Influencing the Long-Term Evolution of Online Communities using Social Norms

Yu Zhang
Electrical Engineering
University of California, Los Angeles
Los Angeles, United States
yuzhang@ucla.edu

Mihaela van der Schaar
Electrical Engineering
University of California, Los Angeles
Los Angeles, United States
mihaela@ee.ucla.edu

***Abstract*— This paper focuses on analyzing the interactions emerging between users in online communities. Network utility maximization and other methods can be used to achieve efficient designs when the communities are composed of compliant users. However, such methods are not effective and efficient when the communities are composed of intelligent and self-interested users (multimedia social communities, social networks etc.), because the interests of the individual users may be in conflict. In our prior work, we designed social reciprocation protocols by assuming a stationary community in which a continuum population interacts. We proved that given these assumptions, users have incentives to voluntarily operate according to pre-determined social norms and provide services. In this paper, we extend this study to analyze the interactions of self-interested users under a social norm in an online community of finite population and without making stationary assumptions about the community. To optimize their long-term performance while operating in the community, users adapt strategies to play their best response based on their knowledge by solving individual stochastic control problems. The best-response dynamic introduces a stochastic dynamic process in the community, in which the strategies of users evolve over time. Understanding how a community responds to incentives in the long-term provides protocol designers with guidelines for designing social norms in which no user will find it into its self-interest to adapt and deviate from the prescribed protocol. This will, in turn, influence the evolution of the community and induce the emergence of cooperative behavior among users, thereby maximizing the optimal social welfare of the community.**

Keywords- Learning in Online Communities, Social Norms, Markov Decision Process, Stochastically Stable Equilibrium.

I. INTRODUCTION

The proliferation of social networking services has permeated our social and economic lives and created online social communities where individuals interact with each other. However, online communities in general rely on the voluntary contribution of services by individual users and are, therefore, vulnerable to intrinsic incentive problems which lead to prevalent free-riding behaviors among users, at the expense of the collective social welfare of the community [1][3].

Various incentive mechanisms have been proposed to encourage cooperation in online communities [2][14][16], with a large body of them relying on the idea of *reciprocity*, in which users are rated and differentially served based on their past behaviors. Reciprocity-based mechanisms can be further classified into *direct reciprocity* and *indirect reciprocity* depending on how the rating score is generated. Existing direct reciprocity protocols [2][3] do not scale well to online communities with large populations of anonymous users, since frequent interactions between two users are required in order to build up accurate mutual ratings in this bilateral reciprocation paradigm. In social reciprocation schemes, individuals obtain some information about other individuals (for example, their ratings) and decide their behaviors toward this individual based on their information about that individual. Hence, an individual can be rewarded or punished by other individuals in the community who have not had direct interactions with it. Since social reciprocation requires neither observable identities nor frequent interactions, it has a potential to form a basis of successful incentive schemes for online communities. As such, we have developed a general framework on social norms in [12] to study incentive schemes based on social reciprocation. Social norms are defined as the rules that are deployed in a group of users to regulate user behaviors. In incentive schemes based on social norms, a label containing information about its past behavior is assigned to each user indicating its reputation, status, etc. Users with different labels are treated differently by other users with which they interact. Hence, a social norm can be easily adopted in various online communities as long as an infrastructure exists for collecting, processing, and delivering information about the users' behaviors, such as a tracker or a portal.

In the analysis of [12], the state of a community is represented by the reputation distribution of users in it. It focuses on the fully-compliant state of a community in which all users comply with the prescribed social norm voluntarily and determines how to design the social norm to prevent users' deviations from this state. When there is a continuum population in the community, the fully-compliant state is stable since the changes in individual users' reputations average out at the community level and the law of large numbers can be applied [18]. In a real community of a finite population where the law of large numbers does not hold, the reputation distribution of users will be disturbed by stochastic permutations and varies over time. Hence, the fully-compliant state is not necessarily stable. Such variations affect the users' incentives to comply with the social norm, and these users may find it in their self-interest to adapt their strategies and deviate from the social norm as pointed in [3]. Therefore, the analysis of [12] no longer applies to communities of finite populations. We extend in this work the study in [12] to analyze the interactions of self-interested users under a social norm in an online community with a finite population. It is of critical importance to understand the users' learning and adaptation behavior and how it influences the long-term evolution of users' strategies, which can provide essential insights to facilitate the design of incentive mechanisms that can improve the efficiency and survivability of online communities.

The evolution of users' strategies can be modeled as a multi-agent learning problem (MAL). There are two prevailing agendas existing in the current MAL works, namely descriptive learning and prescriptive learning respectively.

The descriptive agenda focuses on investigating formal models of learning that agree with people's natural behavior in the context of other learners and analyzes properties of the game, i.e. equilibrium results and convergence of users' learning and

adaptation process [6][20]. This agenda fails to answer the question on how the protocol designers should design protocols in order to construct a multi-agent learning system to achieve some desirable properties, e.g. the highest social efficiency.

The prescriptive agenda exploits how agents should learn and how the protocol designers should construct systems that induct agents to learn certain strategies and influence the evolution of the community to achieve certain goals [10][13][19]. However, the current research in this agenda mostly focuses on the repeated or stochastic "common-payoff" (or "team") games where the interests of the protocol designers and the interests of the learning agents are aligned.

We focus in this work on the prescriptive learning in order to design protocols to induct users' learning behavior. Different from [10][13][19], the interests of the protocol designer in our work, i.e. the social welfare of the community, and the self-interest of an individual user is not aligned but rather in conflict with each other. Instead of restricting users' learning and adaptation rules as in [7][8], we allow users to form their own beliefs depending on their observations from the community, and adapt their service strategies using best responses to maximize their own long-term utility. Given the users' best response dynamics, we design effective protocols based on social norms which induce users learn to cooperate with each other and contribute services voluntarily in the long term. Our design also studies how the social norms need to be adapted to different community characteristics, including the service cost and benefit, the population size of the community, and the users' discount factor for the future utility. In the final part, we extend our work to consider additional issues such as the long-term behavior of the community when users are heterogeneous and have different beliefs.

