Social diversity reduces the complexity and cost of fostering fairness

Theodor Cimpeanu [a], Alessandro Di Stefano [b], Cedric Perret [c], The Anh Han [b,*]

[a] School of Mathematics and Statistics, University of St Andrews, United Kingdom
[b] School Computing, Engineering and Digital Technologies, Teesside University, United Kingdom
[c] College of Life and Environmental Sciences, Exeter University, United Kingdom

ARTICLE INFO

ABSTRACT

Institutions and investors are constantly faced with the challenge of appropriately distributing endowments. No budget is limitless and optimising overall spending without sacrificing positive outcomes has been approached and resolved using several heuristics. To date, prior works have failed to consider how to encourage fairness in a population where social diversity is ubiquitous, and in which investors can only partially observe the population. Herein, by incorporating social diversity in the Ultimatum game through heterogeneous graphs, we investigate the effects of several interference mechanisms which assume incomplete information and flexible standards of fairness. We quantify the role of diversity and show how it reduces the need for information gathering, allowing us to relax a strict, costly interference process. Furthermore, we find that the influence of certain individuals, expressed by different network centrality measures, can be exploited to further reduce spending if minimal fairness requirements are lowered. Our results indicate that diversity changes and opens up novel mechanisms available to institutions wishing to promote fairness. Overall, our analysis provides novel insights to guide institutional policies in socially diverse complex systems.

1. Introduction

Fairness has a deep impact on decision-making and individuals often prefer fair outcomes over payoff-maximising ones [1,2]. For example, fairness concerns emerge and play a crucial role in group interactions, when agents must decide upon outcomes possibly favouring different parts unequally [3]. These concerns arise in many domains — hybrid collectives of humans and machines [4], wildlife management [5], conflict resolution [6] or enforcing global climate change actions [7–9], just to name a few. In this context, several mechanisms have been identified to explain why fairness is widespread in human decision-making, but it is typically assumed to emerge from the actions of individuals within the system. Whether due to risk or safety concerns, there exist several domains, such as the ones listed above, that call for a new set of heuristics aimed at promoting fair outcomes in populations of self-regarding individuals.

Humans have developed considerable machinery used at scale to create policies and to distribute incentives, yet we are always searching for ways in which to improve upon these organisations, often referred to as institutions [10,11]. In these scenarios, external decision-makers must find a trade-off between the cost of the investment and its effectiveness in ensuring high levels of fair behaviour. Several works have provided insights on how best to promote cooperation while also considering the costs of such *interference* [12–16]. Fairness, modelled using the Ultimatum Game [1], has seen relatively little attention in the literature, and previous works have so far only considered an ideal world in which interactions are perfectly homogeneous [15]. Nevertheless, real-world networks of individuals, such as social networks and networks of collaboration, are inherently *heterogeneous* [17]. Moreover, in the context of Evolutionary Game Theory (EGT), scale-free networks imply more than the underlying interaction structure. Heterogeneous graphs can portray social diversity [18], and the inherent inequality that exists between individuals. Individuals vary in influence and accumulated wealth, and this lends diversity particular importance in the quest towards fairness for two main reasons. Firstly, it has been shown to play a key role in the evolution of cooperative behaviours [19,20] and the emergence of fairness [21]. It may enhance the resilience of cooperation, inducing cooperative agents to create assortative clusters, where they reciprocate cooperation [19,22,23]. Secondly, decision-makers can then base their investment strategies not only on the state of the system at a global scale, but also on individual characteristics found only in certain parts of the network. This could potentially allow for novel interference strategies, which would maintain high standards of fairness at a reduced cost. For instance, an institution might decide to focus its efforts on building up deprived neighbourhoods or selectively invest only in very influential samaritans.

---

* Corresponding author.
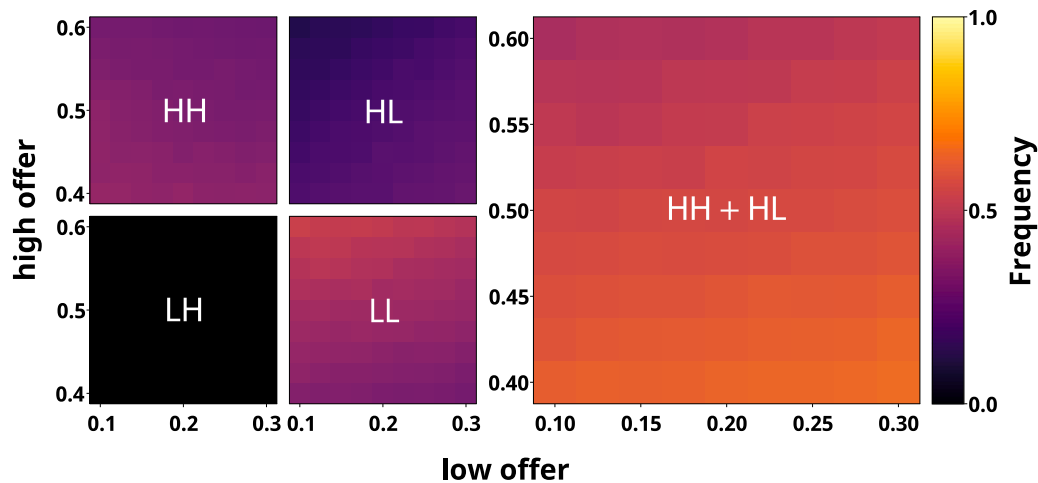 *E-mail address:* T.Han@tees.ac.uk (T.A. Han).

Fig. 1. Baseline frequencies.
The figure shows the baseline frequencies for each strategy for scale-free (BA) networks, with a separate panel for overall fairness (fair offers). Colour corresponds to the frequency, ranging from 'black' (lowest) to 'yellow' (highest) (see colour bar). The left panel shows the frequency of each strategy, which is divided into four quadrants representing the different strategies: HH (upper-left), HL, (upper-right), LH (lower-left), and LL (lower-right). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Herein we systematically investigate several interference mechanisms for promoting the evolution of fairness in heterogeneous network settings assuming incomplete information and flexible standards of fairness. We resort to the Ultimatum Game [1] as a suitable mathematical approach to model fair decision making (see further details in Section 2.1). In the Ultimatum Game, one of the players can decide how to split a sum of money. Offers close to an even split are considered fair. An uneven split, in which the proposer gets to keep most of the money, is considered unfair. Because the proposer has asymmetric power in the interaction, the only way in which they can be "punished" is if the responder declines an unfair offer, thus causing neither individual to receive anything from the original sum. Fair individuals would always propose an even split, and always decline unfair offers to prevent selfish players from receiving part of the endowment. In contrast, very unfair individuals would propose to keep most if not all of the endowment while accepting anything they are given. We determine how these heterogeneous network characteristics can be exploited to reduce costs while maintaining high standards of fairness, providing insights on how the presence of *social diversity* alters the complexity of engineering fairness and how it can be done efficiently.

