety of sampling techniques have been proposed (e.g. [14,16,19]). To evaluate our approach, we use Metropolis–Hastings RandomWalk (MHRW) [14], which starts with a randomly chosen node in a seed set. It then performs a random walk with biased transition probabilities, with the aim of producing a uniform sample results. The sampled networks used here have been used by Franks et al. [12] for studying influence in social network.

Figure 24 shows the result obtained from applying the model on a sampled 1000 nodes of the Gnutella network, with similar results obtained from samples of the other two networks. The results are similar to those of scale-free networks, which is reasonable given that scale-free network are usually considered to be the closest to real world network. This results show that the developed adaptive policy learning technique allows norms to be established in real world settings.

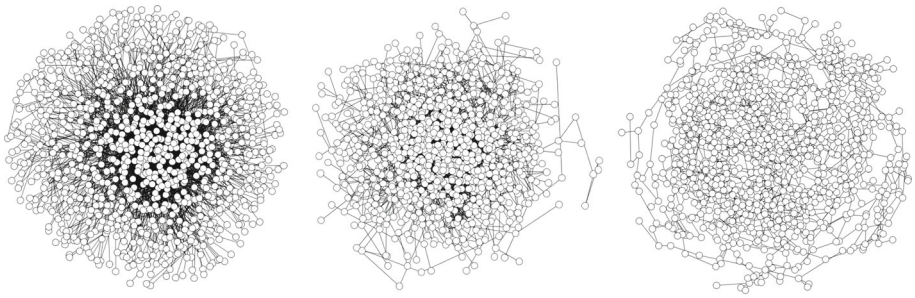

Fig. 23 Real world topologies. **a** EuAll—1000 nodes. **b** Epinions—1000 nodes. **c** Gnutella—1000 nodes

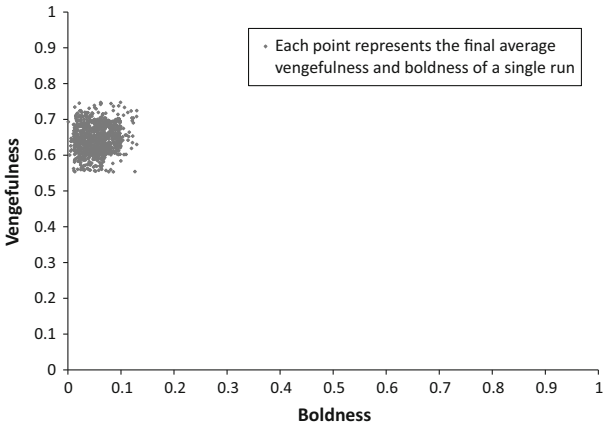

Fig. 24 Gnutella network: 1000 Nodes and 1,000,000 timesteps

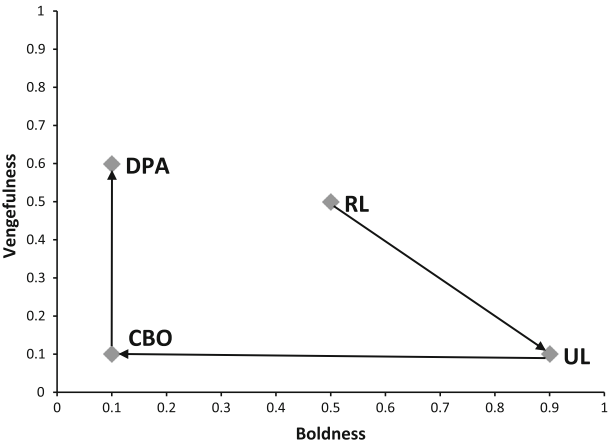

Fig. 25 Evolution of results with scale-free networks

9 Conclusion and discussion

Meta-punishment has been shown to be effective in the real world, for example in denunciation in communist societies, or by enforcing penalties on those who fail to report child abuse. In that context, this paper is an attempt to improve the metanorm model suggested by Axelrod to bring it a step closer to being applicable for distributed computational systems, such as P2P systems or wireless sensor networks, where free riding is a common phenomenon. While addressing the problems arising from the fundamental nature of Axelrod's model, we have also sought to address its limitations arising from the unrealistic assumption that all agents are fully connected so that they can all interact and observe every other agent in the population. Such an assumption is not possible in general for computational systems for many reasons. First, the number of agents in such systems can be so large that it is possible neither to manage the traffic generated from their interactions nor to observe all other agents. Second, the connections between agents may be subject to different network topologies not considered by Axelrod. To address these issues, in this paper we adapted the model to be

effective with different interaction topologies and their impact on norm emergence. We were able to show that our model achieves norm emergence over both lattices and small world networks, but not in scale-free networks. More precisely, the refined model is very effective for lattices, but its effectiveness varies with the rewiring probability in small world networks. Moreover, we demonstrated that, given fixed penalties, for lattices, the effectiveness of the proposed approach depends only on the number of neighbours of each agent, *not* on the total population size. For small world networks, increasing the population size with a high rewiring probability decreases vengefulness, constraining norm emergence significantly.

