Chaos

The combination of social reward and punishment is conducive to the cooperation and heterogeneity of social relations

Cite as: Chaos **32**, 103104 (2022); https://doi.org/10.1063/5.0102483
Submitted: 10 June 2022 • Accepted: 06 September 2022 • Published Online: 06 October 2022

Ming Zhang, Xu Zhang, Cunquan Qu, et al.

The combination of social reward and punishment is conducive to the cooperation and heterogeneity of social relations

Cite as: Chaos **32**, 103104 (2022); doi: 10.1063/5.0102483
Submitted: 10 June 2022 · Accepted: 6 September 2022 ·
Published Online: 6 October 2022

View Online Export Citation CrossMark

Ming Zhang,[1] Xu Zhang,[2] Cunquan Qu,[2,a)] Guanghui Wang,[1,2] and Xin Lu[3]

AFFILIATIONS

[1]School of Mathematics, Shandong University, Jinan 250100, People's Republic of China
[2]Data Science Institute, Shandong University, Jinan 250100, People's Republic of China
[3]College of Systems Engineering, National University of Defense Technology, Changsha 410073, People's Republic of China

Note: This article is part of the Focus Issue on Complex Systems and Inter/Transdisciplinary Research.
[a)]**Author to whom correspondence should be addressed:** cqqu@sdu.edu.cn

ABSTRACT

Individual behaviors and social relations influence each other. However, understanding the underlying mechanism remains challenging. From social norms controlling human behavior to individual management of interpersonal relationships, rewards and punishments are some of the most commonly used measures. Through simulating the weak prisoner's dilemma in finite populations, we find that neither a simple reward measure nor a pure punishment mechanism can extensively promote cooperation. Instead, a combination of appropriate punishment and reward mechanisms can promote cooperation's prosperity regardless of how large or small the temptation to defect is. In addition, the combination spontaneously produces inhomogeneities in social relations and individual influence, which support the continued existence of cooperative behavior. Finally, we further explain how cooperators establish a sustainable existence under the combination by investigating the social relations at different moments in a small system. These results demonstrate that dispensing rewards and punishments impartially in society is essential to social harmony.

Cooperation plays an important role in the development of human civilization and society. How cooperation emerges in human society, however, is both an evolutionary puzzle and a practical one that has real implications for social harmony. Recently, scientists have revealed multiple mechanisms for promoting cooperation, one of which is population structure, as it enables localized reciprocity. However, this explanation assumes static social interactions, whereas human interactions are often dynamic. To this end, we combine network reciprocity with a reward–punishment mechanism to achieve coevolution of network structure and individual behavior. The results show that an appropriate combination of rewards and punishments can greatly promote the prosperity of cooperation and maintain social order and development.

I. INTRODUCTION

Cooperation is at the core of the success of human society.[1–4] However, cooperation is costly: Cooperators contribute to the collective at personal cost, whereas defectors only enjoy social welfare. According to the principle of "survival of the fittest" in the biological evolutionism, the low-payoff cooperative behavior will be eliminated by the high-payoff defective behavior. Therefore, evolution of cooperation has become a dilemma.[5] The evolutionary game theory established by Smith and Price[6] and enriched by Smith, Gintis, and Nowak[7–9] provides a practical theoretical framework for exploring the resource allocation behavior of participants with bounded rationality. A large number of research works has extensively explored the methods of promoting the defectors in a dilemma to become cooperators under this framework.[10–14]

Inspired by spatial games, many works have focused on exploring whether structured groups can promote the evolution of cooperation.[15] Among them, complex networks have significant advantages in exploring the influence of different structures on cooperation.[15–22] They are efficient abstractions of the spatial structure of groups. Specifically, individuals are likened to nodes in a network, and their relationships are described by edges between nodes. In a well-mixed group, based on the basic rule of Darwinian evolutionary selection, the extinction of cooperators is inevitable.[23]

The spatial structure of a group offers a solution to this problem.[5] It is introduced by the interaction of neighbors, and the group is formed in the process of evolution, which protects cooperators from being exploited by defectors.[24] The complex spatial structures (also called social diversity) has been proven to play a significant role in promoting the evolution of cooperation.[25,26]

To date, most research on network games has aimed to describe the evolution of strategies in static networks with different structural characteristics.[24] However, in real-world social networks, the network structure is dynamically changing.[27] It has been recognized that social relationships affect individual decision-making in social networks and that personal behavior affects social relations.[11,28–31] For example, in the work of Hanaki *et al.*,[32] players not only learned to change their strategies but also to choose to interact with other players by creating and/or severing ties. The results showed that cooperation significantly improves when connections are costly. Rand *et al.*[11] compared networks with different update speeds and concluded that dynamic networks promote cooperation better than static networks in frequently updated networks. Therefore, enriching and developing evolutionary games require coevolution between strategies and the population structure. Just as in social life, people are not isolated individuals, and various interactions continue to occur. Under multiple needs, each person develops a unique personalized network over the advancement of time. The phenomenon of individuals interacting to generate connections in social groups has attracted widespread attention in economic societies.[33–35]

Apart from social diversity, attributes, such as reputation, teaching activities, pure reward mechanism, and pure punishment mechanism in social life, also affect the emergence of cooperative behavior.[36–39] In particular, teaching activities[40] are considered to be a developmental attribute of participants. Simple rules of coevolution may cause highly dispersed teaching activities from the initial nonpreferential setting, leading to heterogeneous leadership in the evolutionary process,[40–44] thereby promoting cooperation in social dilemmas. A more efficient approach to promoting cooperation is to establish reward or punishment mechanisms.[45–51] For example, Yang *et al.*[52] proposed a reward mechanism in which group members determine the allocation of the bonus raised by the tax in the public goods game experiment, which showed that endogenous rewards effectively promote the contribution of public goods. However, an enormous amount of research has focused on fines for defectors.[50,53,54] It has been found that under these mechanisms, a single punishment cannot promote cooperation under all circumstances. Wu *et al.*[55] analyzed four comparative experiments and concluded that expensive punishments will not promote cooperation. However, rewards and punishments coexist in real social management. The above works only consider the reward mechanism or punishment mechanism respectively, lacking the synergy between the two processes. It has even been concluded that rewards or punishments do not always promote cooperation.[50,55] In contrast, we focus on the evolution process of cooperation under the synergy of reward mechanism and punishment mechanism, so as to explore the optimal reward and punishment intensity in different social environments.

In this work, we propose a new combined reward–punishment mechanism (RPM) model to regulate individual behaviors for maximizing collective benefits. The simulation results suggest that the pure reward mechanism (RM) hinders the prosperity of cooperation, and the pure punishment mechanism (PM) promotes cooperation only when the temptation to defect is great, but the RPM makes up for this shortcoming and significantly improves the level of cooperation. These phenomena indicate that our RPM is successful, and fair rewards and punishments are significant to social harmony, which is in line with the general law of social governance. The government requires strict laws; additionally, relevant preferential or encouraging policies are also necessary to promote the standard progress of society. In addition, the RPM builds a bridge between social relationships and individual behaviors, realizing their coevolution. Under this coevolution, the heterogeneity of social relations and individual influence occurs spontaneously. This finding provides insights for revealing the origin of heterogeneous networks from the perspective of evolutionary game theory, which is widely explored.[40,56–60]

II. MODEL

A. Reward–punishment mechanism

Human interactions are not random but are structured in social networks. Importantly, ties in these networks are often dynamic. These changes may be due to the break in friendships caused by the betrayal between friends or due to the gradual expansion of personal social relations. We merge the two cases by introducing an RPM on social relations to investigate the coevolution of network structure and individual behavior.

The PM is in line with the common saying, "Three strikes and you're out." When a friend betrays you many times beyond your patience, you may sever ties with him or her. Ignoring the differences in individual patience, a person's patience can be generalized to the tolerance of society. If society values or is sensitive to trust, one betrayal will lead to the breakdown of a friendship. However, if a society has a high tolerance for betrayal, even if a person has been betrayed many times, he or she will keep in touch with the betrayer. Formally, consider a social network $G = (\mathscr{V}, \mathscr{E})$ where the nodes $\mathscr{V} = \{1, \ldots, N\}$ correspond to individuals and each edge in the set $\mathscr{E} \subseteq \mathscr{V} \times \mathscr{V}$ represents a 2-player game between neighboring individuals. Define the weight C^d_{xy} of each link as the cumulative number of times that individual x is betrayed by neighbor y, which is asymmetric and time dependent. Define $k_1 \in [0, 1]$ as the penalty or tolerance factor, which indicates that the larger the k_1 is, the lower the society's tolerance, and vice versa. Assume that the probability of x severing contact with y is as follows:

$$P^1_{x \nrightarrow y} = \tanh\left(k_1 C^d_{xy}\right) = \frac{e^{k_1 C^d_{xy}} - e^{-k_1 C^d_{xy}}}{e^{k_1 C^d_{xy}} + e^{-k_1 C^d_{xy}}}, \tag{1}$$

which means that the more times y betrays x, the larger the probability that x severs ties from y.

The RM is used to model the expansion process of an individual friendship network. When a person's influence accumulates to a certain level, his or her social circle will expand accordingly. For example, if you are influential in a village, the villagers are likely to choose you as the village chief. As a result, you will know the mayor or other village chiefs when you attend a town committee meeting. This phenomenon involves society's evaluation criteria for excellence. Low social standards mean that individuals need to be only generally good (influential) to activate reward behaviors to expand their social circle. In contrast, high social standards indicate that the threshold for triggering rewards is high, and an individual may need to be extremely excellent to develop his or her contacts.