The remainder of this paper is organized as follows. In Section II, we introduce our proposed social norm based framework for indirect reciprocity in online communities and propose the design problem for the protocol designers. In Section III, we study users' learning behavior and the protocol design to influence the long-term evolution of the community. Section IV presents our experimental results and the conclusions are drawn in Section V where directions for future research are also outlined.

II. System Model and Problem Formulation

A. Repeated game formulation

We consider an online community consisting of N users. Each user can offer some valuable services to others, such as information, knowledge, data files, computation resources, storage space, etc. The community is modeled as a discrete-time system with time divided into periods. In each period, each user selects an idle user who is not serving others to request a service [3]. The selection is uniformly random such that all users in the community have an equal probability to be chosen by a particular user. Therefore, each user also receives one service request from other users per period [1].

We model the interaction between a pair of matched users as a one-period asymmetric gift-giving game to characterize the asymmetry of interests among users [4]. The user who requests services is called a *client* and the user whose services are requested is called a *server*. Upon receiving the request, the server selects its level of contribution $z \in \mathcal{Z} = \{0,1\}$ to the client, where $z = 1$ indicates that the server provides the requested service to the client and $z = 0$ indicates that the server refuses to provide the service. When the server provides a service by choosing $z = 1$, it incurs a cost of c, and the client gains a benefit of b by receiving the requested service; whereas both users receive a utility of 0 when the server chooses $z = 0$. The utility matrix of a one-period gift-giving game is presented in Table 1.

Table1. Utility matrix of a gift-giving game.

		Server	
		$z = 1$	$z = 0$
Client		b , $-c$	0 , 0

The social welfare of the community is quantified by the average one-period utility received by all users in the community, denoted as U. It is maximized at $U = b - c$ if all users are cooperative and choose $z = 1$ when being requested. On the contrary, $z = 0$ is always chosen by a self-interested server if it expects to maximize its one-period utility myopically. As a result, the social welfare is minimized as $U = 0$ at the Nash equilibrium of the one-period game with no user providing services to others. To improve the inefficiency of the myopic Nash equilibrium, we design a new set of incentive protocols based on social norms to exploit the repeated nature of the users' interactions and provide incentives for cooperation using the threat of future punishments. A social norm proposes a set of rules to regulate user behaviors. Compliance to these rules is positively rewarded (i.e. an increased level of service is provided to such users) and failure to comply with these rules can result in (severe) punishments (i.e. a decreased level of service is provided).

The regulation of a social norm takes effect through its manipulation over the users' social status, as in [9]. In the repeated game, each user is tagged with a reputation $\theta \in \boldsymbol{\Theta} = \{0,1,2,\cdots,L\}$ representing its social status, where L is the reputation length and also the highest reputation that can be obtained by a user. The behavioral strategy that a user adopts in the repeated game is reputation-based and represented as $\sigma = \{\sigma(\tilde{\theta})\}_{\tilde{\theta}=0}^{L}$, where each term $\sigma(\tilde{\theta}) \in \mathcal{Z}$ is the contribution level of this user when it is matched with a client of reputation $\tilde{\theta}$. The set of strategies that can be chosen by a user is finite and denoted as Γ. Different from [12] where we focused on determining the complete set of feasible strategies independent of their characteristics or implementation simplicity, in this work we explicitly consider the implementation and designing requirements and complexity of a strategy. In particular, each strategy $\sigma \in \Gamma$ in this work can be characterized by a service threshold $h_\sigma \in \{0,1,\cdots,L+1\}$. By adopting σ, a user provides services only to clients whose reputations are no less than h_σ [2]. Formally, a threshold-based strategy can be represented as follows:

$$\sigma(\tilde{\theta}) = \begin{cases} 1 & if\ \tilde{\theta} \geq h_\sigma \\ 0 & if\ \tilde{\theta} < h_\sigma \end{cases}. \qquad (1)$$

As particular cases, the fully cooperative strategy with $h_\sigma = 0$ that provides services to all users unconditionally is denoted as σ_C, whereas the fully defective strategy with $h_\sigma = L+1$ that provides no service to any of the users is denoted as σ_D.

We consider a social norm $\kappa = (\phi, \tau)$ that consists of a social rule and a reputation scheme, as shown in Figure 1. A social rule $\phi = \{\sigma_\theta \,|\theta \in \{0,\ldots,L\}\}$ is a set of behavioral strategies that specifies the approved behaviors of users within the community. For a user of reputation θ, its reputation is increased when it

[1] The analysis can be extended to a general case when the service request rate per period is a positive real number.

[2] The existing social norm based protocols such as [5] are in fact special cases of the threshold-based strategy proposed here.

follows σ_θ and is decreased otherwise. Alternatively, the social rule can be represented as a mapping $\phi : \Theta \times \Theta \to \mathcal{Z}$, with $\phi(\theta, \tilde{\theta}) = \sigma_\theta(\tilde{\theta})$ for all θ and $\tilde{\theta}$. For illustration, we consider social rules which satisfy the following property in the design of protocols:

$$\phi(\theta,\tilde{\theta}) = \begin{cases} 1 & if\ \theta \geq h\ and\ \tilde{\theta} \geq h \\ 1 & if\ \theta < h \\ 0 & otherwise \end{cases}. \qquad (2)$$

Hence a social rule instructs a user of reputation $\theta < h$ to provide services to all users in the community, i.e. $h_{\sigma_\theta} = 0$ if $\theta < h$; while a user of reputation $\theta \geq h$ only have to provide service to users whose reputation is also above h, i.e. $h_{\sigma_\theta} = h$ if $\theta \geq h$. For the convenience of illustration, we refer to h as the *social threshold*, which differentiates the services that users of different reputations receive and provide. Users of reputations lower than the social threshold are regarded as "bad users", and users of reputations no less than the social threshold are regarded as "good users". It should be noted that when $h = 0$ or $h = L+1$, $\sigma_\theta = \sigma_C$ for all θ, which cannot provide differential services to users and thus cannot be enforced among self-interested users, as shown in [12]. Therefore, we restrict our attention in the design on social rules whose social threshold $h \in \{1, \cdots, L\}$ without loss of generality.

