Proceedings of the
47th IEEE Conference on Decision and Control
Cancun, Mexico, Dec. 9-11, 2008

Efficient Network Formation by Distributed Reinforcement

Georgios C. Chasparis and Jeff S. Shamma

Abstract— **We consider the problem of efficient network formation in a distributed fashion. Network formation is modeled as an evolutionary process, where agents can form and sever unidirectional links and derive direct and indirect benefits from these links. We formulate and analyze an evolutionary model in which each agent's choices depend on its own previous links and benefits, and link selections are subject to random perturbations. Agents reinforce the establishment of a link if it was beneficial in the past, and suppress it otherwise. We illustrate the flexibility of the model to incorporate various design criteria, including dynamic cost functions that reflect link establishment and maintenance, and distance-dependent benefit functions. We show that the evolutionary process assigns positive probability to the emergence of multiple stable configurations (i.e., strict Nash networks), which need not emerge under alternative processes such as best-reply dynamics. We analyze the specific case of so-called frictionless benefit flow, and show that a single agent can reinforce the emergence of an efficient network through an enhanced evolutionary process known as dynamic reinforcement.**

I. Introduction

Several studies in social networks concern how the emergence of specific network formations is associated with the strategic framework of local interactions [2]. Likewise, a challenge in sensor networks is to design protocols that guarantee energy efficient network formation, where the energy of transmitting signals is the major part of the energy consumption [3], [4]. In this paper, we wish to provide a dynamic framework that will serve both as a design procedure for distributed convergence to a desirable network and as a justification for the emergence of certain networks.

Several models for social network formation have been proposed that are based on game theory. These include *static models*, [5], [6], [7], [8], [9], [10], where agents play a one-stage game, with actions corresponding to network links. These studies characterize networks in terms of the Nash equilibria of the associated game. The processes under which such equilibria can emerge are proposed via *dynamic* or *evolutionary* models [11], [12], [13], [14], [15], [16], [17]. In these models, players adaptively form and sever links in reaction to an evolving network, and in some models, their decisions are subject to random perturbations.

For the sake of brevity, all proofs have been omitted and can be found in [1].

This work was supported by AFOSR grants #FA9550-05-1-0239, #FA9550-08-1-0375 and NSF grant #ECS-0501394.

G.C. Chasparis is a post-doctoral scholar in the Department of Mechanical and Aerospace Engineering, University of California Los Angeles, Los Angeles, CA 90095. E-mail: gchas@seas.ucla.edu. URL: http://www.seas.ucla.edu/~gchas.

J.S. Shamma is with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332. E-mail: shamma@gatech.edu. URL: http://www.prism.gatech.edu/~jshamma3.

This paper is also concerned with dynamic or evolutionary models, and is mostly related to the papers of [9], [13], [16], [17]. In particular, we consider self-interested players that have the discretion of establishing or severing unidirectional links with neighboring agents based on myopic considerations. However, we drop the typical assumption that players are aware of the current network structure, usually employed by processes such as best-reply. Rather, our model is *payoff based*. Agents can only measure derived benefits from past decisions of forming or severing links. Players will reinforce a link if it was beneficial in the past and suppress it otherwise. These dynamics belong to the general class of learning automata [18], [19] and are motivated by related models of human-like decision making [20].

The main difference with both [16] and [17] is in the reward function. In [16], [17] the reward function is based on the principle of reciprocity which models social relationships such as friendship. Here instead we use the *connections model* of [7]. According to this model, the benefits received from each agent can be viewed as the information available from its direct and indirect links. In other words, agents are rewarded for being connected to other agents, either directly or indirectly. Additional features of our model are a *state-dependent* cost function for the establishment and maintenance of links and a *distance-dependent* reward function for information benefits. This framework can model various economic and social contexts, such as the production and transmission of information, consumer products, etc. Models of this form (that also assume the consent of both parties) include the static model of [7] and the dynamic models of [11], [12].

We will show that our model (i.e., the combined evolutionary process, reward functions, and cost functions) assigns positive probability to the emergence of multiple stable configurations (Nash networks). When the aforementioned *state-dependent* cost function is considered, we show that the set of strict Nash networks emerging may be larger than the one arising from best-reply dynamics considered in [13].

A specific case of our reward functions is "frictionless information flow", i.e., where benefits are derived from being connected to other agents and are not distance dependent. For this special case, we demonstrate the utility of an enhanced evolutionary process known as *dynamic reinforcement*. In particular, we will show how a single individual can reinforce the emergence of an efficient network through a simple "dynamic" processing of its own available information that uses the *rate* of observed reward changes [21]. This has the effect of reinforcing efficient networks while destabilizing the non-efficient networks.

978-1-4244-3124-3/08/$25.00 ©2008 IEEE

The remainder of the paper is organized as follows. Section II presents the model for network formation. Section III analyzes its asymptotic stability properties. Section IV specifies the possible steady-state configurations, distinguishing between the cases of frictionless and decaying flow of benefits. Section V analyzes the effect of the enhanced dynamic reinforcement scheme on the emergence of efficient network configurations. Finally, Section VI presents concluding remarks.

