Sharing in Networks of Strategic Agents

Jie Xu, Yangbo Song, and Mihaela van der Schaar, *Fellow, IEEE*

***Abstract*—In social, economic and engineering networks, connected agents need to cooperate by repeatedly sharing information and/or goods. Typically, sharing is costly and there are no immediate benefits for agents who share. Hence, agents who strategically aim to maximize their own individual utilities will "free-ride" because they lack incentives to cooperate/share, thereby leading to inefficient operation or even collapse of networks. To incentivize the strategic agents to cooperate with each other, we design *distributed* rating protocols which exploit the ongoing nature of the agents' interactions to assign ratings and through them, determine future rewards and punishments: agents that have behaved as directed enjoy high ratings—and hence greater future access to the information/goods of others; agents that have not behaved as directed enjoy low ratings—and hence less future access to the information/goods of others. Unlike existing rating protocols, the proposed protocol operates in a distributed manner and takes into consideration the underlying interconnectivity of agents as well as their heterogeneity. We prove that in many networks, the price of anarchy (PoA) obtained by adopting the proposed rating protocols is 1, that is, the optimal social welfare is attained. In networks where PoA is larger than 1, we show that the proposed rating protocol significantly outperforms existing incentive mechanisms. Last but not least, the proposed rating protocols can also operate efficiently in dynamic networks, where new agents enter the network over time.**

***Index Terms*—Cooperative networks, distributed protocols, economics networks, imperfect monitoring, incentive design, indirect reciprocity, ratings, repeated games, social networks, social reciprocation.**

I. Introduction

In recent years, extensive research effort has been devoted to studying cooperative networks where autonomous agents interact repeatedly with each other over an exogenously given network by sharing information (such as measurements, estimates, beliefs, or opinions) or goods (such as endowments or production). These networks require various levels of coordinated behavior and cooperation among the autonomous agents. However, in many scenarios, participating in the cooperative process entails costs to the agents, such as the cost of producing, processing and transferring information/goods to their neighbors. If agents are strategic, they will choose to cooperate with other agents in the network only if cooperation maximizes their own long-term utilities, which take into account both current and future benefits. Absent incentives for cooperation, agents will free-ride and the networks will work inefficiently or even collapse [2]. If a central authority existed in the network, which was omniscient about agents' utilities and actions as well as capable of computing and enforcing an efficient behavior profile for all the agents, the social optimum could be attained; but, in practice, such central authorities do not exist. On the contrary, agents usually possess only local information, and they act selfishly to maximize their own payoff. Hence, incentives are needed to compel the strategic agents to act in a socially optimal manner. Designing incentives for networks of strategic agents is significantly more challenging than in scenarios where agents are randomly matched [10], [19]–[21] or interact as independent pairs [7], [8], since the incentives of agents are complexly coupled based on the connectivity of agents. Moreover, effective implementation of an incentive scheme requires that it be distributed, which represents another key challenge. In this paper we present the first scheme that solves these problems.

To better motivate this work, we provide two concrete application scenarios. Establishing a secure cyber environment requires investments on security technologies (e.g. firewalls, access control etc.) from autonomous systems (ASes). Improved security can be achieved if ASes deploy proactive protection technologies (e.g. outbound traffic control) which are more effective because ASes have better control over their own devices and traffic originating from their own users [29]. However, ASes are self-interested and are reluctant to make security investment on these proactive technologies since doing that is not directly beneficial to themselves [29]. The similar incentive problem also exists in joint spectrum sensing problems in cognitive radio systems [9]. To enable dynamic spectrum access, the preliminary requirement is the ability to accurately identify the presence of primary users over a wide range of spectrum. With joint spectrum sensing, each secondary user senses the spectrum individually and then shares the raw sensing results to their neighbors at the beginning of each transmission slot to improve the detection probability in this slot. However, secondary users are self-interested and lack the incentives to send their sensing raw results to their neighbors which will cost extra resources such as energy and transmission time.

We resolve the above incentive problem by deploying a *distributed* rating protocol. The rating protocol consists of three components: a set of ratings, recommended strategies (for each agent) and rating update rules (for each agent). In each period, each agent is assigned a rating, which is maintained and updated according to the rating protocol. The actions recommended to

Manuscript received September 15, 2013; revised January 22, 2014 and April 02, 2014; accepted April 02, 2014. Date of publication April 11, 2014; date of current version July 16, 2014. This work was supported by the National Science Foundation (NSF) under Grant CCF-1218136. The guest editor coordinating the review of this manuscript and approving it for publication was Prof. H. Vincent Poor.

J. Xu and M. van der Schaar are with the Department of Electrical Engineering, University of California at Los Angeles (UCLA), Los Angeles, CA 90095 USA (e-mail: jiexu@ucla.edu; mihaela@ee.ucla.edu).

Y. Song is with the Department of Economics, University of California at Los Angeles (UCLA), Los Angeles, CA 90095 USA (e-mail: darcy07@ucla.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTSP.2014.2316597

the agents by the system (e.g. how much to share) depend on the ratings of their neighboring agents (i.e. the agents with whom they are directly connected). The recommendations can be determined in a distributed manner by the system. For example, each agent could be interacting with other agents through a software client (similar to BitTorrent). Each agent's software is then preprogrammed to recommend actions based on the local network structure it observes, information that is received from the software of neighboring agents and the current ratings of neighboring agents. Since agents are strategic and want to maximize their own utility, they have the freedom to decide their own actions and they may comply or not with the recommended actions. Based on the agent's current rating and whether it has followed/deviated from the recommended strategy, the software increases/decreases an agent's rating. We design rating protocols (i.e. recommended strategies and rating update rules) that are incentive-compatible (i.e. agents have incentives to follow the recommended strategies) and maximize the social welfare (i.e. the sum of the utility of all agents).

There are two central challenges. The first arises from the fact that agents interact over a network. In particular, the agents' interactions are subject to network constraints, i.e. agents can only interact with their neighbors. This is in stark contrast with existing works in repeated games relying on social reciprocation which assume that the agents are randomly matched [10], [19], [20] or interact on a complete graph [9]. Due to the network constraints, agents' incentives are coupled in a much more complex manner since they depend directly on the behavior of their immediate neighbors and indirectly on the behavior of more distant remaining agents. Because of the different network constraints, there is not a universal rating protocol that can work efficiently in all networks. Instead, the rating protocol design must explicitly take into account the specific coupling among agents arising from the specific network.

The second arises because we insist on protocols that are distributed and informationally decentralized. We do not need to assume the existence of any central entity that can monitor the entire network (i.e. network topology, all agents' utility functions and actions) and communicate to all individual agents about all other agents' behavior. Decentralization rules out protocols proposed in prior works [19],[20] since they are designed and implemented in a centralized manner, requiring the knowledge of the entire network at a central entity. In this paper, the rating protocol is designed and implemented in a distributed manner, requiring only limited message exchange (i.e. Lagrangian multipliers during configuration and agents' ratings during interaction) among the software of neighboring agents.

The main contributions of this paper are:

1) We develop a framework for providing incentives in networks where heterogeneous agents interact repeatedly over a network. This framework is very general and can be employed for a variety of applications, including in networks where bilateral interest may not exist between agents and hence, existing works based on direct reciprocation such as Tit-for-Tat [7], [8] do not work.
2) We rigorously analyze the incentives (Theorem 1) of agents operating under the rating protocol framework using a novel repeated game with imperfect monitoring formalism, which explicitly considers the network structure, agents' utility functions etc. With these constraints and using the dual decomposition method, we propose a novel and fully distributed algorithm to compute the optimal recommended strategy of the rating protocol that maximizes the social welfare.
3) We show how different networks may affect agents' incentives in different ways and how to design rating protocols that are tailored to different networks. Modified rating protocols that apply to various dynamic networks are also proposed and analyzed.

The remaining part of this paper is organized as follows. In Section II, we review related works and existing solutions, and highlight the key differences to this work. Section III outlines the system model and formulates the protocol design problem. In Section IV, we design the optimal rating protocol to maximize the social welfare. The performance of the optimal design is then analyzed in Section V. Section VI studies the rating protocol design in a class of dynamic networks. Section VII provides numerical results to highlight the features of the proposed protocol. Finally, we conclude this paper in Section VIII.

II. Related Works

Cooperation among the agents (e.g. repeated sharing) is critical for the enhanced performance and robustness of various types of social, economic and engineering networks [1]. The main focus of this literature is on determining the resulting network performance if agents repeatedly share and process information/goods. However, absent incentives and in the presence of strategic agents, these networks will work inefficiently or even collapse [2]. Thus, the main focus of the current paper is how to incentivize strategic agents to cooperate such that networks can operate efficiently.

A variety of incentive schemes has been proposed to encourage cooperation among agents (see e.g. [4] for a review of different game theoretic solutions). Two popular incentive schemes are pricing and differential service. Pricing schemes [5], [6] use payments to reward and punish individuals for their behavior. However, they often require complex accounting and monitoring infrastructure, which introduces substantial communication and computation overhead. Differential service schemes, on the other hand, reward and punish individuals by providing differential services depending on their behavior. Differential services can be provided by the network operator [10], [16], [19]. However, in many networks of autonomous agents, such a centralized network operator does not exist. Alternatively, differential services can also be provided by the other agents participating in the network since agents in the considered applications derive their utility from their interactions with other agents [7]–[12], [19]–[21]. Such incentive schemes are based on the principle of reciprocity and can be classified into direct (personal) reciprocation and social reciprocation. In direct (personal) reciprocation schemes (e.g. the widely adopted Tit-for-Tat strategy [7], [8]), the behavior of an individual agent toward another is based on its personal experience with that agent. However, they only work when

TABLE I
COMPARISON WITH EXISTING WORKS

	Social Learning [1]	**Direct Reciprocation** [7][8]	**Social Reciprocation** [10][19][20]	**This paper**
Information/goods exchange	Costless	Costly	Costly	Costly
Asymmetric interests	No	No	Yes	Yes
Objective of study	Convergence of agents' beliefs and actions	Incentives for agents to cooperate	Incentives for agents to cooperate	Incentives for agents to cooperate
Game type	One-shot game/ Bayesian game	Repeated game	Repeated game	Repeated game
Robust to monitoring errors	No	Yes & No	Yes & No	Yes
Equilibrium concept	Bayesian equilibrium	Subgame perfect equilibrium	Public perfect equilibrium	Perfect local equilibrium
Network topology	Arbitrary	Arbitrary	Fully connected/ Random matching	Arbitrary
Agents actions	Belief update	Cooperation level	Cooperation level	Cooperation level
Agents' utility depends on	Self belief/action	Own actions and others actions	Own actions and others actions	Own actions and others (joint) actions
Utility function	Homogeneous & Heterogeneous	Homogeneous	Homogenous	Heterogeneous
Distributed design	Yes	Yes	No	Yes

two interacting agents have bilateral interests. In social reciprocation schemes [9]–[12], [19]–[21], individual agents obtain some (public) information about other individuals (e.g. their ratings) and decide their behavior toward other agents based on this information.

Incentive mechanisms based on social reciprocation are often studied using the familiar framework of repeated games. In [9], the sharing game is studied in a narrower context of cooperative spectrum sensing and various simple strategies are investigated. Agents are assumed to be able to communicate and share sensing results with all other agents, effectively forming a complete graph where the agents' knowledge of the network is complete and symmetric. However, such an assumption rarely holds in distributed networks where, instead, agents may interact over arbitrary topologies and have incomplete and asymmetric knowledge of the entire network. In such scenarios, simple strategies proposed in [9] will fail to work and the incentives design becomes significantly more challenging.

Contagion strategies on networks [10]–[12] are proposed as a simple method to provide incentives for agents to cooperate. However, such methods do not perform well if monitoring is imperfect since any single error can lead to a network collapse. Even if certain forms of forgiveness are introduced, contagion strategies are shown to be effective only in very specific networks [11], [12]. It is still extremely difficult, if not impossible, to design efficient forgiving schemes in arbitrary distributed networks since agents will have difficulty in conditioning their actions on history, e.g. whether they are in the contagion phase or the forgiving phase, due to the asymmetric and incomplete knowledge.

Rating/reputation mechanisms are proposed as another promising solution to implement social reciprocation. Much of the existing work on reputation mechanism is concerned with practical implementation details such as effective information gathering techniques [13] or determining the impact of reputation on a seller's prices and sales [14], [15]. The few works providing theoretical results on rating protocol design consider either one (or a few) long-lived agent(s) interacting with many short-lived agents [16]–[18] or anonymous, homogeneous and unconnected agents selected to interact with each other using random matching [10], [19], [20]. Importantly, few of the prior works consider the design of such rating protocols for networks where agents interact over a network, which leads to extremely complex and coupled interactions among agents. Moreover, the distributed nature of the considered sharing networks imposes unique challenges for the rating protocol design and implementation which are not addressed in prior works [19], [20].

In Table I, we compare the current paper with existing works on social learning and incentive schemes based on direct reciprocation and social reciprocation.