A reputation scheme τ updates a user's reputation based on its past behavior as a server. After a server plays, its client reports the contribution level to some trustworthy third-party managing device in the community (e.g. the tracker in P2P networks), and the managing device updates the server's reputation according to τ and the client's report. Formally, a reputation scheme τ updates a server's reputation based on the reputations of the matched users and the reported contribution level of the server. It is represented as a mapping $\tau : \Theta \times \Theta \times \mathcal{Z} \to \Theta$, in which $\tau(\theta, \tilde{\theta}, z)$ is the reputation of the server in the next period given its current reputation θ, the client's reputation $\tilde{\theta}$, and the server's reported contribution level $z \in \mathcal{Z}$. As an example, we consider the following simple reputation scheme in this paper:

$$\tau(\theta,\tilde{\theta},z) = \begin{cases} \min\{L, \theta+1\} & if\ z = \phi(\theta,\tilde{\theta}) \\ 0 & if\ z \neq \phi(\theta,\tilde{\theta}) \end{cases}. \qquad (3)$$

In this scheme, the server's reputation is increased by 1 while not exceeding L if the reported contribution level is the same as that specified by the social rule ϕ. Otherwise, the server's reputation is set to 0 any time it deviates from the social rule. In practice, a system is continually being subjected to small stochastic perturbations that arise due to various types of operation errors in the community. To formalize the effect of such perturbations on the evolution of the community, we assume that the client's report is subject to a small error probability ε of being reversed. That is, $z = 0$ is reported to the managing device with probability ε while the server actually plays $z = 1$, and vice versa. It should be noted that the value of ε is known to all users in the community.

B. Utility function and the designer problem

In general, a user's expected utility in one period depends on its own strategy as well as other users' reputations and strategies in the community. In the beginning of each period, the tracker broadcasts the reputation distribution of the community, which is also called as a *community configuration*, to all users. The community configuration is denoted as $\mu = \{n(\theta)\}_{\theta=0}^{L}$, in which $n(\theta)$ represents the number of users of reputation θ in the community at this time. Due to their limited processing capabilities and interactions with others, users can only form simple beliefs about the strategies deployed by other users [15]. We assume that each user maintains a belief that a user of reputation higher than 0 will comply with the social rule ϕ in the current period with the probability $1-\varepsilon$, and play σ_D with the probability ε. In contrast, a user of reputation 0 will play σ_D with the probability $1-\varepsilon$ and comply with ϕ with the probability ε since it deviated in the previous period and has been punished.

Since a user can never be matched with itself, it is more convenient to compute a user's expected one period utility by employing the reputation distribution of all users except itself, which is called as the *opponent configuration*. The opponent configuration of a user of reputation θ is denoted as $\eta_\theta = \{m_\theta(\theta')\}_{\theta'=0}^{L}$, which preserves the relationship with the community configuration as $m_\theta(\theta') = n(\theta')$ for all $\theta' \neq \theta$ and $m_\theta(\theta) = n(\theta) - 1$. Let $v_\kappa(\sigma, \theta, \mu)$ denote the expected one-period utility for a user of reputation θ playing the strategy σ when the community configuration is μ, we can compute it as:

$$\begin{aligned} & v_\kappa(\sigma,\theta,\mu) = \\ & \frac{1-\varepsilon}{N-1}\sum_{\theta'\neq 0} m_\theta(\theta') b(\theta',\theta,\phi\ \ \theta',\theta\ ) + \frac{\varepsilon}{N-1} m_\theta(0) b(0,\theta,\phi(0,\theta)) \\ & + \frac{\varepsilon}{N-1}\sum_{\theta'\neq 0} m_\theta(\theta') b(0,\theta,\sigma_D(\theta)) + \frac{1-\varepsilon}{N-1} m_\theta(0) b(0,\theta,\sigma_D(\theta)) \\ & - \frac{1}{N-1}\sum_{\theta'} m_\theta(\theta') c(\theta,\theta',\sigma(\theta')) \end{aligned} \qquad .(4)$$

Here $b(\theta', \theta, \phi(\theta', \theta))$ is the one-period benefit which this user can receive when its matched server has a reputation θ' and complies with ϕ; $c(\theta, \theta', \sigma(\theta'))$ is the one-period cost of this user when its matched client has a reputation θ'. The expected utility of a user complying with ϕ is compactly denoted as $v_\kappa(\theta, \mu) = v_\kappa(\sigma_\theta, \theta, \mu)$. A user's long-term utility in the repeated game is evaluated with the infinite-horizon discounted sum criterion. Starting from any period t_0, the user's expected long-term utility is expressed as

$$v_\kappa^\infty(\sigma^{(t_0)}, \theta^{(t_0)}, \mu^{(t_0)}) = E\{\sum_{t=t_0}^{\infty} \delta^{t-t_0} v_\kappa(\sigma^{(t)}, \theta^{(t)}, \mu^{(t)})\}, \quad (5)$$

where $\delta \in [0,1)$ is the discount factor which represents a user's preference of the future utility.