II. The Model

A. One-way benefit flow

Let $\mathcal{I} = \{1, ..., n\}$ denote a finite set of agents. The network relations among agents are represented by a graph, whose nodes are identified with the agents and whose edges capture the pairwise relations.

We will consider a *one-way* flow model, where a *network* $\mathcal{G}$ is defined as a collection of pairwise directed links, (i, s), $i, s \in \mathcal{I}$, where benefits flow from s to i. More precisely, $\mathcal{G} \subseteq \{(i, s) : i, s \in \mathcal{I}\}$. For example, the network $\mathcal{G} = \{(1, 2), (1, 3), (3, 1)\}$ is illustrated in Fig. 1.

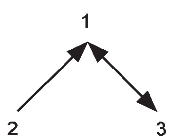

Fig. 1. A network of three players and one-way flow of benefits.

Define a *path* from s to i in $\mathcal{G}$, as $(i \leftarrow s) = \bigcup_{k=0}^{m-1}(s_{k+1}, s_k)$ for some positive integer, m, where $\{s_k\}_{k=0}^{m}$ is a sequence in $\mathcal{I}$ that satisfies $s_0 = s$, $s_m = i$, $s_k \neq s_{k+1}$ and $(s_{k+1}, s_k) \in \mathcal{G}$ for any $k = 0, 1, ..., m-1$.

Definition 2.1 (Connectivity): A node $i \in \mathcal{I}$ is connected to a node $s \in \mathcal{I}\backslash i$ if there is a path from s to i. A network is connected if any $i \in \mathcal{I}$ is connected to any $s \in \mathcal{I}\backslash i$.

We further assume that each agent is able to establish links only with "neighboring agents." The set of neighbors of agent i is denoted as $\mathcal{N}_i$ with cardinality $|\mathcal{N}_i|$. In the unconstrained neighbors case, $\mathcal{N}_i = \mathcal{I}\backslash i$.

B. The network formation model

We will model network formation as an evolutionary process, where at each stage agents decide which links to form. Based on agents' decisions, a graph is being formed, and a reward is assigned to each agent based on the information it receives through its links and its neighbors' links. In detail, the network formation model is described as follows.

1) Action space: The set of actions of agent i, denoted $\mathcal{A}_i$, contains all the possible combinations of neighbors with which a link can be established including the case of not establishing any link, i.e., $\mathcal{A}_i = 2^{\mathcal{N}_i}$.

By enumerating the elements in $\mathcal{A}_i$, we can associate the j^{th} element of $\mathcal{A}_i$ with a vertex, e_j, of the probability simplex of dimension $|\mathcal{A}_i|$, i.e., $\Delta(|\mathcal{A}_i|)$. Accordingly, we will use the same notation, $a_i \in \mathcal{A}_i$, to refer to an element of $\mathcal{A}_i$ either in terms of an index over $\mathcal{A}_i$, a vertex of $\Delta(|\mathcal{A}_i|)$, or an element of $2^{\mathcal{N}_i}$. Finally, let $|a_i|$ denote the cardinality of a_i viewed as an element of $2^{\mathcal{N}_i}$.

2) Learning Algorithm: At each stage $k \in \mathbb{N}$, each agent i selects an action $a_i(k) \in \mathcal{A}_i$ according to the probability distribution over $\mathcal{A}_i$

$$(1-\lambda)x_i(k) + \frac{\lambda}{|\mathcal{A}_i|}\mathbf{1},$$

where i) $x_i(k) \in \Delta(|\mathcal{A}_i|)$ is the *strategy* of agent i at stage k; ii) $\mathbf{1}$ is a vector of appropriate dimension with each element equal to 1; and iii) $\lambda \geq 0$ is a parameter used to model possible perturbations in the decision making process, also called *mutations* [22], [23].

The strategy of agent i is updated according to:

$$x_i(k+1) = x_i(k) + \epsilon(k) \cdot (R_i(a(k)) - C_i(a_i(k), x_i(k))) \cdot (a_i(k) - x_i(k)), \quad (1)$$

which is a stochastic recursion with step size sequence $\epsilon(k) \triangleq 1/(k+1)$.

In the above recursion, $R_i : \mathcal{A} \to \mathbb{R}_+$ denotes the reward of agent i, which generally depends on the choices of all agents $a = (a_1, ..., a_n)$ (i.e., the current network) defined in the product set $\mathcal{A} \triangleq \times_{i \in \mathcal{I}} \mathcal{A}_i$.

We will assume that the rewards are bounded and nonnegative, i.e., $0 \leq R_i(\cdot) < R_{\max} < \infty$, for some $R_{\max} > 0$.

We also assume that the establishment and maintenance of a link is costly. In recursion (1), $C_i : \mathcal{A}_i \times \Delta(|\mathcal{A}_i|) \to \mathbb{R}_+$ denotes the cost of establishing and maintaining a link. This cost is assumed to depend on both the current and previously established links. Dependence of previously established links is implicit through the strategy $x_i(k)$.

