Network Formation: Neighborhood Structures, Establishment Costs, and Distributed Learning

Georgios C. Chasparis, *Member, IEEE*, and Jeff S. Shamma

***Abstract*—We consider the problem of network formation in a distributed fashion. Network formation is modeled as a strategic-form game, where agents represent nodes that form and sever unidirectional links with other nodes and derive utilities from these links. Furthermore, agents can form links only with a limited set of neighbors. Agents trade off the benefit from links, which is determined by a distance-dependent reward function, and the cost of maintaining links. When each agent acts independently, trying to maximize its own utility function, we can characterize "stable" networks through the notion of Nash equilibrium. In fact, the introduced reward and cost functions lead to Nash equilibria (networks), which exhibit several desirable properties such as connectivity, bounded-hop diameter, and efficiency (i.e., minimum number of links). Since Nash networks may not necessarily be efficient, we also explore the possibility of "shaping" the set of Nash networks through the introduction of state-based utility functions. Such utility functions may represent dynamic phenomena such as establishment costs (either positive or negative). Finally, we show how Nash networks can be the outcome of a distributed learning process. In particular, we extend previous learning processes to so-called "state-based" weakly acyclic games, and we show that the proposed network formation games belong to this class of games.**

***Index Terms*—Ad hoc networks, distributed algorithms, distributed network formation, game theory, learning automata, wireless networks.**

I. Introduction

RECENT advances in ad hoc network technologies have demanded the development of efficient *overlay routing* or *network formation* protocols over complex physical network structures, such as Internet, cellular, and wireless networks. The objective of such overlay routing schemes is to achieve certain routing properties, for example, small network diameter, small congestion, and minimum communication cost, without the need to standardize or deploy a new routing protocol [2]. The advantage of overlay routing in such complex infrastructures can be significant, e.g., to divert congested traffic in cellular networks [3]. Other scenarios where overlay routing can be advantageous are, for example, peer-to-peer file transferring and end-host multicasting [4].

The approaches that have been proposed for overlay routing include mostly *centralized* optimization schemes, where the information needed for each node to calculate an optimal routing path may involve the collection of information from the whole network (see, e.g., [3]–[5]). However, centralized schemes may suffer from several issues related to energy conservation and information and computational complexity. Thus, recent trends in wireless networks technology (not necessarily restricted to overlay routing) have focused more on *decentralized* schemes [6], when information and computational capacity available to each node are limited.

To this end, more recent approaches have utilized distributed optimization techniques to address the problem of efficient overlay routing. In particular, several game-theoretic approaches have been considered, where each node acts independently, trying to maximize its own *utility function* or *performance measure*. The definition of such utilities is open ended, e.g., end-to-end delays in the node's connections [7], and constitutes one of the main challenges for such approaches. The main question emerging is whether such utility function exists that 1) assumes minimum available information to each node (preferably local) and 2) guarantees the establishment of desirable network configurations when each node myopically maximizes its own performance measure. Since robust solutions in such a distributed optimization framework can be described with respect to *Nash equilibria*, naturally, a secondary question emerging is *whether routing layouts, which correspond to Nash equilibria, exhibit desirable properties* (e.g., connectivity, small diameter, or small number of links).

This paper presents a distributed optimization approach to the establishment of efficient networks for overlay routing. Motivated by the current research on social networks [8], [9], we model the problem as a strategic-form game. In this framework, each agent represents a node and decides which links to form with its neighboring nodes so that its own utility is maximized. Our goal is to *explore how decision rules at the node level can justify the emergence of various network configurations*. This paper can also serve as a design tool for network formation, e.g., for overlay routing and topology control in ad hoc networks [10], [11], where minimum communication cost is a common requirement.

The remainder of this paper is organized as follows. Section II discusses related work on the subject of efficient network formation and states the contributions of this paper. Section III presents the necessary terminology and introduces the framework of state-based utility functions. Section IV

Manuscript received April 27, 2011; revised June 12, 2012 and September 19, 2012; accepted November 24, 2012. Date of publication February 1, 2013; date of current version November 18, 2013. This work was supported by the Air Force Office of Scientific Research under Grant FA9550-09-1-0420 and Grant FA9550-10-1-0573. The work of G. C. Chasparis was supported by the Swedish Research Council through the Linnaeus Center. A preliminary version of this paper was presented in part at the 47th Conference on Decision and Control, Fiesta Americana Grand Coral Beach, Cancun, Mexico, December 2008. This paper was recommended by Associate Editor T. Vasilakos.

G. C. Chasparis is with the Department of Automatic Control, Lund University, 221 00 Lund, Sweden (e-mail: georgios.chasparis@control.lth.se).

J. S. Shamma is with the School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332 USA (e-mail: shamma@gatech.edu).

Digital Object Identifier 10.1109/TSMCB.2012.2236553

presents the network formation model and the different versions of reward/cost functions. Section V analyzes the set of Nash and efficient networks of the network formation games. Section VI presents two different learning processes and analyzes their convergence properties when applied to network formation games. Section VI also presents selected illustrative simulations. Finally, Section VII presents concluding remarks.

II. Related Work

Several models for distributed network formation have been proposed that are based on game-theoretic formulations. These include *static models*, where the problem of network formation is usually modeled as a strategic-form game of several agents (or nodes), where agents' actions correspond to network links. Such approach was first considered by [12], where agents propose links sequentially, which are then accepted or rejected, forming an extensive-form game. These studies characterize networks in terms of the Nash equilibria of the associated game, which are called *Nash networks*. An improvement over such model was presented in [13], where agents simultaneously announce their preferences for links, and a link is formed only if both agents agree. One of the issues emerging in such model is the multiplicity of Nash equilibria. Reference [14] further discussed the relationship of the emerging equilibria with the concept of *pairwise stability*. Another static model, which is closer to the work of cooperative games [15], was presented in [8], where agents benefit from direct or indirect connections to other nodes (*connections model*). Reference [8] discusses one of the main issues present in network formation games, that is, the discrepancy between *efficient* and *stable networks*. Some extensions of these models include [14], which deals with the problem of constructing utility functions for which efficient networks are pairwise stable, and [16], which extends [8] to the case of directed links.

Although *static models* capture the stability properties of certain networks based on node (agent) objectives, they do not capture *how these networks emerge*. Such questions can be answered by designing *dynamic models* to capture how agents make decisions and how these decisions contribute over time to network formation. For example, [17] models network formation as a dynamic process, according to which, at each time instant, a pair of nodes is randomly selected, and a link is added between them if both agents benefit from it.

One of the main issues of dynamic formulations is the fact that the dynamic process may cycle. To avoid such cycles, [18] introduced random perturbations to the formation process. Somewhat parallel to [18], [9] develops models of network formation that use tools from *noncooperative game theory*, according to which agents can form and sever links unilaterally, i.e., no mutual consent is needed to form a link between two agents. A more recent work [19] has extended the model of [9] to the case where agents may establish links only with a subset of other agents. A similar extension has been analyzed independently in both the current and an earlier version of this paper [1].