In scale-free networks, however, the results of this new model are not as good as for the other topologies due to the nature of the connection distribution between hub and outlier nodes. Because of the vast number of connections of hubs in scale-free networks, their interactions are much more frequent than outliers, causing them to be the only agents that learn as a result. To address this, the model was modified to remove the restrictions on learning which constrained it to apply only to poorly performing agents, and to limit knowledge of the performance of others, which is unreasonable in computational systems. While this new *universal learning* (UL) technique allows all agents to improve performance, it unfortunately leads only to norm collapse due to the constraints imposed by the limited observability of outlier behaviour. In fact, this limited observability is another inadequacy of Axelrod's original formulation of the model with a uniform probability of being seen; in real systems, observability is restricted by network connections rather than some arbitrary probability distribution. In response, a *connection-based observation* (CBO) technique reduces the tendency of agents to defect, without maintaining their tendency to punish defectors. Yet all of these improvements do not fundamentally impact on the problems arising from the asymmetric nature of scale-free networks with hubs and outliers, largely because the use of a uniform learning rate to modify strategies is ineffective. Our final refinement, therefore adjusts the amount of learning in relation to performance through *dynamic policy adaptation* (DPA), bringing about the desired behaviour and norm emergence. In particular, this adaptation mechanism is effective when applied to samples of networks from real world applications. This pattern of development of mechanism is illustrated clearly in Fig. 25, which shows the progress from the reinforcement learning (RL) technique that is the starting point early in this paper through to DPA via UL and CBO.

As stated by Axelrod [3], the role of agent-based simulations is not to provide an accurate representation of the real world, but to raise awareness of phenomena that can occur in the real world. Moreover, concepts used in building a simulation and results obtained from such simulations can provide a certain understanding, which helps when dealing with comparable questions in the real world. From the analysis presented in this paper, we can observe that the underlying structure of any system is clearly an important factor, and has many direct and indirect influences on how the overall system functions. More specifically, the effect of a homogenous structure such as a lattice is clearly less significant than it is for a complex structure like a scale-free network. A small change to the structure, such as moving from a lattice to a small world network, where agents still have the same number of connections but the locality of these connections is less, may also lead to considerable change in the achieved results. Hubs in complex networks are clearly more influential than other nodes, and should be treated with special care. A second observation is related to the ability of agents to monitor the interactions that occur among others. This is clearly important as shown when we move from a specific observation probability to the idea of neighbourhood based observation. The greater the observation capabilities that exist in the population, the better the chance that violations will be detected and dealt with. However, observation is not cost-free and, as discussed above, some domains may have privacy issues, which restrict observability.

Overall, we clearly see that the idea of metapunishment works very well in motivating a high level of responsibility among agents to defend the norms of the society.

Despite the successful results obtained from the substantially refined metanorm model presented in this paper, there are still some limitations and assumptions that may prevent the model from being directly applied in some real world applications. While observation of the interaction of others may be considered a valid assumption in some domains such as social networks (for example, a person is able to observe the interactions of their friends with other people), this may be invalid in other domains (for example, communication over the internet between nodes may involve some form of encryption, which can prevent agents even from detecting a violation). In such domains, it will be difficult for the metanorm model to function, since it is highly dependent on observation for the metapunishment to occur. In addition, while the current static punishments, set at design time, seem to produce satisfactory results, identifying a value that will always work for these static punishments can prove difficult in complex unpredictable environments such as the internet. So, having an adaptive decision making mechanism, by which agents can decide on the punishment value based on the context might be more suitable. We believe that the dynamic policy adaptation technique introduced in this paper provides a solid grounding for such an adaptive punishment approach. Finally, resources are an important factor that are currently ignored in our model, and in the literature on norm emergence in general. It has been generally assumed that agents have access to unlimited resources they can use in enforcement, but this is clearly not a real-world property. Adaptive punishment and limited resources are an important line of research, which we plan to investigate as future work. In addition, we plan to conduct a more detailed analysis of the effect of different levels of the probability of observation on the results of the model. On a different line of attack, but an interesting phenomenon that we aim to explore in the future, is the existence of agents that identify the fixed punishment strategies followed by other agents, and try to exploit them by developing strategies against them. One way to avoid this might be to combine agents with different formulae and different observational capabilities in the same network. While we believe this will make the model much stronger for real applications, the fundamental mechanisms proposed and demonstrated in this paper already present the core aspects of a more general functional approach to establishing norms and metanorms in complex topological structures.