III. SYSTEM MODEL

A. Network Environment

We consider a network of N agents, indexed by $\{1, 2, \ldots, N\} = \mathcal{N}$. Agents are connected subject to an underlying topology $G = \{g_{ij}\}_{i,j\in\mathcal{N}}$ with $g_{ij} = g_{ji} = 1$ (here we consider undirected connection) representing agent i and j being connected (e.g. there is a communication channel between them) and $g_{ij} = g_{ji} = 0$ otherwise. Moreover, we set $g_{ii} = 0$. We say that agent i and agent j are neighbors if they are connected. For now we assume a static network G but dynamic networks are also allowed in our framework and this will be discussed in detail in Section VI.

Time is infinite and divided into discrete periods. In each time period, each agent i decides its action (e.g. information/goods sharing) towards each of its neighbors j, denoted by $a_{ij} \in \mathbb{R}_+$.[1] For example, a_{ij} can represent the effort spent (e.g. information/goods shared) by agent i when interacting with agent j. We

[1]More general action space is also allowed, e.g. a_{ij} is upper bounded.

collect the actions of agent i towards all its neighbors in the notation $\boldsymbol{a}_i = \{a_{ij}\}_{j:g_{ij}=1}$. Denote $\boldsymbol{a} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N)$ as the action profile of all agents and $\boldsymbol{a}_{-i} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{i-1}, \boldsymbol{a}_{i+1}, \ldots, \boldsymbol{a}_N)$ as the action profile of agents except i. Let $\mathcal{A}_i = \mathbb{R}_+^{d_i}$ be the action space of agent i where $d_i = \sum_j g_{ij}$ is the number of agent i's neighbors. Let $\mathcal{A} = \times_{i\in\mathcal{N}} \mathcal{A}_i$ be the action space of all agents.

Agents obtain benefits from the information/goods shared by neighbors. We denote the actions of agent i's neighbors towards agent i by $\hat{\boldsymbol{a}}_i = \{a_{ji}\}_{j:g_{ij}=1}$ and let $b_i(\hat{\boldsymbol{a}}_i)$ be the benefit that agent i obtains from these actions. Spending effort (e.g. sharing information/goods) is costly and the cost $c_i(\boldsymbol{a}_i)$ depends on an agent i's own actions $\boldsymbol{a}_i$. Hence, given the action profile $\boldsymbol{a}$ of all agents, the utility of agent i is

$$u_i(\boldsymbol{a}) = b_i(\hat{\boldsymbol{a}}_i) - c_i(\boldsymbol{a}_i) \tag{1}$$

We impose some constraints on the benefit and cost functions.

Assumption 1: For each i, the benefit $b_i(\hat{\boldsymbol{a}}_i)$ is non-decreasing in each $a_{ji}, \forall j : g_{ij} = 1$ and is strictly concave in $\hat{\boldsymbol{a}}_i$ (in other words, jointly strictly concave in $a_{ji}, \forall j : g_{ij} = 1$).

Assumption 2: For each i, the cost is linear in its sum action, i.e. $c_i(\boldsymbol{a}_i) = \sum_{j:g_{ij}} a_{ij}$.

The above assumptions state that (1) agents receive decreasing marginal benefits of information/goods acquisition, which captures the fact that agents become more or less "satiated" when they possess sufficient information/goods, in the sense that additional information/goods would only generate little additional payoff; (2) the cost incurred by an agent is equal (or proportional) to the sum effort spent to cooperate with all its neighbors. We note that the utility model is general enough to account for the heterogeneity of the value of information/goods to different users since $b_i(\hat{\boldsymbol{a}}_i)$ is agent-specific and depends on the action vector of all agent i's neighbors. For a concrete example, the benefit function can be the widely-adopted Dixit-Stiglitz utility function [26] which captures the information/goods heterogeneity and diversity produced by different agents, i.e.

$$b_i(\hat{\boldsymbol{a}}_i) = f\left(\left(\sum_{j\in\mathcal{N}_i} (w_{ji}a_{ji})^{\gamma_i}\right)^{\frac{1}{\gamma_i}}\right) \tag{2}$$

where $w_{ji} \geq 0$ describes the relative importance of agent j's information/goods to agent i, $\gamma_i \in (0, 1)$ measures agent i' appreciation for information/goods diversity and $f(\cdot)$ is a concave and increasing function.

B. Rating Protocol

Each agent i is associated with a rating $\theta_i(t) \in \Theta = \{1, 2, \ldots, K\}$ in each period t which is maintained and updated according to the rating protocol. The rating of agent i is maintained by the software client of agent i. We collect agent i's neighbors' ratings in $\hat{\boldsymbol{\theta}}_i = \{\theta_j\}_{j:g_{ij}=1} \in \Theta^{d_i}$. The rating protocol recommends actions to an agent depending on neighbors' ratings $\boldsymbol{\sigma}_i : \Theta^{d_i} \rightarrow \mathcal{A}_i$. We refer to this recommendation as the *recommended strategy*. For agent i, $\boldsymbol{\sigma}_i = \{\sigma_{ij}\}_{j\in\mathcal{N}_i}$ consists of d_i elements with $\sigma_{ij}(\theta_j)$ representing the recommended sharing action of agent i towards agent j if agent j's rating is θ_j. We collect the strategies of agent i's neighbors towards agent i in $\hat{\boldsymbol{\sigma}}_i(\theta_i) = \{\sigma_{ji}(\theta_i)\}_{j:g_{ij}=1}$. These recommendations are done in a distributed manner by the system, through the software clients of the agents.

Depending on whether or not agent i followed the recommended strategy, its software client updates agent i's rating at the end of each period. Let $y_i \in Y = \{0, 1\}$ be the public monitoring signal of agent i with $y_i = 1$ if $\boldsymbol{a}_i = \boldsymbol{\sigma}_i$ and $y_i = 0$ if $\boldsymbol{a}_i \neq \boldsymbol{\sigma}_i$ which is generated by the software of agent i. However, monitoring may not be perfect and hence it is possible that even if $\boldsymbol{a}_i = \boldsymbol{\sigma}_i$, it can still be $y_i = 0$ (and if $\boldsymbol{a}_i \neq \boldsymbol{\sigma}_i$, $y_i = 1$). The rating update rule for agent i is a function $\tau_i : \Theta \times Y \rightarrow \Delta(\Theta)$ where $\Delta(\Theta)$ is the probability simplex of the rating set and $\tau_i(\theta_i^+; \theta_i, y_i)$ is the probability that the updated rating is θ_i^+ if agent i's current rating is θ_i and the public signal is y_i. In particular, we consider the following parameterized rating update rule, for agent i,

$$\tau_i\left(\theta_i^+; \theta_i, y_i p\right) = \begin{cases} \alpha_{i,k}, & \text{if } \theta_i^+ = \max\{1, k-1\}, y_i = 0 \\ 1 - \alpha_{i,k}, & \text{if } \theta_i^+ = k, y_i = 0 \\ \beta_{i,k}, & \text{if } \theta_i^+ = \min\{K, k+1\}, y_i = 1 \\ 1 - \beta_{i,k}, & \text{if } \theta_i^+ = k, y_i = 1 \end{cases} \tag{3}$$

In words, compliant agents are rewarded with a higher rating with some probability while deviating agents are punished with a lower rating with some (other) probability. These probabilities $\alpha_{i,k}, \beta_{i,k}$ are in the range of [0,1]. Note that when $\alpha_{i,k} = 0$, the rating set of agent i effectively reduces to a subset $\{k, k+1, \ldots, K\}$ since its rating will never drop below k (if its initial rating is higher than k). Note also that agents remain at the highest rating $\theta = K$ if they always follow the recommended strategy regardless of the choice of $\beta_{i,K}$.

To sum up, the rating protocol is uniquely determined by the recommended (public) strategies $\boldsymbol{\sigma}_i(\hat{\boldsymbol{\theta}}_i), \forall i, \forall \hat{\boldsymbol{\theta}}_i$ and the rating update probabilities $\alpha_{i,k}, \beta_{i,k}$ for every i and k. These will be our design parameters. We denote the rating protocol by $\pi = (\Theta, \boldsymbol{\sigma}, \boldsymbol{\alpha}, \boldsymbol{\beta})$. The rating protocol is configured (i.e. the values of the design parameters are determined) at the beginning of the system by the software clients of the agents. The configuration is carried out in a distributed way, requiring the software clients to exchange with neighbors limited messages (i.e. Lagrangian multipliers etc.). When the network is static, the rating protocol is configured only once at the beginning. When the network is dynamic, the rating protocol is reconfigured once in a while, to adapt to the varying network. We assume that all agents are synchronized and enter the reconfiguration period simultaneously. This synchronization can be coordinated by an exogenous stochastic process (not controlled by any central planner), for instance a random sequence generator with the same seed for each agent. Alternatively, the reconfiguration can also be initiated by a particular agent and then this signal is spread over the entire network. We also note that agents will not have incentives not to perform reconfiguration since the protocol is designed in such a way that participation in this period produces a higher utility for the agent than not participating. Table II summarizes the operation of the rating protocol.

TABLE II
OPERATION OF THE RATING PROTOCOL

		*Agent (**Strategic**)*	*Software client (**Non-strategic**)*
Configuration	*Information*	---	Own agents' utility function and local connectivity
	Action/ Functionality	---	Determine the rating protocol in a distributed manner
Interaction Periods	*Information*	The instantiated rating protocol; Neighbors' ratings.	Whether or not the agent followed the recommended strategy
	Action/ Functionality	Choose sharing actions aiming to maximize own utility	Update the agent's rating; Broadcast it in the neighborhood

C. Problem Formulation

The objective of the protocol designer is to maximize the social welfare of the network, which is defined as the time-average sum utility of all agents, i.e.

$$V = \lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T} \sum_{i} u_i\left(\boldsymbol{a}(t)\right) \tag{4}$$

If agents are obedient, then the system designer can assign socially optimal actions, denoted by $\boldsymbol{a}^{opt}(t), \forall t$, to agents and then agents will simply take the actions prescribed by the system designer. Determining the socially optimal actions involves solving the following utility maximization problem:

$$\begin{aligned} \underset{\boldsymbol{a}}{\text{maximize}} \quad & V \\ \text{subject to} \quad & a_{ij}(t) \geq 0,\ \forall i, j : g_{ij} = 1, \forall t \end{aligned} \tag{5}$$

This problem can be easily solved and any action profile $\boldsymbol{a}^{opt}$ that satisfies

$$\hat{\boldsymbol{a}}_i^{opt}(t) \in \arg\max_{\hat{a}} b_i\left(\hat{\boldsymbol{a}}_i(t)\right) - c_i\left(\hat{\boldsymbol{a}}_i(t)\right) p \tag{6}$$

is a solution. We denote the optimal social welfare by V^{opt}.

The network cooperation (e.g. information/goods sharing) problem becomes much more difficult in the presence of strategic agents: strategic agents may not want to take the prescribed actions because these actions do not maximize their own utilities.

Definition 1: A (one-shot) *network sharing game* is a tuple $\mathcal{G} = \langle \mathcal{N}, \mathcal{A}, \{u_i(\cdot)\}_{i\in\mathcal{N}}; G\rangle$ where $\mathcal{N}$ is the set of players, $\mathcal{A}$ is the action space of all players, $u_i(\cdot)$ is the utility function of player i (defined by (1)) and G is the underlying network.

Consider the utility of an agent i in (1). In order to maximize its own utility, agent i will take the action $\boldsymbol{a}_i = 0$ regardless of other agents' actions $\boldsymbol{a}_{-i}$. Therefore, there exists a unique Nash equilibrium (NE) $\boldsymbol{a}^{NE} = 0$ in the network sharing game in any period.

In this paper, we exploit the repeated interactions among agents to provide agents with incentives to cooperate. In the following, we introduce the equilibrium concept used in this paper.