A central implication of unilateral link formation [9] is that it leads to the concept of *Nash equilibrium*. Naturally, *best reply dynamics* have been utilized to study convergence to Nash equilibria. Such dynamics have been also employed in models with bilateral link formation [20]. Under such dynamics, agents are usually aware of all other agents' actions, implying large information and computational complexity. To avoid the resulting complexity, several dynamic models assume that agents learn how to play the game through time by adaptively reacting to measurements of their own performance (utility). A few models that belong to this category include [21] and [22], where agents adaptively form and sever links in reaction to an evolving network, and in some models, their decisions are subject to small random perturbations. The rewards received from each agent determine which interactions will be reinforced, and the network structure emerges as a consequence of the agents' learning behavior.

Recently, distributed network formation have been also considered as a way for distributed *overlay routing* over complex physical network structures, e.g., Internet, cellular, and wireless networks. The approaches that have been proposed for overlay routing include mostly *centralized* optimization schemes. For example, in [4], a centralized shortest path algorithm is used to find overlay links that satisfy a quality of service requirement. Centralized information is also necessary in [3] for overlay routing in cellular networks, where each base station needs to exchange information with all the available sources. In addition, in [5], a centralized scheme for bandwidth-aware routing is introduced, where each node periodically measures bandwidth capacity to every node in the network.

Game-theoretic approaches have been also considered to address the problem of overlay routing in order to avoid issues due to information and communication complexity of centralized schemes. For example, [23] models the problem of overlay routing in large-scale content sharing applications as a strategic-form game. However, the resulting game may not exhibit Nash equilibria, whereas Nash equilibria (if exist) cannot be explicitly characterized. A noncooperative game formulation for overlay routing is also considered in [7], where the cost function of each node accounts for the end-to-end delays in its connections. Reference [7] characterizes the best reply strategy for each node and computes pure Nash equilibria through iterative best response search.

The aforementioned game-theoretic formulations for distributed network formation or overlay routing reveal some of the main issues present in such approaches, that is, 1) *information complexity for best reply computation*, i.e., computation of best reply assumes that each player is aware of the previous actions of all other agents (see, e.g., [7] and [9]); 2) *existence* and *characterization of Nash networks*, i.e., showing that Nash networks exist and characterizing those networks may not be possible (see, e.g., [23]); 3) *nonefficiency of Nash networks* (see, e.g., [9]); and 4) *distributed convergence to desirable Nash networks*, i.e., most proposed schemes assume best reply dynamics for convergence to Nash networks, which might be infeasible (see, e.g., the discounted connections model in [9]).

This paper presents a novel game-theoretic approach to distributed network formation that addresses most of the aforementioned issues. Our approach is mostly related to dynamic and evolutionary models, such as [9], [16], [21], and [22]. In particular, we consider agents that have freedom over establishing or severing *unidirectional* links with *neighboring* agents based on myopic considerations. Unidirectional links model phenomena such as web links, observations of others, citations, etc. [24].

Specifically, here are our contributions.

1) We discuss the case where nodes can form links only with a subset of the other nodes (i.e., neighborhood structures), as opposed to the entire network.
2) We introduce utility functions that are distance-dependent variations of the *connections model* of [8] and guarantee that Nash networks *exist* and exhibit desirable properties, e.g., connectivity, efficiency, and bounded-hop diameter.
3) We introduce *state-dependent utility functions* that can model dynamic phenomena such as *establishment costs* and can be used as an equilibrium selection mechanism in favor of efficient Nash networks.
4) We derive a learning process that guarantees convergence to Nash equilibria for the state-based extension of *weakly acyclic* games, among which, the proposed network formation games.
5) We employ *payoff-based* dynamics for distributed convergence to Nash networks based on a modified reinforcement learning scheme [25] and drop the typical assumptions that nodes have knowledge of the full network structure and can compute optimal link decisions.

III. Terminology and Background

A. Games With State-Based Utilities

A game involves a finite set of *agents* (or *players*), $\mathcal{I} = \{1, 2, \ldots, n\}$. Each agent i has a finite number of *actions*, which is denoted by $\mathcal{A}_i$. We will enumerate the actions of each agent i and let α_i denote the index of agent i's action. The total number of actions of agent i is denoted by $|\mathcal{A}_i|$. Define $\mathcal{A}$ to be the Cartesian product of the action spaces of all agents, i.e., $\mathcal{A} \triangleq \mathcal{A}_1 \times \cdots \times \mathcal{A}_n$. In addition, let $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n) \in \mathcal{A}$ be the *action profile* of all agents.

For any positive integer m, $\Delta(m)$ denotes the probability simplex in $\mathbb{R}^m$, i.e., $\Delta(m) \triangleq \{x \in \mathbb{R}^m : x \geq 0, \mathbf{1}^T x = 1\}$, where $\mathbf{1}$ is a vector of ones of appropriate size. The vectors e_j, $j = 1, 2, \ldots, m$, denote the vertices of $\Delta(m)$. In some cases and by abusing notation, we will identify actions by vertices of the simplex (instead of indices), i.e., $\alpha_i = e_j$ implies that agent i selected action $j \in \mathcal{A}_i$.

In strategic-form games, after each agent $i \in \mathcal{I}$ has selected an action $\alpha_i \in \mathcal{A}_i$, agents are assigned *utilities*, i.e., evaluations of their own performance. Usually, such utilities are represented as instances of a function $v_i : \mathcal{A} \to \mathbb{R}_+$, which is called a *utility function*, that takes values in the set of action profiles. In other words, the utility of each agent i depends, in general, on the actions of all players $(\alpha_1, \alpha_2, \ldots, \alpha_n) \in \mathcal{A}$.

In order to also incorporate dynamic phenomena in the utilities such as establishment costs, we will assume that agents measure a utility or a payoff that depends on two variables, namely, the action profile, i.e., α, and an internal agent-specific state, i.e., x_i. The definition of this state variable is open ended (cf., "state-based" utility functions in [26] and "sometimes weakly acyclic" games in [27]). In the present setting, we restrict x_i to the simplex $\Delta(|\mathcal{A}_i|)$ and interpret the jth entry of x_i, $x_{ij} \in [0, 1]$, as the "familiarity" weighting of agent i with action $j \in \mathcal{A}_i$. Since x_i is in the simplex, these relative weights sum to one.

We now define *state-based utility functions* as follows.

Definition 3.1 (State-Based Utility Function): A state-based utility function maps $v_i : \mathcal{A} \times \Delta(|\mathcal{A}_i|) \to \mathbb{R}_+$ with $v_i(\alpha, x_i)$ being the payoff of agent i at joint action α and at (familiarity) state x_i.

The expression $v_i(\alpha, \alpha_i)$ is taken to mean $v_i(\alpha, x_i)$ evaluated at $x_i = \alpha_i$. In that case, the familiarity vector x_i corresponds to a unit vector, for example, e_j, i.e., the "familiarity" weighting of agent i with action j is 1, whereas the corresponding weighting with other actions is 0.

In several cases, we will need to compare joint action profiles based on *efficiency*. To this end, we define *efficient action profiles* as follows.