The recursion (1) is a modified version of the linear reward-inaction scheme that was first considered in mathematical psychology by [24] and introduced in engineering by [25]. The main difference here is the introduction of the mutation parameter, λ, which will be essential for the equilibrium selection analysis under dynamic reinforcement models defined in Section V.

3) Learning algorithm with dynamic reinforcement: Further insights into the possible emergence of efficient network structures can be derived by considering a dynamic processing of the local information available to each agent. In particular, agents might be more satisfied with links that increased their benefits in the *recent* history than with links that have provided large benefits throughout the *whole* history.

Based on similar reasoning, we will utilize a modified action selection probability distribution of the form

$$\Pi_{\Delta}\{(1-\lambda)[x_i(k) + \gamma_i \cdot (x_i(k) - y_i(k))] + \frac{\lambda}{|\mathcal{A}_i|}\mathbf{1}\}, \quad (2)$$

for some $\gamma_i \geq 0$, where $\Pi_{\Delta}\{\cdot\}$ is the projection to the probability simplex, $\Delta(|\mathcal{A}_i|)$, and the new state variable, y_i, is updated according to the recursion

$$y_i(k+1) = y_i(k) + \epsilon(k) \cdot (x_i(k) - y_i(k)).$$

A standard controls interpretation of this dynamic reinforcement scheme is that agents use it to "predict" more rewarding outcomes [21]. A similar approach was investigated by [26] as an approach to enable stabilization of mixed equilibria in learning in games. The intention here instead is to use dynamic reinforcement to enforce convergence to an efficient pure equilibrium.

C. Reward and cost function

The reward function will be defined as in the network formation models of [7], [11], [13]. More specifically, we assume that a link with another agent allows access to the benefits available to the latter via its own links. Define:

$$R_i(a) \triangleq \sum_{s\in\mathcal{I}, s\neq i} \delta^{d_{is}(a)} \tag{3}$$

where i) $\delta \in (0,1]$ measures the level of information decay and ii) $d_{ij} : \mathcal{A} \to \mathbb{N}$ is defined as the minimum distance from j to i given the current action profile $a \in \mathcal{A}$. We adopt the convention that $d_{ij}(\cdot) = \infty$, when $(i,j) \notin \mathcal{G}$.

For each agent $i \in \mathcal{I}$, we define the cost function $C_i : \mathcal{A}_i \times \Delta(|\mathcal{A}_i|) \to \mathbb{R}_+$ to be:

$$C_i(a_i, x_i) \triangleq \kappa_0 \, |a_i| + \kappa_1 \varphi_i(a_i)^{\mathrm{T}}(\mathbf{1} - \varphi_i(x_i)), \tag{4}$$

for some $\kappa_0, \kappa_1 \geq 0$. The parameter κ_0 corresponds to the cost of maintaining an existing link, while κ_1 corresponds to the cost of establishing a new link. The function $\varphi_i : \Delta(|\mathcal{A}_i|) \to \mathbb{R}^{|\mathcal{N}_i|}$ is defined by

$$[\varphi_i(x_i)]_j = \sum_{\{\alpha\in\mathcal{A}_i : j\in\alpha\}} x_{i\alpha}.$$

By abuse of notation, we are using α as both an index, as in $x_{i\alpha}$, and a set, as in $j \in \alpha \in 2^{\mathcal{N}_i}$. In words, the $[\varphi_i(x_i)]_j$ denotes the probability that agent i will form a link to neighbor j based on the distribution x_i. The term $\varphi_i(a_i)^{\mathrm{T}}(\mathbf{1} - \varphi_i(x_i))$ penalizes misalignment of the action a_i with the distribution x_i. In the perfectly aligned case, for any $a_i \in \mathcal{A}_i$ (viewed as a vertex of $\Delta(|\mathcal{A}_i|)$), $\varphi_i(a_i)^{\mathrm{T}}(\mathbf{1} - \varphi_i(a_i)) = 0$ whereas in the worst case,

$$\max_{x_i} \varphi_i(a_i)^{\mathrm{T}}(\mathbf{1} - \varphi_i(x_i)) = |a_i| \,.$$

We make the following assumptions for the ***remainder of the paper***:

Assumption 2.1: $0 \leq \kappa_0 + \kappa_1 < \delta$.
This assumption assures that agents always have an incentive to form at least one link.

Assumption 2.2: The neighbor sets $\{\mathcal{N}_1, \mathcal{N}_2, ..., \mathcal{N}_n\}$ are such that a connected network is feasible.

D. Efficiency

We also need to characterize the *efficiency* of a network structure. To this end, we borrow the definition of the *value* of a network from [7].