Analogously, the teaching activity (or influence) C_x^s of individual x is defined as the cumulative number of times that his or her behavior is learned (or imitated) by neighbors with different strategies. The teaching activities of individuals vary from person to person, change over time, and represent the abilities of individuals to make opponents learn their strategies. The higher the individual's teaching ability, the higher the probability of receiving a reward. Like the PM, the RM considers a random reward rule. Individual x randomly selects individual $z \notin \tilde{\mathcal{N}_x}$, where $\tilde{\mathcal{N}_x}$ is the set of neighbors, including himself/herself, and creates a new link between them according to the following probability:

$$P^2_{x \to z} = \tanh(k_2 C_x^s) = \frac{e^{k_2 C_x^s} - e^{-k_2 C_x^s}}{e^{k_2 C_x^s} + e^{-k_2 C_x^s}}, \tag{2}$$

where $k_2 \in [0, 1]$, called the reward or influence factor, reflects society's standard of excellence. The larger the k_2 is, the lower the society's standard; therefore, it is easy to activate the reward behavior. Conversely, the smaller the k_2 is, the higher the threshold to trigger the reward.

Finally, we argue that punishment (trust detection) and reward (teaching) go hand in hand, not sequentially. It can be the same neighbor who is selected for trust detection and teaching. For example, a friend betrays you many times beyond your patience, you sever ties with him or her. At the same time, he or she feels that your behavior is successful and imitates your behavior.

B. Coevolution of structure and behavior

It should be noted that our mechanism can be applied to almost all existing network evolutionary game models. In what follows, we take the classic prisoner's dilemma (PD) game, which has been widely studied to explore complex cooperation mechanisms,[1] as an example to discuss the impact of the RPM on the evolution of cooperation. We choose the homogeneous Erdős–Rényi (ER) random graph[61] as the initial underlying network structure and consider the stochastic strategy evolution rule, the reversed Fermi rule.[62] Under this rule, the behaviors of successful individuals are more likely to be imitated by others.

In classic game theory, each game player can choose one of two behaviors: cooperation (C) or defection (D). Once the behaviors are determined in each game shot, they can get a payoff according to the following payoff matrix:

$$A = \begin{bmatrix} R & S \\ T & P \end{bmatrix},$$

in which R, P, S, and T represent the reward for mutual cooperation, the punishment for mutual defection, the sucker's payoff, and the temptation to defect, respectively. The game is a PD if $T > R > P > S$. For the weak PD game,[24] the payoff matrix can be simplified to $R = 1$, $P = S = 0$, and $T = b$. Generally speaking, b takes values from 1 to 2, which characterizes the extent of the advantage of D against C. For the completeness of the experiment, in our simulations, we expand the upper bound of b to 2.5. Note that when $k_1 = k_2 = 0.0$, our model will degenerate into a traditional PD game.

The social relations among the population are abstracted by undirected networks. Initially, we characterize the network structure as ER random graphs.[61] In addition, in light of the resulting heterogeneous distribution of social relations and individual influence, we compare the Barabási–Albert (BA) scale-free network[56] as the underlying network, see Sec. I in the supplementary material for details. Finally, in order to verify that the results are in line with the real world, we apply our RPM model to the real networks[63] in Sec. G of the supplementary material.

In many social environments, individuals imitate more successful behaviors. We attempt to describe this phenomenon in the language of evolutionary dynamics. After playing a round of a PD game and gaining cumulative profits, individuals seek more successful strategies by comparing their own earnings with those of their neighbors. Although the behaviors of successful individuals are more accessible for other individuals to imitate, such imitations are not deterministic. Thus, a stochastic strategy evolution rule, the reversed Fermi rule,[62] is adopted to mimic the strategy evolution. That is, each individual x randomly opts a neighbor y, and if they have different strategies, then y adopts x's strategy with a probability of

$$W_{S_y \leftarrow S_x} = \frac{1}{1 + \exp[(\Pi_y - \Pi_x)/K]}, \tag{3}$$

where $K = 0.1$ denotes the amplitude of noise and S_x and Π_x are the strategy and the payoff of x in the current round, respectively.

Initially, each agent evenly chooses cooperation or defection with the same probability. Then, by repeating the above evolution mechanism of structure and behavior [see Fig. 1(a) and Algorithm 1], the system is pushed to the final steady state described by the average proportion of cooperators (i.e., average cooperation frequency) ρ_C and the stable distributions of the degree (i.e., number of neighbors) *Deg* and the teaching activity C^s. The steady-state average cooperation frequency is an important indicator to measure the degree of system cooperation emergence. Specifically, the stationary average frequency of cooperation ρ_C is determined within 10^4 time steps after discarding sufficiently long transient time steps. Moreover, because the random initial state and RPM may introduce additional disturbances, the final results are averaged over 50 independent realizations for each set of parameter values to ensure accuracy. Unless otherwise noted, simulation results are obtained on populations comprising $N = 1000$ individuals, each with six connections on average (see Sec. F in the supplementary material for other average number of connections).

III. RESULTS

A. Cooperative behavior under the reward–punishment mechanism

In contrast to the traditional PD game (i.e., $k_1 = k_2 = 0.0$) in ER networks, where cooperators disappear when a large temptation to defect exists, cooperative behavior persists under the RPM with cluster formation to defend against the invasion of defectors. See Figs. 1(b)–1(d) as an example of an intuitive illustration when the temptation to defect is in the middle ($b = 1.4$). Furthermore, for

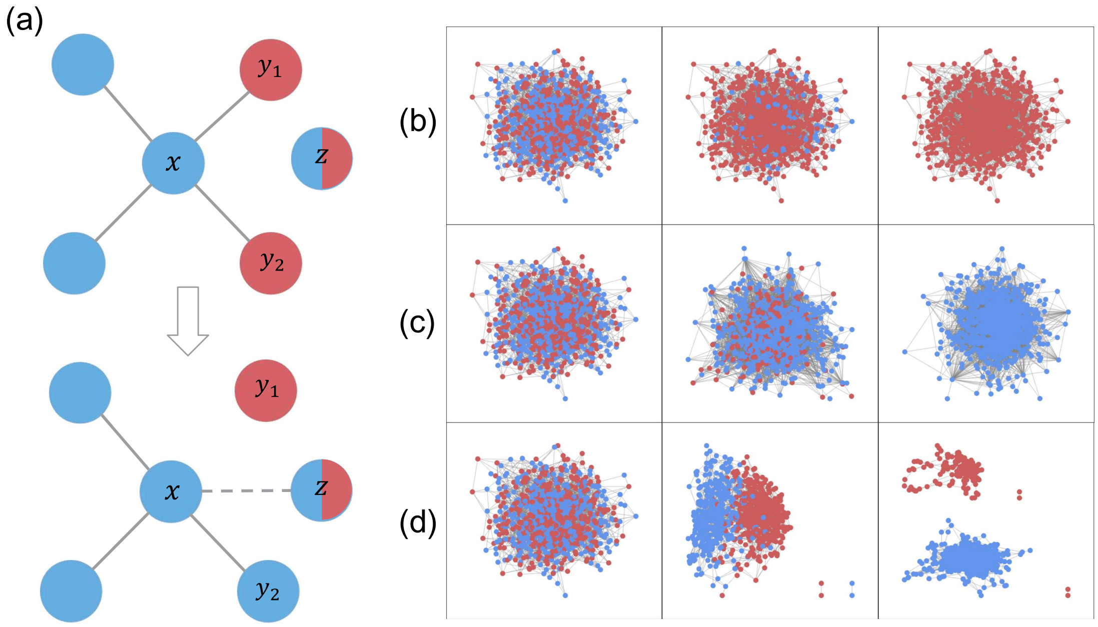


FIG. 1. Coevolution of network structure and behavior under the RPM. Blue and red represent cooperation and defection, respectively. (a) The microprocess of RPM. The currently selected individual is regarded as an x-type individual; y_1-type and y_2-type individuals represent the selected neighbors for trust detection and teaching, respectively; the z-type individual represents the selected individual for expansion of contacts. Note that y_1 and y_2 can be the same neighbor. (b)–(d) Macro snapshots of the PD game under the RPM at $b = 1.4$ with $k_1 = k_2 = 0.0$ (b), $k_1 = k_2 = 0.6$ (c), and $k_1 = 1.0$ and $k_2 = 0.2$ (d).

different temptation to defect, Fig. 2 compares the optimal average cooperation frequency in the steady state under our model (ρ_C^{opt}) and the traditional PD game (ρ_C^{Tra}). Compared to the traditional PD game, our model can significantly improve cooperation, especially under a large temptation. The corresponding optimal solutions (k_1^*, k_2^*) are all greater than 0, reflecting the significance of the combination of social reward and punishment in promoting social cooperation. In addition, not only the optimal solution, for any $k_1, k_2 > 0$ [see Figs. 3(a)–3(d)], compared with the traditional PD game, in which the steady-state average cooperation frequency keeps high for a short period of time with the increase of the temptation of defect and then rapidly declines to 0, under our mechanism, it first remains high for a long time, then rapidly drops, and finally slowly drops to a steady state. Cooperative behavior persists in populations with fairly high temptation to defect.