Optimising fairness becomes especially challenging when studying evolving populations that incorporate diverse stochastic effects and uncertainty factors, such as a non-deterministic behaviour update. Undesired behaviours can reoccur over time and in order to hinder them, external decision-makers must repeatedly interfere in the system. Our focus is therefore on *cost-efficient interference* problems, namely exploring how efficiently to interfere in a spatially heterogeneous population to achieve high levels of fairness while minimising the cost of interference. Given its relevance, this cost-efficient interference problem has naturally attracted significant attention in both evolution of collective behaviours and computational modelling research. It can be formulated as a bi-objective optimisation problem, where the target is dual — finding a cost-efficient interference scheme leading to the desired goal, while minimising the total cost of interference [12,15,16,24].

Our results show that social diversity reduces the complexity and cost of promoting the evolution of fairness in several respects. Positive outcomes can be achieved without extensive information gathering, by reducing the strictness of eligibility for a player to receive an incentive. Moreover, lowering the required standards of fairness allows decision-makers to capitalise on the influence of hubs to reduce costs. Finally, we find that heterogeneous networks have a tendency towards polarisation, and suggest that this can be exploited to induce pro-sociality more efficiently.

2. Models and methods

2.1. Ultimatum game (UG)

In our work, the interaction between agents is modelled using the one-shot Ultimatum Game (UG) [1,25], which serves as a useful framework for addressing and exploring the evolution of fairness. In this game, two players are asked to divide a sum and possibly win a certain amount of money. One player (the *proposer*) suggests how to split the sum. The other player (the *responder*) decides whether to accept or reject the offer. Only if the responder accepts, will the sum be shared as proposed. Otherwise, neither player receives any part of the initial sum. Coherently with [1,25], we assume that a player is equally likely to act in either role. A player's strategy is defined by a pair of parameters, p and q, where p is the proposer's offer, while q is the acceptance threshold. That is, when acting as proposer, the player offers p, whereas in a receiver's role, the player rejects any offer smaller than q.

In this paper, we focus on how the presence of multiple roles in the interactions affects decision making in the investment process. As such, we consider a baseline UG model where proposers have two possible strategic offers, a low (L, with $p = l$) and a high one (fair) (H, with $p = h$), where $l < h$, with $l, h \in [0, 1]$. On the other hand, receivers have two options, a low threshold (L, with $q = l$) and a high threshold (H, with $q = h$). Thus, overall, there are four possible strategies HH, HL, LH and LL (e.g., HL denotes a strategy that offers high and accepting any offers).

In our model of the UG, *fairness* is measured by calculating what percentage of the population is representative for either the HH or HL strategies (i.e., fair proposers). This allows us to have a clear comparison with previous works – in terms of fairness achieved at the level of population – that have investigated the evolution of fairness in the UG, such as in [1,2,25].

Particularly, we set $h = 0.6$ and $l = 0.1$, as this represents the environment with (roughly) the lowest frequency of fair proposals (see Fig. 1; for DMS networks, see Fig. S1 in Supporting Information (**SI**)). Furthermore, in line with evidence from behavioural experiments [2,26], which suggests that people (almost) never offer more than half of the sum. We also explore the typical case of $h = 0.5$ and validate our findings (see Figs. S2 and S3 in SI).

The payoff matrix for the four strategies HH, HL, LH, and LL reads (for row player):

	HH	HL	LH	LL
HH	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1-h}{2}$	$\frac{1-h}{2}$
HL	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1-h+l}{2}$	$\frac{1-h+l}{2}$
LH	$\frac{h}{2}$	$\frac{1+h-l}{2}$	0	$\frac{1-l}{2}$
LL	$\frac{h}{2}$	$\frac{1+h-l}{2}$	$\frac{l}{2}$	$\frac{1}{2}$

For example, an HH player encountering an HL player results in the payoff $\frac{1}{2}$ for either player, as both of them propose and accept a fair split (i.e., one interaction results in the payoff $1 - h$ for the proposer, and h for the receiver, and vice-versa for when the roles are reversed).

2.2. Population structure and evolutionary dynamics

Initially each agent is designated as one of the four strategies (i.e., HH, HL, LH, HH), with equal probability. At each time step, each agent plays the UG with its immediate neighbours. The score for each agent is the sum of the payoffs in these encounters. At the end of each step, an agent A with fitness f_A chooses to copy the strategy of a randomly selected neighbour agent B with score f_B, with a probability given by the Fermi function (i.e., *stochastic update*) [27]:

$$W(S_B \rightarrow S_A) = (1 + e^{(f_A - f_B)/K})^{-1},$$

where K denotes the amplitude of noise in the imitation process [27]. In line with previous works and lab experiments [2,27], we set $K = 0.1$ in our simulations.

We simulate this evolutionary process until a stationary state or a cyclic pattern is reached. For the sake of a clear and fair comparison, all simulations are run for 500000 generations. Moreover, for each simulation, the results are averaged over the final 25000 generations, in order to account for the fluctuations characteristic of these stable states. Furthermore, to improve accuracy, for each set of parameter values, the final results are obtained from averaging 20 independent realisations. When shown in figures, the error bars represent the standard error of the mean between replicates.

2.3. Network topologies

The Barabási and Albert (**BA**) model [17] is one of the most famous models used in the study of highly heterogeneous, complex networks. The main features of the BA model are that it follows a *preferential attachment* rule, has a small clustering coefficient, and a typical *power-law degree distribution*. In order to explain preferential attachment, let us describe the construction of a BA network. Starting from a small set of m_0 interconnected nodes, each new node selects and creates a link with m older nodes according to a probability proportional to their degree. The procedure stops when the required network size of N is reached. This will produce a network characterised by a power-law distribution, $p_k \sim k^{-\gamma}$, where the exponent γ is its *degree exponent* [28]. There is a high degree correlation between nodes, and the degree distribution is typically skewed with a long tail. There are few hubs in the network that attract an increasing number of new nodes which attach as the network grows (in a typical "*rich-get-richer*" scenario). The power-law distribution exhibited by BA networks resembles the heterogeneity present in many real-world networks, however, they are also defined by low clustering coefficients, which means they cannot always be used to approximate realistic settings [29].

To build heterogeneous networks with a large clustering coefficient, Dorogovtsev et al. [30] have proposed the eponymous Dorogovtsev–Mendes–Samukhin (**DMS**) model. This model follows a similar method of construction as the BA model. The crucial difference is that each new node connects with the two extremities of m ($m \geq 2$) randomly chosen edges, instead, therefore forming characteristic triangular motifs whenever a new node is added to the network. Since the number of edges arriving to any node reflects its degree, the probability of attaching the new node to an old node is proportional to its degree and preferential attachment is recovered. The degree distribution is therefore the same as the one of a BA model, and the degree–degree correlations are also equal [31]. However, the clustering coefficient is large, and more accurately mimics many realistic social networks [29,32]. The average connectivity for both types of scale-free networks is $z = 2m$. For all of our experiments, we seed 10 different networks (of each type) of size $N = 2000$, with an average connectivity of $z = 4$.