First, define the agent utility function $v_i : \mathcal{A} \times \Delta(\mathcal{A}_i) \to \mathbb{R}_+$ as

$$v_i(a, x_i) = R_i(a) - C_i(a_i, x_i), \tag{5}$$

i.e., the combined reward minus cost in the update equation (1). Note that unlike typical utility functions in network formation games, this utility function depends explicitly on both collective actions, a, and an agent's strategy, x_i. In the special case where $x_i = a_i$, the cost term only reflects maintenance costs (i.e., the κ_0 term), whereas establishment costs (the κ_1 term) are zero.

Definition 2.2 (Network value): The value of the network $V : \mathcal{A} \to \mathbb{R}_+$, is the sum of agent rewards minus maintenance costs at an action profile, $a \in \mathcal{A}$, i.e.,

$$V(a) = \sum_{i\in\mathcal{I}} v_i(a, a_i). \tag{6}$$

Definition 2.3 (Efficient network): An efficient network is a joint action profile $a \in \mathcal{A}$ with the maximum value.

Claim 2.1: An efficient network is connected. In the special case of $\delta = 1$, an efficient network is a connected network with a minimal number of links.

III. Stability Analysis

A. Asymptotic stability analysis

We formulate the stochastic recursion (1) so that the ODE method for stochastic approximations [27] can be applied. In order to characterize the stability properties of the stochastic iteration (1), we rewrite it as

$$x_i(k+1) = x_i(k) + \epsilon(k) \cdot (\bar{g}_i(x(k)) + \xi_i(k)), \tag{7}$$

where the original observation sequence has been decomposed into a deterministic sequence $\bar{g}_i : \times_{i\in\mathcal{I}}\Delta(|\mathcal{A}_i|) \to \Delta(|\mathcal{A}_i|)$ such that

$$\bar{g}_i(x) = E\{(R_i(a) - C_i(a, x_i))(a_i - x_i)\big|x\},$$

and a zero-mean noise sequence

$$\xi_i = (R_i(a) - C_i(a_i, x_i))(a_i - x_i) - \bar{g}_i(x).$$

A more compact form of (7) is

$$x(k+1) = x(k) + \epsilon(k) \cdot (\bar{g}(x(k)) + \xi(k)), \tag{8}$$

where $\bar{g}(\cdot) \triangleq \mathrm{col}\{\bar{g}_i(\cdot)\}_{i\in\mathcal{I}}$, $\xi(\cdot) \triangleq \mathrm{col}\{\xi_i(\cdot)\}_{i\in\mathcal{I}}$, and $\mathrm{col}\{\cdot\}$ denotes the column vector.

Proposition 3.1 (Convergence): For $\lambda > 0$, the stochastic iteration (8) is such that the sequence $\{x(k)\}$ converges to an invariant set of the ODE

$$\dot{x} = \bar{g}(x). \tag{9}$$

Furthermore, let $A \subset \times_{i\in\mathcal{I}}\Delta(|\mathcal{A}_i|)$ be a locally asymptotically stable set in the sense of Lyapunov for (9). Then $\mathrm{Prob}\{\lim_{k\to\infty} x(k) \in A\} > 0$.

Thus, according to Proposition 3.1, the sequence will converge to A with some positive probability.

The characterization of *nonconvergence* properties of the stochastic iteration (8) about a stationary point is also of interest.

Proposition 3.2 (Nonconvergence): For $\lambda > 0$, the stochastic iteration (8) is such that, if x^* is a linearly unstable stationary point of the ODE (9), then $\mathrm{Prob}\{\lim_{k\to\infty} x(k) = x^*\} = 0$.

B. Stationary points

It has been shown by Proposition 3.4 in [21] that for $\lambda = 0$, any pure strategy profile $a^* = (a_1^*, ..., a_n^*)$ is a stationary point of the stochastic iteration (8).

Moreover, by Proposition 3.5 of [21], for sufficiently small $\lambda > 0$, there exists a unique continuously differentiable function $w^* : \mathbb{R}_+ \to \times_i \mathbb{R}^{|\mathcal{A}_i|}$, such that $\lim_{\lambda\to 0} \lambda w^*(\lambda) = 0$, and

$$x^* = a^* + \lambda w^*(\lambda) \tag{10}$$

is a stationary point of the ODE (9).

C. Local asymptotic stability (LAS)

In order to characterize locally the stability properties of the stationary points of ODE (9), we define the *conditional expected utility* $\overline{v}_i(a_i, x)$ of agent i as[1]

$$\overline{v}_i(a_i, x) = E\{v_i(a, x_i)\big|a_i, x_{-i}\},$$

where $v_i(\cdot,\cdot)$ is defined in (5).

Proposition 3.3 (LAS of Standard Reinforcement): For sufficiently small $\lambda > 0$, let x^* be a stationary point of the ODE (9) corresponding to some $a^* \in \mathcal{A}$ according to (10). The stationary point x^* is a locally asymptotically stable point of the ODE (9) for sufficiently small $\lambda > 0$ if and only if, for each $i \in \mathcal{I}$,

$$\overline{v}_i(a_i^*, x^*) > \overline{v}_i(a_i', x^*) \tag{11}$$

for all $a_i' \in \mathcal{A}_i\backslash a_i^*$.