Under a social norm κ, adaptive self-interested users play the best response in order to optimizes their expected long-term utility. The best response σ^* for a user in period t_0 is a strategy which satisfies

$$\sigma^* = \arg\max_\sigma v_\kappa^\infty(\sigma, \theta^{(t_0)}, \mu^{(t_0)}). \qquad (6)$$

The best-response dynamic of users thus introduces a dynamic process in the community which will be analyzed in the next section using a Markov chain analysis. Users' strategies in this process evolve over time [11]. We are interested in whether this

process may converge to an equilibrium in the long term, i.e. each user holds a fixed reputation and plays a fixed strategy after a sufficiently long period of time, and if an equilibrium exists, how the protocol designer can design the social norm in order to enforce all users to play cooperation in their best responses. This provides the protocol designer with guidelines for selecting the correct social norm based on the community characteristics, e.g. the utility structure (b,c) and the discount factor δ, in order to optimize the sharing efficiency in the community. To formalize the effect of stochastic permutations, we consider operation errors with the probability of occurrence approaches to 0, i.e. $\varepsilon \to 0$.The resulting equilibrium is defined as a *stochastically stable equilibrium* in [11]. Indexing all users in the community by $i \in \{1,\ldots,N\}$ and denote the reputation profile and the strategy profile of all users as $\boldsymbol{\theta} = (\theta^1,\ldots,\theta^i,\ldots,\theta^N)$ and $\boldsymbol{\sigma} = (\sigma^1,\ldots,\sigma^i,\ldots,\sigma^N)$, we formally define a stochastically stable equilibrium and the protocol designer's design problem as follows.

Definition 1 [11]**.** A strategy profile $\boldsymbol{\sigma}$ together with a community configuration μ is a stochastically stable equilibrium if and only if when $\varepsilon \to 0$

(1) $\boldsymbol{\sigma}$ is the best response of users against μ, i.e.

$$\sigma^i = \arg\max_{\sigma} v_{\kappa}^{\infty}(\sigma^i,\theta^i,\mu),\ \forall i \in \{1,\ldots,N\}; \qquad (7)$$

(2) μ is time invariant under the best response dynamics introduced by $\boldsymbol{\sigma}$;

(3) The community stays at μ with a positive fraction of time in the long term.

The protocol designer tries to optimize the social welfare U whose maximum value is achieved when all users choose to cooperate with their clients and provide services. As we will show in the next section, the community converges to configurations belonging to the set of stochastically stable equilibria with probability 1 in the long term. Hence, the protocol designer should find a social norm under which users are mutually cooperative in all stochastically stable equilibria.

III. Users' Strategic Learning and The Community's Long Term Evolution

A. The MDP formulation and the structure of the optimal policy

In this section, we analyze the evolution of the community and the design of protocols to induce cooperation among users. We first model a user's learning and adaptation as a MDP, and characterize the structure of a user's optimal policy of strategy adaptation. With the structure, we prove the properties of stochastically stable equilibria in Theorem 1, which helps us to design the optimal protocol in Theorem 2 in order to induce users to play cooperatively in the long term.

At the beginning of each period, we assume that each user plays its service strategy used in the previous period with probability $1-\gamma$, and adapts to play its best response which satisfies (7) with probability γ. The probability $\gamma \in (0,1)$ is called the adaptation rate of a user. Formally, the adaptation of a user could be represented as an optimization problem over all possible selections of its sharing strategy Γ in order to maximize its expected long-term utility $v_{\kappa}^{\infty}(\sigma,\theta,\mu)$, given its current knowledge and beliefs on the community. This is summarized below for readers' convenience.

Knowledge: At the beginning of each period, a user learns its current reputation θ as well as the current community configuration μ accurately from the managing device, e.g. the tracker. Hence, a user also learns its current opponent configuration η_θ.

Belief: Each user maintains a belief about the strategies of other users as described in Section II.B. That is, a user whose reputation is higher than 0 will comply with the social rule ϕ with the probability $1-\varepsilon$ in the current period, while a user whose reputation is 0 adopts the service strategy σ_D with the probability $1-\varepsilon$.

Given the observations and beliefs, a user's optimization problem can be formulated as a Markov Decision Process (MDP) [17] as below.

State: The experienced dynamic of a user in each period is its current reputation θ and the opponent configuration η_θ, which is defined as its state $s = (\theta,\eta_\theta) \in \mathcal{S}$.

Action: The action of a user is represented by the threshold of its serving strategy $a \in \mathcal{A} = \{0,1,\cdots,L+1\}$, which is determined at the beginning of a period based on its state in this period.

Transition probability: The transition of s across periods is Markovian with the transition probability $p(s' \mid s,a)$, which is determined by the action a.

Reward function: The one-period reward function $r(s,a)$ is defined as the expected one-period utility as in (4). Similarly, the long-term reward function $R(s)$ of a user is defined as the following discounted sum

$$R(s) = \sum_{t=0}^{\infty} \delta^t r(s^{(t)},a^{(t)}). \qquad (8)$$

Policy and Value function: The solution of the MDP is a policy $\pi : \mathcal{S} \to \{0,1,\cdots,L+1\}$, which maps each state to a service threshold. The value function is thus defined as the expected long-term utility $v_{\kappa}^{\infty}(\sigma,\theta,\mu)$, under a policy π, which is formally represented as

$$\begin{aligned} V^{\pi}(s) &= E\{R(s)\big|\pi\} = E\{\sum_{t=0}^{\infty} \delta^t r(s^{(t)},a^{(t)})\Big| s^{(0)} = s,\pi\} \\ &= E\{r(s,a)\big|\pi\} + \delta \sum_{s'} p(s' \mid s,\pi) V^{\pi}(s') \end{aligned} . \qquad (9)$$

The above MDP can be solved using common computation methods such as value iteration and Q-learning [17], with the resulting optimal policy and value function being $\{\pi^*(s)\}$ and $\{V^*(s)\}$. In any period, the optimal policy instructs a user of reputation θ who faces an opponent configuration η_θ to play $\pi^*(\theta,\eta_\theta)$, which is the best response of this user that satisfies (6). In the rest of this section, we characterize the structure of $\{\pi^*(s)\}$. First, it can be shown that the best response of a user in any period is always above the corresponding service threshold of the social rule, regardless of the user's reputation θ and the opponent configuration η. That is, $\pi^*(\theta,\eta) \geq h_{\sigma_\theta}$ for any θ and η. For the bad users, this conclusion is obvious. Suppose there is a good user with $\theta \geq h$ and $\pi^*(\theta,\eta) < h$, the best response of this user is to provide more services to the community than what is required by the social rule. This leads to a higher expected service cost for this user in the current period. However, this user also gets a higher probability to be punished by the social norm for being deviated

from ϕ [3], and hence a lower expected future utility compared to what it could receive by complying with ϕ. Therefore, it can be concluded that this user will receive a strictly lower expected long-term utility by playing $\pi^*(\theta,\eta)$ than what it could receive by complying with ϕ, which contradicts the fact that $\pi^*(\theta,\eta)$ is the best response which satisfies (6).