It is worth pointing out that there is an interesting phenomenon in Fig. 3(b), where there is a small "convex"; that is, near the inflection point, when the temptation to defect increases, the cooperation frequency increases. For example, when $k_1 = 0.4$, the cooperation frequency of $b = 1.3$ is less than that of $b = 1.4$. A survey of the results of 50 independent realizations [Fig. A2(ce) in the supplementary material] found that this is because the steady state in some cases at the inflection point ($b = 1.3$) is all defection. To better explain this phenomenon, we need to know how the intensity of reward and punishment affects the evolution of cooperation.

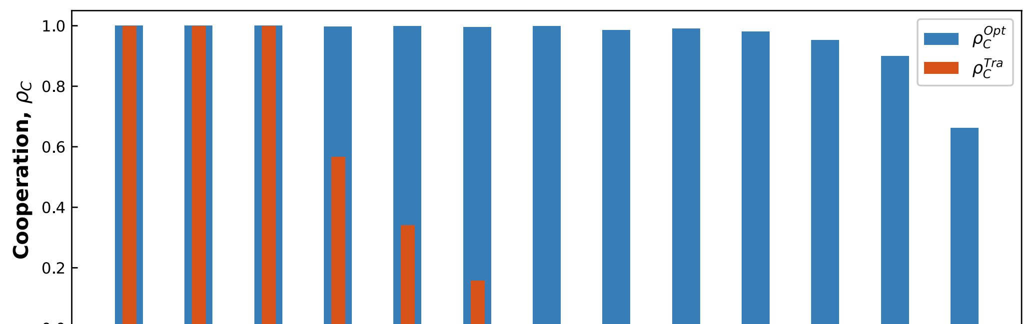
Figure 3(e) shows that when there is no RM ($k_2 = 0.0$), punishment reduces the polarization of cooperation caused by different temptations to defect, which characterizes the discrete degree of cooperation frequency and is quantified by the standard deviation *std*. Specifically, when there is no punishment ($k_1 = 0.0$), a small change in the temptation of defect can make the cooperation frequency oscillate from 1 to 0, and the polarization (i.e., standard deviation) of cooperation at this time is 0.4308. This polarization indicates social and economic unrest and instability. As the intensity of punishment increases, the mean level of cooperation (black solid line with triangles) remains almost unchanged, while its polarization decreases from 0.4308 ($k_1 = 0.0$) to 0.1656 ($k_1 = 1.0$), which greatly stabilizes the social unrest caused by the different temptations of defect. In addition, the existence of punishment ($k_1 > 0.0$) weakens the influence of the large temptation to defect ($b \geq 1.15$) to support cooperation (compared to the case when $k_1 = 0.0$). Especially when temptation is quite large, cooperators do not disappear, in contrast to the effect in the traditional PD game. However, when temptation to defect is small ($b < 1.15$), a pure PM does not promote cooperation. Therefore, it is important to dispense rewards and punishments impartially. Now, we consider the case when the RM exists, taking $k_2 = 0.8$ as an example [Fig. 3(f)]. Regardless of the value of temptation, cooperation first increases significantly and then declines slowly with increasing punishment. The difference is that the punishment required for the cooperation peak increases with the increase in temptation to defect. This indicates that "excessive punishment" is not desirable, and the greater the temptation to defect, the greater the punishment required. Furthermore, from a horizontal perspective, as the intensity of reward increases [see Figs. A1(n)–A1(r) in the supplementary material], greater

ALGORITHM 1. RPM.

Input: Network $G(N, \langle Deg \rangle)$, parameters (b, k_1, k_2);
Initialize $C^s, C^d = \mathbf{0}$;
Initialize the strategy S by randomly allocating cooperation (1) or defection (0) with uniform probability for each node;
repeat
    Randomly select node x, and neighbors $y_1, y_2 \in \mathcal{N}_x$;
    // Punishment (Trust Detection) & Reward (Teaching) Are Simultaneous
    // Trust Detection & Punishment
    if $S_x == 1$ and $S_{y_1} == 0$ **then**
        $C^d_{xy_1} + +$;
        Calculate the punishment probability $P^1_{x \nrightarrow y_1}$ with Eq. (1);
        if *Reach the punishment threshold* **then**
            Break the link between x and y_1;
        end
    end
    // Teaching & Reward
    if $S_x \neq S_{y_2}$ **then**
        Calculate the cumulative payoffs Π_x, Π_{y_2} of the game between x, y_2 and their neighbors, respectively;
        Calculate the strategy transition probability $W_{S_{y_2} \leftarrow S_x}$ with Eq. (3);
        if *Successful strategy transfer/teaching* **then**
            $S_{y_2} = S_x$;
            $C^s_x + +$;
            Calculate the reward probability P^2_x with Eq. (2);
            if *Reach the reward standard and* $Deg_x < N - 1$ **then**
                Establish a link between x and random $z \notin \tilde{\mathcal{N}_x}$;
            end
        end
    end
until *convergence*;

punishment is needed to promote the emergence of cooperation. See Table I for details.

The same analysis is applicable to the influence of reward on the evolution of cooperation. The existence of reward alone [Fig. 3(g)] will annihilate cooperation because the RM enhances the temptation that cooperators defect for more benefits. This finding indicates that "pure reward education" is a failure. However, as punishment increases [Figs. A1(t)–A1(x) in the



	$b = 1.0$	1.05	1.1	1.15	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
	0.9983	0.9985	0.9984	0.5665	0.3407	0.1571	0.0028	0.0021	0.0013	0.0011	0.0012	0.0014	0.0012
	0.9999	1.0000	0.9999	0.9981	0.9994	0.9957	0.9983	0.9859	0.9908	0.9813	0.9524	0.9007	0.6632
(k_1^*, k_2^*)	(0.2, 1.0)	(0.2, 1.0)	(0.2, 0.2)	(0.4, 1.0)	(0.4, 1.0)	(0.4, 0.2)	(0.6, 1.0)	(0.8, 1.0)	(0.8, 1.0)	(0.8, 0.4)	(1.0, 1.0)	(1.0, 0.8)	(1.0, 1.0)

FIG. 2. The global maximum value (optimum value) ρ_C^{Opt} of function: $(k_1, k_2) \rightarrow \rho_C$ for different values of b. (k_1^*, k_2^*) is the corresponding global maximum point (optimum solution). In addition, ρ_C^{Tra} is the frequency of cooperation when $k_1 = k_2 = 0.0$.

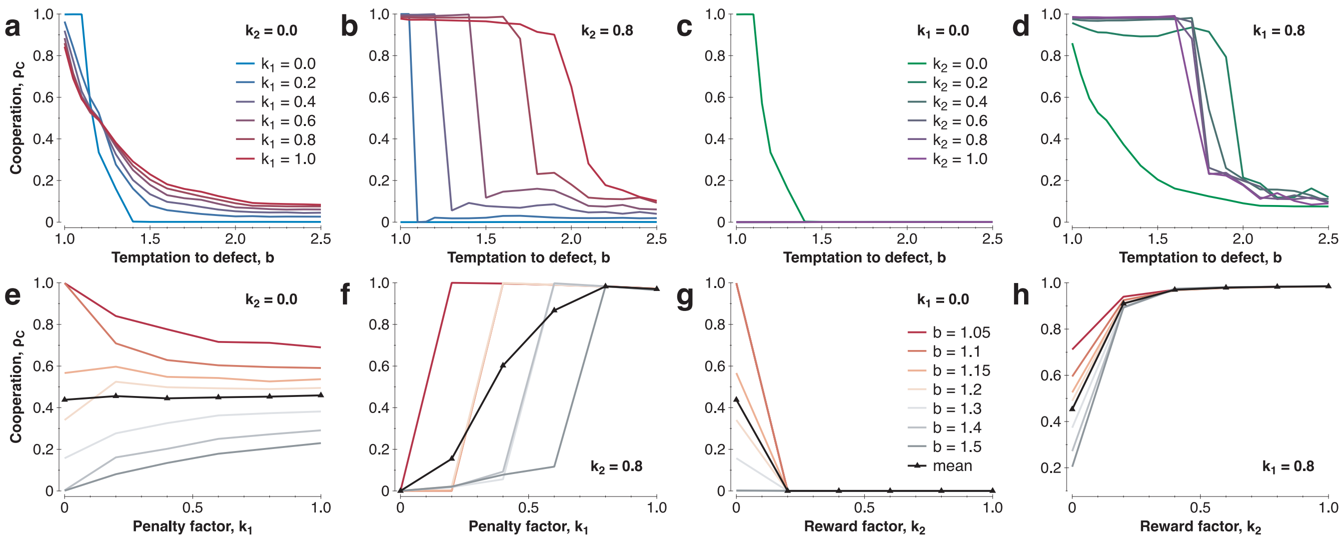


FIG. 3. The average frequency of cooperation in the steady state, ρ_C, as a function of the temptation to defect b (a)–(d), the penalty factor k_1 (e)–(f), and the reward factor k_2 (g)–(h). See Fig. A1 in the supplementary material for complete parameters (b, k_1, k_2).

supplementary material], an increasing number of polylines are "pulled up" ($\rho_C \gg 0$). From this perspective, it also indicates that punishment acts as a counterbalance to reward and temptation to defect and that sufficient punishment can offset the betrayal caused by reward and temptation. After a polyline is "pulled up" [Fig. 3(h)], cooperation increases with increasing reward, indicating that punishment and reward are also interdependent. The combination of reward and punishment not only reduces the polarization of cooperation, but it also increases the mean to obtain the optimal frequency of cooperation (e.g., when $b = 1.1$ and $k_1 = k_2 = 0.2$, $\rho_C = 0.9999$). As illustrated in Fig. 2, for different values of temptation to defect, the optimal frequency of cooperation can be obtained by adjusting the intensity of reward and punishment.