Definition 3.2 (Efficient Action Profile): Define the value function $V(\alpha) \triangleq \sum_{i \in \mathcal{I}} v_i(\alpha, \alpha_i)$. An efficient action profile is an action profile $\alpha^* \in \mathcal{A}$ such that $V(\alpha^*) = \max_{\alpha \in \mathcal{A}} V(\alpha)$.

Let $-i$ denote the complementary set $\mathcal{I} \backslash \{i\}$. We will often split the argument of a function this way, e.g., $F(\alpha) = F(\alpha_i, \alpha_{-i})$. The following extends the notion of *better reply* to state-based utility functions.

Definition 3.3 (Better Reply): The better reply set of agent $i \in \mathcal{I}$ to an action profile $\alpha = (\alpha_i, \alpha_{-i}) \in \mathcal{A}$ is a function $\mathrm{BR}_i : \mathcal{A} \to \mathcal{A}_i$ such that for any $\alpha_i^* \in \mathrm{BR}_i(\alpha)$ we have

$$v_i\left((\alpha_i^*, \alpha_{-i}), \alpha_i\right) > v_i\left((\alpha_i, \alpha_{-i}), \alpha_i\right). \tag{1}$$

Note that a better reply is a set-valued function and might be empty. Furthermore, when we evaluate the better reply set of an agent i to an action profile $\alpha = (\alpha_i, \alpha_{-i})$, the underlying familiarity state in the agent's state-based utility is assumed to be the corresponding action of that agent, i.e., α_i.

Based on this definition of a better reply, we introduce the notion of a "stable" action profile by extending the definition of a Nash equilibrium to state-based utility functions.

Definition 3.4 (State-Based Nash Equilibrium): An action profile α^* is a (state-based) Nash equilibrium if $\mathrm{BR}_i(\alpha^*) = \varnothing$ for every $i \in \mathcal{I}$, i.e.,

$$v_i\left(\left(\alpha_i^*, \alpha_{-i}^*\right), \alpha_i^*\right) \geq v_i\left(\left(\alpha_i', \alpha_{-i}^*\right), \alpha_i^*\right) \tag{2}$$

for all $\alpha_i' \in \mathcal{A}_i \backslash \{\alpha_i^*\}$ and $i \in \mathcal{I}$. Likewise, a strict (state-based) Nash equilibrium satisfies the strict inequality in (2).

For the remainder of this paper, we will refer to a state-based Nash equilibrium as a Nash equilibrium.

B. Coordination Games

Here, we introduce sufficient conditions for the existence of state-based Nash equilibria. To this end, it is necessary to first introduce the notion of an *improvement step* defined as follows.

Definition 3.5 (Improvement Step): The improvement step function of agent $i \in \mathcal{I}$ to an action profile $\alpha = (\alpha_i, \alpha_{-i}) \in \mathcal{A}$

is a function $\mathrm{IS}_i : \mathcal{A} \to \mathcal{A}_i$ such that, for any $\alpha_i^* \in \mathrm{IS}_i(\alpha)$, the following two conditions are satisfied:

1) $v_i((\alpha_i^*, \alpha_{-i}), \alpha_i) > v_i((\alpha_i, \alpha_{-i}), \alpha_i)$;
2) $v_i((\alpha_i^*, \alpha_{-i}), \alpha_i^*) > v_i((\alpha_i, \alpha_{-i}), \alpha_i)$.

Note that, if α_i^* is an improvement step for agent i to the action profile α, i.e., $\alpha_i^* \in \mathrm{IS}_i(\alpha)$, then α_i^* is also a better reply to α. Accordingly, if α is a Nash equilibrium profile, i.e., $\mathrm{BR}_i(\alpha) = \varnothing$ for all $i \in \mathcal{I}$, then we also have $\mathrm{IS}_i(\alpha) = \varnothing$. The converse is not necessarily true, i.e., there might not be an improvement step from an "unstable" action profile.

We utilize the notion of an improvement step to derive sufficient conditions for the existence of Nash equilibria. To this end, we introduce the following class of games.

Definition 3.6 (Coordination Game): A coordination game is such that there exists a function $\phi : \mathcal{A} \to \mathbb{R}$ with the following property: for any action profile $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathcal{A}$ other than a Nash equilibrium, there exists an agent $i \in \mathcal{I}$ such that $\mathrm{IS}_i(\alpha) \neq \varnothing$ and an action $\alpha_i' \in \mathrm{IS}_i(\alpha)$ such that

$$\phi(\alpha_i', \alpha_{-i}) > \phi(\alpha_i, \alpha_{-i}).$$

We will refer to this property as *coordination property* and the function ϕ as *coordination function*.

In words, a coordination game is such that, at any action profile other than a Nash equilibrium, there exist an agent and an action which can improve both its own payoff and the value of the coordination function ϕ. Such a feature introduces a weak form of *ordinal potential games* (cf., [28]).

Such a feature is also related to a form of "*coincidence of interests*." Note that prior definitions of coordination games, e.g., by [29] and [30], assume stronger conditions. For example, in [30], a coordination game is defined such that *any* better reply (for nonstate-based utilities) makes no other agent worse off.

The following is a straightforward consequence of the definition of coordination games.

Claim 3.1: For coordination games, any action profile that maximizes the coordination function ϕ is a Nash equilibrium.

This claim can be used to relate efficient action profiles with Nash equilibria when ϕ corresponds to the value function.

Claim 3.2: For any coordination game such that the coordination function ϕ corresponds to the value function, i.e., $\phi = V$, any efficient action profile is a Nash equilibrium.

Another straightforward consequence of the definition of a coordination game is the following.

Claim 3.3: For coordination games, starting from any action profile $\alpha \in \mathcal{A}$, there exists a finite sequence of action profiles $\{\alpha^0, \alpha^1, \ldots, \alpha^m\}$, such that $\alpha^0 = \alpha$, $\alpha_i^k \in \mathrm{IS}_i(\alpha^{k-1})$ for some i, and α^m is a Nash equilibrium.

Setting aside the state-based aspect, this claim indicates that coordination games defined here resemble weakly acyclic games (cf., [31]).

The coordination property introduced here will be particularly useful in 1) showing the existence of Nash equilibria (due to Claim 3.3) in network formation games and 2) designing distributed learning schemes for convergence to Nash equilibria, as we will discuss in the forthcoming sections.

IV. Network Formation Model

A. One-Way Benefit Flow

We will consider a *one-way* (directed) benefit flow model, where a directed *network* G consists of the nonempty finite set of agents (or *nodes*) $\mathcal{I}$ and a finite set of pairwise directed links (or *edges*) $\mathcal{E}$.