At the end of each interaction period, each agent i observes the (imperfect) monitoring signal $y_j \in Y = \{0,1\}$ of the action of each of its neighbor j. Write $\mathcal{Y}_i$ for the space of signals observed by agent i and $\mathcal{Y} = \times_{i\in\mathcal{N}}\mathcal{Y}_i$ for the space of signal profiles. A profile of actions $\boldsymbol{a} \in \mathcal{A}$ determines a distribution of signals $\mu_{\boldsymbol{a}} \in \Delta(\mathcal{Y})$; agents observe a realization drawn at random from this distribution. In our network setting, the signal distribution is *local* in the sense that agent i's observed signal depends only on the actions of i' neighbors. Fig. 1 illustrates the local signals observed by agents. A signal history of length T is an element $\boldsymbol{y} = (\boldsymbol{y}^1, \ldots, \boldsymbol{y}^T) \in \mathcal{Y}^T$; $\boldsymbol{y}^t$ is the signal profile at time t and $\boldsymbol{y}_i^t$ is the signal profile observed by agent i at time t. In addition to signals, agents know their own actions and their realized own utilities, so a private history of length T for agent i is an element $h \in (\mathcal{A}_i \times \mathbb{R} \times \mathcal{Y}_i)^T = \mathcal{H}_i^T$ and a private history of length T is a profile of private histories for each agent. A strategy for agent i is a function $\boldsymbol{\sigma}_i : \mathcal{H}_i \to \mathcal{A}_i$, prescribing an action following each history. The strategy $\boldsymbol{\sigma}_i$ is a *local strategy* (or a *local signal strategy*) if it depends only on the history of local signals observed by i (and not on the history of i's actions or realized utilities.) An infinite history for agent i is an element of $(\mathcal{A}_i \times \mathbb{R} \times \mathcal{Y}_i)^\infty = \mathcal{H}_i^\infty$. Note that a strategy profile $\boldsymbol{\sigma}$ defines, for each agent i, a probability distribution $\zeta_i(\boldsymbol{\sigma})$ on the infinite histories $\mathcal{H}_i^\infty$ and hence a probability distribution $\nu_i(\boldsymbol{\sigma})$ on infinite utility streams $\mathbb{R}^\infty$. Agents discount future utilities, so the utility agent i derives from the infinite utility stream $\boldsymbol{u}_i = (u_i^1, u_i^2, \ldots)$ is

$$W_i(\boldsymbol{u}_i) = \sum_{t=0}^{\infty} \delta^t u_i^t \tag{7}$$

where $\delta \in (0,1)$ is the discount factor. Hence the (expected) utility agent i derives if agents follow the strategy profile σ is

$$U_i(\boldsymbol{\sigma}) = \mathbb{E}W_i(\boldsymbol{u}_i) = \int_{\boldsymbol{u}_i} W_i(\boldsymbol{u}_i) d\nu_i(\boldsymbol{\sigma})(\boldsymbol{u}_i) \tag{8}$$

A strategy profile $\boldsymbol{\sigma}$ is a Nash equilibrium if for each agent i, the strategy $\boldsymbol{\sigma}_i$ is a best response to other agents' strategy profile $\boldsymbol{\sigma}_{-i}$; that is

$$U_i(\boldsymbol{\sigma}_i, \boldsymbol{\sigma}_{-i}) \geq U_i(\hat{\boldsymbol{\sigma}}_i, \boldsymbol{\sigma}_{-i}) \tag{9}$$

for every strategy $\hat{\boldsymbol{\sigma}}_i$. The profile σ is a *local equilibrium* if it is a Nash equilibrium and every agent uses a local strategy; it is a *perfect local equilibrium* (PLE) if in addition it is a Nash equilibrium following every history.

The proposed rating protocol assigns each agent a rating that summarizes the public signal history of the action of that agent.

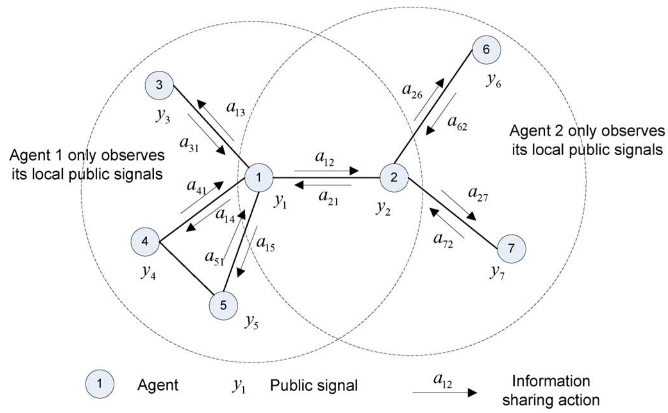


Fig. 1. Illustration of local public signals.

Hence, $\boldsymbol{\sigma}_i : \mathcal{H}_i \rightarrow \mathcal{A}_i$ is reduced to $\boldsymbol{\sigma}_i : \Theta^{d_i} \rightarrow \mathcal{A}_i$. This significantly reduces the implementation complexity since agents need to keep only the current ratings of their neighbors instead of the entire signal history of their neighbors. If a recommended strategy profile constitutes a PLE, then agents have incentives to follow their recommended strategies. Denote the achievable social welfare by adopting the rating protocol by $V(\pi)$. The rating protocol design problem thus is

$$\begin{aligned} &\underset{\pi=(\Theta,\boldsymbol{\sigma},\boldsymbol{\alpha},\boldsymbol{\beta})}{\text{maximize}} && V(\pi) \\ &\text{subject to} && \boldsymbol{\sigma} \text{ constitutes a PLE} \end{aligned} \tag{10}$$

D. Illustrative Example: Cooperative Estimation

We illustrate the generality of our formalism by showing how well-studied joint estimation problems [9], [28] can be cast into it. Our proposed framework can also be used to solve problems such as distributed cybersecurity investment [29] and cooperation in economics networks [30] etc.

Suppose that each agent observes in each period a noisy version of a time-varying underlying system parameter $s(t)$ of interest. Denote the observation of agent i by $o_i(t)$. We assume that $o_i(t) = s(t) + \epsilon_i(t)$, where the observation error $\epsilon_i(t)$ is i.i.d. Gaussian across agents and time with mean zero and variance r^2. Agents can exchange observations with their neighbors to obtain better estimations of the system parameter. Let $a_{ij}(t)$ be the transmission power spent by agent i. The higher the transmission power the larger probability that agent j receives this additional observation from agent i. Agents can use various combination rules [9] to obtain the final estimations. The expected mean square error (MSE) of agent i's final estimation will depend on the actions of its neighbors, denoted by $MSE_i(\hat{\boldsymbol{a}}_i(t))$. If we define the MSE improvement as the benefit of agents, i.e. $b_i(\hat{\boldsymbol{a}}_i(t)) = r^2 - MSE(\hat{\boldsymbol{a}}_i(t))$, then the utility of agent i in period t given the received benefit and its incurred cost is $u_i(\boldsymbol{a}(t)) = r^2 - MSE_i(\hat{\boldsymbol{a}}_i(t)) - \|\boldsymbol{a}_i(t)\|_1$.

IV. Distributed Optimal Rating Protocol Design

If a rating protocol constitutes a PLE, then all agents will find it in their self-interest to follow the recommended strategies. If the rating update rule updates the ratings of compliant agents upward with positive probabilities, then eventually all agents will have the highest ratings forever (assuming no update errors). Therefore, the social welfare, which is the time-average sum utility, is asymptotically the same as the sum utility of all agents when they have the highest ratings and follow the recommended strategy, i.e.

$$V = \sum_i \left(b_i\left(\hat{\boldsymbol{\sigma}}_i(K)\right) - c_i\left(\boldsymbol{\sigma}_i(\boldsymbol{K})\right)\right) \tag{11}$$

This means that the recommended strategy profile $\boldsymbol{\sigma}(\boldsymbol{K})$ for the highest ratings determines the social welfare that can be achieved by the rating protocol. If this strategy profile can be arbitrarily chosen, then we can solve a similar problem as (5) for the obedient agent case. However, in the presence of self-interested agents, this strategy profile, together with the other components of a rating protocol, need to satisfy the equilibrium constraint such that self-interested agents have incentives to follow the recommended strategies. In Theorem 1, we identify a sufficient and necessary condition on $\boldsymbol{\sigma}(\boldsymbol{K})$ such that an equilibrium rating protocol can be constructed. With this, we are able to determine the optimal rating protocol in a distributed way in order to maximize the social welfare. We denote the social welfare that can be achieved by the optimal rating protocol as V^* and use *the price of anarchy* (PoA), defined as $PoA = V^{opt}/V^*$, as the performance measure of the rating protocol.

A. Sufficient and Necessary Condition

To see whether a rating protocol can constitute a PLE, it suffices to check whether agents can improve their long-term utilities by one-shot unilateral deviation from the recommended strategy after any history (according to the one-shot deviation principle in repeated game theory [24]). Since in the rating protocol, the history is summarized by the ratings, this reduces to checking the long-term utility in any state (i.e. any rating profile $\boldsymbol{\theta}$ of agents). Agent i's long-term utility when agents choose the action profile $\boldsymbol{a}$ is

$$U_i(\boldsymbol{\theta}; \boldsymbol{a}) = u_i(\boldsymbol{\theta}; \boldsymbol{a}) + \delta \sum_{\boldsymbol{\theta}'} p(\boldsymbol{\theta}'|\boldsymbol{\theta}; \boldsymbol{a}) U_i^*(\boldsymbol{\theta}'), \tag{12}$$

where $p(\boldsymbol{\theta}'|\boldsymbol{\theta}; \boldsymbol{a})$ is the rating profile transition probability which can be fully determined by the rating update rule based on agents' actions and $U_i^*(\boldsymbol{\theta}')$ is the optimal value of agent i at the rating profile $\boldsymbol{\theta}'$, i.e. $U_i^*(\boldsymbol{\theta}') = \max_{\boldsymbol{a}} U_i(\boldsymbol{\theta}; \boldsymbol{a})$. PLE requires that the recommended actions for any rating profile are the optimal actions that maximize agents' long-term utilities. Before we proceed to the proof of Theorem 1, we prove the following Lemma, whose proof is deferred to the Appendix.

Lemma 1:
1) $\forall \boldsymbol{\theta}$, the optimal action of agent i is either $\boldsymbol{a}_i^*(\boldsymbol{\theta}) = \mathbf{0}$ or $\boldsymbol{a}_i^*(\boldsymbol{\theta}) = \boldsymbol{\sigma}_i(\hat{\boldsymbol{\theta}}_i)$.
2) $\forall \theta_i$, if for $\hat{\boldsymbol{\theta}}_i = K$, $\boldsymbol{a}_i^*(\boldsymbol{\theta}) = \boldsymbol{\sigma}_i(\hat{\boldsymbol{\theta}}_i)$, then for any other $\hat{\boldsymbol{\theta}}_i$, $\boldsymbol{a}_i^*(\boldsymbol{\theta}) = \boldsymbol{\sigma}_i(\hat{\boldsymbol{\theta}}_i)$.
3) Let $\hat{\boldsymbol{\theta}}_i = \boldsymbol{K}$, suppose $\forall \theta_i$, $\boldsymbol{a}_i^*(\boldsymbol{\theta}) = \boldsymbol{\sigma}_i(\hat{\boldsymbol{\theta}}_i)$, then $\theta_i < \theta_i' \Leftrightarrow U_i^*(\theta_i, \hat{\boldsymbol{\theta}}_i) \leq U_i^*(\theta_i', \hat{\boldsymbol{\theta}}_i)$

Lemma 1.1 characterizes the set of possible optimal actions. That is, self-interested agents choose to either share nothing or the recommended amount of information/goods with their neighbors. Lemma 1.2 states that if an agent has an incentive to follow the recommended strategy when all its neighbors have the highest ratings, then it will also have an incentive to follow the recommended strategy in all other cases. Lemma 1.3 shows that the optimal long-term utility of an agent is monotonic in its rating when all its neighbors have the highest ratings—the higher the rating the larger the long-term utility the agent obtains. With these results in hand, we are ready to present and prove Theorem 1.

Theorem 1: Given the rating protocol structure and the network structure, at least one rating protocol can be constructed to be a PLE if and only if $\delta b_i(\hat{\boldsymbol{\sigma}}_i(K)) \geq c_i(\boldsymbol{\sigma}_i(\boldsymbol{K})), \forall i$.

Proof: (*Sketch*): For the sake of conciseness, we provide only the proof sketch of Theorem 1. The complete proof can be found in Appendix. According to Lemma 1.2, it suffices to ensure that agent i has an incentive to take the recommended strategy when it neighbors' ratings are $\hat{\boldsymbol{\theta}}_i = \boldsymbol{K}$. However, we need to prove that this holds for all ratings of agent i. To prove the "only if" part, we show that if $\delta b_i(\hat{\boldsymbol{\sigma}}_i) < c_i(\boldsymbol{\sigma}_i), \forall i$, then no rating protocol can constitute a PLE by showing a contradiction. To prove the "if" part, we construct a binary rating protocol that can constitute a PLE when $\delta b_i(\hat{\boldsymbol{\sigma}}_i) \geq c_i(\boldsymbol{\sigma}_i)$ is satisfied. In particular, we choose $\alpha_{i,2} = \beta_{i,1} = 1, \forall i$ as the rating update probabilities in such a rating protocol. ■

B. Computing the Recommended Strategy

Theorem 1 provides a sufficient and necessary condition for the existence of a PLE with respect to the recommended strategies when agents have the highest ratings. From (11) we already know that these strategies fully determine the social welfare that can be achieved by the rating protocol. Therefore, the optimal values of $\boldsymbol{\sigma}(\boldsymbol{K})$ can be determined by solving the following *optimal recommended strategy design* problem:

$$\begin{aligned} \underset{\boldsymbol{\sigma}}{\text{maximize}} \quad & \sum_i \left(b_i\left(\hat{\boldsymbol{\sigma}}_i(K)\right) - c_i\left(\boldsymbol{\sigma}_i(\boldsymbol{K})\right)\right) \\ \text{subject to} \quad & c_i\left(\boldsymbol{\sigma}_i(\boldsymbol{K})\right) \leq \delta b_i\left(\hat{\boldsymbol{\sigma}}_i(K)\right), \forall i \\ & \boldsymbol{\sigma} \geq 0 \end{aligned} \tag{13}$$

where the constraint ensures that an equilibrium rating protocol can be constructed. Note that this problem implicitly depends on the network since both $\hat{\boldsymbol{\sigma}}_i(K)$ and $\boldsymbol{\sigma}_i(\boldsymbol{K}), \forall i$ are network-dependent (since for each agent i, the strategy is only towards its neighbors). In this subsection, we will write $\boldsymbol{\sigma}_i(\boldsymbol{K})$ as $\boldsymbol{\sigma}_i$ and $\hat{\boldsymbol{\sigma}}_i(K)$ as $\hat{\boldsymbol{\sigma}}_i$ to keep the notation simple.