TABLE I. $\text{argmax}_{k_1} \rho_C = f(k_1, k_2, b)$, given k_2 and b.

	k_2					
b	0.0	0.2	0.4	0.6	0.8	1.0
1.0						
1.05			$k_1 = 0.2$	$k_1 = 0.2$	$k_1 = 0.2$	$k_1 = 0.2$
1.1	$k_1 = 0.0$	$k_1 = 0.2$				
1.15						
1.2	$k_1 = 0.2$		$k_1 = 0.4$	$k_1 = 0.4$	$k_1 = 0.4$	$k_1 = 0.4$
1.3		$k_1 = 0.4$				
1.4			$k_1 = 0.6$	$k_1 = 0.6$	$k_1 = 0.6$	$k_1 = 0.6$
1.5						
1.6		$k_1 = 0.6$				
1.7			$k_1 = 0.8$	$k_1 = 0.8$	$k_1 = 0.8$	$k_1 = 0.8$
1.8						
1.9						
2.0	$k_1 = 1.0$	$k_1 = 0.8$	$k_1 = 1.0$	$k_1 = 1.0$	$k_1 = 1.0$	$k_1 = 1.0$

We revisit the interesting phenomenon mentioned above. With $k_1 = 0.4$ and $k_2 = 0.8$, when temptation to defect is small, the punishment is sufficient, and with the blessing of the reward, the cooperation is maintained at a high level. When temptation to defect increases to 1.3, the punishment is not enough (see Fig. 2 and Table I), and the reward is dominant at this time, resulting in some situations equivalent to degenerating to only RM, in which the network evolves to almost fully connected and the cooperators are eliminated. However, as temptation to defect continues to increase, for example, $b = 1.4$, punishment and reward are not enough at this time, and the large temptation of defect dominates. Only a few cooperators severed ties with defectors early, forming a cluster to resist they invasion. The phenomenon shows that it is important to formulate different levels of reward and punishment according to different circumstances.

In Sec. A in the supplementary material, by analyzing the time series of cooperative behavior, we provide a more detailed explanation of why the RPM mechanism promotes cooperation.

B. Heterogeneous structure emerging with the reward–punishment mechanism

From the previous discussion, we know that under the influence of the RPM, the network structure is constantly evolving. In this section, we explore how the network structure evolves and attempt to explain why the RPM can structurally promote cooperation.

First, we count a typical time series of the average (/maximum) degree (/teaching activity) for both cooperators and defectors, as illustrated in Fig. 4. Compared with the pure RM or PM, the RPM induces cooperators to become influential leaders and hubs

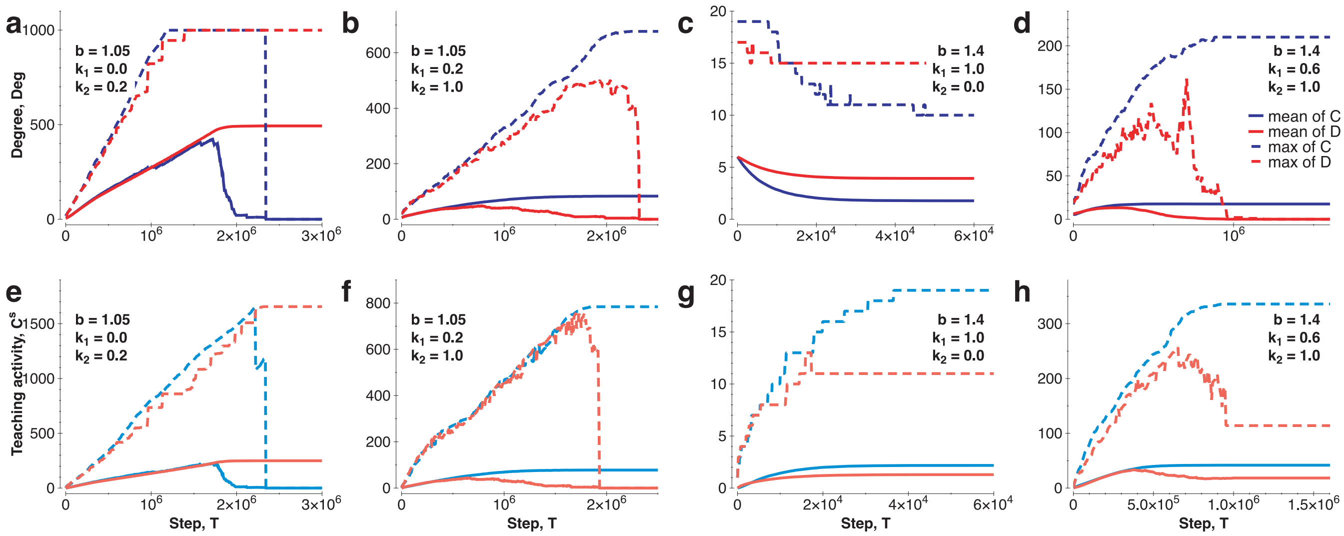


FIG. 4. Plots of a typical time series of the average (/maximum) degree *Deg* (a)–(d) [/teaching activity C^s (e)–(h)] for cooperators *C* and defectors *D* for different values of (b, k_1, k_2).

with many friends. There is a super-hub cooperator (a cooperation node with the maximum degree that is far greater than the average degree), even when cooperators are initially at a disadvantage, which indicates that the RM prefers cooperators, even if it is neutral. Consistent with the assumptions of many theoretical models, individuals seek connections with cooperators and avoid defectors.[11] Therefore, we predict that the existence of the RPM makes homogeneous networks evolve into inhomogeneous networks.

To verify our prediction, we count the degree distribution of the network and the distribution of nodes' teaching activities over time. Figures 5(a)–5(d) show an increasing heterogeneity of degree and nodes' influence with time evolution. It has been shown that hubs in heterogeneous networks play an important role in the emergence of cooperation,[16,64–66] because they tend to adopt a cooperative strategy and can effectively resist intrusion by defectors. Consequently, heterogeneous networks can significantly promote the emergence of cooperative behavior, placing cooperators in dominant positions in the network. In summary, in previous studies, hubs in heterogeneous networks become cooperators to incentivize cooperation. However, in this study, we find an inverse process. Under RPM, cooperators become hubs and leaders to promote cooperation, which changes the networks from a homogeneous to a heterogeneous state. This process also explains why most real-life large-scale social networks present heterogeneity to some extent. At the same time, we also theoretically analyze the expectation of individual degree variation over time; see Sec. B in the supplementary material for details.

Next, we reveal the mechanism that produces the heterogeneity of the degree distribution and nodes' teaching activity distribution in the network. An analysis of our model indicates that the PM is more discriminating than the RM. First, punishment is specifically for defectors. Cooperators disconnect only from defectors and not from other cooperators. In addition, although reward favors cooperation, it is not only for cooperators because the reward received by an individual may connect with a defector. In other words, under the RPM, individuals are willing to give potential new friends the benefit of the doubt, which is reminiscent of a form of forgiveness or leniency that is considered to promote cooperation.[67,68]

For details, the evolution process behind the RPM can more clearly explain why cooperators become influential leaders and hubs. When a cooperator is selected, he or she will successfully teach the connected defector whose payoff is lower and will have a high probability of disconnecting with the defector who always betrays him or her (i.e., who has a high payoff). For the cooperator, first, his or her payoff will increase, and the number of high-payoff defectors around will decrease, resulting in a decrease in the probability that he or she is successfully taught later (i.e., his or her time scales to persist in cooperation become longer). At the same time, his or her influence increases, and additional links are rewarded. Regarding the network, the number of cooperators (defectors) and the *CC* (*CD*) links increase (decrease). The whole process forms a virtuous circle [see Fig. 6(a)], enables cooperators to develop into leaders and hubs, and promotes the emergence of cooperation. However, from the defectors' perspective [Fig. 6(b)], they can successfully teach and receive rewards in the beginning. However, this temporary benefit will have serious consequences. The cooperative neighbors around will gradually turn into defectors, and they will be punished by being disconnected from more cooperators. Thus, their payoffs and the probabilities of successful teaching gradually decrease, which inhibits the reward. As a consequence, they transform into cooperators or into loners who are segregated by cooperators.