2.4. Cost-efficient interference in networks

We aim to study how one can efficiently interfere in a spatially heterogeneous population to achieve high levels of fairness while minimising the cost of interference. As mentioned above, the level of fairness is measured by the fraction of fair offers in the population [2]. An investment decision consists of a cost $\theta > 0$ to the external decision-making agent/investor, and this value θ is added as surplus to the payoff of each suitable candidate. In order to determine cost-efficiency, we vary θ for each proposed interference strategy, measuring the total accumulated costs to the investor. Thus, the most efficient interference schemes will be the ones with the lowest relative total cost.

We examine different approaches to interference, where fairness is advocated for either role or both, leading to different desirable behaviours to be targeted:

 (i) ensure all proposals are fair, thus investing in HH and HL (**Target: HH, HL**);
 (ii) ensure only fair offers are accepted, thus investing in HH and LH (**Target: HH, LH**);
 (iii) ensure both (i) and (ii), i.e., investing in HH only (**Target: HH**).

Moreover, in line with previous works on network interference [12,24, 33,34], we compare global interference strategies where investments are triggered based on network-wide information, local neighbourhood information, and, lastly, node centrality information.

In the *population-based* (**POP**) approach, a decision to invest in desirable behaviours is based on the current composition of the population. We denote x_f the fraction of individuals in the population with a desirable behaviour, given a targeting approach, i.e., (i), (ii) or (iii) as defined above. Namely, an investment is made if x_f is less or at most equal to a threshold p_f (i.e., $x_f \leq p_f$), for $0 \leq p_f \leq 1$. They do not invest otherwise (i.e., $x_f > p_f$). The value p_f describes how rare the desirable behaviours should be to trigger external support.

In the *neighbourhood-based* (**NEB**) approach, a decision to invest is based on the fraction x_f of neighbours of a focal individual with the desirable behaviours, calculated at the local level. Investment happens if x_f is less or at most equal to a threshold n_f (i.e., $x_f \leq n_f$), for $0 \leq n_f \leq 1$; otherwise, no investment is made.

As the presence of structural heterogeneity in scale-free networks introduces a level of inequality between nodes in terms of influence, we also examine a *node-influence-based* (NI) approach. Here, we build upon the literature by introducing two measures for defining a node's influence, *degree centrality* (NI-deg) and *eigenvector centrality* (NI-eig).

Firstly, degree centrality is the oldest measure of influence used in network science [35]. It denotes the number of neighbours of the node i (i.e., its number of incoming edges). By definition, degree centrality is normalised using the total number of nodes, or the maximal possible degree, $n - 1$, to obtain a number between 0 and 1. *Degree centrality*, denoted by **NI-deg** or x_i^{deg}, is defined as follows:

$$x_i^{deg} = deg_i = \frac{k_i}{n-1}, \tag{1}$$

where k_i is the degree of the node i and $n - 1$ is the total number of nodes. The degree k_i of a node i is given by: $k_i = \sum_{j=1}^{n} A_{ij}$.

Despite its simple definition, degree centrality is often a highly effective measure of the influence or importance of a node, since

people with more connections tend to be more influential in a social network [36]. However, degree centrality lacks potentially important aspects of the network's architecture or the node's position in the network. It can be considered a 'local' centrality measure, measuring only the quantity and not the quality of connections [22,23].

Secondly, eigenvector centrality represents a related measure of prestige, since the importance of a node i depends on the prestige of its neighbours [36,37]. In other words, this centrality measure acknowledges that not all connections are equal, but connections to nodes who are themselves influential will make a node more influential [37]. Eigenvector centrality is computed by assuming that the centrality of node i is proportional to the sum of the centralities of node i's neighbours. Thus, *eigenvector centrality*, denoted by **NI-eig** or x_i^{eig}, is defined as follows:

$$x_i^{eig} = eig_i = \frac{1}{\lambda} \sum_{j=1}^{n} A_{ij} x_j, \qquad (2)$$

where λ is a positive constant or proportionality factor.

Defining the vector of centralities $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, we can rewrite the previous definition in matrix form as $\lambda \cdot \mathbf{x} = \mathbf{A} \cdot \mathbf{x}$. Hence $\mathbf{x}$ is an eigenvector of the adjacency matrix with eigenvalue λ. Since the centralities must be non-negative, it can be shown (using the Perron–Frobenius theorem) that λ must be the largest eigenvalue of the adjacency matrix and $\mathbf{x}$ the corresponding eigenvector. Thus, even if a large number of connections increases the centrality measure, a node with a smaller number of high-quality nodes may still outrank one with a larger number of low-influential nodes.

We denote by x_i the node's centrality measure (e.g., for NI-deg: $x_i = deg_i = x_i^{deg}$). The nodes are sorted in ascending order based on their influence x_i, and the threshold i_f denotes the fraction of nodes that will be selected for interference, if their behaviour satisfies the given targeting approach. For instance, given a network of size 1000 and a threshold $i_f = 0.001$ would mean selecting only the most influential node in the network for investment.

Irrespective of the interference scheme or the targeted behaviour, the threshold signifies an increase in the number of nodes that satisfy the requirements for investment. In other words, a threshold of 1 means investing in all nodes which follow the desired strategy. Conversely, a lower threshold implies a more careful approach to investment, whereby the exogenous agent is stricter in their selection of suitable candidates. Moreover, the target selection also affects the number of candidate nodes. Stricter schemes, such as targeting individuals who are fair when proposing, and also when responding (HH), narrow the search for nodes which satisfy the requirements.

3. Results

Several factors must be taken into consideration before deciding to invest in a population of individuals in order to adequately promote fairness. Among these, we take into account and resolve questions related to the importance of different roles of players, the size of the endowments, the threshold at which to resume investment, and the amount and quality of information that is available to the external decision-maker. We also consider that there exists a *minimal level of fairness* which the external decision-maker aims to enforce [12,15,33], and we study the most efficient strategies according to these varying standards of fairness.

We structure the subsections that follow according to key insights derived from the results, and refer to previous results on structured populations [15]. We will highlight the key differences that arise in the presence of diversity (in the form of spatial heterogeneity), and mention similarities where appropriate.

Table 1

Most cost-efficient interference schemes to reach a minimum fairness of proposals in BA networks. For each minimal standard of fairness, we highlight (in bold) the least costly options across schemes.

Scheme	Minimum fairness	Target	Threshold	θ	Cost (mean ± se)
POP	75%	HH	0.2	56.23	168655 ± 14592
POP	90%	HH	0.4	74.98	176377 ± 14389
POP	99%	HH LH	0.8	56.23	293956 ± 20785
NEB	75%	HH LH	0.7	56.23	112870 ± 944
NEB	90%	HH LH	0.7	56.23	**112870 ± 944**
NEB	99%	HH LH	0.7	56.23	**112870 ± 944**
NI-DEG	75%	HH	0.005	17.78	**66891 ± 2185**
NI-DEG	90%	HH LH	0.007	23.71	143260 ± 2000
NI-DEG	99%	HH	0.017	31.62	512727 ± 1885
NI-EIG	75%	HH	0.003	31.62	**66252 ± 2623**
NI-EIG	90%	HH	0.003	74.98	190862 ± 3974
NI-EIG	99%	HH	0.017	42.16	705906 ± 2352

3.1. Social diversity reduces interference complexity

Social diversity introduces several challenges which must be overcome by an external decision-maker, but these bring with them opportunities to exploit the inherent mix of strategies that can be successful, according to different initial network conditions. If we consider a hierarchy of complexity based on the inherent costs of gathering information as explained above, we show that two targets have the potential to be optimal in a wide range of schemes and fairness requirements (see Table 1). Ensuring both offers and responses are fair is the strict, but also intuitive approach to investment. Nevertheless, several configurations in which rewarding fair responders (i.e., HH or LH) succeed as the most cost-effective avenues towards fairness.