Firstly, we show the strong duality holds for the problem (13) under mild conditions.

Proposition 1: Strong duality holds for (13) if the following condition on the benefit function holds: $\forall i \in \mathcal{N}$

$$\max_j \left. \frac{\partial b_i(\hat{\boldsymbol{x}}_i)}{\partial x_{ji}} \right|_{\hat{\boldsymbol{x}}_i = \boldsymbol{0}} > \frac{d_i}{\delta} \tag{14}$$

Proof: It is easy to see that the problem in (13) is a convex optimization problem. According to the Slater's condition [25], strong duality holds if there exists a strictly feasible solution $\boldsymbol{\sigma}$ such that

$$\begin{aligned} & c_i\left(\boldsymbol{\sigma}_i(K)\right) < \delta b_i(\hat{\boldsymbol{\sigma}}_i), \forall i \\ & \boldsymbol{\sigma} \geq \boldsymbol{0} \end{aligned} \tag{15}$$

i.e. a solution such that the non-linear constraints are strictly satisfied.

Consider agent i, we find its neighbor j^* such that $j^* = \arg\max_j \left(\partial b_i(\hat{\boldsymbol{x}}_i)/\partial x_{ji}\right)|_{\hat{\boldsymbol{x}}_i=\boldsymbol{0}}$. We construct a strategy $\hat{\boldsymbol{\sigma}}_i$ such that $\sigma_{j^*i} = \epsilon$ and $\sigma_{ji} = 0, \forall j \neq j^*$. Because $(\partial b_i(\hat{\boldsymbol{x}}_i)/\partial x_{j^*i})|_{\hat{\boldsymbol{x}}_i=\boldsymbol{0}} > (d_i/\delta)$ and the $b_i(\hat{\boldsymbol{x}}_i)$ is concave, we can always find an $\bar{\epsilon} > \boldsymbol{0}$ such that $(\partial b_i(\hat{\boldsymbol{x}}_i)/\partial x_{j^*i})|_{\hat{\boldsymbol{x}}_i=\hat{\boldsymbol{\sigma}}_i} = (d_i/\delta)$. Hence, for any $\epsilon \in (0, \bar{\epsilon})$, $b_i(\hat{\boldsymbol{\sigma}}_i) > (d_i/\delta) \geq (c(\boldsymbol{\sigma}_i)/\delta)$. The last inequality is because the cost of agent i is at most $c_i(\boldsymbol{\sigma}_i) \leq d_i$. If we do this for all agents, then we find a strategy profile $\boldsymbol{\sigma}$ that is a strictly feasible solution. ■

The condition in the above proposition requires that each agent can obtain a sufficiently large marginal benefit at $\boldsymbol{0}$ from at least one of its neighbors. This is a mild condition and holds for numerous benefit functions such as the Dixit-Stiglitz utility function in (2). Moreover, (15) is rather conservative: in many problems, the right-hand side of (15) can be much smaller.

Now, we propose a distributed algorithm to compute these recommended strategies using the dual decomposition method [23], [25]. The idea is that we decompose the Optimal Recommended Strategy Design problem (13) into N sub-problems each of which is locally solved for each agent. Note that unlike the case with obedient agents, these sub-problems have coupled constraints. Therefore, the software of agents will need to go through an iterative process to exchange messages (i.e. the Lagrangian multipliers) with their neighbors such that their local solutions converge to the global optimal solution. We describe the algorithm in detail below.

We perform dual decomposition on (13) and form the partial Lagrangian,

$$\begin{aligned} L(\boldsymbol{\sigma}, \boldsymbol{\lambda}) &= \sum_i \left(b_i(\hat{\boldsymbol{\sigma}}_i) - c_i(\boldsymbol{\sigma}_i)\right) + \sum_i \lambda_i \left(c_i(\boldsymbol{\sigma}_i) - \delta b_i(\hat{\boldsymbol{\sigma}}_i)\right) \\ &= \sum_i \left[(1 + \lambda_i \delta) b_i(\hat{\boldsymbol{\sigma}}_i) - \sum_{j: g_{ij}=1} (1 + \lambda_j)\sigma_{ji}\right] \\ &\triangleq \sum_i L_i(\hat{\boldsymbol{\sigma}}_i, \boldsymbol{\lambda}) \end{aligned} \tag{16}$$

where $\lambda_i \geq 0$ is the Lagrange multiplier associated with the incentive constraint of agent i. The second equality is due to the linearity of the cost function. The master dual problem is,

$$\begin{aligned} \underset{\boldsymbol{\lambda}}{\text{minimize}} \quad & g(\boldsymbol{\lambda}) = \sum_i g_i(\boldsymbol{\lambda}) \\ \text{subject to} \quad & \lambda_i \geq 0, \forall i \end{aligned} \tag{17}$$

where $g(\boldsymbol{\lambda}) = \max_{\boldsymbol{\lambda}} L(\boldsymbol{\sigma}, \boldsymbol{\lambda})$. When strong duality holds, the optimal value $g^*(\boldsymbol{\lambda})$ equals the optimal value of the original primal problem (13). Next, we solve $g^*(\boldsymbol{\lambda})$ using the subgradient method. A subgradient of $-g$ is as follows: for λ_i, the subgradient is $c_i(\boldsymbol{\sigma}_i^*(\boldsymbol{\lambda})) - \delta b_i(\hat{\boldsymbol{\sigma}}_i^*(\boldsymbol{\lambda}))$. Therefore, we need to solve the optimal $\boldsymbol{\sigma}^*(\boldsymbol{\lambda})$ for a given $\boldsymbol{\lambda}$ to get the subgradient. Notice that the Lagragian $L(\boldsymbol{\sigma}, \boldsymbol{\lambda})$ can be separated into N sub-Lagrangians $L_i(\hat{\boldsymbol{\sigma}}_i, \boldsymbol{\lambda}_i)$, we can obtain $\hat{\boldsymbol{\sigma}}_i^*, \forall i$ by solving each subproblem individually,

$$\underset{\hat{\boldsymbol{\sigma}}_i}{\text{maximize}} \quad [1+\lambda_i\delta]b_i(\hat{\boldsymbol{\sigma}}_i) - \sum_{j:g_{ij}=1}(1+\lambda_j)\sigma_{ji} \tag{18}$$

The above problem is a convex optimization problem and hence is easy to solve. Now we have found the subgradient, that master algorithm updates the dual variable $\boldsymbol{\lambda}$ based on this subgradient,

$$\lambda_i(q+1)=\left[\lambda_i(q)+w\left(c_i\left(\boldsymbol{\sigma}_i^*\left(\boldsymbol{\lambda}(q)\right)\right)-\delta b_i\left(\hat{\boldsymbol{\sigma}}_i^*\left(\boldsymbol{\lambda}(q)\right)\right)\right)\right]^+, \forall i \tag{19}$$

where q is the iteration index, $w > 0$ is a sufficiently small positive step-size. Because (13) is a convex optimization, it is well known [25] that such an iterative algorithm will converge to the dual optimal $\boldsymbol{\lambda}^*$ as $q \to \infty$ and the primal variable $\boldsymbol{\sigma}^*(\boldsymbol{\lambda}(q))$ will also converge to the primal optimal $\boldsymbol{\sigma}^*$.

This iterative process can be made fully distributed which requires only limited message exchange between the software clients of neighboring agents. We present the Distributed Computation of the Recommended Strategy (DCRS) Algorithm below which is run locally by the software client of each agent.

Algorithm: Distributed Computation of the Recommended Strategy (DCRS)

(Run by the software client of agent i)

Input: Connectivity and utility function of agent i.

Output: $\boldsymbol{\sigma}_i(\boldsymbol{K}) = \{\sigma_{ij}(K)\}_{j:g_{ij}=1}$ (denoted by $\boldsymbol{\sigma}_i = \{\sigma_{ij}\}_{j:g_{ij}=1}$ for simplification)

Initialization:, $q = 0$; $\lambda_i(q) = 0$

Repeat:

Send $\lambda_i(q)$ to neighbor $j, \forall j : g_{ij} = 1$.

(Obtain $\lambda_j(q)$ from $j, \forall j$)

Solve (18) using $\lambda_i(q)$, $\{\lambda_j(q)\}_{j:g_{ij}=1}$ to obtain $\hat{\boldsymbol{\sigma}}_i(\boldsymbol{\lambda}(q))$.

Send $\sigma_{ji}(\boldsymbol{\lambda}(q))$ to neighbor $j, \forall j : g_{ij} = 1$.

(Obtain $\sigma_{ij}(\boldsymbol{\lambda}(q))$ from $j, \forall j$)

Update $\lambda_i(q+1)$ according to (19).

Stop until $\|\lambda_{ji}(q+1) - \lambda_{ji}(q)\|_2 < \varepsilon_\lambda$

The above DCRS algorithm has the following interpretation. In each configuration slot, the software client of each agent computes the sharing actions of the agent's neighbors that maximize the social surplus with respect to its own agent (i.e. the benefit obtained by its own agent minus the cost incurred by its neighbors). However, this computation has to take into account whether neighboring agents' incentive constraints are satisfied, which are reflected by the Lagrangian multipliers. The larger λ_i is, the more likely it is that agent i's incentive constraint is violated. Hence, the neighbors of agent i should acquire less information/goods from it. We note that the DCRS algorithm needs to be run to compute the optimal strategy only once at the beginning if the network is static.

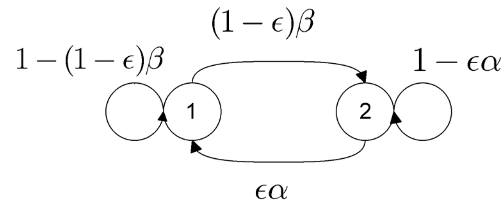


Fig. 2. Markov chain of the rating transition.

C. Computing the Remaining Components of the Rating Protocol

Even though the DCRS algorithm provides a distributed way to compute the recommended strategy when agents have the highest ratings, the other elements of the rating protocol remain to be determined. There are many possible rating protocols that can constitute a PLE given the obtained recommended strategies. In fact, we already provided one way to compute these remaining elements when we determined the sufficient condition in Theorem 1 by using a constructive method. However, this is not the most efficient design in the imperfect monitoring scenario where ratings will occasionally drop due to monitoring errors. Therefore, the remaining components of the rating protocol should still be smartly chosen in the presence of monitoring errors. In this subsection, we consider a rating protocol with a binary rating set $\Theta = \{1, 2\}$ and $\sigma_{ij}(\theta = 1) = 0, \forall i, j : g_{ij} = 1$. We design the rating update probabilities $\alpha_{i,2}, \beta_{i,1}, \forall i$ to maximize the social welfare when monitoring error exists.

Proposition 2: Given a binary rating protocol $\Theta = \{1, 2\}$, $\sigma_{ij}(2), \forall i, j : g_{ij} = 1$ determined by the DCRS algorithm and $\sigma_{ij}(1) = 0, \forall i, j : g_{ij} = 1$, when the monitoring error >0, the optimal rating update probability that maximize the social welfare is, $\forall i$, $\beta_{i,1}^* = 1$, $\alpha_{i,2}^* = c_i(\boldsymbol{\sigma}_i(\mathbf{2}))/\delta b_i(\hat{\boldsymbol{\sigma}}_i(2))$

Proof: The social welfare is the time-average sum utility of all agents. Therefore, we need to maximize the expected utility for each individual agent. Since we consider a binary rating protocol, let η_i^1, η_i^2 be the probability that agent i has rating 1 and rating 2, respectively. Note that $\eta_i^1 + \eta_i^2 = 1$. The expected time-average utility of agent i can be written as $\mathbb{E}V_i = \eta_i^1 u_i(1) + \eta_i^2 u_i(2)$. Since the utility of having a higher rating is larger than that of having a lower rating, $u_i(2) \geq u_i(1)$. Therefore, in order to maximize $\mathbb{E}V_i$, we need to maximize π_i^2. Given $\alpha_{i,2}, \beta_{i,1}$, we can determine η_i^2 by solving the stationary distribution of a two-state Markov chain. In this Markov chain, the states are the ratings and the transition probabilities are depicted in Fig. 2. A simple calculation of this Markov chain yields the solution of $\eta_i^2 = (1-)\beta_{i,1}/(\alpha_{i,2} + (1-)\beta_{i,1})$.