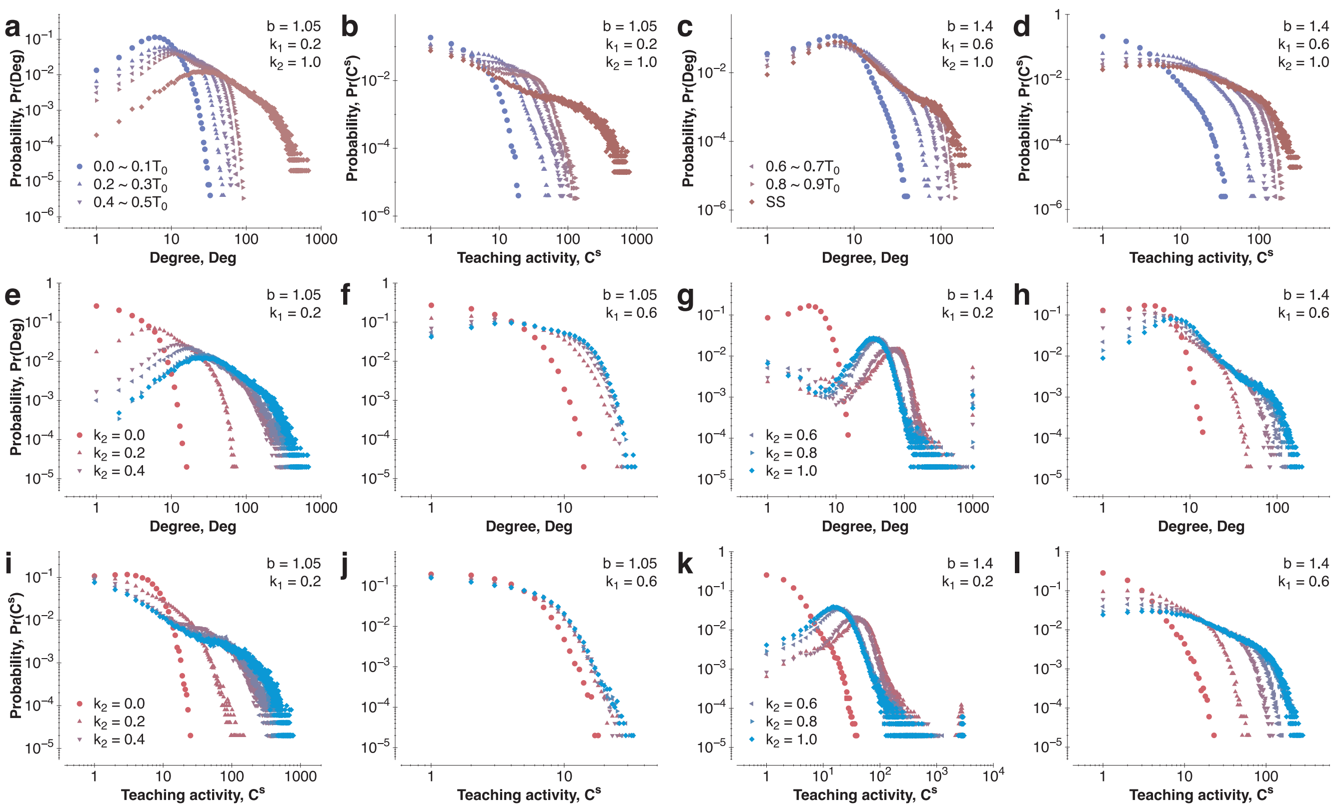


FIG. 5. Log–log plot of the distributions of degree $Pr(Deg)$ and teaching activity $Pr(C^s)$ over time (a)–(d) and in the steady state (e)–(l) for different values of (b, k_1, k_2). Here, T_0 is the transient time, and SS represents the steady state.

The above analysis shows that the structural evolution mode behind the RPM guarantees the emergence of cooperation. However, it remains unclear about how the parameters influence the group structure at the final steady state and how the final population structure correlates with the proportion of cooperation. Next, we report simulations for which the evolution dynamics with varying temptation to defect and the intensity of reward and punishment are shown in Figs. 5(e)–5(l) (see Sec. C in the supplementary material for more complete parameters).

First, when the punishment is insufficient [Fig. 5(g)], the overall node degree of the network is relatively larger. The reason is that, in this case, the number of new links established due to rewards is far greater than that of the links disconnected due to punishments, which is equivalent to the case where only the RM works. However, when the punishment is sufficient, as punishment increases, the maximum degree of the network decreases, and the proportion of nodes with a large degree decreases (i.e., the degree heterogeneity decreases). In contrast, with the increase in the reward, the maximum degree of the network increases, and the proportion of hub nodes increases (i.e., the heterogeneity of the degree distribution increases).

In turn, based on the final degree distribution, we can judge whether the punishment is sufficient. On the double logarithmic scale, the bell-shaped degree distribution corresponds to insufficient punishment, while the heavy-tailed degree distribution means that the punishment is sufficient. From the above discussion, we know that when the punishment is not severe enough, it is equivalent to the pure RM case, and the cooperative behavior will be eliminated. Therefore, we can obtain the relationship between the degree distribution and the cooperation ratio in the steady state under the RPM. On the double logarithmic scale, the bell-shaped degree distribution corresponds to a cooperation ratio of approximately 0, and the heavier the tail of the degree distribution, the greater the cooperation ratio.

Figures 5(i)–5(l) show that the distribution of teaching ability is consistent with the degree distribution. Therefore, regardless of the structure of the initial interactive network, the introduction of RPM makes the degree distribution of the network and the distribution of nodes' teaching activities exceedingly inhomogeneous. Moreover, this spontaneous inhomogeneity is an essential factor in maintaining the emergence of cooperative behavior in the PD game.

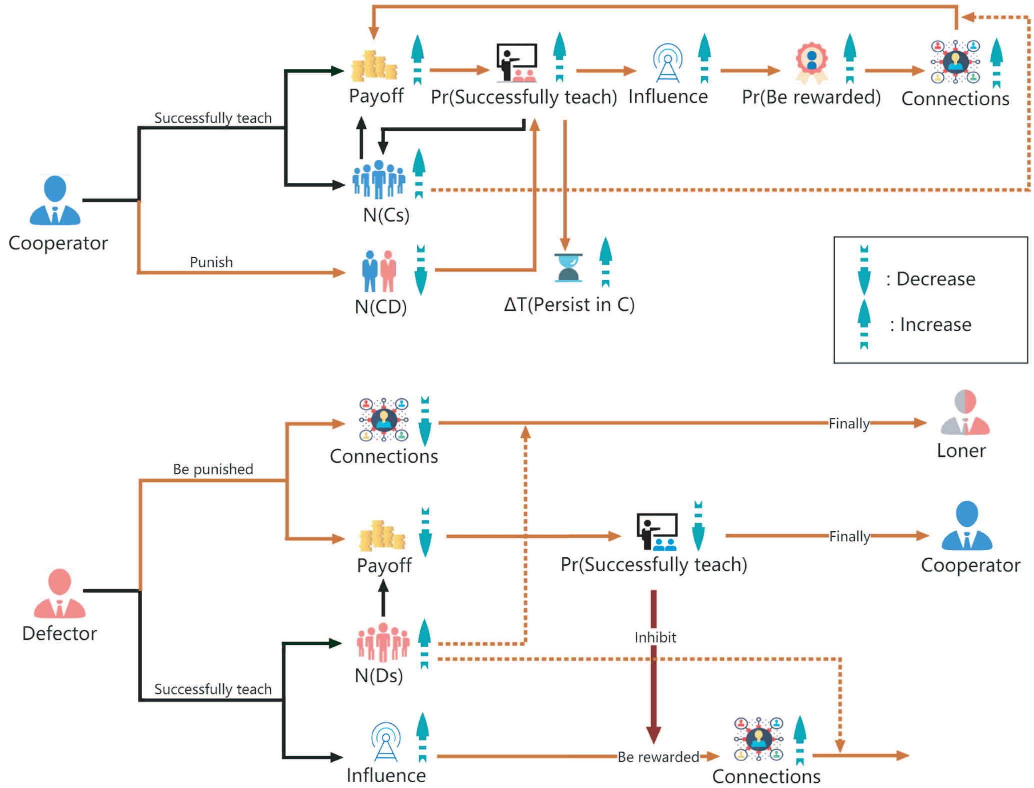


FIG. 6. Flow charts from the perspective of a cooperator (a) and a defector (b). *Pr*, *N*, and ΔT indicate the probability, number, and time scale, respectively. Black flow arrows indicate the processes of the traditional game, and orange and red ones are special for the RPM. Unless otherwise noted, a solid line indicates leading to, and a dotted line indicates blessing.

In addition, we observe a highly positive correlation between degree *Deg* and teaching ability C^s , which will be discussed further in Sec. D of the supplementary material.

C. Learning patterns of cooperators and defectors

To further dissect how cooperators maintain sustainability under the RPM, in this part, we explore the different learning patterns of cooperators and defectors, microcosmically providing solid experimental evidence for our two observations. First, the RPM can incentivize cooperation, and it is vital to dispense rewards and punishments impartially. Second, under the RPM, cooperators are both influential leaders and hubs with many friends, and there is a super-hub cooperator who has the maximum degree.

To provide an intuitive illustration of how the evolution of individual behavior and population structure are related, we visualize a series of PD game snapshots in which the initial interaction network is an ER network with 100 nodes and an average degree of 3 (see Fig. 7 and Figs. E11–E18 in the supplementary material). Note that in the following simulations, the initial network structure and strategy remain unchanged to avoid introducing unnecessary randomness.

First, for a traditional PD game with $b = 1.5$ [$k_1 = k_2 = 0.0$; see Figs. 7(b)–7(e)], immense temptation gives an absolute advantage to defecting, and the proportion of cooperation decreases sharply over time. Specifically, defectors quickly occupy most nodes to become leaders and then disperse and invade operations until cooperators are eliminated.