In the presence of diversity, we see that strictness, while generally effective, is not necessarily optimal. Heterogeneity allows for the co-existence of several strategies in a cluster, and relaxing the eligibility conditions for investment can allow an external decision-maker to reinforce positive behaviour, ultimately producing the optimal outcomes shown in Table 1. These results contrast starkly with previous observations in structured populations, for which stringent information gathering (i.e., targeting HH) leads to the lowest total costs and the highest levels of fairness, in almost all cases. Previously, for structured populations, relaxing targeting conditions was only desirable in the presence of high mutation rates. Diversity acts in a similar fashion to the noise associated with high mutation rates, which also allows the coexistence of several strategies. This is because heterogeneity allows for the coexistence of several strategies in a cluster, and relaxing the eligibility conditions for investment can allow an external decision-maker to reinforce positive behaviour, ultimately producing the optimal outcomes shown in Table 1.

3.2. Standards of fairness stipulate divergent approaches

Fundamentally, institutional incentive schemes, in the context of asymmetric interactions and the Ultimatum Game, are diametrically different to cooperative dilemmas where heterogeneity naturally promotes pro-sociality [18,34,38]. As the UG allows for a larger room for improvement, due to relatively low baseline fairness levels (see again Section 2.1 and Fig. 1), an investor can modify their goals and opt for modest improvements. They could, for instance, decide to invest only in large hubs, and ignore any potential outliers. Whether due to budget constraints, lack of information, or even uncertainty of network characteristics, they could adjust their margin for improvement and prevent unnecessary spending. Fig. 2 shows the average total cost required for optimal investment schemes across a wide range of goals. We note the differences in the scales of the y-axes, which imply that on average, the strictest target (HH) is also the cheapest, followed by ensuring fair responses (HH and LH) and finally only ensuring fair proposals (HH and HL), which is significantly more expensive on
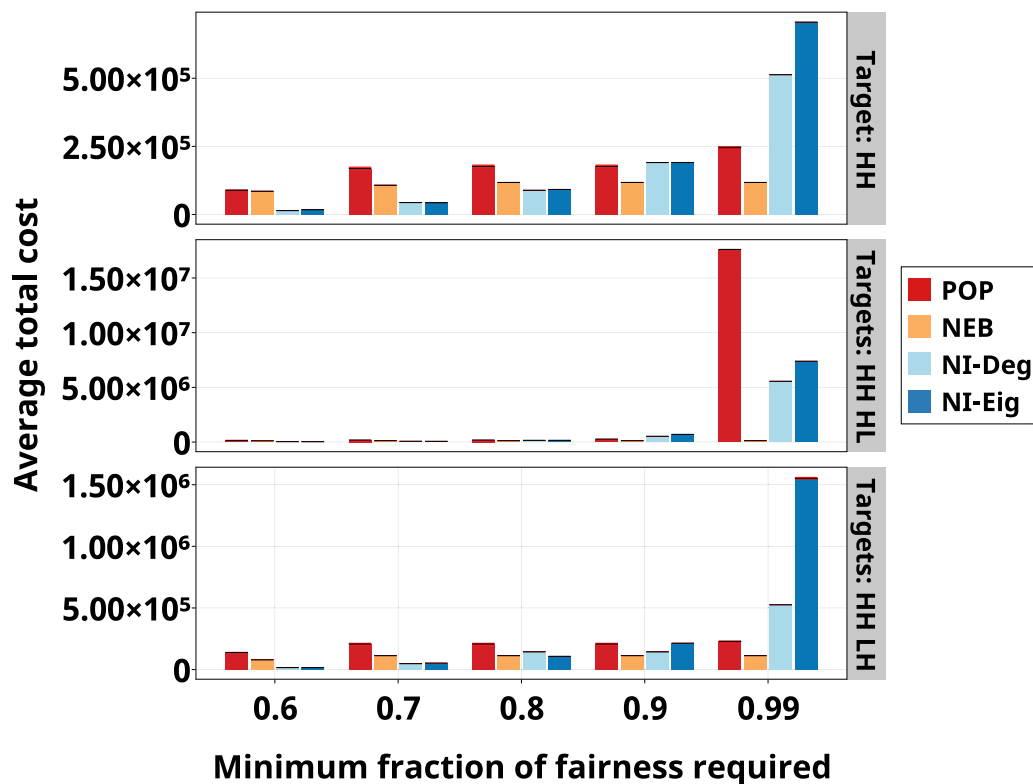
Fig. 2. Mean total costs for each possible target and scheme.
The figure illustrates the mean total costs for the most efficient combinations of threshold and investment amount θ for each possible target and scheme in BA networks. Error bars are shown in light red. Colours correspond to four different approaches: POP — population-based (red colour), NEB — neighbourhood-based (orange), and node influence-based by considering respectively degree centrality (NI-deg) (light blue) and eigenvector centrality (NI-eig) (blue). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

average. This variance increases substantially with higher requirements for fairness.

Fig. 2 shows that targeting hubs (highly connected nodes) prevents over-spending as long as a small fraction of unfairness is accepted. Targeting hubs is the most cost-efficient strategy when the minimum fraction of fairness is low but this benefit decreases as the minimum fraction of fairness increase. In the case where less than 1% of unfair individuals is accepted, the cost of targeting hubs is much higher than other interference strategies. This is because the hubs' spheres of influence do not extend far enough towards the leaves of the graph for them to make a marked improvement. Local observations, while comparatively expensive for most minimal fairness requirements, have the benefit of being able to extend their reach to lowly connected nodes.

Given a low enough minimum fairness requirement, targeting hubs (highly connected nodes) prevents over-spending. As these requirements increase, we observe that the hubs' spheres of influence do not extend far enough towards the leaves of the graph for them to make a marked improvement. Local observations, while comparatively expensive for most minimal fairness requirements, have the benefit of being able to extend their reach to lowly connected nodes. We further exemplify the effects of such an approach in Fig. S4 in the SI, where we select one of the most cost-effective values of individual investment θ and show outcomes for all possible combinations of targets and centrality measures.

Moreover, the differences between the two schemes (NI and NEB) are exacerbated for more demanding targets. Ensuring only fair responses implies the suitability of a swathe of nodes, thus being comparatively more effective at promoting fairness than the stricter target (HH), which restricts node candidacy. An external decision maker should be strict about which hubs to invest into, but can be more lenient in the selection of sparsely connected individuals. This is a promising result, as the cost of information gathering is assumed to

scale with the number of subjects. Investors could potentially afford to spend the implied additional costs to scrutinise influential individuals, while allowing for a much broader classification if they opted for local observations, instead.