Now, in order to maximize η_i^2, it is equivalent to maximize $\beta_{i,1}/\alpha_{i,2}$. However, $\alpha_{i,2}$ and $\beta_{i,1}$ are subject to the incentive constraints and we can derive the feasible values of $\alpha_{i,2}, \beta_{i,1}, \forall i$ as follows,

$$\begin{aligned}\beta_{i,1} &\geq \frac{1-\delta}{\delta}\frac{c_i\left(\boldsymbol{\sigma}_i(\mathbf{2})\right)}{b_i\left(\hat{\boldsymbol{\sigma}}_i(2)\right) - c_i\left(\boldsymbol{\sigma}_i(\mathbf{2})\right)},\\ \alpha_{i,2} &\geq \frac{1-\delta(1-\beta_{i,1})}{\delta}\frac{c_i\left(\boldsymbol{\sigma}_i(\mathbf{2})\right)}{b_i\left(\hat{\boldsymbol{\sigma}}_i(2)\right)}\end{aligned} \tag{20}$$

For any $\beta_{i,1}$, the optimal value of $\alpha_{i,2}$ is the binding value of second inequality in (20) and hence, we need minimize $[1 - \delta(1-\beta_{i,1})]/\beta_{i,1}$. Because $[1 - \delta(1-\beta_{i,1})]/\beta_{i,1}$ is decreasing

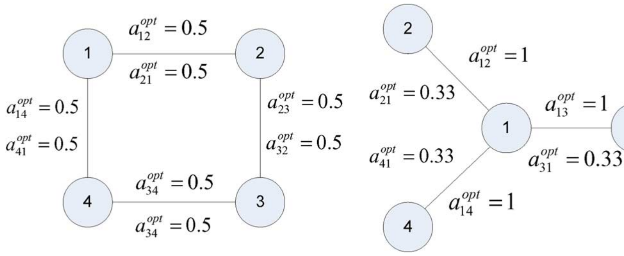


Fig. 3. Optimal strategies for obedient agents.

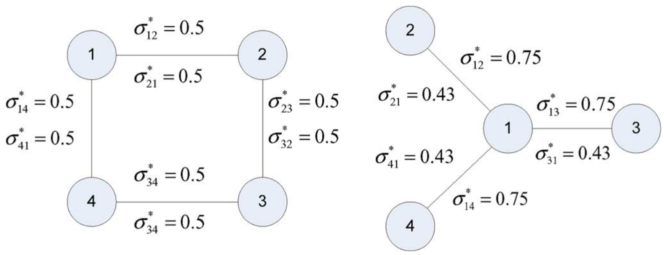


Fig. 4. Optimal strategies for strategic agents.

in $\beta_{i,1}$, the optimal value of $\beta_{i,1}$ is $\beta_{i,1}^* = 1$. Using (20) again, the optimal value of $\alpha_{i,2}^* = c_i(\boldsymbol{\sigma}_i(\mathbf{2}))/\delta b_i(\hat{\boldsymbol{\sigma}}_i(2))$. ■

It is worth noting that these probabilities can be computed locally by the software of the agents which do not require any information from other agents.

D. Illustrative Rating Protocols

In this section we show how the rating protocol can be determined in a distributed manner given the network structure. Specifically, we consider a set of 4 agents performing cooperative estimation (as in Section III-D) over two fixed networks—a ring and a star. A possible approximation of the utility function of each agent i when the uniform combination rule is used is $u_i(\boldsymbol{a}(t)) = [r^2-(r^2/(1+\sum_{j:g_{ij}} a_{ji}))]-\sum_{j:g_{ij}} a_{ij}$. We assume that the noise variance $r^2 = 4$. Fig. 3 illustrates the optimal actions in different networks by solving (5). In both networks, the optimal social welfare is $V^{opt} = 4$. Fig. 4 illustrates the optimal recommended strategies computed using the method developed in this section for these two topologies (assuming $\epsilon_i \to 0, \forall i$).

In the ring network, agents are homogeneous and links are symmetric. As we can see, the optimal recommended strategy $\boldsymbol{\sigma}^*$ is exactly the same as the socially optimal action profile $\boldsymbol{a}^{opt}$ for obedient agent case because $\boldsymbol{a}^{opt}$ already provides sufficient incentive for strategic agents to follow. Therefore, we can easily determine that $PoA = 1$. However, the strategic behavior of agents indeed degrades the social welfare in other cases, especially when the network becomes more heterogeneous and asymmetric, e.g. the star network. Even though taking $\boldsymbol{a}^{opt}$ maximizes the social welfare $V^{opt} = 4$ in the star network, these actions are not incentive-compatible for all agents. In particular, the maximum welfare $V^{opt} = 4$ is achieved by sacrificing the individual utility of the center agent (i.e. agent 1 needs to contribute much more than it obtains). However, when agents are strategic, the center agent will not follow these actions $\boldsymbol{a}^{opt}$ and hence, $V^{opt} = 4$ cannot be achieved. More problematically, since the center agent will choose not to participate in the sharing process, the periphery agents do not obtain benefits and hence, they will also choose not to participate in the sharing process. This leads to a network collapse. In the proposed rating protocol, the recommended strategies satisfy all agents' incentive constraints, namely $\delta b_i(\hat{\boldsymbol{\sigma}}_i(K)) \geq c_i(\boldsymbol{\sigma}_i(\boldsymbol{K})), \forall i$. By comparing $\boldsymbol{a}^{opt}$ and $\boldsymbol{\sigma}^*$, we can see that the rating protocol recommends more sharing from the periphery agents to the center agent and less sharing from the center agent to the periphery agents than the obedient agent case. In this way, the center agent will obtain sufficient benefits from participating in the sharing. However, due to this compensation for the center agent, the PoA is increased to $PoA = 1.036$.

V. Performance Analysis

In this section, we analyze the performance of the rating protocol and try to answer two important questions: (1) What is the performance loss induced by the strategic behavior of agents? (2) What is the performance improvement compared to other (simple) incentive mechanisms?

A. Price of Anarchy

Consider the social welfare maximization problems (5) and (13) for obedient agents and strategic agents (by using rating protocols), respectively. It is clear that the social welfare achieved by the rating system is always no larger than that obtained when agents are obedient due to the equilibrium constraint; hence, i.e. $PoA \geq 1$. The exact value of PoA will, in general, depend on the specific network structure (topology and individual utility functions). In this subsection, we identify a sufficient condition for the connectivity degree of the network such that PoA is one. To simplify the analysis, we assume that agents' benefit functions are homogeneous and depend only on the sum sharing action of the neighboring agents, i.e. $b_i(\hat{\boldsymbol{a}}_i) = b(\sum_{j:g_{ij}=1} a_{ji})$. Recall that $d_i = \sum_j g_{ij}$ is the number of neighbors of agent i. The degree of network G is defined as $d = \max_i d_i$.

Proposition 3: If the benefit function satisfies $b_i(\hat{\boldsymbol{a}}_i) = b(\sum_{j:g_{ij}=1} a_{ji}), \forall i$ and the sharing action is upper-bounded $a_{ij} \leq 1, \forall i, j$, then there exists a $\bar{d}$ such that if $d \leq \bar{d}$, $PoA = 1$.

Proof: Due to the concavity of the benefit function (Assumption 1), there exists d^* such that if $d > d^*$, $b(d) - d$ is increasing and if $d \leq d^*$, $b(d) - d$ is decreasing. If the connectivity degree satisfies $d < d^*$, then the optimal solution of (5) is $a_{ij} = 1, \forall i, j : g_{ij} = 1$. That is, optimality is achieved when all agents share the maximal amount of information/goods with all their neighbors. Therefore, $\forall d < d^*$, the agent i's benefit is $b(d_i)$ and its cost is d_i in the optimal solution. Moreover due to the concavity of the benefit function, there exists d^{**} such that if $d > d^{**}$, $\delta b(d) - d < 0$ and if $d \leq d^{**}$, $\delta b(d) - d \geq 0$. Therefore, if $d \leq d^{**}$, then agents' incentives are satisfied. Therefore if we let $\bar{d} = \min\{d^*, d^{**}\}$, then $\forall d < \bar{d}$, all agents have incentives to share the maximal amount of information/goods with their neighbors in which case the social optimum is also obtained. Hence, $PoA = 1$. ■

Proposition 3 states that when the connectivity degree is low, the proposed rating protocol can achieve the optimal performance even when agents are strategic.

B. Comparison With Direct Reciprocation

The proposed rating protocol is not the only incentive mechanism that can incentivize agents to share information/goods with other agents. A well-known direct reciprocation based incentive mechanism is the Tit-for-Tat strategy, which is widely adopted in many networking applications [7], [8]. The main feature of the Tit-for-Tat strategy is that it exploits the repeated *bilateral* interactions between connected agents, which can be utilized to incentivize agents to *directly* reciprocate to each other. However, when agents have asymmetric interests, such mechanisms fail to provide such incentives and direct reciprocity algorithms cannot be applied.

Moreover, even if we assume that interests are symmetric between agents, our proposed rating protocol is still guaranteed to outperform the Tit-for-Tat strategy when the utility function takes a concave form as assumed in this paper. Intuitively, because the marginal benefit from acquiring information/goods from one neighbor is decreasing in the total number of neighbors, agents become less incentivized to cooperate when their deviation towards some neighboring agent would not affect future information/goods acquisition from others, as is the case with the Tit-for-Tat strategy. In the following, we formally compare our proposed rating protocol with the Tit-for-Tat strategy. We assume that an agent i has two possible actions towards its neighboring agent j from: either no cooperation at all, or a fixed sharing action, i.e. $\{0, \bar{a}_{ij}\}$ where $\bar{a}_{ij} \in \mathbb{R}_+$. The Tit-for-Tat strategy prescribes the action for each agent i as follows, $\forall j : g_{ij} = 1$,

$$\begin{aligned} a_{ij}(0) &= \bar{a}_{ij} \\ a_{ij}(t+1) &= \begin{cases} \bar{a}_{ij}, & \text{if } a_{ji}(t) = \bar{a}_{ji} \\ 0, & \text{if } a_{ji}(t) = 0 \end{cases}, \quad \forall t \geq 0 \end{aligned} \tag{21}$$

Proposition 4: Given the network structure and the discount factor, any action profile $\bar{\boldsymbol{a}}$ that can be sustained by the Tit-for-Tat strategy can also be sustained by the rating protocol.

Proof: Consider the interactions between any pair of agents i, j. In the Tit-for-Tat strategy, the long-term utility of agent i by following the strategy when agent j played $\bar{a}_{ji}$ in the previous period is $U_i = (\tilde{b}_i(\bar{a}_{ji}) - \bar{a}_{ij})/(1-\delta)$ where $\tilde{b}_i(x) = b_i(\hat{a}_i | a_{ki} = \bar{a}_{ki}, a_{ji} = x)$. If agent i deviates in the current period, Tit-for-Tat induces a continuation history $(\{\bar{a}_{ij}, 0\}, \{0, \bar{a}_{ji}\}, \{\bar{a}_{ij}, 0\} \ldots)$ where the first components are agent i's actions and the second components is agent j's actions. The long-term utility of agent i by one-shot deviation is thus

$$\begin{aligned} U_i' &= \tilde{b}_i(\bar{a}_{ji}) + \delta \left[\frac{\tilde{b}_i(0) - \bar{a}_{ij}}{1-\delta^2} + \delta \frac{\tilde{b}_i(\bar{a}_{ji})}{1-\delta^2} \right] \\ &= \frac{\tilde{b}_i(\bar{a}_{ji})}{1-\delta^2} + \delta \frac{\tilde{b}_i(0) - \bar{a}_{ij}}{1-\delta^2} \end{aligned} \tag{22}$$

Incentive-compatibility requires that $U_i \geq U_i'$ and therefore

$$\delta \left(\tilde{b}_i(\bar{a}_{ji}) - \tilde{b}_i(0) \right) \geq \bar{a}_{ij} \tag{23}$$

Due to the concavity of the benefit function, it is easy to see that (23) leads to $\delta b_i(\hat{\boldsymbol{a}}_i) \geq c_i(\boldsymbol{a}_i)$ which is a sufficient condition for the rating protocol to be an equilibrium. ■

Proposition 4 proves that the social welfare achievable by the rating protocol equals or exceeds that of the Tit-for-Tat strategy, which confirms the intuitive argument before that diminishing marginal benefit from information/goods acquisition would result in less incentives to cooperate in an environment with only direct reciprocation than in one allowing indirect reciprocation. We note that different action profiles $\bar{\boldsymbol{a}}$ will generate different social welfare. However, computing the best $\bar{\boldsymbol{a}}$ among the incentive-compatible Tit-for-Tat strategies is often intractable since (23) is a non-convex constraint. Hence, implementing the best Tit-for-Tat strategy to maximize the social welfare is often intractable. In contrast, the proposed rating protocol does not have this problem since the equilibrium constraint established in Theorem 1 is convex and hence, the optimal recommended strategy can be solved in a distributed manner by the proposed DCRS algorithm.