In the case of the dynamic reinforcement scheme of (2), the relevant ODE is now

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} \bar{g}(x, y) \\ x - y \end{pmatrix}, \tag{12}$$

and the condition for stability takes the following form:

Proposition 3.4 (LAS of Dynamic Reinforcement): Assume the hypotheses of Proposition 3.3 under stability condition (11). Assume that each agent i applies dynamic reinforcement (2) for some $\gamma_i \geq 0$. The strategy profile x^* is a locally asymptotically stable stationary point of the ODE (12) for sufficiently small $\lambda > 0$ if and only if, for each agent $i \in \mathcal{I}$, the derivative feedback coefficient satisfies

$$0 \leq \gamma_i < \frac{\overline{v}_i(a_i^*, x^*) - \overline{v}_i(a_i', x^*) + 1}{\overline{v}_i(a_i', x^*)} \tag{13}$$

for all $a_i' \in \mathcal{A}_i\backslash a_i^*$.

IV. Nash Networks

In the literature of network formation, Nash equilibria are usually called *Nash networks*, [13]. In the framework of our network formation model, where decisions are state-dependent, we define:

Definition 4.1 (Nash network): An action profile $a^* \in \mathcal{A}$ is a Nash network if and only if

$$v_i((a_i^*, a_{-i}^*), a_i^*) \geq v_i((a_i', a_{-i}^*), a_i^*), \tag{14}$$

for all $a_i' \in \mathcal{A}_i\backslash a_i^*$ and $i \in \mathcal{I}$. Likewise, a strict Nash network satisfies the strict inequality in (14).

Claim 4.1: Nash networks are connected.

The Nash networks for $n = 3$ agents and no decay are shown in Fig. 2. Assuming that $\kappa_0 > 0$ and $\kappa_1 = 0$, Fig. 2(a)

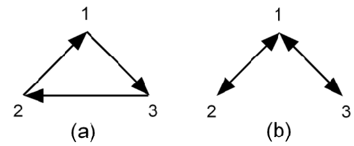

Fig. 2. Nash networks in case of $n = 3$ agents and $\kappa_0 > 0$, $\kappa_1 \geq 0$ and $\kappa_0 + \kappa_1 < 1$.

corresponds to a strict Nash network, since each agent can only be worse off by unilaterally changing its links. Fig. 2(b), instead, corresponds to a non-strict Nash network since no player can increase its payoff by unilaterally deviating.

However, in case $\kappa_1 > 0$, both Nash networks in Fig. 2 are strict, since each deviation from the equilibrium play is charged by an extra cost of order κ_1.

According to the definition of a Nash network and local stability analysis of Proposition 3.3, we conclude that:

Proposition 4.1: Under the hypotheses of Proposition 3.3, a stationary point $x^* = a^* + \lambda w^*(\lambda)$, such that a^* is a strict Nash network, is a locally asymptotically stable point of the ODE (9) for sufficiently small $\lambda > 0$.

Therefore, finding the set of strict Nash networks a^* reveals the set of stationary points x^* that are locally stable.

Note that according to Propositions 3.1–3.2, convergence to non-strict Nash network need not be excluded. Figs. 3–4, simulate two characteristic responses of the stochastic recursion (1) where we consider the following action spaces $\mathcal{A}_1 = \{\varnothing, \{2\}, \{3\}, \{2,3\}\}$, $\mathcal{A}_2 = \{\varnothing, \{1\}, \{3\}, \{1,3\}\}$, $\mathcal{A}_3 = \{\varnothing, \{1\}, \{2\}, \{1,2\}\}$, denoted by $\mathcal{A}_i = \{\mathtt{A}, \mathtt{B}, \mathtt{C}, \mathtt{D}\}$, $i = 1, 2, 3$. In Fig. 3, the recursion converges to the efficient formation of Fig. 2(a), while in Fig. 4 the recursion converges to the non-efficient formation of Fig. 2(b).

A. Frictionless benefit flow ($\delta = 1$)

In order to characterize the Nash networks in the general case of $n > 3$ agents, we need to define a general class of networks called *critically linked networks*.

Definition 4.2 (Critically linked network): A network, $\mathcal{G}$, is critically linked if i) it is connected and ii) for all $(i, j) \in \mathcal{G}$, the unique path $(i \leftarrow j)$ is (i, j).

In words, a critically linked network is such that if agent i drops a direct link to (neighboring) agent j, then i loses connectivity to j by any means.

Proposition 4.2 (Nash networks): For $\delta = 1$, $n > 2$, and $\kappa_0, \kappa_1 > 0$, a network is a strict Nash network if and only if it is a critically linked network.

A special class of critically linked networks are so-called *flower networks*, defined in [13]. Contrary to best-reply dynamics, where *not* all flower networks are Nash networks [13], here we see that, due to the dynamic establishment cost function, *all* flower networks are Nash networks.