3.3. Polarisation towards fairness

Departing significantly from previously discussed results on homogeneous populations, we observe a tendency towards polarisation (see Figs. 3 and 4). Across a wide spectrum of both interference schemes and various targets, we show that a very large θ can propel the system towards fairness, while also minimising cost. After having reached close to 100% fairness, it is difficult for unfair strategists to invade the network, due to the heterogeneous dynamics at play, so no further investment is required.

By varying the threshold for investment, the amount of funding that is required for the system to shift towards fairness also changes. In a phenomenon akin to energy landscapes encountered in physical systems, the "energy" required to push fairness across the local maxima increases the further away it gets from the global minimum. Lowering the threshold can be beneficial, as overspending is avoided. On the other hand, the obstacle that arises is two-fold. As the goal is ultimately to reach a high level of fairness, a low threshold allows unfair individuals to thrive before fair individuals are eligible for investment. In other words, the system dips further away from the global minimum, and the investment amount required increases appropriately.

Theoretically, this would allow for a number of viable investment approaches in such a system, but there are also several practical concerns of note. Institutions wishing to employ the practice of increasing their endowment amounts would be expected to have access to a considerable amount of initial funding, as opposed to spreading out the costs over multiple investment rounds. Moreover, there exists an
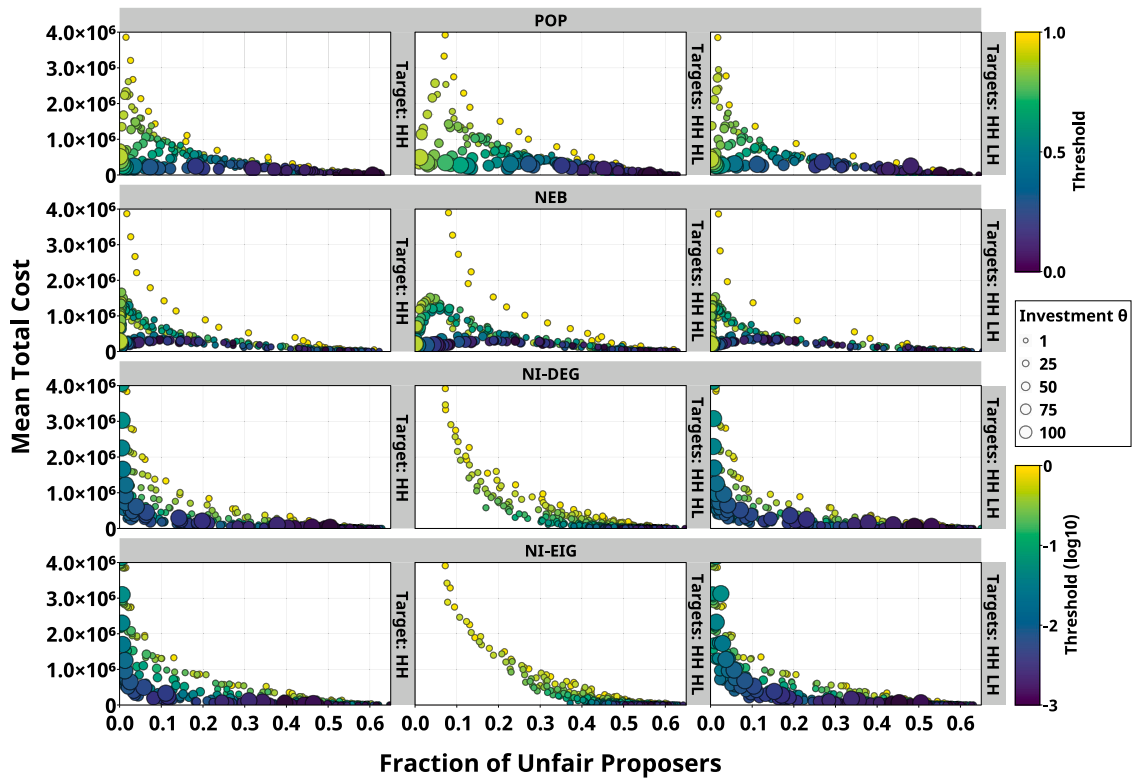
Fig. 3. Pareto fronts for each scheme in BA networks.
We show the proportion of unfair proposers as a function of the average interference cost for each scheme and target combination. The markers' size is determined by the individual investment θ (grouped to the nearest value), whereas the colour is determined by the threshold (ranging from 'blue' (lowest) to 'yellow' (highest) value). Markers near the origin indicate the optimal solutions. Note that we only show the most cost-effective solutions, by limiting the maximum total cost. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Most cost-efficient interference schemes to reach a minimum fairness of proposals in DMS networks.

Scheme	Minimum fairness	Target	Threshold	θ	Cost (mean ± se)
POP	75%	HH HL	0.8	74.98	294102 ± 34137
POP	90%	HH HL	0.8	74.98	294102 ± 34137
POP	99%	HH HL	0.9	74.98	342927 ± 30915
NEB	75%	HH LH	0.3	74.98	123497 ± 7776
NEB	90%	HH LH	0.3	74.98	123497 ± 7776
NEB	99%	HH LH	0.3	74.98	123497 ± 7776
NI-DEG	75%	HH HL	0.004	10.00	25173 ± 1060
NI-DEG	90%	HH HL	0.004	42.16	158095 ± 1113
NI-DEG	99%	HH	0.031	42.16	1260950 ± 5993
NI-EIG	75%	HH HL	0.001	17.78	14995 ± 376
NI-EIG	90%	HH HL	0.004	42.16	160555 ± 1343
NI-EIG	99%	HH LH	0.177	23.71	4026960 ± 32543

intermediary region where overspending becomes problematic (see Fig. 4). This issue occurs when the investment amount is large enough to be effective at inducing the change towards fairness, but not at the level where it does so rapidly. Given a relatively high threshold, many candidates become suitable for assistance, thus leading to excessive funds being deployed in locations of the network which are ineffective.

3.4. Clustering further reduces the burden on investors

Real-world scale-free networks often have high transitivity (i.e., clustering), a feature missing in BA networks [29,32], so it is crucial to measure the effects of network transitivity on the choice of investment policies. In the absence of investment, high transitivity positively influences fairness, thereby lowering the total amount of costs required to reach minimum standards of fairness (see Fig. 5).

Moreover, eigenvector centrality (NI-Eig) can be employed to reduce overall spending for moderate fairness requirements, and hubs can be exploited to reduce spending for all but the most strict fairness regimes, unlike in lowly clustered networks. Comparing optimal investment schemes, we show that costs remain similar across all but the minimal desired standards of fairness (see Fig. 5 and Table 2). Strict fairness regimes can be enforced using local information without overspending, similar to BA networks, yet we also show that population-based metrics are less risky in highly clustered networks. Thus, network transitivity acts as an equaliser between the different schemes and targets.