VI. Dynamic Networks

In Section IV, we designed the optimal rating protocol by assuming that the network is static. In practice, the social network can also change over time due to, e.g., new agents entering the network and new links being created. Nevertheless, our framework can easily handle such growing networks by adopting a simple extension which refreshes the rating protocol (i.e. re-computes the recommended strategy, rating update rules and re-initializes the ratings of agents) with a certain probability each period. We call this probability the refreshing rate and denote it by $\rho \in [0, 1]$. When networks are dynamic, the refreshing rate will also be an important design parameter of the rating protocol.

A. Refreshing Rate Design Problem

Denote the network in period t by $G(t)$. We assume that in each period an expected number $n(t)$ of new agents enter the network and stay forever. Therefore, the network $G(t+1)$ will be formed based on $G(t)$ and the new agents. Note that before the next protocol refreshing, these new agents do not create benefits to or obtain benefits from their neighbors due to the incentive problem. Let $V^{opt}(G(T); \rho)$ be the optimal social welfare and $V^*(G(T); \rho)$ be the social welfare achieved by the rating protocol starting from a network G for a refreshing rate ρ. Our objective is to minimize the PoA by choosing a proper ρ. The optimal social welfare $V^{opt}(G(T); \rho)$ can be computed as follows,

$$V^{opt}(G(T); \rho) = \mathbb{E} \sum_{t=0}^{\infty} \rho(1-\rho)^t \frac{1}{t+1} \sum_{\tau=0}^{t} V^{opt}(G(T+\tau)) \tag{24}$$

Due to the refreshing, agents' discount factor effectively becomes $(1-\rho)\delta$. Therefore, the social welfare achieved by the rating protocol $V^*(G(T); \rho)$ can be obtained by solving the following optimization problem

$$\begin{aligned} \underset{\boldsymbol{\sigma}}{\text{maximize}} \quad & \sum_i \left(b_i \left(\hat{\boldsymbol{\sigma}}_i(K) \right) - c_i \left(\boldsymbol{\sigma}_i(K) \right) \right) \\ \text{subject to} \quad & c_i \left(\boldsymbol{\sigma}_i(K) \right) \leq (1-\rho)\delta b_i \left(\hat{\boldsymbol{\sigma}}_i(K) \right), \forall i \\ & \boldsymbol{\sigma} \geq 0 \end{aligned} \tag{25}$$

Formally, the refreshing rate design problem is formulated as the following optimization problem,

$$\begin{aligned} \underset{\rho}{\text{minimize}} \quad & PoA(\rho) \triangleq \frac{V^{opt}\left(G(T);\rho\right)}{V^{*}\left(G(T);\rho\right)} \\ \text{subject to} \quad & V^{opt}\left(G(T);\rho\right) \text{ is computed by } (24) \\ & V^{*}\left(G(T);\rho\right) \text{ is solved by } (25) \end{aligned} \tag{26}$$

B. Impact of the Refreshing Rate

In this subsection, we study the impact of ρ on $V^*(G(T);\rho)$ and $V^{opt}(G(T);\rho)$ separately and then provide guidelines on choosing the optimal ρ^* that minimizes $PoA(\rho)$.

Proposition 5: Both $V^*(G(T);\rho)$ and $V^{opt}(G(T);\rho)$ are non-increasing in ρ.

Proof: Since $V^*(G(T);\rho)$ is the optimal solution of (25), relaxing the constraints by decreasing ρ weakly increases $V^*(G(T);\rho)$. Therefore $V^*(G(T);\rho)$ is non-increasing in ρ.

It is easy to show that $V^{opt}(G(T+\tau))$ is non-decreasing in τ because we can let the new agents share nothing and the existing agents keep their previous strategies. Then according to (24) it is easy to see that $V^{opt}(G(T);\rho)$ is non-increasing in ρ. ■

Proposition 5 shows the monotonicity of $V^*(G(T);\rho)$ and $V^{opt}(G(T);\rho)$ with respect to ρ. If ρ is smaller, then there are more new entering agents and hence, the time-average optimal social welfare is larger. Moreover, since a smaller ρ means a more static rating protocol, the existing agents have more incentives to follow it.

Proposition 6: (1) $\lim_{\rho\to 1} PoA(\rho) \to \infty$. (2) If $\forall t_2 > t_1$, $V^{opt}(G(t_2)) - V^{opt}(G(t_1)) > \kappa > 0$, then $\lim_{\rho\to 0} PoA(\rho) \to \infty$. (3) If $\lim_{t\to\infty} V^{opt}(G(t)) - V^{opt}(G(T)) < \kappa$, then $\lim_{\kappa\to 0} \rho^* \to 0$.

Proof: (1) Because $V^{opt}(G(T);1) = V^{opt}(G(T)) > 0$ and $V^*(G(T);1) = 0$, $\lim_{\rho\to 0} PoA(\rho) \to \infty$. (2) Since in each time the increase of the optimal social welfare is at least a constant positive value, $\lim_{\rho\to 0} V^{opt}(G(T);\rho) \to \infty$. Because $V^*(G(T);0) = V^*(G(T)) > 0$, $\lim_{\rho\to 0} PoA(\rho) \to \infty$. (3) $\kappa \to 0$ implies that $\lim_{\rho\to 0} V^{opt}(G(T);\rho) \to V^{opt}(G(T))$. Since $V^*(G(T);\rho)$ is non-increasing in ρ, $PoA(\rho)$ is non-decreasing in ρ. Therefore $\lim_{\kappa\to 0} \rho^* \to 0$. ■

The first two parts of Proposition 6 reveals the impact of the refreshing rate on the PoA in two different ways. On one hand, a larger refreshing rate provides less incentives for agents to follow the current rating protocol designed in time T. One the other hand, a smaller refreshing rate leads to a worse adaptation of the rating protocol to the changing network. Therefore, the optimal refreshing probability ρ^* should be neither too larger nor too small. The third part states that if the speed of the optimal social welfare increase tends to 0 sufficiently quickly (e.g. the arrival rate of new agent is sufficiently smaller), then the optimal refreshing rate tends to be 0, i.e. the protocol is almost never refreshed. This is intuitive since if the network changes extremely slowly, then we almost do not need to refresh the rating protocol.

C. Exiting Agents

The proposed rating protocol with refreshing can also be applied to the general dynamic networks with both entering and exiting agents. However, when agents are exiting, unlike (25), the social welfare $V^*(G(T);\rho)$ that can be achieved by the rating protocol is difficult to characterize analytically. In particular, agents' incentives can be affected in different ways for different networks and $V^*(G(T);\rho)$ could be 0 in the worst case. Below we provide two examples that illustrate the different impacts.

1) Consider a star network with N periphery agents where at time T each periphery agent shares one unit of information/goods with the center agent and vice versa. The center agent's incentive constraint satisfies $c(N) \leq (1-\rho)\delta b(N)$. Suppose one periphery agent exits the network before the next refreshing update of the rating protocol. The center agent then receives one less unit of information/goods and needs to send one less unit of information/goods. If N is large, the incentive constraint of the center agent is still satisfied $c(N-1) \leq (1-\rho)\delta b(N-1)$ since the benefit function is concave. Because the center agent still has an incentive to follow the recommended strategy with respect to the remaining periphery agents, the remaining periphery agents' incentives to follow the recommended strategy are not affected. Therefore, the rating protocol works efficiently before the next refreshing update.
2) Consider a ring network where at time T each agent has the incentives to follow the recommended strategy which recommends sharing one unit of information/goods to its right-hand side neighbor. Each agent's incentive constraint satisfies $c(1) \leq (1-\rho)\delta b(1)p$. Suppose a single agent exits the network before the next refreshing update of the rating protocol. In this case, the incentive of its right-hand side neighbor to follow the recommended strategy is violated since all its benefit disappears. More problematically, this will cause a "chain effect" which leads top all remaining agents not sharing any information/goods with others. In such scenarios, the rating protocol fails to provide agents with sharing incentives.

From the above two examples, we see that it is significantly more difficult to understand the incentives of agents for the case with agents exiting since the game played by the agents may change in unpredictable ways. In this case, we may require other game theoretical concepts and tools to tackle this problem. One possible solution is making conjectures and using the notion of conjectural equilibrium [31] or using social learning [1]. We leave this as an interesting future research topic.

VII. Illustrative Results

In this section, we provide simulation results to illustrate the performance of the rating protocol. In all simulations, we consider the cooperative estimation problem introduced in Section III-A. Therefore, agents' utility function takes the form of $u_i(\boldsymbol{a}(t)) = [r^2 - MSE_i(\hat{\boldsymbol{a}}_i(t))] - \boldsymbol{a}_i(t)$ [28]. We will investigate different aspects of the rating protocol by varying the underlying topologies and the environment parameters.

A. Impact of Network Connectivity

Now we investigate in more detail how the agents' connectivity shapes their incentives and influences the resulting social

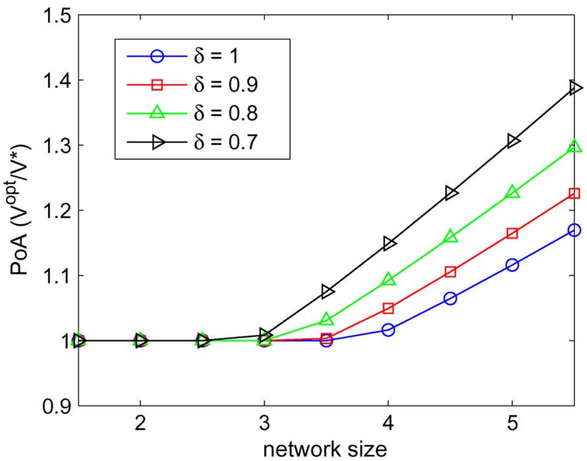


Fig. 5. Performance for different connectivity degrees d of star networks.

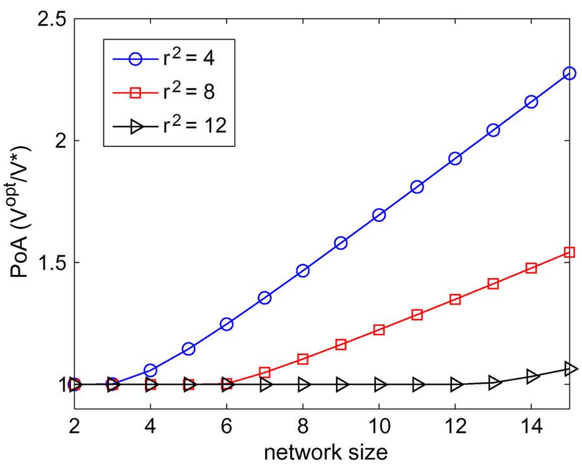


Fig. 6. Performance for different noise variance r^2.

welfare. In the first experiment, we consider the cooperative estimation over star topologies with different sizes (hence, different connectivity degrees). Fig. 5 shows the PoA achieved by the rating protocol for discount factors $\delta = 1, 0.9, 0.8, 0.7$ for the noise variance $r^2 = 8$. As predicted by Proposition 3, when the connectivity degree is small enough, the PoA equals one and hence, the performance gap is zero. As the network size increases (hence the connectivity degree increases in the star network), the socially optimal action requires the center agent to share more with the periphery agents. However, it becomes more difficult for the center agent to have incentives to do so since the sharing cost becomes much larger than the benefit. In order to provide sufficient incentives for the center agent to participate in the sharing process, the rating protocol recommends less sharing from the center agent to each periphery agent. However, incentives are provided at a cost of reduced social welfare. Fig. 5 also reveals that when agents' discount factor is lower (agents value less the future utility), incentives are more difficult to provide and hence, the PoA becomes higher. Since our applies to any benefit function that satisfies the Assumption, we show in Fig. 6 the PoA for different noise variances r^2 for discount factor $\delta = 0.9$. As we can see that the above analysis holds for other values of r^2. Moreover, as the noise variance increases, PoA is smaller for the same network size. This is because the benefit from cooperation increases and hence, agents are more likely to cooperate at the optimal level.

TABLE III
PERFORMANCE FOR VARIOUS d^{SF} IN SCALE-FREE NETWORKS

d^{SF}		$\epsilon = 0$	$\epsilon = 0.05$	$\epsilon = 0.1$
2	*Mean*	1.151	1.174	1.199
	Variance	5.9e-3	6.2e-3	6.4e-3
3	*Mean*	1.154	1.177	1.203
	Variance	8.6e-3	8.8e-3	9.2e-3
4	*Mean*	1.002	1.023	1.046
	Variance	3.1e-5	2.9e-5	2.7e-5
5	*Mean*	1.001	1.022	1.044
	Variance	1.6e-5	1.5e-5	1.4e-5
6	*Mean*	1.000	1.022	1.046
	Variance	~0	5.3e-7	2.5e-6

TABLE IV
PERFORMANCE FOR SCALE-FREE NETWORKS OF DIFFERENT SIZES

d^{SF}	N = 100	N = 200	N = 500
2	1.174	1.173	1.176
3	1.177	1.175	1.179
4	1.023	1.023	1.023
5	1.022	1.021	1.022
6	1.022	1.022	1.020

In the next simulation, we study scale-free networks in the imperfect monitoring scenarios. We used the standard Barbasi-Albert (BA) model to create the networks [27]. In scale-free networks, the number of neighboring agents is distributed as a power law (denote the power law parameter by d^{SF}). Table III shows the mean and variance of PoA achieved by the rating protocol developed for various values of d^{SF} and different monitoring error probabilities ϵ. The noise variance is set to be $r^2 = 4$ and the discount factor is $\delta = 0.8$. Each result is obtained by running 100 random trials. As we can see, the proposed rating protocol achieves close-to-optimal social welfare in all the simulated environments. In Table IV, we further show the achievable PoA by the proposed rating protocol for scale-free networks of different sizes when $\epsilon = 0.05$. Since the considered network is scale-free, the performance is similar for different network sizes.