If there is only the PM [$k_1 = 1.0, k_2 = 0.0$ in Figs. 7(f)–7(i)], cooperation persists and is robust and stable, which is consistent with the previous conclusion that a pure PM supports cooperation by weakening the influence of b. Unlike the classic reciprocity mechanism, such as the TFT strategy, the PM does not change an individual's behavior to respond to the behavior of the interacting neighbors. Indeed, it is extreme to change a person's cooperative behavior just for a single defector in social networks, because other interactive individuals may be more critical cooperators.[69] Under the PM, individuals do not have to give up cooperation to punish free-riders, but avoid contacting them and exclude them from the benefits of future cooperation, thereby inhibiting the occurrence of betrayal. In addition, the PM will not reduce the benefits for cooperators, but defectors will pay for their actions. The costly punishment will deter some individuals who would have defected otherwise from defecting easily. It also makes it more likely that some punished defectors will turn to cooperation in the future. We can see that the

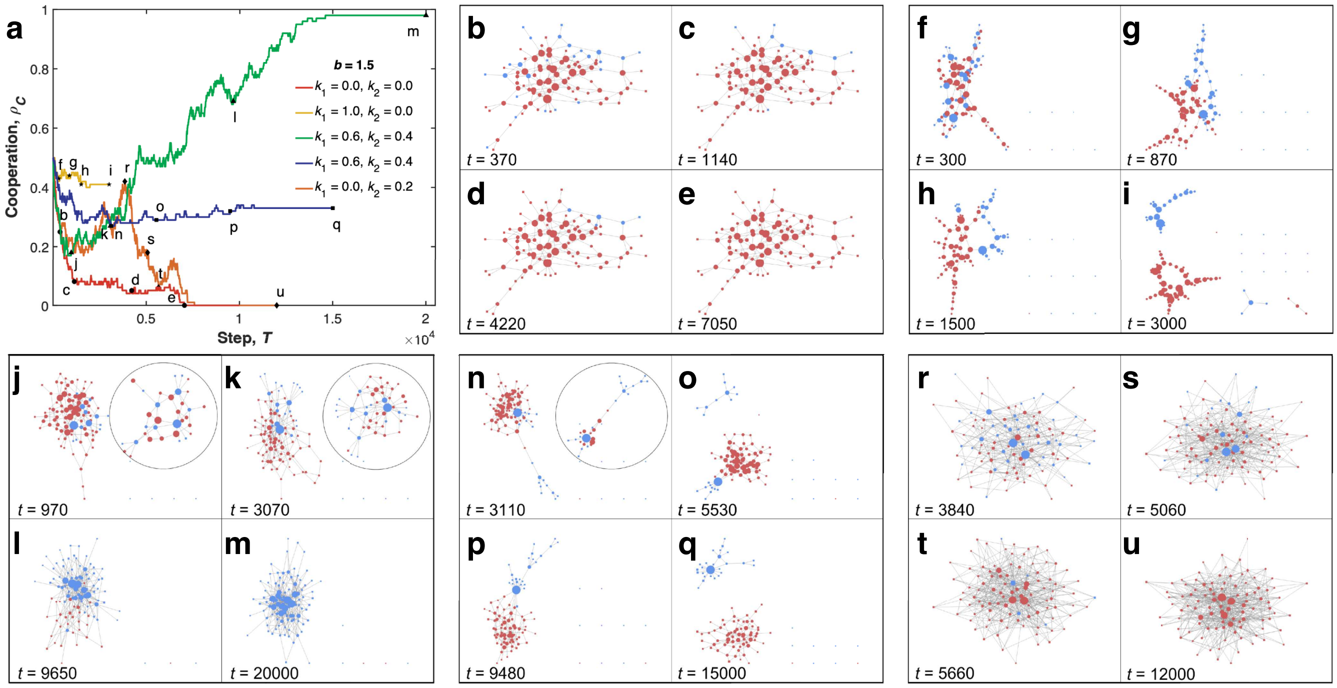


FIG. 7. (a) Plot of a typical time series of cooperation frequency, ρ_C, for different (k_1, k_2). (b)–(u) Snapshots of the PD game under the RPM at different time steps t and the intensity of reward and punishment (k_1, k_2). The blue dots represent cooperators, the red dots represent defectors, and the sizes of the dots correspond to their degrees. Here, $N = 100$, $\langle Deg \rangle = 3$, and $b = 1.5$ with $(k_1, k_2) = (0.0, 0.0)$ (b)–(e), $(k_1, k_2) = (1.0, 0.0)$ (f)–(i), $(k_1, k_2) = (0.6, 0.4)$ (j)–(m) and (n)–(q), and $(k_1, k_2) = (0.0, 0.2)$ (r)–(u).

correlation is positive: The more punished defectors there are, the greater their tendency to cooperate. In addition, the rupture of social relations caused by punishment effectively prevents the invasion of betrayers.

However, when the reward and punishment coexist (here, opting for $k_1 = 0.6$, $k_2 = 0.4$), cooperation is promoted more than under the pure PM [Fig. E19(c) in the supplementary material], and there are two different evolution patterns of cooperation proportion. Figures 7(j)–7(m) show the emergence of a giant component consisting of only cooperators protecting cooperation. Figures 7(n)–7(q) show the two independent connected components, C-cluster and D-cluster, which prevent defectors from invasion. In the early stage of the game, under the enormous temptation to defect (here, $b = 1.5$), the amount of cooperators decreases sharply over time. However, due to the existence of the RPM, the super-hub cooperator (the highest degree) appears for the two situations discussed above. The difference is that for the first case, in addition to the super-hub cooperator, other hub cooperators form a closely connected cluster [Figs. 7(j) and 7(k)]. When the proportion of cooperators decreases to a certain level, the payoff, Π, of defectors located at the boundary of the C-cluster is significantly lower than that of the cooperators [see Figs. 8(a) and 8(b)]. The cooperators can, in turn, successfully teach the defectors. At this moment, teaching is the primary action for cooperators, supplemented by the punishment of disconnection. Over time, the cooperators occupy the hub nodes and finally form a giant connected component in the network [Figs. 7(l) and 7(m)]. For the latter, although there is also a super-hub cooperator, the cooperators present separated star-like or filamentous clusters [Figs. 7(n) and 7(o)]. Only the super-hub cooperator has a higher benefit than the defectors among the individuals on the border between the C-cluster and the D-cluster [Fig. 8(c)]. Currently, punishment for link breaking is the primary action for cooperators, and teaching is supplemented. This is reduced almost to the pure PM case. As shown in Fig. 7(i), the network eventually develops into two independent clusters [Fig. 7(q)].

For the completeness of the discussion, we also analyze the case with only reward [$k_1 = 0.0$, $k_2 = 0.2$, as shown in Figs. 7(r)–7(u)]. As mentioned in Secs. III A and III B, the RM does not promote cooperation. At the beginning of the game, the trend of ρ_C is the same as in Figs. 7(j) and 7(k). However, when the proportion of cooperation rises to a certain level, the payoff, Π, of defectors located on the border between clusters C and D is greater than those of their cooperative neighbors [Figs. 7(r) and 8(d)], and cooperation begins to decline sharply. Therefore, once the cooperator who cannot stop interacting fails in teaching, he or she can only face the invasion of defectors. Additionally, we further verify that the pure RM does not promote cooperation, but only reduces it [Figs. E11, E12, and E19(a) in the supplementary material].

Therefore, it is significant to dispense rewards and punishments impartially in social networks, and the combination of

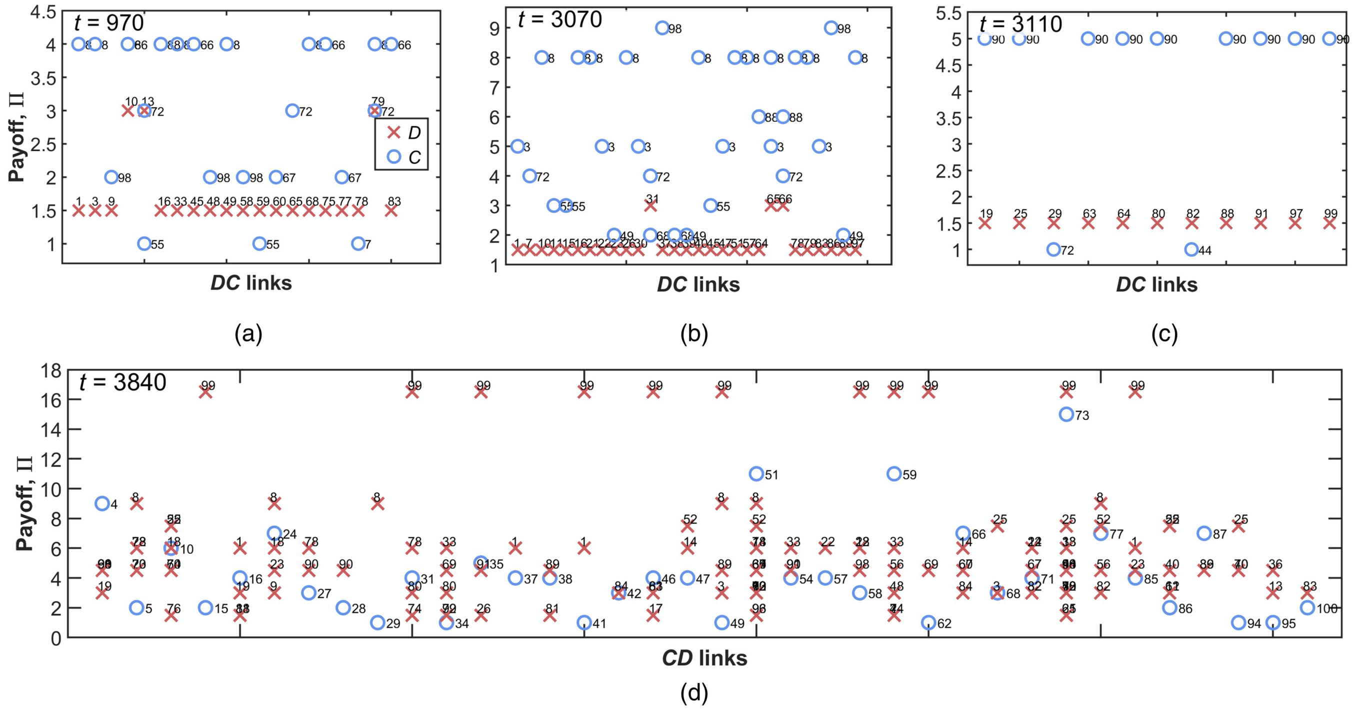


FIG. 8. The payoff, Π, for different *DC* or *CD* links. (a) and (b) correspond to Figs. 7(j) and 7(k), respectively; (c) corresponds to Fig. 7(n); and (d) corresponds to Fig. 7(r). Individual labels are represented by numbers. The labels of defectors are above the corresponding cross signs and that of cooperators are on the right side of the corresponding open circles.

appropriate punishments and rewards can better incentivize cooperation. In the process of the RPM evolution, cooperators become influential hubs, which causes network structural heterogeneity and promotes the emergence of cooperative behavior.