Lemma 1. $\pi^*(\theta,\eta) \geq h_{\sigma_\theta}$, for any $\theta \in \boldsymbol{\Theta}$ and η. ■

The next lemma further shows that confronted with a same opponent configuration, the service threshold of a bad user's best response (weakly) decreases against its reputation, and the same conclusion holds for a good user. We first provide some intuitive explanation. From (4), it can be determined that the expected amount of service that has to be provided by a bad user in one period only depends on its current opponent configuration and its selection of the sharing strategy, but it is independent on its current reputation. However, the average amount of service that a bad user expects to receive (weakly) increases with its reputation. Using similar ideas as in Lemma 1, we can argue that if this user's best response $\pi^*(\theta_1,\eta)$ is smaller than $\pi^*(\theta_2,\eta)$, it could always choose to play $\pi^*(\theta_2,\eta)$ when its reputation is θ_1 in order to receive a higher long-term utility. This contradicts the definition of the best response. The same analysis can be applied to a good user as well.

Lemma 2. $\pi^*(\theta_1,\eta) \geq \pi^*(\theta_2,\eta)$ if $\theta_1 < \theta_2 < h$, and $\pi^*(\theta_1,\eta) \geq \pi^*(\theta_2,\eta)$ for any $h \leq \theta_1 < \theta_2$. ■

As proved by Lemma 1 and 2, a user's best response adaptation will not increase the social welfare of a community, and in most cases, will actually decrease it.

B. The community's long term evolution and the stochastic stability

In this section, we examine the evolution of the community under the best response dynamics. We define the strategy configuration of the community as $\{\pi^*_\mu(\theta)\}_{\theta=0}^L$ in one period when the community configuration is μ. Particularly, $\pi^*_\mu(\theta)$ represents the strategy that users of reputation θ will play as the best response in one period. For $\mu = \{n(\theta)\}_{\theta=0}^L$ and $\eta = \{m(\theta)\}_{\theta=0}^L$, we have that $\pi^*_\mu(\theta') = \pi^*(\theta',\eta)$ if $m(\theta) = n(\theta)$ for all $\theta \neq \theta'$ and $m(\theta') = n(\theta') - 1$.

A Markov chain analysis is employed to analyze the evolution of the community configuration μ. For notational convenience, let $Hist$ denote the history of the community, which includes past community configurations and the users' actions until the current period.

Given the social norm $\kappa = (\phi,\tau)$, it is already shown that the best response of a user is fully determined by its opponent configuration, or alternatively, the user's reputation and the community configuration. Meanwhile, the probability of the community configuration in the next period being μ', specified as $p(\mu' \mid Hist)$, is also determined by μ and $\{\pi^*_\mu(\theta)\}_{\theta=0}^L$. As a result, we have that $p(\mu' \mid Hist) = p(\mu' \mid \mu, \{\pi^*_\mu(\theta)\}_{\theta=0}^L) = p(\mu' \mid \mu)$, which implies that the community configuration evolves as a Markov chain on the finite space

$$\mathcal{U} = \{\mu = (n(0),\ldots,n(\mathrm{L})) \mid n(\theta) \in \mathbb{N} \text{ and } \sum_{\theta=0}^{L} n(\theta) = N\} \quad (10)$$

whose size is $|\mathcal{U}| = \sum_{l=0}^{L-1} C_N^l$. The transition probabilities are given by $p_{\mu\mu'} = p(\mu' \mid \mu)$ between any two configurations μ and $\mu' \in \mathcal{U}$ and $P = [p_{\mu\mu'}]$ is the transition matrix. Note that with $\varepsilon > 0$, all entries in P are positive. It is well-known that this Markov chain is irreducible and aperiodic, which introduces a unique stationary distribution. Let $\Delta_{|\mathcal{U}|} \equiv \{\boldsymbol{q} \in \mathbb{R}^{|\mathcal{U}|} \mid q_i \geq 0 \text{ for } i = 1,2,\cdots,|\mathcal{U}| \text{ and } \sum_i q_i = 1\}$ be the $|\mathcal{U}|$-dimensional simplex. Any $\boldsymbol{q} \in \Delta_{|\mathcal{U}|}$ represents a probability distribution on the set of community configurations. Indexing all configurations in $\mathcal{U}$ from 1 to $|\mathcal{U}|$, q_i then represents the frequency that the community stays at the i-th configuration in the long term. A stationary distribution on the space $\mathcal{U}$ is a row vector $\boldsymbol{\omega} = (\omega_1,\ldots,\omega_{|\mathcal{U}|}) \in \Delta_{|\mathcal{U}|}$ satisfying

$$\boldsymbol{\omega} = \boldsymbol{\omega} P. \quad (11)$$

To emphasize its dependence on ε, we sometimes also write $\boldsymbol{\omega}$ as $\boldsymbol{\omega}(\varepsilon)$. When P is strictly positive, not only $\boldsymbol{\omega}(\varepsilon)$ exists and is unique, it also preserves the following important properties.

Lemma 3. $\boldsymbol{\omega}(\varepsilon)$ preserves the following properties:

(1) Stability: Starting from an arbitrary initial community configuration $\boldsymbol{q}$, we have $\lim_{t\to\infty} \boldsymbol{q}P^t \to \boldsymbol{\omega}(\varepsilon)$.

(2) Ergodicity: Starting from an arbitrary initial community configuration $\boldsymbol{q}$, the fraction of time that the community stays at configuration i is $w_i(\varepsilon)$ in the long term, for any $1 \leq i \leq |\mathcal{U}|$. ■

Next, we analyze the stochastic stability of the community when $\varepsilon \to 0$. To facilitate the analysis, we first define the concept of the limiting configuration distribution.