B. Comparison With Tit-for-Tat

As mentioned in the analysis, incentive mechanisms based on direct reciprocation such as Tit-for-Tat do not work in networks lacking bilateral interests between connected agents and hence, reasons to mutually reciprocate. In this simulation, to make possible a direct comparison with the Tit-for-Tat strategy, we consider a scenario where the connected agents do have bilateral interests and show that the proposed rating protocol significantly outperforms the Tit-for-Tat strategy. In general, computing the optimal action profile $\bar{a}^*$ for the Tit-for-Tat strategy is difficult because it involves the non-convex constraint $\delta(b_i(\{\bar{a}^*_{ki}\}_{k:g_{ik}=1}) - b_i(\{\bar{a}^*_{ki}\}_{k\neq j:g_{ik}=1}, 0)) \geq \bar{a}^*_{ij}$, $\forall i, \forall j \neq i : g_{ij} = 1$; such a difficulty is not presented in our proposed rating protocol because the constraints in our formulated problem are convex. For tractability, here we consider a symmetric and homogeneous network to enable the computation of the optimal action for the Tit-for-Tat strategy. We consider a number $N = 100$ of agents and that the number of neighbors of each agent is the same $d_i = d, \forall i$ and each agent adopts a symmetric action profile $\bar{a}_{ij} = \bar{a}, \forall i, j$. The

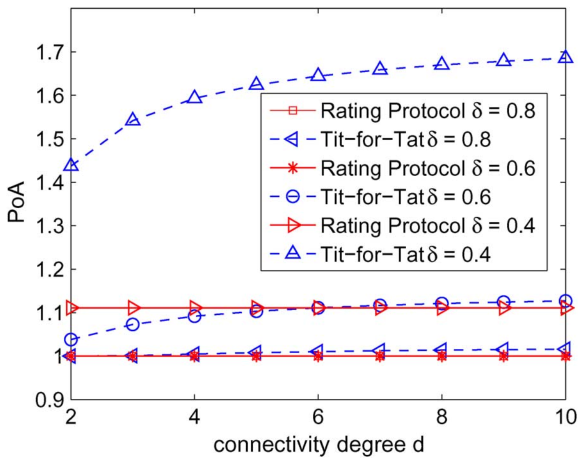


Fig. 7. Performance comparison with Tit-for-Tat.

TABLE V
PoA OF RATING PROTOCOLS WITH DIFFERENT REFRESHING RATES

ρ	0.005	0.02	0.04	0.06	0.08	0.10	0.12	0.14
PoA	1.35	1.20	1.18	1.22	1.25	1.29	1.34	1.41

noise variance is set to be $r^2 = 4$ in this simulation. Fig. 7 illustrates the PoA achieved by the proposed rating protocol and the Tit-for-Tat strategy. As predicted by Proposition 4, any action profile that can be sustained by the Tit-for-Tat strategy can also be sustained by the proposed rating protocol (for the same δ). Hence, the rating protocol yields at least as much social welfare as the Tit-for-Tat strategy (for the same δ). As the discount factor becomes smaller, agents' incentives to cooperate become less and hence, the PoA is larger. Note that for $\delta = 0.6$, 0.8, our rating protocol achieves $PoA = 1$ for all connectivity degrees.

C. *Rating Protocol With Refreshing*

Finally, we consider the optimal choice of the rating protocol refreshing rate ρ when the network is growing as considered in Section VI. In this simulation, the network starts with $N = 50$ agents. In each period, a new agent enters the network with probability 0.1 and stays in the network forever. Any two agents are connected with *a priori* probability 0.2. We vary the refreshing rate from 0.005 to 0.14. Table V records the PoA achieved the rating protocol with refreshing for $\delta = 0.4$. It shows that the optimal refreshing rate needs to be carefully chosen. If ρ is too large, the incentives for agents to cooperate is small hence, the incentive-compatible rating protocol achieves less social welfare. If ρ is too small, the rating protocol is not able to adapt to the changing network well. This introduces more social welfare loss in the long-term as well. The optimal refreshing rate in the simulated network is around 0.04.

VIII. CONCLUSIONS

In this paper, we provided a framework for designing incentives protocols (based on ratings) aimed at maximizing the social welfare of strategic agents which are repeatedly sharing information/goods across a network. Our rating protocols can be implemented in a distributed and informationally decentralized manner and achieve much higher social welfare than existing incentive mechanisms. Our framework and analysis can also be used to provide guidelines for designing and planning social, economic and engineering networks of strategic agents, such that the social welfare of such networks is maximized. The proposed ratings framework can also be used to design protocols for a wide range of engineering networks where strategic agents interact—communications networks, power networks, transportation networks, and computer networks.

APPENDIX A
PROOF OF LEMMA 1

(1) Consider any action $\boldsymbol{a}_i(\boldsymbol{\theta}) \neq \boldsymbol{\sigma}_i(\hat{\boldsymbol{\theta}}_i)$. According to the rating update rule, $p(\boldsymbol{\theta}'|\boldsymbol{\theta}, \boldsymbol{a}_i(\boldsymbol{\theta})) = p(\boldsymbol{\theta}'|\boldsymbol{\theta}, \mathbf{0})$. Since $u_i(\boldsymbol{\theta}, \mathbf{0}) > u_i(\boldsymbol{\theta}, \boldsymbol{a}_i(\boldsymbol{\theta}))$, we can see that $U_i(\boldsymbol{\theta}, \mathbf{0}) > U_i(\boldsymbol{\theta}, \boldsymbol{a}_i(\boldsymbol{\theta})))$. Therefore, there are only two possible actions that can potentially maximize the long-term utility.

(2) According to part (1), there are only two possible actions that can be optimal. First, we note that the continuation utility difference by choosing these two actions is

$$\delta \sum_{\boldsymbol{\theta}'} p\left(\boldsymbol{\theta}'|\boldsymbol{\theta}, \boldsymbol{\sigma}_i(\hat{\boldsymbol{\theta}}_i)\right) U_i^*(\hat{\boldsymbol{\theta}}_i) - \delta \sum_{\boldsymbol{\theta}'} p(\boldsymbol{\theta}'|\boldsymbol{\theta}, \mathbf{0}) U_i^*(\hat{\boldsymbol{\theta}}_i) \quad (27)$$

which is independent of other agents' ratings $\hat{\boldsymbol{\theta}}_i$ when we consider agent i's one-shot unilateral deviation. This is because the benefit that an agent can potentially receive only depends on its own rating while the cost that the agent incurs depends only on its neighbors' ratings. The benefit is determined by agent i's current action since different actions lead to different transitions of only agent i's own rating. The costs are cancelled out because the neighbors' ratings are independent on agent i's actions.

It is obvious that the current period utility different satisfies,

$$\begin{aligned} &u_i\left((\theta_i, \boldsymbol{K}), \mathbf{0}\right) - u_i\left((\theta_i, \boldsymbol{K}), \boldsymbol{\sigma}_i(\hat{\boldsymbol{\theta}}_i)\right) \\ &\qquad \geq u_i\left((\theta_i, \boldsymbol{\theta}_{-i}), \mathbf{0}\right) - u_i\left((\theta_i, \boldsymbol{\theta}_{-i}), \boldsymbol{\sigma}_i(\hat{\boldsymbol{\theta}}_i)\right), \forall \theta_{-i} \end{aligned} \quad (28)$$

If for $\boldsymbol{\theta}_{-i} = \boldsymbol{K}$, the optimal action of agent i is $\boldsymbol{a}_i^* = \boldsymbol{\sigma}_i(\hat{\boldsymbol{\theta}}_i)$, then the following holds,

$$\begin{aligned} &u_i\left((\theta_i, \boldsymbol{K}), \mathbf{0}\right) - u_i\left((\theta_i, \boldsymbol{K}), \boldsymbol{\sigma}_i(\hat{\boldsymbol{\theta}}_i)\right) \\ &\leq \delta \sum_{\boldsymbol{\theta}'} p\left(\boldsymbol{\theta}'|\boldsymbol{\theta}, \boldsymbol{\sigma}_i(\hat{\boldsymbol{\theta}}_i)\right) U_i^*(\hat{\boldsymbol{\theta}}_i) - \delta \sum_{\boldsymbol{\theta}'} p(\boldsymbol{\theta}'|\boldsymbol{\theta}, \mathbf{0}) U_i^*(\hat{\boldsymbol{\theta}}_i) \end{aligned} \quad (29)$$

which means that the following is also true,

$$\begin{aligned} &u_i\left((\theta_i, \boldsymbol{\theta}_{-i}), \mathbf{0}\right) - u_i\left((\theta_i, \boldsymbol{\theta}_{-i}), \boldsymbol{\sigma}_i(\hat{\boldsymbol{\theta}}_i)\right), \forall \theta_{-i} \\ &\leq \delta \sum_{\boldsymbol{\theta}'} p\left(\boldsymbol{\theta}'|\boldsymbol{\theta}, \boldsymbol{\sigma}_i(\hat{\boldsymbol{\theta}}_i)\right) U_i^*(\hat{\boldsymbol{\theta}}_i) - \delta \sum_{\boldsymbol{\theta}'} p(\boldsymbol{\theta}'|\boldsymbol{\theta}, \mathbf{0}) U_i^*(\hat{\boldsymbol{\theta}}_i) \end{aligned} \quad (30)$$

Therefore, for any other $\boldsymbol{\theta}_{-i}$, the optimal action of agent i is also $\boldsymbol{a}_i^* = \boldsymbol{\sigma}_i(\hat{\boldsymbol{\theta}}_i)$.

(3) To simplify notations, we suppress $\boldsymbol{K}$ in the utility and simply write $U_i^*(\theta_i)$ instead of $U_i^*(\theta_i, \hat{\boldsymbol{\theta}}_i)$. We also write $\beta_{i,k}$ as β_k. The value functions can be obtained by solving the following recursive equations,

$$\begin{aligned} U_i^*(K) &= u(K, \boldsymbol{\sigma}_i) + \delta U_i^*(K) \\ U_i^*(K-1) &= u(K-1, \boldsymbol{\sigma}_i) \\ &\quad + \delta\left(\beta_{K-1} U_i^*(K) + (1-\beta_{K-1}) U_i^*(K-1)\right) \\ \ldots U_i^*(1) &= u(1, \boldsymbol{\sigma}_i) + \delta\left(\beta_1 U_i^*(2) + (1-\beta_1) U_i^*(1)\right) \end{aligned} \quad (31)$$

We prove by induction. Suppose $U_i^*(l) \geq U_i^*(l-1), \forall l : K \geq l \geq k+1$. We need to show that $U_i^*(k) \geq U_i^*(k-1)$. The value functions of level k and $k-1$ are

$$\begin{aligned} U_i^*(k) &= u_i(k, \boldsymbol{\sigma}_i) + \delta\left(\beta_k U_i^*(k+1) + (1-\beta_k) U_i^*(k)\right) \\ U_i^*(k-1) &= u_i(k-1, \boldsymbol{\sigma}_i) \\ &\quad + \delta\left(\beta_{k-1} U_i^*(k) + (1-\beta_k) U_i^*(k-1)\right) \end{aligned} \tag{32}$$

To prove $U_i^*(k) \geq U_i^*(k-1)$, we use contradiction. Suppose $U_i^*(k) < U_i^*(k-1)$, then

$$\begin{aligned} U_i^*(k) &\geq u_i(k, \boldsymbol{\sigma}_i) + \delta U_i^*(k) \\ U_i^*(k-1) &< u_i(k-1, \boldsymbol{\sigma}_i) + \delta U_i^*(k-1) \end{aligned} \tag{33}$$

This leads to $u_i(k, \boldsymbol{\sigma}_i) < u_i(k-1, \boldsymbol{\sigma}_i)$ which is a contradiction. Hence, it only remains to prove $U_i^*(K) \geq U_i^*(K-1)$. This can be easily shown by computing $U_i^*(K) - U_i^*(K-1)$, i.e.

$$U_i^*(K) - U_i^*(K-1) = \frac{u_i(K, \boldsymbol{\sigma}_i) - u_i(K-1, \boldsymbol{\sigma}_i)}{1-\delta(1-\beta_{K-1})} > 0 \tag{34}$$

This completes the proof.