[1] The notation $-i$ denotes the complementary set $\mathcal{I}\backslash i$. We will often split a strategy profile x as $x = (x_i, x_{-i})$.

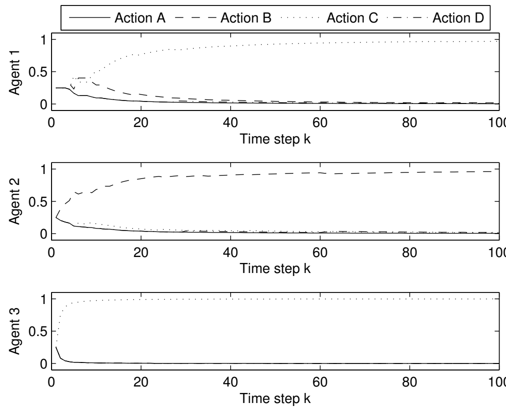

Fig. 3. A typical response of the stochastic recursion (1), for $\delta = 1$, $\kappa = 1/2$, $\kappa_1 = 0$, $\lambda = 0.01$. Convergence to the efficient formation of Fig. 2(a) is observed.

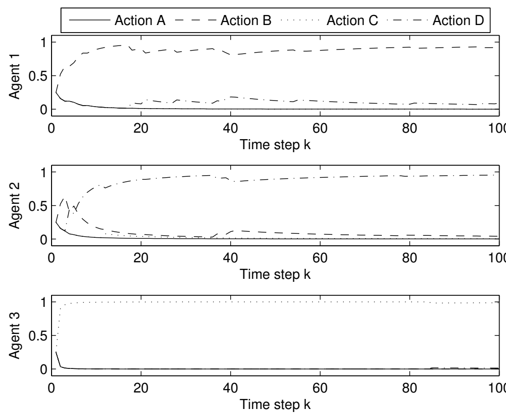

Fig. 4. A typical response of the stochastic recursion (1), for $\delta = 1$, $\kappa = 1/2$, $\kappa_1 = 0$, $\lambda = 0.01$. Convergence to the non-efficient formation of Fig. 2(b) is observed.

In the special case of both i) no establishment cost ($\kappa_1 = 0$) and ii) unconstrained neighbors ($\mathcal{N}_i = \mathcal{I}\backslash i$), we can have a more explicit characterization of strict Nash networks. We first define the following.

Definition 4.3 (Wheel network): A wheel network is a connected network uniquely defined by a path $(i \leftarrow i)$ for some $i \in \mathcal{I}$ where every agent in $\mathcal{I}$ is visited only once.

Proposition 4.3: For $\delta = 1$, $n > 2$, and $1 > \kappa_0 > \kappa_1 = 0$, the wheel network is the unique strict Nash network.

B. Decaying benefit flow ($\delta < 1$)

When the information flow is also subject to decay, the Nash equilibrium condition imposes a structural constraint on the distances between nodes.

Proposition 4.4 (Nash networks with decay): Let $0 < \delta < 1$, $n > 2$, $\kappa_0 > 0$, and $\kappa_1 \geq 0$. Let $\mathcal{G}$ be a Nash network corresponding to the joint action $a \in \mathcal{A}$. For any agent i, if $|a_i| < |\mathcal{N}_i|$, then $\delta - \delta^{d_{ij}(a)} \leq \kappa_0 + \kappa_1$ for all $j \in \mathcal{N}_i$.

The condition $|a_i| < |\mathcal{N}_i|$ means that agent i is not using all of its available links. This could be of interest, for example, in the unconstrained neighbors case with a large number of agents. This theorem can be used to bound distances to neighbors since

$$d_{ij}(a) \leq \left\lfloor \frac{\log(\delta - (\kappa_0 + \kappa_1))}{\log(\delta)} \right\rfloor \triangleq d_{\max}.$$

V. Dynamic Reinforcement

In [21] it was shown that a dynamic reinforcement scheme of the form of (2) can destabilize all non-efficient Nash equilibria in pure coordination games. We wish to answer the question of whether such a reinforcement scheme can be used for distributed reinforcement to desirable networks.

Intuitively, the dynamic reinforcement scheme of (2) reinforces *changes* in strategy. Following the language of Propositions 3.3–3.4, let x^* be a joint equilibrium strategy associated with some joint action $a^* \in \mathcal{A}$ for sufficiently small λ. Dynamic reinforcement effectively skews the perceived payoff benefits of unilateral action deviations. For example, suppose a_i' is an alternative action for agent i. Under dynamic reinforcement, the perceived benefit of a deviation is

$$(1 + \gamma_i)\overline{v}_i(a_i', x^*) - (\overline{v}_i(a_i^*, x^*) + 1),$$

as opposed to the actual benefit in the absence of dynamic reinforcement, which is $\overline{v}_i(a_i', x^*) - \overline{v}_i(a_i^*, x^*)$. If a^* corresponds to a Nash equilibrium strategy, the actual deviation benefit will be negative for all alternatives, a_i'. Under dynamic reinforcement, the perceived benefit can be positive and induce a departure for that agent from the action a_i^*. This departure can, in turn, induce other agents to abandon their Nash equilibrium actions.