Previously, we had seen that targeting proposers and responders (HH) and solely proposers (HH and HL) were both sensible approaches towards leveraging fairness. In the presence of clustering, targeting fair responders (HH and LH) becomes viable and in some cases, optimal (see Table 2). In fact, we see that diversity and transitivity act similarly to noise or behavioural exploration in the choice for investment, but on a much broader scale. Both of these factors open up novel mechanisms of engineering fairness while minimising the risk of choosing inappropriate candidates for endowments. With this reduction in complexity, we also see a very slight increase in the overall costs required to promote fairness. Paradoxically, a higher baseline of fairness also requires more endowments to reach fully positive outcomes. Heterogeneity acts as an equaliser in the truest sense, aiding in the quest towards a fair society, but also fostering a small minority of unfair individuals. We note that all the findings outlined above are consistent across both types of scale-free networks, and across interference schemes. For a detailed view of each scheme, see Figs. S5–S13 in SI.

4. Discussion

In this work, we have introduced social diversity in the Ultimatum game and studied the effects of heterogeneity on external interference,
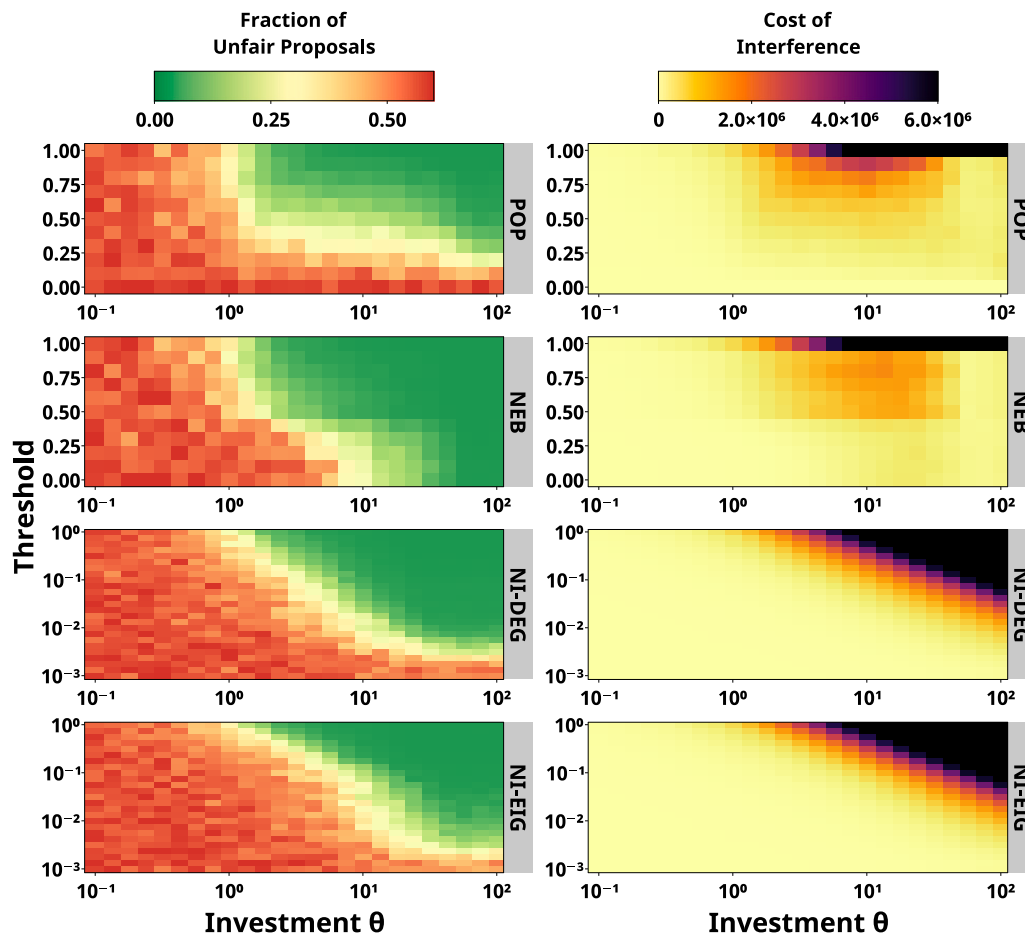
Fig. 4. Targeting fair responders in BA networks.
The figure show the proportion of unfair proposers and total costs of investment for each scheme, while targeting fair responses (HH and LH). In the left column, colour corresponds to the proportion of unfair proposals, ranging from 'green' (lowest) to 'red' (highest) (see colour bar on the top of left column). In the right column, colour corresponds to the interference, ranging from 'yellow' (lowest) to 'black' (highest) (see colour bar on the top of right column). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

taking into account cost optimisation, limited information and standards of fairness. We found that diversity reduces the need for complex information gathering, and allows for less strictness in the eligibility criteria for receiving endowments. Exploiting different measures of centrality has been shown to enable a reduction in spending if standards of fairness are lowered. Our results indicate that the presence of diversity reduces the cost and complexity of promoting the evolution of fairness.

Existing models of institutional incentives aimed at promoting collective behaviours, such as cooperation and fairness [7,39–43], usually ignore the problem of cost-efficiency. These works often consider non-adaptive incentive mechanisms, studying how minimal incentive mechanisms can promote cooperation. The overarching goal of our research is to exploit available information that can be gathered to design cost-efficient, adaptive incentive mechanisms. Our work also differs from EGT literature on optimal control in networked populations, where cost-efficiency is not considered [44,45]. Instead, these works on controllability focus on identifying which individuals or nodes are the most important to control (i.e. where individuals can be assigned strategies as control inputs), for different population structures.

Previous works studying the optimisation of the cost of providing institutional incentives in an evolving population have mainly focused on cooperative behaviours and symmetric games, namely the Prisoner's Dilemma (PD) [12,13,33,34] and Public Goods Games (PGG) [14, 16,24]. Both homogeneous and heterogeneous networks have been analysed, showing that taking into account local network properties, such as detailed neighbourhood information and node degrees, can

significantly reduce the cost required to ensure a certain level of cooperation. There are some recent efforts to extend these analyses to asymmetric games, namely the Ultimatum Game, but they are limited to the simpler setting of homogeneous networks, in the form of well-mixed and lattice graphs [15]. Thus, the present work advances this literature where the external decision maker needs to account for both the heterogeneous characteristics of agents and the hierarchical asymmetry in their interactions. As shown, cost-efficient interference mechanisms that incorporate this combined information can outperform those which only consider global population statistics and neighbourhood properties [15].

In the context of institutional incentives modelling for promoting enhanced collective behaviours, an important issue is how to set up and maintain the incentive budget. The problem of who contributes to the budget is a social dilemma in itself, and how to escape it is a challenging research question. Facilitating solutions include pool incentives with second-order punishments [43,46], democratic decisions [47], or mixed incentives (both positive and negative) [24], just to name a few. In this work, we do not address this issue, focusing instead on how to optimise the spending from a given budget by exploiting network properties and information gathering. However, it would be interesting to study the co-evolutional institutional formation with different interference strategies and individual strategic behaviours.