References

1. Axelrod, R. M. (1984). *The evolution of cooperation*. New York: Basic Books.
2. Axelrod, R. (1986). An evolutionary approach to norms. *The American Political Science Review*, *80*(4), 1095–1111.
3. Axelrod, R. (1997). *The complexity of cooperation: agent-based models of competition and collaboration*. Princeton: Princeton University Press.
4. Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*(5439), 509–512.
5. Boman, M. (1999). Norms in artificial decision making. *Artificial Intelligence and Law*, *7*(1), 17–35.
6. Davis, J., Laughlin, P., & Komorita, S. (1976). The social psychology of small groups: Cooperative and mixed-motive interaction. *Annual Review of Psychology*, *27*(1), 501–541.

7. de Pinninck, A. P., Sierra, C., & Schorlemmer, W. M. (2007). Friends no more: norm enforcement in multiagent systems. In *Proceedings of the sixth international joint conference on autonomous agents and multi-agent systems* (pp. 640–642).
8. Delgado, J. (2002). Emergence of social conventions in complex networks. *Artificial Intelligence*, *141*(1–2), 171–185.
9. Delgado, J., Pujol, J. M., & Sangüesa, R. (2003). Emergence of coordination in scale-free networks. *Web Intelligence and Agent Systems*, *1*, 131–138.
10. Epstein, J. M. (2001). Learning to be thoughtless: Social norms and individual computation. *Computational Economics*, *18*(1), 9–24.
11. Flentge, F., Polani, D., & Uthmann, T. (2001). Modelling the emergence of possession norms using memes. *Journal of Artificial Societies and Social Simulation*, *4*(4). http://jasss.soc.surrey.ac.uk/4/4/3.html.
12. Franks, H., Griffiths, N., & Anand, S. S. (2013). Learning influence in complex social networks. In *Proceedings of the 2013 international conference on autonomous agents and multi-agent systems*, AAMAS '13 (pp. 447–454), Richland, SC, International Foundation for Autonomous Agents and Multiagent Systems.
13. Franks, H., Griffiths, N., & Jhumka, A. (2012). Manipulating convention emergence using influencer agents. *Autonomous Agents and Multi-Agent Systems*, *26*(3), 315–353.
14. Gjoka, M., Kurant, M., Butts, C.T., & Markopoulou, A. (2010). Walking in facebook: A case study of unbiased sampling of osns. In *2010 Proceedings IEEE on INFOCOM* (pp. 1–9).
15. González, M. C., Lind, P. G., & Herrmannc, H. J. (2006). Networks based on collisions among mobile agents. *Physica D-nonlinear Phenomena*, *224*, 137–148.
16. Jin, L., Chen, Y., Hui, P., Ding, C., Wang, T., Vasilakos, A. V., Deng, B., & Li, X. (2011). Albatross sampling: Robust and effective hybrid vertex sampling for social graphs. In *Proceedings of the 3rd ACM international workshop on MobiArch*, HotPlanet '11 (pp. 11–16).
17. Kandori, M., Mailath, G. J., & Rob, R. (1993). Learning, mutation, and long run equilibria in games. *Econometrica*, *61*(1), 29–56.
18. Kittock, J. (1995). Emergent conventions and the structure of multi–agent systems. In *Lectures in complex systems: The proceedings of the 1993 Complex Systems Summer School, Santa Fe Institute Studies in the Sciences of Complexity Lecture Volume VI, Santa Fe Institute*, Addison-Wesley (pp. 507–521).
19. Lee, C., Xu, X., & Eun, D. Y. (2012). Beyond random walk and metropolis-hastings samplers: Why you should not backtrack for unbiased graph sampling. *ACM SIGMETRICS Performance Evaluation*, *40*(1), 319–330.
20. Leskovec, Jure, Lang, Kevin J., Dasgupta, Anirban & Mahoney, Michael W. (2008). Statistical properties of community structure in large social and information networks. In *Proceedings of the 17th international conference on World Wide Web*. WWW '08, New York, NY, USA, ACM. pp. 695–704.
21. Liu, L., Jie, X., Russell, D., Townend, P., & Webster, D. (2009). Efficient and scalable search on scale-free p2p networks. *Peer-to-Peer Networking and Applications*, *2*(2), 98–108.
22. Mahmoud, S., Griffiths, N., Keppens, J., & Luck, M. (2012). Establishing norms for network topologies. *Coordination* (pp. 203–220). Organizations, institutions, and norms in agent system VII, volume 7254 of Lecture notes in computer science. Berlin: Springer.
23. Mahmoud, S., Griffiths, N., Keppens, J., & Luck, M. (2013). Norm emergence through dynamic policy adaptation in scale free networks. *Coordination* (pp. 123–140). Organizations, institutions, and norms in agent systems VIII, volume 7756 of Lecture notes in computer science. Berlin: Springer.
24. Mahmoud, S., Griffiths, N., Keppens, J., Taweel, A., Bench-Capon, T. J. M., & Luck, M. (2015). Establishing norms with metanorms in distributed computational systems. *Artificial Intelligence and Law*, *23*(4), 367–407.
25. Mungovan, D., Howley, E., & Duggan, J. (2011). The influence of random interactions and decision heuristics on norm evolution in social networks. *Computational and Mathematical Organization Theory*, *17*(2), 152–178.
26. Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Review*, *45*(2), 167–256.
27. Newman, M. E. J., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E*, *69*(2), 026113.
28. O'Riordan, C., Cunningham, A., & Sorensen, H. (2008). Emergence of cooperation in n-player games on small world networks. In S. Bullock, J. Noble, R. Watson, & M. A. Bedau (Eds.), *Artificial life XI: Proceedings of the eleventh international conference on the simulation and synthesis of living systems* (pp. 436–442). Cambridge, MA: MIT Press.
29. O'Riordan, C., & Sorensen, H. (2008). Stable cooperation in the n-player prisoner's dilemma: The importance of community structure. In K. Tuyls, A. Nowe, Z. Guessoum, & D. Kudenko (Eds.), *Adaptive agents*