On a cautionary note, excessive dynamic reinforcement can induce deviations from all Nash equilibria. The key to evoking efficient outcomes lies in finding the correct level of dynamic reinforcement, as measure by the coefficients γ_i, to induce deviations from inefficient equilibria while maintaining stability of efficient equilibria. The following claims explicitly carry out this procedure for a special case of network formation.

Claim 5.1: Assume that for each $i \in \mathcal{I}$, $\mathcal{N}_i = \mathcal{I}\backslash i$. Let $\delta = 1$, $n > 2$, and $\kappa_0, \kappa_1 \geq 0$. Let x^{non} be a stationary point corresponding to a non-efficient Nash network configuration, a^{non}, according to (10) for sufficiently small $\lambda > 0$. There exists an agent i and constant $\gamma^{\text{non}} = (1 + \kappa_1)/((n-1) - (\kappa_0 + \kappa_1)) > 0$ such that if agent i applies the dynamic reinforcement scheme of (2) with coefficient $\gamma_i > \gamma^{\text{non}}$ then the non-efficient equilibrium formation, x^{non}, is linearly unstable point for (12).

Claim 5.1 shows that there exists an agent who is able to destabilize a non-efficient network. The process could get attracted to another steady-state configuration that is not efficient. However, if each agent $i \in \mathcal{I}$ applies derivative action with $\gamma_i > \gamma^{\text{non}}$, then all non-efficient networks will be linearly unstable.

The following claim computes an upper bound on the γ_i so that stability of the efficient (wheel) network is maintained.

Claim 5.2: Assume that for each $i \in \mathcal{I}$, $\mathcal{N}_i = \mathcal{I}\backslash i$. Let $\delta = 1$, $n > 2$, and $\kappa_0, \kappa_1 \geq 0$. Let x^{eff} be a stationary

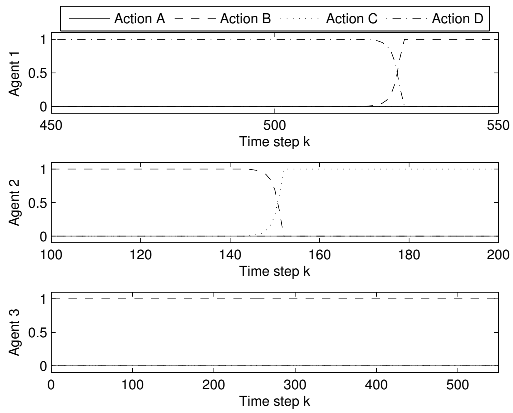

Fig. 5. A typical response of the stochastic iteration (1), for $\delta = 1$, $\kappa_0 = 1/2$, $\kappa_1 = 0$, $\lambda = 0.01$ when all agents apply dynamic reinforcement with $\gamma = 1$ and for an initial condition that is close to the non-efficient formation of Fig. 2(b).

point corresponding to the efficient Nash network wheel configuration, a^{eff}, according to (10) for sufficiently small $\lambda > 0$. There exists a $\gamma^{\text{eff}} > \gamma^{\text{non}}$ such that if any agent i applies the dynamic reinforcement scheme of (2) with coefficient

$$0 < \gamma_i < \frac{1 + \kappa_0 + \kappa_1}{((n-1) - 2\kappa_0 - \kappa_1)} + O(\lambda) \triangleq \gamma^{\text{eff}},$$

then x^{eff} is a locally stable equilibrium for (12).

Based on the previous claims, we are ready to describe convergence and non-convergence properties of the dynamic reinforcement scheme in the case of frictionless benefit flow.

Proposition 5.1: In the framework of Claims 5.1–5.2, if $\gamma_i \in [\gamma^{\text{non}}, \gamma^{\text{eff}})$ for all $i \in \mathcal{I}$, then $\text{Prob}\{\lim_{k\to\infty} x(k) = x^{\text{eff}}\} > 0$, and $\text{Prob}\{\lim_{k\to\infty} x(k) = x^{\text{non}}\} = 0$.

For example, let us consider the case of $n = 3$ agents and $\kappa_0 = 1/2$, $\kappa_1 = 0$, $\lambda = 0.01$ and $\delta = 1$. According to Claims 5.1–5.2, $\gamma^{\text{non}} = 2/3$ and $\gamma^{\text{eff}} = 3/2 + O(\lambda)$. In Fig. 5 we have simulated the stochastic recursion (1) with initial conditions that are close to the non-efficient network of Fig. 2(b) when all agents apply the dynamic reinforcement scheme of (2) with $\gamma = 1$. Since $\gamma \in [\gamma^{\text{non}}, \gamma^{\text{eff}})$, according to Proposition 5.1 the non-efficient network Fig. 2(b) will be linearly unstable. We observe that deviation from the non-efficient network is achieved and the process converges to the efficient configuration.