A directed link from node i to node j, $j \neq i$, is denoted (i, j), which represents the flow of benefits/information from i to j. A *path* from i to j in G, $(i \to j)$ is a sequence of distinct nodes that starts at i and ends at j, i.e., $(i \to j) = \{i = s_0, s_1, \ldots, s_m = j\}$ for some positive integer m, such that $(s_k, s_{k+1}) \in \mathcal{E}$ for $0 \leq k \leq m-1$. The number of links in the path $(i \to j)$ is denoted $|(i \to j)|$. For nodes i and j in G, the *distance from i to j*, which is denoted $\mathrm{dist}_G(i, j)$, is the minimum length of a path $(i \to j)$, if j is reachable from i, i.e., $\mathrm{dist}_G(i, j) \triangleq \min_{(i \to j) \subseteq G} |(i \to j)|$. If there is no path from i to j in G, then, by convention, $\mathrm{dist}_G(i, j) = \infty$. In addition, $\mathrm{dist}_G(i, i) = 0$ for every node $i \in \mathcal{I}$.

Definition 4.1 (Connectivity): A network G is connected if, for all $i \neq j$, $(i \to j) \subseteq G$.

Two useful subclasses of connected networks are *critically connected networks* and *minimally connected networks*.

Definition 4.2 (Critical Connectivity): A network G is critically connected if 1) it is connected and 2) if $(i, j) \in G$, then the unique path $(i \to j)$ is (i, j).

In words, a critically connected network is such that, if the link from agent i to (neighboring) agent j is dropped, then there is no path $(i \to j)$ in the network.

Definition 4.3 (Minimal Connectivity): A network G is minimally connected if 1) it is connected and 2) it has the minimum number of links.

Note that a minimally connected network will be also critically connected. The converse is not necessarily true.

B. Action Spaces and Neighborhood Structures

We assume that each agent i may establish links only with its neighbors, which is denoted by $\mathcal{N}_i$ with cardinality $|\mathcal{N}_i|$. In the unconstrained neighbors case, $\mathcal{N}_i = \mathcal{I} \backslash \{i\}$. ***For the remainder of this paper***, we will assume that *the neighborhoods are such that connectivity is always feasible*.

The set of actions of agent i, which is denoted $\mathcal{A}_i$, contains all possible combinations of neighbors with which a link can be established, including the case of establishing no links, i.e., $\mathcal{A}_i = 2^{\mathcal{N}_i}$. The notation $|\alpha_i|$ denotes the cardinality of $\alpha \in 2^{\mathcal{N}_i}$.

By abuse of notation, we will use $\alpha_i \in \mathcal{A}_i$ to refer to either an element of $\mathcal{A}_i = 2^{\mathcal{N}_i}$ or an index over $\mathcal{A}_i$. Likewise, we will use α to denote the network, i.e., G, which is induced by the collective actions $\alpha \in \mathcal{A}$, and thus, we may write expressions such as $\mathrm{dist}_\alpha(i, j)$ rather than $\mathrm{dist}_G(i, j)$.

C. Reward and Cost Functions

The state-based utility function of agent i is a function of the form $v_i : \mathcal{A} \times \Delta(|\mathcal{A}_i|) \to \mathbb{R}_+$, where

$$v_i(\alpha, x_i) \triangleq R_i(\alpha) - C_i(\alpha_i, x_i).$$

The function $R_i : \mathcal{A} \to \mathbb{R}_+$ is the reward of agent i, and the function $C_i : \mathcal{A}_i \times \Delta(|\mathcal{A}_i|) \to \mathbb{R}_+$ is the cost of the action of agent i. We will consider several forms of the reward and cost function defined in the following sections, specifically tailored for network formation.

1) Connections Reward Model: Assume that each individual is a source of benefits that others can access via the formation of links. In particular, a link with another agent inherits the benefits available to that agent via its own links. Following [8], define the connections reward function

$$R_i(\alpha) \triangleq \sum_{s \in \mathcal{I} \setminus \{i\}} \chi_\alpha(s \to i) \tag{3}$$

where

$$\chi_\alpha(s \to i) \triangleq \begin{cases} \delta^{\mathrm{dist}_\alpha(s,i)}, & \mathrm{dist}_\alpha(s,i) < \infty \\ 0, & \mathrm{dist}_\alpha(s,i) = \infty \end{cases}$$

for some $\delta \in (0, 1]$. We will refer to the case of $\delta = 1$ as the *frictionless benefit flow* and to the case of $\delta < 1$ as the *decaying benefit flow*.

This reward function has been also used in a game-theoretic formulation for topology control in wireless ad hoc networks by [32]. However, in [32], the action space is the neighborhood range, and every two nodes within the same range are always connected with a bidirectional link. Instead, here, links are assumed directional to reduce communication, and agents have the ability to choose over which links to establish, assuming a given neighborhood layout. In [8], [9], and [17], the same reward function has been considered, but without the neighborhood constraints imposed on the action space.

Reference [9] also considers the possibility of decaying benefits. As we shall see later, such a reward function can establish an upper bound in the distance among any two neighboring nodes at any Nash equilibrium. However, for the case of $\delta < 1$, existence of Nash equilibria is not guaranteed. This is a main reason of introducing the forthcoming *limited connections reward model*.

2) Limited Connections Reward Model: We consider here an alternative reward function that keeps track only of those neighbors which are, at most, K hops away, where $K \in \mathbb{N}$. Define the *benefit* function of agent $i \in \mathcal{I}$ as

$$\beta_i(\alpha) \triangleq \sum_{s \in \mathcal{N}_i} \chi_\alpha^K(s \to i) \tag{4}$$

where

$$\chi_\alpha^K(s \to i) \triangleq \begin{cases} 1, & \mathrm{dist}_\alpha(s,i) \le K \\ 0, & \mathrm{dist}_\alpha(s,i) > K. \end{cases}$$

In other words, the benefit function of agent i counts the number of neighbors within distance K. One can think of neighbors in this setting as favorite agents. Although an agent can access its favorites directly, doing so incurs both an establishment and maintenance cost (forthcoming). Therefore, an agent may seek to gain indirect access to its favorites.

Unfortunately, the preceding utility function does not necessarily define a coordination game for a generic neighborhood structure. This implies that Nash equilibria may not exist. To resolve this issue, we modify the reward to include "downstream" effects of an agent's actions.

For agent $i \in \mathcal{I}$, define the set of nodes that are accessing agent i as

$$B_i(\alpha; K) \triangleq \{j \in \mathcal{I} : (i \to j) \in \alpha \text{ and } \mathrm{dist}_\alpha(i,j) \le K-1\}.$$

This set describes the downstream beneficiaries of links made by agent i. Define the downstream deficiency of agent i as

$$d_i(\alpha) \triangleq \sum_{j \in B_i(\alpha;K)} \begin{cases} 1, & \beta_j(\alpha) < |\mathcal{N}_j| \\ 0, & \text{otherwise.} \end{cases}$$

An agent at full benefit capacity has $\beta_j(\alpha) = |\mathcal{N}_j|$; otherwise, the agent has a deficiency. The function $d_i(\alpha)$ counts the number of downstream beneficiaries that are deficient.

We now define the *limited connections* reward model as follows. For $0 < \gamma < 1$, define

$$R_i(\alpha) \triangleq \beta_i(\alpha) \left(1 - \gamma \frac{d_i(\alpha)}{1 + d_i(\alpha)}\right). \tag{5}$$

In words, this function rewards connections to neighbors, but at a reduced rate because of possible downstream deficiencies. In case of no such deficiencies, i.e., $d_i(\alpha) = 0$, the reward is equal to the benefit function.