and multi-agent systems III. Adaptation and multi-agent learning, volume 4865 of lecture notes in computer science (pp. 157–168). Berlin: Springer.
30. Savarimuthu, B. T. R. (2008). Stephen Cranefield, Maryam Purvis, and Martin Purvis. Role model based mechanism for norm emergence in artificial agent societies. In *Coordination, organizations, institutions, and norms in agent systems III, COIN 2007 international workshops* (pp. 203–217), volume 4870 of Lecture notes in computer science. Springer.
31. Savarimuthu, B. T. R., Cranefield, S., Purvis, M. & Purvis, M. (2007). Norm emergence in agent societies formed by dynamically changing networks. In *IAT '07: Proceedings of the 2007 IEEE/WIC/ACM international conference on intelligent agent technology* (pp. 464–470).
32. Savarimuthu, B. T. R., Purvis, M., Purvis, M., & Cranefield, S. (2009). Social norm emergence in virtual agent societies. In M. Baldoni, T. C. Son, M. B. van Riemsdijk, & M. Winikoff (Eds.), *Declarative agent languages and technologies VI* (Vol. 5397, pp. 18–28)., Lecture notes in computer science Berlin: Springer.
33. Sen, S. & Airiau, S. (2007). Emergence of norms through social learning. In *IJCAI 2007: Proceedings of the 20th international joint conference on artificial intelligence* (pp. 1507–1512). Morgan Kaufmann Publishers Inc.
34. Sen, O. & Sen, S. (2010). Effects of social network topology and options on norm emergence. In *Proceedings of the fifth international conference on coordination, organizations, institutions, and norms in agent systems* (pp. 211–222).
35. Shoham, Y. & Tennenholtz, M. (1992). Emergent conventions in multi-agent systems: Initial experimental results and observations (Preliminary Report). In *Proceedings of the 3rd international conference on KR&R* (pp. 225–232)
36. Shoham, Y., & Tennenholtz, M. (1995). On social laws for artificial agent societies: Off-line design. *Artificial Intelligence*, *73*(1–2), 231–252.
37. Urbano, P., Balsa, J., Antunes, L. & Moniz, L. (2008). Force versus majority: A comparison in convention emergence efficiency. In *Coordination, organizations, institutions and norms in agent systems IV: COIN 2008 international workshops,* COIN@AAMAS 2008, Estoril, Portugal, May 12, 2008. COIN@AAAI 2008, Chicago, USA, July 14, 2008. Revised Selected Papers, pp. 48–63.
38. Villatoro, D., Sen, S. & Sabater-Mir, J. (2009). Topology and memory effect on convention emergence. In *Proceedings of the 2009 IEEE/WIC/ACM international conference on web intelligence and intelligent agent technologies* (pp. 233–240). IEEE.
39. Walker, A., & Wooldridge, M. (1995). Understanding the emergence of conventions in multi-agent systems. In V. Lesser (Ed.), *Proceedings of the first international conference on multi-agent systems* (pp. 384–389). Cambridge: MIT Press.
40. Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, *393*(6684), 440–442.
41. Yamashita, T., Izumi, K., & Kurumatani, K. (2005). An investigation into the use of group dynamics for solving social dilemmas. In P. Davidsson, B. Logan, & K. Takadama (Eds.), *Multi-agent and multi-agent-based simulation* (Vol. 3415, pp. 185–194)., Lecture Notes in Artificial Intelligence Berlin: Springer.