IV. DISCUSSION

In many social situations, individuals choose how to interact with others and with whom they interact. These two processes are called behavioral dynamics and structural dynamics, respectively. The coevolution of structure and behavior is a critical and fundamental problem.[32] Furthermore, rewards and punishments are essential factors that affect people's behavioral choices and are widespread in the real world, from biological systems to economic and social systems. In this context, we build a bridge between these two dynamics through a social RPM to realize the coevolution of behavior and interaction.

Taking the PD game as an instance, we report simulations on both constructed and real networks showing that the frequency of cooperation is greatly improved if appropriate punishments and rewards are introduced simultaneously, especially under immense temptation to defect. The results also demonstrate that dispensing rewards and punishments impartially in society is essential to social harmony. For the case on the real networks,[63] see Sec. G in the supplementary material.

Then, we pay attention to the following two questions: Why can the appropriate punishment promote cooperation (rather than the larger, the better) under the RPM? Moreover, why does the optimal punishment increase as the temptation to defect increases? These phenomena are in line with reality. In general, different punishments are set according to the severity of the circumstances. In our model, the level of temptation to defect represents the seriousness of the circumstances. Thus, hierarchical and appropriate punishments can better promote cooperation.

Concurrently, undesirable "excessive punishment" also indicates that excessively frequent network updating is undesirable. Rand *et al.*[11] verified that rapid network updating promotes cooperation, and in viscous dynamic network conditions, cooperation decreases over time and is eventually eliminated. We conclude that it is not advisable to update the network too frequently. In contrast, an appropriate network update frequency can better promote cooperation.

In the RPM, punishment is designed explicitly for defectors since cooperators will break the links only with defectors, not with cooperators. However, the reward is not only for cooperative behavior because the social reward may link individuals with defectors. In other words, individuals never sever ties with their cooperators, but they sometimes establish new connections with defectors, which is reminiscent of forgiveness or leniency that is believed to promote cooperation.[67,68] This raises the question of what will happen if we directly specify the establishment of a new connection with a randomly chosen cooperator when rewarding. We conduct a set of control experiments (see Sec. H in the supplementary material) to prove that establishing contact with defectors is a tolerant behavior that supports cooperation.

In addition, we find that the RPM will cause the network to spontaneously generate heterogeneity in node degrees and teaching activities despite all the mechanisms being homogeneous and unbiased when we set up the model. To some extent, this finding explains why networks with heterogeneous distributions are more common in the real world from the perspective of evolutionary game theory. We believe that this framework can be further generalized to account for other forms of heterogeneity related to individual behaviors.

To prove that heterogeneity occurs spontaneously, we do not introduce any heterogeneity in our experimental settings. In the main text, we consider the homogeneous ER underlying network and make new links randomly (RA). Therefore, we conduct two sets of control experiments reported in Sec. I of the supplementary material, in which we consider a heterogeneous underlying structure: the BA scale-free network and a method for making new contacts, "the rich get richer," preferential attachment (PA).

The two sets of control experiments show that the experimental results in ER/RA also appear in BA and PA. Thus, we can draw certain conclusions. First, neither the underlying network structure nor the method for making new links is the fundamental cause of the RPM promoting cooperation. The influence of these factors can be adjusted through the intensity of rewards and punishments. Second, BA and PA have similar effects for the RPM, both of which promote cooperation, so that their punishments required for optimal cooperation are less. However, the two are not exactly the same. PA, because it acts only on the RM, has no effect on the pure PM or non-RPM. However, the RM, the PM, and traditional PD games are all affected by BA.

In this paper, our social reward–punishment process is simplistic, and we have considered only a classical behavioral rule in which players can either cooperate or defect against all of their partners. Thus, many obvious extensions are easily conceivable. For example, we can take into account that the tolerance varies from person to person or players have different behaviors to different partners. The coevolution of interaction and behavioral dynamics has qualitatively different consequences for collective outcomes and is worth further development.

SUPPLEMENTARY MATERIAL

See the supplementary material for complete studies on the combination mechanism of social reward and punishment. In addition to providing experimental supplements with different parameters and theoretical analysis in Secs. A–E, we also provide some control experiments, such as different initial average number of connections ($\langle Deg \rangle = 4, 8$ in Sec. F), real-world social networks (Sec. G), a discriminatory reward mechanism (RMWC in Sec. H), heterogeneous BA scale-free network (Sec. I), and preferential attachment in making a new link (PA in Sec. I).

ACKNOWLEDGMENTS

M. Zhang, X. Zhang, C. Qu, and G. Wang were supported in part by the National Natural Science Foundation of China under Grant Nos. 12001324, 11631014, and 11871311, in part by the China Postdoctoral Science Foundation under Grant Nos. 2019TQ0188 and 2019M662315, and in part by the Shandong University multidisciplinary research and innovation team of young scholars under Grant No. 2020QNQT017. The funder had no role in study design, data collection and analysis, and decision to publish or preparation of the manuscript.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Ming Zhang: Conceptualization (equal); Formal analysis (lead); Methodology (equal); Software (equal); Visualization (equal); Writing – original draft (lead). **Xu Zhang:** Data curation (lead); Software (equal); Visualization (equal); Writing – original draft (supporting). **Cunquan Qu:** Conceptualization (equal); Funding acquisition (equal); Methodology (equal); Supervision (equal); Visualization (equal); Writing – original draft (supporting). **Guanghui Wang:** Funding acquisition (equal); Supervision (equal); Writing – original draft (supporting). **Xin Lu:** Supervision (supporting); Writing – original draft (supporting).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] R. Axelrod and W. D. Hamilton, "The evolution of cooperation," Science **211**, 1390–1396 (1981).
[2] J. Henrich, J. Ensminger, R. McElreath, A. Barr, C. Barrett, A. Bolyanatz, J. C. Cardenas, M. Gurven, E. Gwako, N. Henrich *et al.*, "Markets, religion, community size, and the evolution of fairness and punishment," Science **327**, 1480–1484 (2010).
[3] E. Ostrom, J. Burger, C. B. Field, R. B. Norgaard, and D. Policansky, "Revisiting the commons: Local lessons, global challenges," Science **284**, 278–282 (1999).
[4] E. Fehr and U. Fischbacher, "The nature of human altruism," Nature **425**, 785–791 (2003).
[5] M. A. Nowak, "Five rules for the evolution of cooperation," Science **314**, 1560–1563 (2006).
[6] J. M. Smith and G. R. Price, "The logic of animal conflict," Nature **246**, 15–18 (1973).
[7] J. M. Smith, *Evolution and the Theory of Games* (Cambridge University Press, 1982).
[8] H. Gintis, *Game Theory Evolving* (Princeton University Press, 2009).
[9] M. A. Nowak, *Evolutionary Dynamics: Exploring the Equations of Life* (Harvard University Press, 2006).
[10] D. G. Rand, A. Dreber, T. Ellingsen, D. Fudenberg, and M. A. Nowak, "Positive interactions promote public cooperation," Science **325**, 1272–1275 (2009).
[11] D. G. Rand, S. Arbesman, and N. A. Christakis, "Dynamic social networks promote cooperation in experiments with humans," Proc. Natl. Acad. Sci. U.S.A. **108**, 19193–19198 (2011).