3) State-Based Cost: The cost function of agent i, i.e., $C_i : \mathcal{A}_i \times \Delta(|\mathcal{A}_i|) \to \mathbb{R}_+$, is defined as

$$C_i(\alpha_i, x_i) \triangleq \kappa_0 |\alpha_i| + \kappa_1 \psi_i(\alpha_i)^T (\mathbf{1} - \psi_i(x_i)) \tag{6}$$

with $\kappa_0 \ge 0$, $\kappa_1 \in \mathbb{R}$, and $\psi_i : \Delta(|\mathcal{A}_i|) \to \mathbb{R}^{|\mathcal{N}_i|}$ defined by $[\psi_i(x_i)]_j \triangleq \sum_{\{a \in \mathcal{A}_i : j \in a\}} x_{ia}$. (Recall that we are using a as both an index, as in x_{ia}, and a set, as in $j \in a \in 2^{\mathcal{N}_i}$.) In words, $[\psi_i(x_i)]_j$ denotes the probability that agent i will form a link to neighbor j based on the distribution x_i. The term $(\mathbf{1} - \psi_i(x_i))^T \psi_i(\alpha_i)$ grows with misalignment of the action α_i with the distribution x_i. In the perfectly aligned case, for any $\alpha_i \in \mathcal{A}_i$ (viewed as a vertex of $\Delta(|\mathcal{A}_i|)$)

$$\psi_i(\alpha_i)^T (\mathbf{1} - \psi_i(\alpha_i)) = 0$$

whereas in the worst case

$$\max_{x_i} \psi_i(\alpha_i)^T (\mathbf{1} - \psi_i(x_i)) = |\alpha_i|.$$

The first part of the cost function (6) corresponds to the cost of maintaining the currently established links. The state x_i reflects familiarity with a particular set of links. Accordingly, the second part corresponds to an establishment cost. The establishment cost models possible inertia of the system. When $\kappa_1 > 0$, this term represents the effort necessary to establish a new link, whereas in the case $\kappa_1 < 0$, it represents incentives to explore.

V. Nash Equilibrium Networks

A. Existence, Connectivity, and Efficiency

We begin by establishing existence of Nash equilibria by deriving sufficient conditions for the network formation games defined here to be coordination games. For the remainder of this paper, we will use the shorthand notation here to represent the specific network formation game under discussion.

– $\mathfrak{C}$: The connections reward function (3) and state-based cost function (6).
– $\mathfrak{L}$: The limited connections reward function (5) and state-based cost function (6).

Proposition 5.1 (Coordination Property): Let $\kappa_1 \geq 0$.

1) $\mathfrak{C}$ is a coordination game for $\delta = 1$ and $\kappa_0 + \kappa_1 < 1$.
2) $\mathfrak{L}$ is a coordination game for

$$\kappa_0 + \kappa_1 < 1 - \gamma \quad \& \quad \kappa_0 < \frac{|\mathcal{N}_i|}{|\mathcal{N}_i| - 1} \cdot \frac{\gamma}{2} \quad \text{for all } i \in \mathcal{I}.$$

Proof: See Appendix. ■

In other words, Proposition 5.1 derives conditions for the cost parameters κ_0 and κ_1 under which the resulting network formation game is a coordination game. Due to this property and Claim 3.3, the resulting network formation games admit Nash equilibria. More specifically, the following properties are direct consequence of Proposition 5.1 and its proof.

Proposition 5.2 (Nash Equilibrium Connectivity): Under the hypotheses of Proposition 5.1:

1) both $\mathfrak{C}$ and $\mathfrak{L}$ admit Nash equilibria;
2) if α^* is a Nash equilibrium in $\mathfrak{C}$, then α^* is connected;
3) if α^* is a Nash equilibrium in $\mathfrak{L}$, then $\text{dist}_{\alpha^*}(j, i) \leq K$, for all $i \in \mathcal{I}$ and $j \in \mathcal{N}_i$.

We comment that condition 2 remains true in the $\mathfrak{C}$ framework for $\kappa_0 + \kappa_1 < \delta < 1$. However, in the case of decaying benefit flow (i.e., $\delta < 1$), the existence of a Nash equilibrium is not guaranteed under neighborhood structures.

We will refer to the Nash equilibria of these network games as *Nash networks*. Note that, because of the state-based utility functions, these Nash networks need not coincide with Nash networks from prior studies.

Finally, the following proposition relates efficiency with Nash equilibria.

Proposition 5.3 (Nash Network Efficiency): Under the hypotheses of Proposition 5.1, for both $\mathfrak{C}$ and $\mathfrak{L}$, efficient networks are Nash networks with a minimum number of links.

Proof: As we showed in the proof of Proposition 5.1, both $\mathfrak{C}$ and $\mathfrak{L}$ satisfy the coordination property, where the coordination function ϕ is defined as the network value (15). Due to the coordination property and Claim 3.2, efficient action profiles are Nash networks. In the case of $\mathfrak{C}$, the value at a Nash network, i.e., α^*, is $V(\alpha^*) = n(n-1) - \kappa_0 \sum_{i \in \mathcal{I}} |\alpha_i^*|$ due to the connectivity property (Proposition 5.2). Likewise, in the case of $\mathfrak{L}$, the value of the network at a Nash network, i.e., α^*, is $V(\alpha^*) = \sum_{i \in \mathcal{I}} |\mathcal{N}_i| - \kappa_0 \sum_{i \in \mathcal{I}} |\alpha_i^*|$. In either case, the value is maximized at a Nash network with a minimum number of links. ■

Propositions 5.2 and 5.3 address one of the main issues related to designing network formation games, that is, 1) *showing existence of Nash equilibria* and 2) *showing efficiency of Nash equilibria*. In particular, the introduced notion of coordination game provides a test criterion for the existence of Nash equilibria in network formation games. Furthermore, due to the coordination property of the designed utility functions in $\mathfrak{C}$ and $\mathfrak{L}$, the efficient networks (under the hypotheses of Proposition 5.1) are also Nash equilibria.

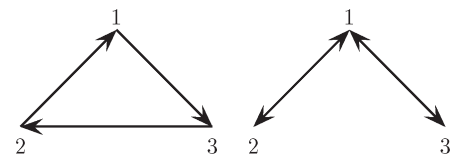


Fig. 1. Two Nash networks for $n = 3$ agents in $\mathfrak{C}$ with $\delta = 1$ and $\kappa_1 = 0$.

B. Special Case: $\kappa_1 = 0$

Nash networks have a special structure in case $\kappa_1 = 0$.

Proposition 5.4 ($\mathfrak{C}$ Nash Networks for $\delta = 1$ and $\kappa_1 = 0$): Under the hypotheses of Proposition 5.1 and for $\kappa_1 = 0$, a network in $\mathfrak{C}$ is a Nash network if and only if it is critically connected.