Definition 2. The limiting configuration distribution of the community is defined by

$$\bar{\boldsymbol{\omega}} = \lim_{\varepsilon\to 0} \boldsymbol{\omega}(\varepsilon). \quad (12)$$

The existence and the uniqueness $\bar{\boldsymbol{\omega}}$ can be proved as in [9]. Here we simply list the result.

Lemma 4. There exists a unique limiting distribution for the community configurations. ■

Correspondingly, we define the concept of stochastically stable configuration as follows.

Definition 3. A community configuration μ_i is called as a *stochastically stable configuration* if and only if $\bar{\omega}_i > 0$.

Therefore, when time goes to infinity, the fraction of time that the community stays at a stochastically stable configuration is bounded away from zero [11]. On the contrary, if a configuration is not stochastically stable, then it has zero probability to emerge in the long term.

We prove in the following theorem that a stochastically stable configuration always belongs to a stochastically stable equilibrium. Moreover, we also prove that a stochastically stable configuration

[3] It should be noted that the social norm does not only punish users who do not provide services as required. If a user provides service to another user who is supposed to be punished by the social norm, this user itself will also be punished.

is solely composed of populations of reputations 0 and L. Alternatively speaking, users in a stochastically stable equilibrium will either always follow σ_D and hold a reputation of 0, or will always comply with the social rule ϕ and hold a reputation of L.

Theorem 1. A community configuration $\mu = \{n(0), \cdots, n(L)\}$ is a stochastically stable configuration if and only if there is a stochastically stable equilibrium with μ being its community configuration. μ also preserves the following property

$$n(\theta) = 0, \text{ for all } \theta \in \{1, \cdots L-1\}. \quad (13)$$

■

Theorem 1 provides characterizations of the stochastically stable configurations. Hence, there are $N+1$ community configurations which could possibly be stochastically stable configurations. With the slight abuse of notations, we use μ_k to denote the community configuration with $n(L) = k$ and $n(0) = N - k$ for $k \in \{0, \ldots N\}$. Since users are not mutually cooperative in any community configuration μ_k with $k \in \{0, \ldots N-1\}$, the protocol designer should design a social norm κ under which there is a unique stochastically stable equilibrium with the community configuration being μ_N. Since the reputations of all users remain at L, it is obvious that all users are complying with the social norm and are mutually cooperative with each other. Hence, the social welfare is maximized. The design problem of the protocol designer could then be presented as follows

$$\begin{aligned} &\textit{select } \kappa \\ &\textit{s.t. } \mu_N \textit{ is the unique stochastically stable configuration} \end{aligned} \quad (14)$$

The following theorem solves the design problem (14) by determining the condition when μ_N is the unique stochastically stable configuration. To do this, we examine the "basin of attraction" of each community configuration [9]. Generally speaking, the configuration of a stochastically stable equilibrium has the largest basin of attraction among all configurations. Therefore, if there is a unique configuration that has the largest basin of attraction, we can then conclude that there is a unique stochastically stable equilibrium [11].

Theorem 2. When the following inequality holds

$$h < \min\{\frac{\ln(c/b)}{\ln \delta}, \frac{\ln(c/b)}{\ln[(\frac{N-1}{N}) + (\frac{N+1}{N} - \frac{1}{\delta})\frac{c}{b}]}\}, \quad (15)$$

the community converges to a unique stochastically stable configuration μ_N in the long term as $n(L) = N$ and $n(\theta) = 0$ for all $\theta < L$. ■

Theorem 2 can be used as a guideline for a protocol designer to select desirable social norms for an online community to induce mutual cooperation among users. In particular, it provides three sufficient conditions that make it in the self-interest of a user to comply with ϕ: (1) the service cost is sufficiently low to ensure a small service cost to benefit ratio c/b; (2) users are sufficiently patient; (3) the social rule provides sufficient services to users with a low service threshold.

We briefly explain the intuitions behind the above three conditions. Condition (1) highlights that by deviating from the social norm, the maximum gain on a user's one-period utility is the saving of its service cost c, at the loss of its future utility which is proportional to b. Hence, the decreasing on c/b provides larger incentives for users to comply with the social norm. A similar analysis can be applied to Condition (2). The discount factor δ adjusts the weights that a user places on its current and future utilities. With a larger δ, a user puts a higher weight on its future utility, and thus becomes more interested in increasing its reputation rather than deviating to saving its current service cost. As a result, the incentive for a user to comply with the social norm increases. Condition (3) contradicts the traditional opinion that a user's incentive will increase when the punishment imposed by the social norm protocol becomes more severe. When the punishment is too severe, which is represented here by a high service threshold in the social norm, bad users with their reputations below h will lose their incentives to comply with the social norm. Hence, this prohibits the protocol designer from increasing h arbitrarily.

IV. EXPERIMENTS

In our experiments, we simulate the learning behavior of $N = 1000$ users. We run the experiment for 10^8 periods and measure the average reputation distribution over every 2.5×10^7 periods. This can be used as an approximation on the community configuration in the long term and the results are illustrated in Figure 2. when $\varepsilon = 0.2$ and 0.05, respectively. The community configuration oscillates when $\varepsilon = 0.2$ and cannot converge to a unique stationary point. Hence, the average reputation distribution changes after every 2.5×10^7 periods, with each reputation taking a positive fraction in the population in the long term. This highlights the difficulties of sustaining cooperation in communities with large errors. When $\varepsilon = 0.05$, the community configuration converges to the stochastically stable equilibrium. As a result, almost all users are of reputations 0 and L in the long term, with users of reputation 0 providing no service and users of reputation L acting in a mutually cooperative manner with each other.