Real-world networks, such as networks of collaboration and social networks, are often inherently *heterogeneous*, whereby individuals differ in the number of connections, capacities, etc. [48]. The presence of
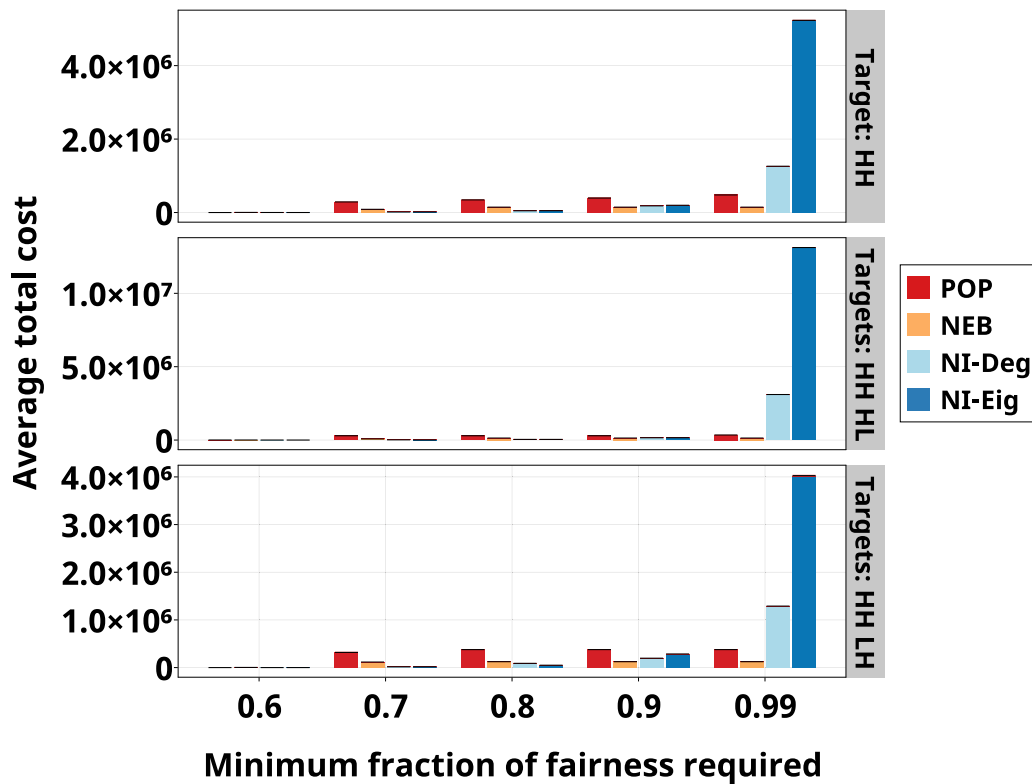
Fig. 5. Most efficient schemes in DMS networks.
Mean total costs (scaled by log10) for the most efficient combinations of threshold and investment amount θ for each possible target and scheme. Colours correspond to four different approaches: POP — population-based ('red' colour), NEB — neighbourhood-based ('orange' colour), and node influence-based by considering respectively degree centrality (NI-deg) ('light' blue colour) and eigenvector centrality (NI-eig) ('blue' colour). Error bars are shown in light red. If a certain scheme is missing, no investment was triggered for each desired standard of fairness. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

individual and network structural heterogeneity has been demonstrated to be crucial in the evolution of various collective behaviours (e.g. co-operation, coordination, trust, or AI safety) in networked games [18, 49]. However, little attention has been given to how this heterogeneity affects external or institutional decision-making, e.g. to optimise the institutional cost of providing incentives to achieve high levels of such collective behaviours. The present paper's analysis aims to shed light on this issue, providing insights into how fairness can be efficiently engineered in real-world, dynamical settings (such as hybrid networked societies of human and autonomous systems) [50,51].

Works on collective behaviour in social networks usually assume that changes are initiated from inside the system [52,53]. Even when the role of influencers was explored [53], external interference mechanisms and incomplete information are not taken into account. Mechanism design schemes have been employed to successfully resolve problems of incentivisation and taxation to enforce desired behaviours [54], but these assume that the decision-maker has complete control over the agents within the systems. Our approach assumes that the decision-maker has little or no control over the agents, relying only on rewarding schemes, and nudging agents towards positive outcomes. Moreover, the literature on mechanism design typically does not focus on the costs of maintaining positive outcomes, whereas cost optimisation is one of our main objectives.

In summary, we have shown that it is crucial to consider the roles' asymmetry to provide cost-efficient investment strategies, an important feature which was not possible to identify in previous works where symmetric games were studied [12–14,16,24,33,34]. We have identified several key features that are required to minimise costs while ensuring positive outcomes. We found that diversity reduces the need for complex information gathering, and allows for less strictness in the eligibility criteria for receiving endowments. These results, regardless of the underlying interaction structure, stand out in sharp contrast

with previous works on cooperation dilemmas, in which interference schemes require an exceptionally strict investment approach [12,33].

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.chaos.2022.113051.

References

[1] Nowak MA, Page KM, Sigmund K. Fairness versus reason in the ultimatum game. Science 2000;289(5485):1773–5.
[2] Rand DG, Tarnita CE, Ohtsuki H, Nowak MA. Evolution of fairness in the one-shot anonymous ultimatum game. Proc Natl Acad Sci 2013;110(7):2581–6.
[3] Teixeira AS, Santos FPFCFP, Francisco AP, Santos FPFCFP. Eliciting fairness in N-player network games through degree-based role assignment. Complexity 2021;2021.
[4] Paiva A, Santos FPFC, Santos FPFC. Engineering pro-sociality with autonomous agents. In: Proceedings of the AAAI conference on artificial intelligence, vol. 32, 2018, p. 7994–9.
[5] Levin SA. Multiple scales and the maintenance of biodiversity. Ecosystems 2000;3(6):498–506.
[6] Pritchett AR, Genton A. Negotiated decentralized aircraft conflict resolution. IEEE Trans Intell Transp Syst 2017;19(1):81–91.

[7] Góis AR, Santos FP, Pacheco JM, Santos FC. Reward and punishment in climate change dilemmas. Sci Rep 2019;9(1):1–9.

[8] Ostrom E. Polycentric systems for coping with collective action and global environmental change. Global Environ Change 2010;20(4):550–7.

[9] Pacheco JM, Vasconcelos VV, Santos FC. Climate change governance, cooperation and self-organization. Phys Life Rev 2014;11(4):573–86.

[10] Ostrom E. Governing the commons: The evolution of institutions for collective action. Cambridge University Press; 1990.

[11] North DC. Institutions, institutional change and economic performance. Cambridge University Press; 1990.

[12] Han TA, Tran-Thanh L. Cost-effective external interference for promoting the evolution of cooperation. Sci Rep 2018;8(15997).

[13] Chen X, Perc M. Optimal distribution of incentives for public cooperation in heterogeneous interaction environments. Front Behav Neurosci 2014;8:248.