Proof: See Appendix. ■

For the connections model $\mathfrak{C}$ with $\delta = 1$ and $\kappa_1 = 0$, the Nash networks for $n = 3$ agents are shown in Fig. 1. Both networks are critically connected.

In other words, Proposition 5.4 revealed that Nash networks in $\mathfrak{C}$, when $\delta = 1$ and $\kappa_1 = 0$, are networks that are not only connected but also critically connected, i.e., there exists, at most, one direct link between any two nodes. Such property indirectly implies that the number of links for each node at a Nash network is limited.

Proposition 5.4 (which was first derived in an earlier version of this paper [1]), extends [9, Prop. 3.1], according to which Nash networks are critically connected under unconstrained neighbors.

An appropriate generalization of a critically connected network is also a Nash network in the $\mathfrak{L}$ framework. Define a *K-critically connected network* to be a critically connected network with the additional property that $\text{dist}_G(j, i) \leq K$ for all $i, j \in \mathcal{I}$ and $j \in \mathcal{N}_i$.

Proposition 5.5 ($\mathfrak{L}$ Nash Networks for $\kappa_1 = 0$): Under the hypotheses of Proposition 5.1 and for $\kappa_1 = 0$, a network in $\mathfrak{L}$ is a Nash network if it is K-critically connected.

Proof: Following the proof of Proposition 5.4, let α^* be a K-critically connected network, and let α_i' be a better reply. From the proof of Proposition 5.1, we can assume that α_i' maintains a radius of K for all of $\mathcal{N}_i$. Furthermore, the assumption on κ_0 implies that α_i' does not induce any downstream deficiencies in $B_i(\alpha^*; K)$. Therefore, as in the proof of Proposition 5.4, $|N_{\text{drop}}| - |N_{\text{add}}| > 0$, and thus, one can apply the same arguments. ■

Note that the reverse implication may not hold.

In the $\mathfrak{C}$ framework with decaying benefits ($\delta < 1$), the Nash equilibrium condition imposes a structural constraint on the distances between nodes.

Proposition 5.6 ($\mathfrak{C}$ Nash Networks for $\delta < 1$ and $\kappa_1 = 0$): For $\mathfrak{C}$ with $\delta < 1$, $0 < \kappa_0 < \delta$, and $\kappa_1 = 0$, let α^* be a Nash

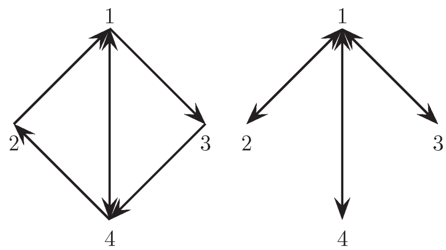


Fig. 2. Two Nash networks for $n = 4$ agents under $\mathfrak{C}$ with $\delta - \delta^2 \leq \kappa_0 < \delta - \delta^3$ and $\kappa_1 = 0$.

network corresponding to the joint action $\alpha^* \in \mathcal{A}$. For any agent i, if $|\alpha_i^*| < |\mathcal{N}_i|$, then

$$\delta^{\text{dist}_{\alpha^*}(j,i)} \geq \delta - \kappa_0 \text{ for all } j \in \mathcal{N}_i. \tag{7}$$

Proof: Let α_i^* satisfy the assumptions of Proposition 5.6, and compare an alternative action $\alpha_i' \in \mathcal{A}_i$ that consists of adding a direct link to neighbor j, i.e., $\alpha_i' = \alpha_i \cup \{j\}$. The resulting utility to agent i can be bounded by

$$v_i\left(\left(\alpha_i', \alpha_{-i}^*\right), \alpha_i^*\right) \geq v_i\left(\left(\alpha_i^*, \alpha_{-i}^*\right), \alpha_i^*\right) + \left(\delta - \delta^{\text{dist}_{\alpha^*}(j,i)}\right) - \kappa_0.$$

That is, the consequence of adding a link to j shortens the distances to other links, adds the direct benefit of a link to j, loses the indirect benefit of a link to j, and incurs additional maintenance cost. Therefore, if $(\delta - \delta^{\text{dist}_{\alpha^*}(j,i)}) - \kappa_0 > 0$, then there is an incentive to add a link to j, and thus, α^* cannot be a Nash network. Conversely, asserting that α^* is a Nash network implies the desired result. ■

The condition $|\alpha_i^*| < |\mathcal{N}_i|$ means that agent i is not using all of its available links. Inequality (7) is revealing only for neighbors of i, for which there is no direct link. 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 as follows. Inequality (7) is equivalent to

$$\text{dist}_{\alpha^*}(j,i) \leq \left\lfloor \frac{\log(\delta - \kappa_0)}{\log(\delta)} \right\rfloor \triangleq d.$$

A sufficient condition to bound the distance to neighbors by d is then $\kappa_0 \leq \delta - \delta^d$.

For example, consider δ and κ_0 such that $\delta - \delta^2 \leq \kappa_0 < \delta - \delta^3$. According to Proposition 5.6, this condition implies that, in any Nash network, the maximum distance that can be supported is $d = 2$. Under these conditions, Fig. 2 shows two Nash networks. It is straightforward to show that both networks in Fig. 2 are also Nash networks for the $\mathfrak{L}$ framework when $K = 2$ and for unconstrained neighbors.

Note, finally, that Proposition 5.6 does not address whether or not Nash equilibria exist in $\mathfrak{C}$ for $\delta < 1$.

C. *Strict Nash Networks and Small* κ_1

A forthcoming section deals with a distributed learning process based on reinforcement learning. It turns out that, under certain conditions, this process can converge to strict Nash equilibria, but not to action profiles that are not Nash equilibria. The following propositions relate strict Nash equlibria for small κ_1 to Nash equilibria for $\kappa_1 = 0$.

We start with considering positive establishment cost.

Proposition 5.7 (Nash Networks for Small $\kappa_1 > 0$*):* Under the hypotheses of Proposition 5.1, for both $\mathfrak{C}$ and $\mathfrak{L}$, there exists $\bar{\kappa}_1 > 0$ such that:

1) if α is not a Nash network for $\kappa_1 = 0$, then α is not a Nash network for $\kappa_1 \in (0, \bar{\kappa}_1)$;
2) if α is a Nash network for $\kappa_1 = 0$, then α is a strict Nash network for $\kappa_1 \in (0, \bar{\kappa}_1)$.

Proof:
Part 1: Suppose α is not a Nash network for $\kappa_1 = 0$. Then, there exists a better reply, $\alpha_i' \neq \alpha_i$, such that

$$R_i(\alpha') - \kappa_0 |\alpha_i'| > R_i(\alpha) - \kappa_0 |\alpha_i|$$

where $\alpha' = (\alpha_i', \alpha_{-i})$. This α_i' remains a better reply for nonzero κ_1 as long as

$$R_i(\alpha') - \kappa_0 |\alpha_i'| - \kappa_1 \psi\left(\alpha_i'\right)^T \left(\mathbf{1} - \psi(\alpha_i)\right) > R_i(\alpha) - \kappa_0 |\alpha_i|.$$

Define

$$\bar{\kappa}_1 \triangleq \min_{\substack{i \in \mathcal{I} \\ \alpha, \alpha' \in \mathcal{A}}} \frac{R_i(\alpha') + \kappa_0\left(|\alpha_i| - |\alpha_i'|\right) - R_i(\alpha)}{\psi\left(\alpha_i'\right)^T \left(\mathbf{1} - \psi(\alpha_i)\right)}$$

subject to α not a Nash network and $\alpha_i' \in \text{BR}_i(\alpha)$. This minimization involves strictly positive values over a finite set. Therefore, the minimum is also strictly positive.