[12]D. Baldassarri and G. Grossman, "Centralized sanctioning and legitimate authority promote cooperation in humans," Proc. Natl. Acad. Sci. U.S.A. **108**, 11023–11027 (2011).
[13]M. Feinberg, R. Willer, and M. Schultz, "Gossip and ostracism promote cooperation in groups," Psychol. Sci. **25**, 656–664 (2014).
[14]D. Melamed and B. Simpson, "Strong ties promote the evolution of cooperation in dynamic networks," Soc. Netw. **45**, 32–44 (2016).
[15]G. Szabó and G. Fáth, "Evolutionary games on graphs," Phys. Rep. **446**(4–6), 97–216 (2007).
[16]F. C. Santos and J. M. Pacheco, "Scale-free networks provide a unifying framework for the emergence of cooperation," Phys. Rev. Lett. **95**, 098104 (2005).
[17]M. A. Nowak, S. Bonhoeffer, and R. M. May, "More spatial games," Int. J. Bifurc. Chaos **4**, 33–56 (1994).
[18]M. H. Vainstein and J. J. Arenzon, "Disordered environments in spatial games," Phys. Rev. E **64**, 051905 (2001).
[19]J. Vukov and G. Szabó, "Evolutionary prisoner's dilemma game on hierarchical lattices," Phys. Rev. E **71**, 036133 (2005).
[20]G. Abramson and M. Kuperman, "Social games in a social network," Phys. Rev. E **63**, 030901 (2001).
[21]M. Kuperman and S. Risau-Gusman, "The effect of the topology on the spatial ultimatum game," Eur. Phys. J. B **62**, 233–238 (2008).
[22]P. Holme, A. Trusina, B. J. Kim, and P. Minnhagen, "Prisoners' dilemma in real-world acquaintance networks: Spikes and quasiequilibria induced by the interplay between structure and dynamics," Phys. Rev. E **68**, 030901 (2003).
[23]J. Hofbauer and K. Sigmund, *Evolutionary Games and Population Dynamics* (Cambridge University Press, 1998).
[24]M. A. Nowak and R. M. May, "Evolutionary games and spatial chaos," Nature **359**, 826–829 (1992).
[25]M. Perc and A. Szolnoki, "Social diversity and promotion of cooperation in the spatial prisoner's dilemma game," Phys. Rev. E **77**, 011904 (2008).
[26]F. C. Santos, M. D. Santos, and J. M. Pacheco, "Social diversity promotes the emergence of cooperation in public goods games," Nature **454**, 213–216 (2008).
[27]S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications* (Cambridge University Press, 1994).
[28]M. Perc and A. Szolnoki, "Coevolutionary games—A mini review," BioSystems **99**, 109–125 (2010).
[29]F. C. Santos, J. M. Pacheco, and T. Lenaerts, "Cooperation prevails when individuals adjust their social ties," PLoS Comput. Biol. **2**, e140 (2006).
[30]Z. Wang, A. Szolnoki, and M. Perc, "Rewarding evolutionary fitness with links between populations promotes cooperation," J. Theor. Biol. **349**, 50–56 (2014).
[31]M. G. Zimmermann, V. M. Eguíluz, and M. San Miguel, "Coevolution of dynamical states and interactions in dynamic networks," Phys. Rev. E **69**, 065102 (2004).
[32]N. Hanaki, A. Peterhansl, P. S. Dodds, and D. J. Watts, "Cooperation in evolving social networks," Manag. Sci. **53**, 1036–1050 (2007).
[33]J. S. Hurlbert and A. C. Acock, "The effects of marital status on the form and composition of social networks," Soc. Sci. Q. **71**(1), 163 (1990), available at https://www.proquest.com/scholarly-journals/effects-marital-status-on-form-composition-social/docview/1291642904/se-2?accountid=13813.
[34]C. Licoppe and Z. Smoreda, "Are social networks technologically embedded?: How networks are changing today with changes in communication technology," Soc. Netw. **27**, 317–335 (2005).
[35]H. Knox, M. Savage, and P. Harvey, "Social networks and the study of relations: Networks as method, metaphor and form," Econ. Soc. **35**, 113–140 (2006).
[36]A. Szolnoki and M. Perc, "Coevolution of teaching activity promotes cooperation," New J. Phys. **10**, 043036 (2008).
[37]F. Fu, C. Hauert, M. A. Nowak, and L. Wang, "Reputation-based partner choice promotes cooperation in social networks," Phys. Rev. E **78**, 026117 (2008).
[38]D. Balliet, L. B. Mulder, and P. A. Van Lange, "Reward, punishment, and cooperation: A meta-analysis," Psychol. Bull. **137**, 594 (2011).
[39]D. Nosenzo and M. Sefton, "Promoting cooperation: The distribution of reward and punishment power," Technical Report, CeDEx Discussion Paper Series, 2012.
[40]Z. Rong, Z.-X. Wu, X. Li, P. Holme, and G. Chen, "Heterogeneous cooperative leadership structure emerging from random regular graphs," Chaos **29**, 103103 (2019).
[41]A. Szolnoki and X. Chen, "Strategy dependent learning activity in cyclic dominant systems," Chaos, Solitons Fractals **138**, 109935 (2020).
[42]A. Szolnoki and X. Chen, "Blocking defector invasion by focusing on the most successful partner," Appl. Math. Comput. **385**, 125430 (2020).
[43]J. He, J. Wang, F. Yu, and L. Zheng, "Reputation-based strategy persistence promotes cooperation in spatial social dilemma," Phys. Lett. A **384**, 126703 (2020).
[44]Q. Wang, X. Ren, B. Gao, and J. Wang, "Heterogeneity reproductive ability promotes cooperation in spatial prisoner's dilemma game," Chaos, Solitons Fractals **134**, 109715 (2020).
[45]D. G. Rand and M. A. Nowak, "Human cooperation," Trends Cognit. Sci. **17**, 413–425 (2013).
[46]Y. Wu, Z. Zhang, and S. Chang, "Reciprocal reward promotes the evolution of cooperation in structured populations," Chaos, Solitons Fractals **119**, 230–236 (2019).
[47]H. S. Lee and D. A. Griffith, "Reward strategy spillover effects on observer cooperation in business networks," Mark. Lett. **32**, 47–59 (2021).
[48]J. H. Fowler, "Altruistic punishment and the origin of cooperation," Proc. Natl. Acad. Sci. U.S.A. **102**, 7047–7049 (2005).
[49]N. J. Raihani, A. Thornton, and R. Bshary, "Punishment and cooperation in nature," Trends Ecol. Evol. **27**, 288–295 (2012).
[50]X. Li, M. Jusup, Z. Wang, H. Li, L. Shi, B. Podobnik, H. E. Stanley, S. Havlin, and S. Boccaletti, "Punishment diminishes the benefits of network reciprocity in social dilemma experiments," Proc. Natl. Acad. Sci. U.S.A. **115**, 30–35 (2018).
[51]J. García and A. Traulsen, "Evolution of coordinated punishment to enforce cooperation from an unbiased strategy space," J. R. Soc. Interface **16**(156), 20190127 (2019).
[52]C.-L. Yang, B. Zhang, G. Charness, C. Li, and J. W. Lien, "Endogenous rewards promote cooperation," Proc. Natl. Acad. Sci. U.S.A. **115**, 9968–9973 (2018).
[53]L. Balafoutas, N. Nikiforakis, and B. Rockenbach, "Direct and indirect punishment among strangers in the field," Proc. Natl. Acad. Sci. U.S.A. **111**, 15924–15927 (2014).
[54]J. Vukov, F. L. Pinheiro, F. C. Santos, and J. M. Pacheco, "Reward from punishment does not emerge at all costs," PLoS Comput. Biol. **9**, e1002868 (2013).
[55]J.-J. Wu, B.-Y. Zhang, Z.-X. Zhou, Q.-Q. He, X.-D. Zheng, R. Cressman, and Y. Tao, "Costly punishment does not always increase cooperation," Proc. Natl. Acad. Sci. U.S.A. **106**, 17448–17451 (2009).
[56]A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," Science **286**, 509–512 (1999).
[57]Z. Toroczkai, B. Kozma, K. E. Bassler, N. Hengartner, and G. Korniss, "Gradient networks," J. Phys. A: Math. Theor. **41**, 155103 (2008).
[58]D.-H. Kim and H. Jeong, "Inhomogeneous substructures hidden in random networks," Phys. Rev. E **73**, 037102 (2006).
[59]S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, "Complex networks: Structure and dynamics," Phys. Rep. **424**, 175–308 (2006).
[60]M. Nagy, Z. Ákos, D. Biro, and T. Vicsek, "Hierarchical group dynamics in pigeon flocks," Nature **464**, 890–893 (2010).
[61]P. Erdős and A. Rényi, "On random graphs I," Publ. Math. Debr. **6**, 290–297 (1959).
[62]G. Szabó and C. Tőke, "Evolutionary prisoner's dilemma game on a square lattice," Phys. Rev. E **58**, 69 (1998).
[63]J. Leskovec and A. Krevl, "SNAP Datasets: Stanford large network dataset collection," (2014); see http://snap.stanford.edu/data.
[64]F. C. Santos, J. M. Pacheco, and T. Lenaerts, "Evolutionary dynamics of social dilemmas in structured heterogeneous populations," Proc. Natl. Acad. Sci. U.S.A. **103**, 3490–3494 (2006).
[65]F. C. Santos and J. M. Pacheco, "A new route to the evolution of cooperation," J. Evol. Biol. **19**, 726–733 (2006).
[66]J. Gómez-Gardenes, M. Campillo, L. M. Floría, and Y. Moreno, "Dynamical organization of cooperation in complex topologies," Phys. Rev. Lett. **98**(10), 108103 (2007).
[67]M. A. Nowak and K. Sigmund, "Tit for tat in heterogeneous populations," Nature **355**, 250–253 (1992).
[68]D. Fudenberg, D. G. Rand, and A. Dreber, "Slow to anger and fast to forgive: Cooperation in an uncertain world," Am. Econ. Rev. **102**, 720–49 (2012).
[69]R. Boyd and P. J. Richerson, "The evolution of reciprocity in sizable groups," J. Theor. Biol. **132**, 337–356 (1988).