[14] Duong MH, Han TA. Cost efficiency of institutional incentives in finite populations. Proc R Soc A 2021.

[15] Cimpeanu T, Perret C, Han TA. Cost-efficient interventions for promoting fairness in the ultimatum game. Knowl-Based Syst 2021;233:107545.

[16] Wang S, Chen X, Szolnoki A. Exploring optimal institutional incentives for public cooperation. Commun Nonlinear Sci Numer Simul 2019;79:104914.

[17] Barabási A-L, Albert R. Emergence of scaling in random networks. Science 1999;286(5439):509–12.

[18] Santos FC, Santos MD, Pacheco JM. Social diversity promotes the emergence of cooperation in public goods games. Nature 2008;454:214–6.

[19] Santos FC, Pacheco JM, Lenaerts T. Evolutionary dynamics of social dilemmas in structured heterogeneous populations. Proc Natl Acad Sci USA 2006;103:3490–4.

[20] Poncela J, Gómez-Gardenes J, Floría LM, Moreno Y. Robustness of cooperation in the evolutionary prisoner's dilemma on complex networks. New J Phys 2007;9(6):184.

[21] Sinatra R, Iranzo J, Gómez-Gardeñes J, Floría LM, Latora V, Moreno Y. The ultimatum game in complex networks. J Stat Mech Theory Exp 2009;2009(09):P09012.

[22] Di Stefano A, Scatà M, Attanasio B, La Corte A, Lió P, Das SK. A novel methodology for designing policies in mobile crowdsensing systems. Pervasive Mob Comput 2020;67:101230.

[23] Di Stefano A, Scatà M, La Corte A, Liò P, Catania E, Guardo E, et al. Quantifying the role of homophily in human cooperation using multiplex evolutionary game theory. PLoS One 2015;10(10):e0140646.

[24] Chen X, Sasaki T, Brännström Å, Dieckmann U. First carrot, then stick: How the adaptive hybridization of incentives promotes cooperation. J R Soc Interface 2015;12(102):20140935.

[25] Page KM, Nowak MA, Sigmund K. The spatial ultimatum game. Proc R Soc B 2000;267(1458):2177–82.

[26] Güth W, Schmittberger R, Schwarze B. An experimental analysis of ultimatum bargaining. J Econ Behav Organ 1982;3(4):367–88.

[27] Szabó G, Fáth G. Evolutionary games on graphs. Phys Rep 2007;446(4–6):97–216, arXiv:0607344.

[28] Barabási A-L. Network science. Cambridge University Press; 2016, p. 474.

[29] Su J, Sharma A, Goel S. The effect of recommendations on network structure. In: 25th international world wide web conference, WWW 2016. International World Wide Web Conferences Steering Committee; 2016, p. 1157–67.

[30] Dorogovtsev SN, Mendes JFF, Samukhin AN. Structure of growing networks with preferential linking. Phys Rev Lett 2000;85(21):4633.

[31] Dall'Asta L, Baronchelli A, Barrat A, Loreto V. Nonequilibrium dynamics of language games on complex networks. Phys Rev E 2006;74(3):36105.

[32] Barrat A, Pastor-Satorras R. Rate equation approach for correlations in growing network models. Phys Rev E 2005;71(3):36127.

[33] Han TA, Lynch S, Tran-Thanh L, Santos FC. Fostering cooperation in structured populations through local and global interference strategies. In: IJCAI-ECAI'2018. AAAI Press; 2018, p. 289–95.

[34] Cimpeanu T, Han TA, Santos FC. Exogenous rewards for promoting cooperation in scale-free networks. In: ALIFE 2019. MIT Press; 2019, p. 316–23.

[35] Boldi P, Vigna S. Axioms for centrality. Internet Math 2014;10(3–4):222–62.

[36] Bloch F, Jackson MO, Tebaldi P. Centrality measures in networks. 2019, Available At SSRN 2749124.

[37] Perra N, Fortunato S. Spectral centrality measures in complex networks. Phys Rev E 2008;78(3):36107.

[38] Santos FC, Pacheco JM. Scale-free networks provide a unifying framework for the emergence of cooperation. Phys Rev Lett 2005;95:98104.

[39] Sigmund K, Hauert C, Nowak M. Reward and punishment. Proc Natl Acad Sci 2001;98(19):10757–62.

[40] Han TA. Institutional incentives for the evolution of committed cooperation: Ensuring participation is as important as enhancing compliance. J R Soc Interface 2022;19(188):20220036.

[41] Sasaki T, Brännström Å, Dieckmann U, Sigmund K. The take-it-or-leave-it option allows small penalties to overcome social dilemmas. Proc Natl Acad Sci 2012;109(4):1165–9.

[42] Sun W, Liu L, Chen X, Szolnoki A, Vasconcelos VV. Combination of institutional incentives for cooperative governance of risky commons. Iscience 2021;24(8):102844.

[43] Sigmund K, De Silva H, Traulsen A, Hauert C. Social learning promotes institutions for governing the commons. Nature 2010;466:7308.

[44] Ramazi P, Cao M. Analysis and control of strategic interactions in finite heterogeneous populations under best-response update rule. In: Decision and control, 2015 IEEE 54th annual conference on. IEEE; 2015, p. 4537–42.

[45] Riehl JR, Cao M. Towards optimal control of evolutionary games on networks. IEEE Trans Automat Control 2016;62(1):458–62.

[46] Perc M. Sustainable institutionalized punishment requires elimination of second-order free-riders. Sci Rep 2012;2(1):1–6.

[47] Hilbe C, Traulsen A, Röhl T, Milinski M. Democratic decisions establish stable authorities that overcome the paradox of second-order punishment. Proc Natl Acad Sci USA 2014;111(2):752–6.

[48] Barabasi A-L. Linked-how everything is connected to everything else and what it means F. Perseus Books Group; 2014.

[49] Kumar A, Capraro V, Perc M. The evolution of trust and trustworthiness. J R Soc Interface 2020;17(169):20200041.

[50] Andras P, Esterle L, Guckert M, Han TA, Lewis PR, Milanovic K, et al. Trusting intelligent machines: Deepening trust within socio-technical systems. IEEE Technol Soc Mag 2018;37(4):76–83.

[51] Dafoe A, Bachrach Y, Hadfield G, Horvitz E, Larson K, Graepel T, et al. Cooperative AI: Machines must learn to find common ground. Nature 2021;593(7857):33–6.

[52] Ranjbar-Sahraei B, Bou Ammar H, Bloembergen D, Tuyls K, Weiss G, Sahraei BR, et al. Evolution of cooperation in arbitrary complex networks. In: AAMAS '14, Paris, France, May 5–9, 2014. International Foundation for Autonomous Agents and Multiagent Systems; 2014, p. 677–84.

[53] Franks H, Griffiths N, Anand SS. Learning agent influence in MAS with complex social networks. J Auton Agents Multi-Agent Syst 2014;28(5):836–66.

[54] Endriss U, Kraus S, Lang J, Wooldridge M. Incentive engineering for Boolean games. In: IJCAI '11. 2011, p. 2602–7.