Part 2: Suppose α is a Nash network for $\kappa_1 = 0$. Then, for all $i \in \mathcal{I}$ and $\alpha_i \in \mathcal{A}_i$

$$R_i(\alpha) - \kappa_0 |\alpha_i| \geq R_i\left(\alpha_i', \alpha_{-i}\right) - \kappa_0 \left|\alpha_i'\right|.$$

Therefore, for positive κ_1

$$R_i(\alpha) - \kappa_0 |\alpha_i| \geq R_i(\alpha_i', \alpha_{-i}) - \kappa_0 \left|\alpha_i'\right| - \kappa_1 \psi\left(\alpha_i'\right)^T \left(\mathbf{1} - \psi(\alpha_i)\right).$$

The question is whether the preceding inequality is strict. Recall the distinct sets N_{keep}, N_{drop}, and N_{add} defined in (16). Clearly, if $N_{\text{add}} \neq \varnothing$, the preceding inequality is strict. Now, suppose that $N_{\text{add}} = \varnothing$ while $N_{\text{drop}} \neq \varnothing$ and

$$R_i(\alpha) - \kappa_0\left(|N_{\text{keep}}| + |N_{\text{drop}}|\right) = R_i(\alpha') - \kappa_0 |N_{\text{keep}}|.$$

This equality means that α' is also a Nash network for $\kappa_1 = 0$, but with α_i' fewer links than α_i. This conclusion violates the derived connectivity properties of Nash networks. ■

In other words, Proposition 5.7 states that, when we increase κ_1 from zero to a positive value, the set of Nash networks remains identical with the case of $\kappa_1 = 0$; however, all Nash networks become strict. This observation has several implications when we discuss distributed learning processes in network formation games since strict Nash networks are potential attractors of the learning process, whereas nonstrict Nash networks may not be. Thus, by increasing κ_1, we are able to shape the set of strict Nash networks to all critically connected networks (due to Proposition 5.4).

The case of negative establishment cost, i.e., $\kappa_1 < 0$, can be viewed as rewarding exploration. The consequences are as follows.

Proposition 5.8 (Nash Networks for Small $\kappa_1 < 0$): Assume the hypotheses of Proposition 5.1 with $\kappa_1 = 0$. For both $\mathfrak{C}$ and $\mathfrak{L}$, there exists a $\underline{\kappa}_1 < 0$ such that:

1) if α is not a Nash network for $\kappa_1 = 0$, then α is not a Nash network for $\kappa_1 \in (\underline{\kappa}_1, 0)$;
2) if α is a nonstrict Nash network for $\kappa_1 = 0$, then α is not a Nash network for $\kappa_1 \in (\underline{\kappa}_1, 0)$;
3) if α is a strict Nash network for $\kappa_1 = 0$, then α is a strict Nash network $\kappa_1 \in (\underline{\kappa}_1, 0)$.

Proof:

Part 1: Since κ_1 is negative, this automatically preserves that α is not a Nash network.

Part 2: Suppose that α is a nonstrict Nash network for $\kappa_1 = 0$, and let α_i' satisfy $R_i(\alpha) - \kappa_0|\alpha_i| = R_i(\alpha') - \kappa_0|\alpha_i'|$. Then, α' is also a nonstrict Nash network for $\kappa_1 = 0$. As argued in the proof of Proposition 5.7, $|\alpha_i| = |\alpha_i'|$. Therefore, there are links in α_i' not in α_i. For $\kappa_1 < 0$, $\alpha_i' \in \mathrm{BR}_i(\alpha)$.

Part 3: If α is a strict Nash network for $\kappa_1 = 0$, then for all $i \in \mathcal{I}$, we have $R_i(\alpha) - \kappa_0|\alpha_i| > R_i(\alpha_i', \alpha_{-i}) - \kappa_0|\alpha_i'|$. This remains a strict Nash network as long as

$$R_i(\alpha) - \kappa_0|\alpha_i| > R_i(\alpha_i', \alpha_{-i}) - \kappa_0|\alpha_i'| - \kappa_1 \psi(\alpha_i')^T (\mathbf{1} - \psi(\alpha)).$$

As in Proposition 5.7, the preceding inequality can be used to extract a lower bound on κ_1 that preserves strictness. ■

In other words, Proposition 5.8 states that, by decreasing the value of κ_1 from zero to a negative value, we can make the strict Nash networks of the case $\kappa_1 = 0$ to be the only Nash networks. This has the opposite effect compared with Proposition 5.7. In fact, and as we will also explain in the following section, we can exclude convergence to any Nash network other than the strict Nash networks of the case $\kappa_1 = 0$. This can be desirable in certain cases. For example, in the unconstrained neighbors case (i.e., when $\mathcal{N}_i \equiv \mathcal{I}$ for all $i \in \mathcal{I}$), and when $\kappa_1 = 0$, the only strict Nash equilibria are the *wheel networks*, where each node has exactly one link (see, e.g., Fig. 1 for the case of three nodes). In this case, the strict Nash networks are minimally connected, and they are the only Nash networks for small negative values of κ_1.

Both Propositions 5.7 and 5.8 reveal the potential of state-based utility functions in shaping the set of Nash equilibria toward ones that exhibit more desirable properties.

VI. Learning Dynamics

Thus far, the "state" in the state-based utility has only served to shape the set of strict Nash networks. Here, the value of this state will be inherited from the state of a learning process. In this dynamic setting, the interpretation of the state reflecting "familiarity" will be apparent. We present two forms of learning dynamics. The first is "action-based" dynamics, i.e., each player can observe the actions of other players. These dynamics will resemble a state-based variation of *adaptive play* defined in [33]. We will show that these dynamics globally converge to a Nash network. The second form of learning dynamics is based on reinforcement learning. A desirable characteristic is that these dynamics are "payoff-based" dynamics. Agents cannot observe the overall network. Rather, agents only measure their utility received from the network. We will show that these dynamics locally converge to a strict Nash network.

A. Adaptive Play

The "state" in the state-based utility will evolve over stages, i.e., $t = 0, 1, 2, \ldots$. Let $M \geq 1$ be an integer, denoting "memory length." For each $i \in \mathcal{I}$, define

$$x_i(t+1) = \frac{1}{M} \sum_{\tau=0}^{M-1} \alpha_i(t-\tau) \tag{8}$$

where we associate each action $\alpha_i(t)$ as a vertex of $\Delta(\mathcal{A}_i)$. In words, $x_i(t+1)$ is the empirical frequency of the actions of agent i over the previous M stages.