VI. Concluding Remarks

We presented a method for distributed network formation and reinforcement of efficient networks by dynamic reinforcement. Some key distinguishing features of this work include: i) payoff based dynamics, in which each agent adapts according to realized link rewards and costs; ii) incorporation of state dependent link establishment costs in addition to link maintenance costs; and iii) reinforcement of efficient networks through dynamic reinforcement. We presented various characterizations and properties of Nash network configurations, in terms of the structure of their connectivity or the distances between nodes. Finally, we provided accompanying convergence results that show how these network configurations can be the outcome of a learning process.

References

[1] G. C. Chasparis, "Distributed learning and efficient outcomes in uncertain and dynamic environments," Ph.D. dissertation, University of California Los Angeles, Los Angeles, May 2008.

[2] M. Jackson, "A survey of models of network formation: Stability and efficiency," Jan. 2003.

[3] K. Sohrabi, J. Gao, V. Ailawadhi, and G. J. Pottie, "Protocols for self-organization of a wireless sensor network," *Personal Communications, IEEE*, vol. 7, no. 5, pp. 16–27, 2000.

[4] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Communications Magazine*, pp. 102–114, Aug. 2002.

[5] R. Aumann and R. Myerson, *Endogenous formation of links between players and coalitions: An application of the Shapley value*. In Roth, A. (ed.), *The Shapley Value*, Cambridge University Press, 1988, pp. 175–191.

[6] R. Myerson, *Game Theory: Analysis of Conflict*. Cambridge, MA: Harvard University Press, 1991.

[7] M. Jackson and A. Wolinsky, "A strategic model of social and economic networks," *Journal of Economic Theory*, vol. 71, pp. 44–74, 1996.

[8] B. Dutta and S. Mutuswami, "Stable networks," *Journal of Economic Theory*, vol. 76, pp. 322–344, 1997.

[9] B. Dutta and M. Jackson, "The stability and efficiency of directed communication networks," *Review of Economic Design*, vol. 5, pp. 251–272, 2000.

[10] S. Currarini and M. Morelli, "Network formation with sequential demands," *Review of Economic Design*, vol. 5, pp. 229–249, 2000.

[11] A. Watts, "A dynamic model of network formation," *Games and Economic Behavior*, vol. 34, pp. 331–341, 2001.

[12] M. Jackson and A. Watts, "The evolution of social and economic networks," *Journal of Economic Theory*, vol. 106, no. 2, pp. 265–295, 2002.

[13] V. Bala and S. Goyal, "A noncooperative model of network formation," *Econometrica*, vol. 68, no. 5, pp. 1181–1229, 2000.

[14] E. Zeggelink, "Dynamics of structure: an individual oriented approach," *Social Networks*, vol. 16, pp. 295–333, 1994.

[15] ——, "Evolving friendship networks: an individual-oriented approach implementing similarity," *Social Networks*, vol. 17, pp. 83–110, 1995.

[16] B. Skyrms and R. Pemantle, "A dynamic model of social network formation," *Proc. of the National Academy of Sciences of the USA*, vol. 97, pp. 9340–9346, 2000.

[17] P. Bonacich and T. M. Liggett, "Asymptotics of a matrix valued markov chain arising in sociology," *Stochastic Processes and their Applications*, vol. 104, pp. 155–171, 2003.

[18] K. Narendra and M. Thathachar, *Learning Automata: An introduction*. Prentice-Hall, 1989.

[19] K. Najim and A. S. Poznyak, *Learning Automata: Theory and Applications*. Pergamon Press Inc, 1994.

[20] W. B. Arthur, "On designing economic agents that behave like human agents," *Journal of Evolutionary Economics*, vol. 3, pp. 1–22, 1993.

[21] G. C. Chasparis and J. S. Shamma, "Distributed dynamic reinforcement of efficient outcomes in multiagent coordination," in *Proc. European Control Conference*, Kos, Greece, July 2007.

[22] M. Kandori, G. Mailath, and R. Rob, "Learning, mutation, and long-run equilibria in games," *Econometrica*, vol. 61, pp. 29–56, 1993.

[23] H. P. Young, "The evolution of conventions," *Econometrica*, vol. 61, pp. 57–84, 1993.

[24] M. F. Norman, "On linear models with two absorbing states," *Journal of Mathematical Psychology*, vol. 5, pp. 225–241, 1968.

[25] I. J. Shapiro and K. S. Narendra, "Use of stochastic automata for parameter self-organization with multi-modal performance criteria," *IEEE Transactions on Systems Science and Cybernetics*, vol. 5, pp. 352–360, 1969.

[26] J. S. Shamma and G. Arslan, "Dynamic fictitious play, dynamic gradient play, and distributed convergence to Nash equilibria," *IEEE Transactions on Automatic Control*, vol. 50, no. 3, pp. 312–327, 2005.

[27] H. J. Kushner and G. G. Yin, *Stochastic Approximation Algorithms and Applications*. Springer-Verlag New York, Inc., 1997.