In the remainder of the experiments, we mainly focus on the stochastic stability of the community by considering a small operation error $\varepsilon = 0.05$. Figure 3. shows how the fraction of users of reputation L, i.e. $n(L)$, changes in the long term against the discount factor δ, and how the service benefit b as well as the social threshold h impacts the stochastically stable equilibrium. Since users become more patient during their adaptations with a larger discount factor, the social norm providers larger threats to users through its punishments. Hence, users have more incentives to comply with the social norm with $n(L)$ monotonically increasing. When $b = 3$, the instant saving of the service cost outweighs the future benefit of obtaining a high reputation and hence, more users maintain the lowest reputation 0 compared to the case with $b = 5$.Similar results are obtained when h is adjusted, with users being easier to incentivize to comply with the social norm and thus, having the social community converge in the long-term to reputation L when h is smaller.

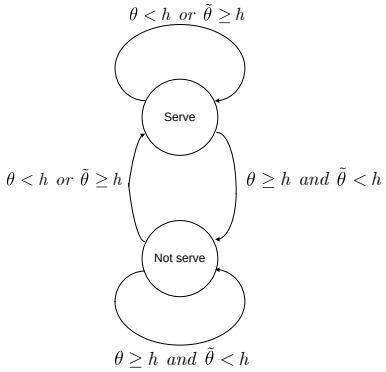
In the final part of the experiment, we assume that the community characteristics, e.g. the utility structure (b, c) and the discount factor δ are not fixed but vary over time and consider how such variation will impact the long term evolution of the community. As an example, we use b as the representative variable to plot the result and assume it varying over time following a Gaussian distribution with the mean $\overline{b} = 3$ and the variance $\sigma^2 = 0.01$. Figure 4. depicts the social welfare of the community over time. We consider two selections on the social

rule as $h=1$ and $h=2$. In both cases, the social welfare when b is variable (solid lines) is smaller than the social welfare when b is constant (dotted lines). This is due to the fact that our designed protocol only guarantees users sufficient incentives to comply with the social norm when b is at its mean value 3. When b deviates from its mean value, users might have incentives to adapt their best responses and deviate from the social norm. In addition, since most users maintain reputation L in the stochastically stable equilibrium when $\bar{b}=3$ and $\delta=0.5$, the social norm with $h=2$ can provide larger incentives for users of reputation L to follow the social strategy. Therefore, it is more robust against the variation on b, which maintains $n(L)$ at a higher level than the social norm with $h=1$. As a result, the social norm with $h=2$ delivers higher social welfare for both $N=500$ and $N=1000$ when b is variable.

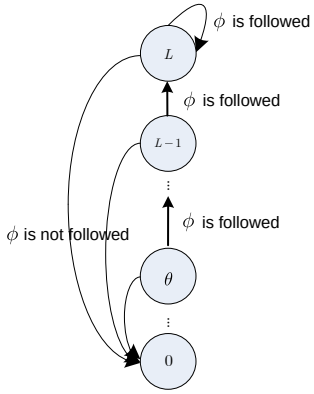
V. Conclusion and Future Research Directions

We have studied the problem of designing social norm based protocols for online communities and analyzed the strategic learning behavior of users under such protocols. Knowledge on the evolution of the community in the long term can enable the protocol designers to select and implement protocols which maximize the social welfare of the community. Our analysis can be extended in several directions, among which we mention four. First, users in the community do not necessarily need to be homogeneous as discussed in Section II. Different users can have different benefits and costs for the service received/provided. Also, they can choose different discount factors δ when evaluating the long-term utility. The discount factor that a user chooses can be dynamically adjusted over time depending on its own expected lifetime in the community. Second, clients can use more complicated decision rules while reporting the servers' actions to the community manager in order to maximize their own long-term utility, instead of always reporting truthfully. Third, online communities may be subject to practical constraints such as topological constraints, in which users can only observe the local information and different users at different locations do not necessarily share the same community information. Hence, the analysis in this paper needs to be extended to scenarios where users adapt based on partial and heterogeneous information. Fourth, users adopt a simple belief model that other users will always comply with the social norm. However, a more sophisticated belief model can be introduced into our framework such that users can update their beliefs on others based on their observation. For example, the formation of user beliefs and opinions in social networks are extensively studied in [21] and [22]. Understanding how the users' beliefs and strategies evolve also forms an important future research direction.

Appendix A (Figures)

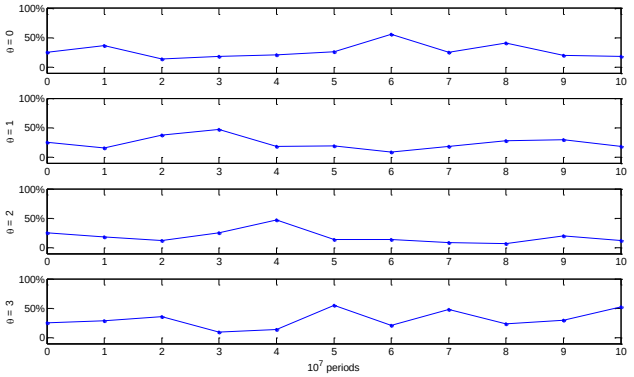


(a)

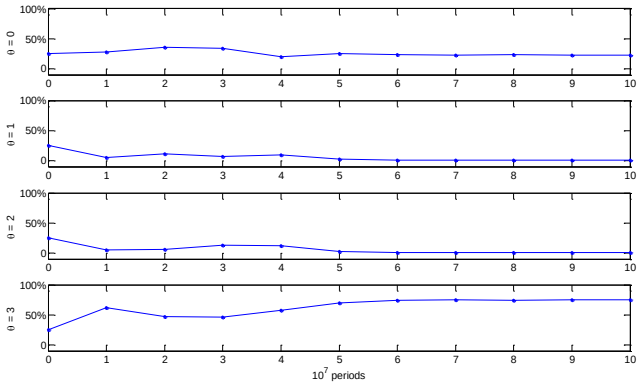


(b)

Figure 1. The schematic representation of a social norm: (a) a representative social rule; (b) a representative reputation scheme



(a) $\varepsilon=0.2$ and $b=3$



(b) $\varepsilon=0.05$ and $b=3$

Figure 2. The evolution of the community configuration in 10^8 periods. ($c=1, \delta=0.5$)