Appendix B
Proof of Theorem 1

According to Lemma 1, it suffices to ensure that agent i has an incentive to take the recommended strategy when it neighbors' ratings are $\hat{\theta}_i = \boldsymbol{K}$. However, we need to prove that this holds for all ratings of agent i. Therefore, we suppress $\hat{\theta}_i = \boldsymbol{K}$ and only write out θ_i whenever it is clear.

We prove the "only if" part first. We need to show that for all rating protocol that is an equilibrium, $\delta b_i(\hat{\boldsymbol{\sigma}}_i(K)) \geq c(\boldsymbol{\sigma}(\boldsymbol{K})), \forall i$ must be satisfied. Consider any rating level k of agent i, following the recommended strategy gives it the following long-term utility,

$$U_i(k, \boldsymbol{\sigma}) = u_i(k, \boldsymbol{\sigma}) + \delta\left(\beta_k U_i^*(k+1) + (1-\beta_k) U_i^*(k)\right) \tag{35}$$

Deviating to $\mathbf{0}$ gives the following long-term utility,

$$U_i(k, \mathbf{0}) = u_i(k, \mathbf{0}) + \delta\left(\alpha_k U_i^*(k-1) + (1-\alpha_k) U_i^*(k)\right) \tag{36}$$

Equilibrium requires that $U_i(k, \boldsymbol{\sigma}) \geq U_i(k, \mathbf{0})$. Therefore, the following must hold,

$$\begin{aligned} u_i(k, \mathbf{0}) - u_i(k, \boldsymbol{\sigma}) \leq \delta\,[&\beta_k U_i^*(k+1) + (1-\beta_k) U_i^*(k) \\ &-\alpha_k U_i^*(k-1) - (1-\alpha_k) U_i^*(k)] \end{aligned} \tag{37}$$

According to Lemma 1.3, $U_i^*(K) \geq U_i^*(k), \forall k$ in an equilibrium. Therefore, the following must hold,

$$u_i(k, \mathbf{0}) - u_i(k, \boldsymbol{\sigma}) \leq \delta U_i^*(K) \tag{38}$$

The left-hand side is $u_i(k, \mathbf{0}) - u_i(k, \boldsymbol{\sigma}) = c(\boldsymbol{\sigma}_i)$. Using the recursive equation of the optimal long-term utilities (32), we can compute the right-hand side as

$$U_i^*(K) = \frac{1}{1-\delta} u_i(1, \boldsymbol{\sigma}_i) = \frac{1}{1-\delta}\left(b_i\left(\hat{\boldsymbol{\sigma}}_i(1)\right) - c_i\left(\boldsymbol{\sigma}_i(\boldsymbol{K})\right)\right). \tag{39}$$

Substituting this into (38), we can obtain the desired result after simple manipulations.

Next, we prove the "if" part by constructing a binary rating protocol. According to the one-shot deviation principle, for agent i to follow the recommended strategy at $\theta_i = 2$, we need

$$u_i(2, \mathbf{0}) - u_i(2, \boldsymbol{\sigma}_i) \leq \delta \alpha_2 \left(U_i^*(2) - U_i^*(1)\right) \tag{40}$$

for agent i to follow the recommended strategy at $\theta_i = 1$, we need

$$u_i(1, \mathbf{0}) - u_i(1, \boldsymbol{\sigma}_i) \leq \delta \beta_1 \left(U_i^*(2) - U_i^*(1)\right) \tag{41}$$

Using the value function (32), we can compute $U_i^*(2) - U_i^*(1)$ which is

$$U_i^*(2) - U_i^*(1) = \frac{u_i(2, \boldsymbol{\sigma}_i) - u_i(1, \boldsymbol{\sigma}_i)}{1-\delta(1-\beta_1)} \tag{42}$$

Moreover, $u_i(2, \mathbf{0}) - u_i(2, \boldsymbol{\sigma}_i) = u_i(1, \mathbf{0}) - u_i(1, \boldsymbol{\sigma}_i) = c_i(\boldsymbol{\sigma}_i)$. For the rating protocol to be an equilibrium, we need to choose α_2, β_1 such that

$$\begin{aligned} c_i(\boldsymbol{\sigma}_i) &\leq \{\alpha_2, \beta_1\} \frac{u_i(2, \boldsymbol{\sigma}_i) - u_i(1, \boldsymbol{\sigma}_i)}{1-\delta(1-\beta_1)} \\ &= \{\alpha_2, \beta_1\} \frac{\delta b_i(\hat{\boldsymbol{\sigma}}_i(2)}{1-\delta(1-\beta_1)} \end{aligned} \tag{43}$$

If we choose $\alpha_2 = \beta_1 = 1$, then the above inequality holds. This means that such a binary rating protocol is a PLE.

Appendix C
Discussion on the DCRS Algorithm

When developing the DCRS algorithm, we used the widely-adopted dual decomposition method. However, there are several significant differences from existing problems.

First, in most existing problems [23], [25], the constraint in the optimization problem comes from the system resource constraints. Our problem is not a NUM problem since we do not have such resource constraints. Instead, the constraints are derived based on the incentive-compatibility of agents, i.e. the incentive condition under which the agents follow the recommended strategy. More specifically, they are derived in Theorem 1 (in the revised manuscript).

Second, in many standard dual decomposition problems [23], [25], the objective functions are directly separable, in the sense that an agent's utility depends on its own action. The coupling among agents only comes from the optimization constraints. For example, the objective function can have the form $\sum_i f_i(\boldsymbol{x}_i)$ where $\boldsymbol{x}_i$ is agent i's action and $f_i(\boldsymbol{x}_i)$ is its utility. The actions of all agents need to satisfy some resource constraints $\sum_i h_i(\boldsymbol{x}_i) \leq 0$. In our problem, an agent's utility depends not only on its own action but also on the neighboring agents' actions, i.e. $\sum_i (b_i(\hat{\boldsymbol{\sigma}}_i) - c_i(\boldsymbol{\sigma}_i))$ where $\boldsymbol{\sigma}_i$ is agent i's strategy and $\hat{\boldsymbol{\sigma}}_i$ is agent i's neighbors strategies towards agent i.

Third, even though dual decomposition allows distributed implementation, in many existing works [23], [25], agents still need to exchange messages with all other agents (e.g. by broadcasting). This requires intensive message exchanges among agents if broadcasting is not available and is even impossible if agents' interactions are subject to underlying topologies. However, our solution enables a completely distributed architecture and message exchange only occurs between connected agents.

ACKNOWLEDGMENT

The authors would like to thank Prof. William Zame and Simpson Zhang from the Dept. of Economics, University of California, Los Angeles (UCLA) for helpful discussions. The authors also would like to thank the associate editor and three anonymous reviewers for valuable comments that helped improve this paper.

REFERENCES

[1] V. Krishnamurthy and H. V. Poor, "Social learning and Bayesian games in multiagent signal processing: How do local and global decision makers interact?," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 43–57, May 2013.

[2] R. Lucky, "Tragedy of the commons," *IEEE Spectrum*, vol. 43, no. 1, p. 88, Jan. 2006.

[3] A. Galeotti, S. Goyal, M. Jackson, and F. Vega-Redondo, "Network games," *Rev. Econ. Stud.*, vol. 77, no. 1, pp. 218–244, 2010.

[4] J. Park and M. van der Schaar, "A game theoretic analysis of incentives in content production and sharing over peer-to-peer networks," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 4, pp. 704–717, Aug. 2010.

[5] D. Bergemann and D. Ozman, "Optimal pricing with recommender systems," in *Proc. 7th ACM Conf. Elect. Commerce*, 2006, pp. 43–51.

[6] J. K. MacKie-Mason and H. R. Varian, "Pricing congestible network resources," *IEEE J. Sel. Areas Commun.*, vol. 13, no. 7, pp. 1141–1149, Jul. 1995.

[7] R. Axelrod, "The emergence of cooperation among egoists," *Amer. Pol. Sci. Rev.*, vol. 75, no. 2, pp. 306–318, 1981.

[8] F. Milan, J. J. Jaramillo, and R. Srikant, "Achieving cooperation in multihop wireless networks of selfish nodes," in *Proc. ACM Workshop Game Theory for Commun. Netw.*, 2006.

[9] C. Song and Q. Zhang, "Achieving cooperative spectrum sensing in wireless cognitive radio networks," *ACM SIGMOBILE Mobile Comput. Commun. Rev.*, vol. 13, no. 2, pp. 14–25, 2009.

[10] M. Kandori, "Social norms and community enforcement," *Rev. Econ. Stud.*, vol. 59, pp. 63–80, 1992.

[11] S. N. Ali and D. A. Miller, *Enforcing Cooperation in Networked Soc.*, working paper 2013.

[12] M. Jackson, T. Rodriguez-Barraquer, and X. Tan, "Social capital and social quilts: Networks patterns of favor exchange," *Amer. Econ. Rev.*, vol. 102, no. 5, pp. 1857–1897, 2012.

[13] S. D. Kamvar, M. T. Schlosser, and H. Garcia-Molina, "The eigentrust algorithm for reputation management in P2P networks," in *Proc. 12th Int. Conf. World Wide Web*, 2003, pp. 640–651.

[14] S. Ba and P. Pavlou, Evidence of the Effect of Trust Building Technology in Electronic Markets: Price Premiums and Buyer Behavior [Online]. Available: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=951734 2006

[15] P. Resnick and R. Zeckhaagent, "Trust among strangers in internet transactions: Empirical analysis of eBay's reputation system," *Adv. Appl. Microeconom.*, pp. 127–157, 2002.

[16] C. Dellarocas, "Reputation mechanism design in online trading environments with pure moral hazard," *Inf. Syst. Res.*, vol. 16, no. 2, pp. 209–230, 2005.

[17] M. Fan, Y. Tan, and A. B. Whinston, "Evaluation and design of online cooperative feedback mechanism for reputation management," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 2, pp. 244–254, 2005.

[18] G. Zacharia, A. Moukas, and P. Maes, "Collaborative reputation mechanism in electronic marketplaces," *Decision Support Syst.*, vol. 29, no. 4, pp. 371–388, 2000.

[19] Y. Zhang, J. Park, and M. van der Schaar, "Rating protocols for online communities," *ACM Trans. Econ. Comput.*, vol. 2, no. 1, pp. 4:1–34, 2013.

[20] J. Xu and M. van der Schaar, "Social norm design for information exchange systems with limited observations," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 11, pp. 2126–2135, Nov. 2012.

[21] M. van der Schaar, J. Xu, and W. Zame, "Efficient online exchange via fiat money," *Econ. Theory*, vol. 5, no. 2, pp. 211–248, Oct. 2013.

[22] J. Xu, W. Zame, and M. van der Schaar, "Token economy for online exchange systems," in *Proc. Int. Conf. Auton. Agents Multiagent Syst. (AAMAS)*, 2012.

[23] D. P. Palomar and M. Chiang, "Alternative distributed algorithms for network utility maximization: Framework and applications," *IEEE Trans. Autom. Control*, vol. 52, no. 12, pp. 2254–2269, Dec. 2007.

[24] G. J. Mailath and L. Samuelson, *Repeated Games and Reputations: Long-Run Relationships*. Oxford, CA, USA: Oxford Univ. Press, 2006.

[25] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. press, 2004.

[26] A. Dixit and J. Stiglitz, "Monopolistic competition and optimum product diversity," *Amer. Econ. Rev.*, vol. 67, no. 3, pp. 297–308, 1977.

[27] A. Albert and A.-L. Barabasi, "Statistical mechanics of complex networks," *Rev. Modern Phys.*, vol. 74, no. 1, p. 47, 2002.

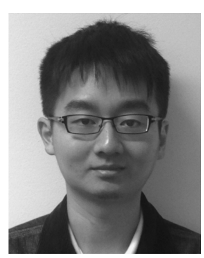
[28] C. K. Yu, M. van der Schaar, and A. H. Sayed, "Distributed spectrum sensing in the presence of selfish users," in *Proc. CAMSAP '13*, 2013.

[29] J. Xu, Y. Zhang, and M. van der Schaar, "Rating systems for enhanced cyber-security investments," in *Proc. ICASSP '13*, pp. 2915–2919.

[30] M. Elliott and B. Golub, "A network approach to public goods," in *Proc. ACM Conf. Electron. Commerce*, 2013.

[31] 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, Sep. 2009.

Jie Xu received the B.S. and M.S. degrees in electronic engineering from Tsinghua University, Beijing, China, in 2008 and 2010, respectively. He is currently a Ph.D. student with the Electrical Engineering Department, UCLA. His primary research interests include incentive design and learning in multi-agent networks.



Yangbo Song received the B.A. in economics and finance from the University of Hong Kong in 2010. He is currently a Ph.D. student with the Economics Department, UCLA. His research interests include network economics, contract theory and mechanism design.

Mihaela van der Schaar (F'10) is currently a Chancellors Professor in the Electrical Engineering Department at UCLA. Her research interests include communications, engineering economics and game theory, strategic design, online reputation and social media, dynamic multi-user networks, and system designs.