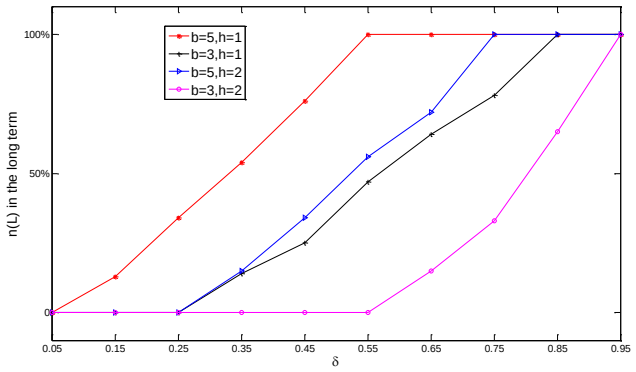


Figure 3. The fraction $n(L)$ of users of reputation L after 10^8 periods ($L=3, \varepsilon=0.05$, $c=1$)

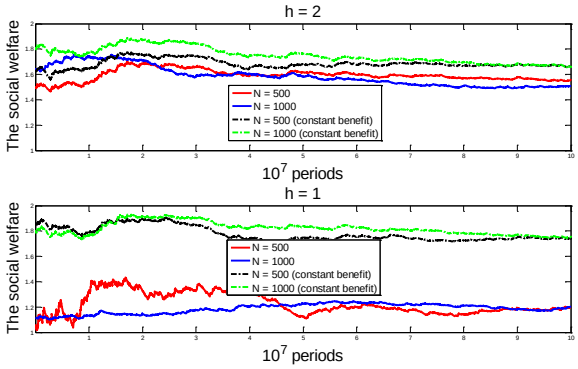


Figure 4. The evolution of the social welfare in the long term when the stage game benefit varies over time ($L=3, \delta=0.5, \varepsilon=0.05$, $c=1$)

REFERENCES

[1] E. Adar, B. A. Huberman, "Free Riding on Gnutella,"*First Monday*, vol. 5, No. 10, Oct. 2000.

[2] B. Cohen, "IncentivesBuilding Robustness in BitTorrent,"*Proc. P2P Econ. Workshop*, Berkeley, CU, 2003.

[3] M. Feldman, K. Lai, I. Stoica, and J. Chuang. "Robust Incentive Techniques for User-to-UserNetworks,"*Proc. ACM Conf. on Elec. Commerce*, no. 4, pp. 102 – 111, 2004.

[4] P. Johnson, D. Levine, W. Pesendorfer, "Evolution and Information in a Gift-Giving Game,"*Journal of Econ. Theory*, vol. 100, no. 1, pp 1-21, 2001.

[5] A. Blanc, Y. Liu, A. Vahdat, "Designing Incentives for Peer-to-Peer Routing,"*Proc. of INFOCOM*, 2005.

[6] J. Hu and M. Wellman, "Nash Q-learning for General-sum Stochastic Games," *Journal of Machine Learning Research*, vol. 4, pp. 1039 – 1069, 2003.

[7] J. Shamma and G. Arslan, "Dimensions of Cooperative Control," *Cooperative Control of Distributed Multiagent Systems*, Wiley, 2008.

[8] Y. Su and M. van der Schaar, "Conjectural Equilibrium in Multi-user Power Control Games," *IEEE Trans. Signal Process.*, vol. 57, no. 9, pp. 3638 – 3650, 2009.

[9] M. Kandori, G. Mailath, and R. Rob, "Learning, Mutation, and Long term Equilibria in Games,"*Econometrica*, 1993, vol. 61, no. 1, pp. 29 – 56, 1993.

[10] C. Camerer, T. Ho, and J. Chong, "Sophisticated EWA Learning and Strategic Teaching in Repeated Games," *Journal of Economic Theory*, vol. 104, pp. 137 – 188, 2002.

[11] D. Foster, P. Young, "Stochastic Evolutionary Game Dynamics,"*Theoretical Population Biology*, vol. 38, no. 2, pp. 219 – 232, 1990.

[12] Y. Zhang, J. Park, M. van der Schaar, "Designing Social Norm Based Incentive Schemes to Sustain Cooperation in a Large Community," *Proc. 2nd International ICST Conf. on Game Theory for Networks*, 2011.

[13] C. Claus and C. Boutilier, "The Dynamics of Reinforcement Learning in Cooperative Multiagent Systems," *Proc. 5th National Conf. on Artificial Intelligence*, pp. 746 – 752, 1998.

[14] KaZaA User-to-User File Sharing CU, [Online]. Available: http://www.kazaa.com/, [Online]. Available

[15] M. Nowak, K. Sigmund, "Evolution of Indirect Reciprocity,"*Nature* 437, pp. 1291 – 1298, 2005.

[16] B. Skyrms, R. Pemantle, "A Dynamic Model of Social Network Formation,"*Adaptive Networks (Understanding Complex Systems)*, pp. 231 – 251, 2009.

[17] R. Sutton, A. Barto, "Reinforcement Learning: An Introduction,"*MIT Press*, Cambridge, MA, 1998.

[18] G. Weinstraub, C. Benkard, B. Van Roy, "Markov Perfect Industry Dynamics with Many Firms,"*Econometrica*, vol. 76, no. 6, pp. 1375 – 1411, 2008.

[19] C. Guestrin, D. Koller, and R. Parr, "Multiagent Planning with Factored MDPs," *Advances in Neural Info. Proc. Systems* (NIPS-14), 2001.

[20] M. Littman, "Friend-or-foe Q-learning in General-sum Games," *Proc. of the 8th Int'l Conf. on Machine Learning*, 2001.

[21] D. Acemoglu, A. Ozdaglar, "Opinion Dynamics and Learning in Social Networks," MIT LIDS report 2851.

[22] V. Krishnamurthy, "Quickest Time Herding and Detection for Optimal Social Learning," arXiv:1003.4972 [cs.IT].