We will need to extend the definition of a better reply. Define the set-valued function

$$\mathrm{BR}_i(\alpha; x_i) \triangleq \{\alpha_i^* \in \mathcal{A}_i : v_i((\alpha_i^*, \alpha_{-i}), x_i) > v_i((\alpha_i', \alpha_{-i}), x_i)\}.$$

(In the previous definition, x_i was set to α_i.)

Let $p \in (0, 1)$. Actions evolve according to the following (nondeterministic) rule:

$$\alpha_i(t) = \begin{cases} \alpha_i(t-1), & \text{if } \mathrm{BR}_i(\alpha(t-1); x_i(t)) = \varnothing \\ \alpha_i'(t), & \text{otherwise} \end{cases} \tag{9a}$$

where

$$\alpha_i'(t) \in \begin{cases} \alpha_i(t-1), & \text{with probability } p \\ \mathrm{BR}_i(\alpha(t-1); x_i(t)), & \text{with probability } 1-p. \end{cases} \tag{9b}$$

Proposition 6.1: Assume the hypotheses of Proposition 5.1, state dynamics (8), and action selection rule (9). In both $\mathfrak{C}$ and $\mathfrak{L}$ frameworks, $x(t)$ converges to a Nash network with probability one for any integer $M \geq 1$ and initialization $\alpha(\tau)$, $\tau = 0, 1, \ldots, M-1$.

Proof: (Sketch) Consider the following chain of events. With positive probability, all agents repeat their actions for M stages prior to stage T. Then, $x(T) = \alpha(t-1)$. At this stage, if $x(T)$ is a Nash network, then the dynamics have converged. Otherwise, there exists a single agent with a better reply. Let this be the only agent that updates its strategy, whereas all others repeat. Now, let all agents again repeat their actions for M stages. According to Claim 3.3, this process can repeat until the state converges to a Nash equilibrium. The probability of such a chain of events, for example, ε^*, is strictly positive (however small). Therefore, by the Borel–Cantelli Lemma (cf., [34, Lemma 3.14]), the process eventually converges to a Nash network. ■

B. State-Based Reinforcement Learning

Our reinforcement learning scheme assumes that, at each stage $t = 0, 1, 2, \ldots$, each agent i selects an action $\alpha_i(t) \in \mathcal{A}_i$ according to the probability distribution

$$(1-\lambda)x_i(t) + \frac{\lambda}{|\mathcal{A}_i|}\mathbf{1} \tag{10}$$

where 1) $x_i(t) \in \Delta(|\mathcal{A}_i|)$ is the *strategy* of agent i at stage t; 2) $\mathbf{1}$ is a vector of appropriate dimension, with each element equal to 1; and 3) $\lambda \geq 0$ is a parameter used to model possible perturbations in the decision-making process, which are also called *mutations* [33], [35].

The strategy of agent i is updated according to the recursion

$$x_i(t+1)=x_i(t)+\epsilon(t)\cdot v_i\left(\alpha(t), x_i(t)\right)\cdot(\alpha_i(t)-x_i(t)). \tag{11}$$

In this recursion, the jth entry of the reinforcement state, i.e., x_{ij}, can naturally capture "familiarity" weighting of agent i with action $j \in \mathcal{A}_i$ since x_{ij} increases if action j is selected and decreases otherwise. Accordingly, we have selected x_i to be the familiarity state in the reward function v_i. Such selection also significantly simplifies the stability analysis of the recursion.

Note that, in standard reinforcement learning, e.g., the models of [36]–[38], the reward v_i is a function of the current action profile $\alpha(t)$ and *not* a function of the reinforcement state $x_i(t)$.

We will generally consider the step-size sequence

$$\epsilon(t) \triangleq 1/(t^{\nu}+1)$$

where $\nu \in (1/2, 1]$. The parameter ν affects the rate of convergence. It is straightforward to show that, for sufficiently large t, the vector $x_i(\cdot)$ evolves within the probability simplex, which is sufficient for the stability analysis considered here.

The convergence properties of (11) can be characterized via the ordinary differential equation method for stochastic approximations (cf., [39]). Before proceeding, first define Ω to be the canonical path space with an element $\omega \in \Omega$ being a sequence $\{x(0), x(1), \ldots\}$, where $x(t) = (x_1(t), \ldots, x_n(t)) \in \mathbf{\Delta}$ is generated by the process, and $\mathbf{\Delta} \triangleq \Delta(|\mathcal{A}_1|) \times \cdots \times \Delta(|\mathcal{A}_n|)$. Define also the random variable $\chi_\tau : \Omega \to \mathbf{\Delta}$ such that $\chi_\tau(\omega) = x(\tau)$. In several cases, we will abuse notation by writing $x(\tau)$ instead of $\chi_\tau(\omega)$. In addition, let $\mathcal{F}$ be a σ-algebra of subsets in Ω and $\mathbb{P}$ a probability measure on $(\Omega, \mathcal{F})$ induced by the recursion (11). The σ-algebra $\mathcal{F}$ will be appropriately generated to allow computation of the probabilities of interest. Finally, let $\mathbb{E}$ denote the expectation with respect to measure $\mathbb{P}$. Define

$$\bar{g}_i\left(x(t)\right) \triangleq \mathbb{E}\left[v_i\left(\alpha(t), x_i(t)\right)\cdot\left(\alpha_i(t)-x_i(t)\right)\middle| x(t)\right]$$

and the ODE

$$\dot{x} = \bar{g}(x) \tag{12}$$

where $\bar{g}(\cdot) \triangleq [\bar{g}_i(\cdot)]_{i\in\mathcal{I}}$. The asymptotic behavior of the recursion (11) can be described through the invariant sets of (12). It has been shown by [25, Prop. 3.4] that, for $\lambda = 0$, any pure strategy profile $\alpha^* = (\alpha_1^*, \ldots, \alpha_n^*)$ is a stationary point of the ODE (12), i.e., $\bar{g}(\alpha^*) = 0$. The sensitivity of stationary points when $\lambda > 0$ is as follows.

Proposition 6.2 (Sensitivity of Stationary Points): For any pure strategy profile α^*, which is a strict Nash equilibrium, and for sufficiently small $\lambda > 0$, there exists a unique continuously differentiable function $\nu^* : \mathbb{R}_+ \to \mathbb{R}^{|\mathcal{A}|}$, such that $\lim_{\lambda\to 0}\nu^*(\lambda) = \nu^*(0) = 0$, and

$$x^* = \alpha^* + \nu^*(\lambda) \in \mathrm{Int}(\mathbf{\Delta}) \tag{13}$$

is a stationary point of the ODE (12). If, instead, α^* is not a Nash equilibrium, then there exist $\varepsilon > 0$ and $\lambda_0 > 0$, such that the ε-neighborhood of α^* in $\mathbf{\Delta}$, $O_\varepsilon(\alpha^*)$, does not contain any stationary point of the ODE (12) for any $0 < \lambda < \lambda_0$.

Proof: The proof follows similar reasoning with [25, Prop. 3.5]. ■

Note that Proposition 6.2 does not discuss the sensitivity of stationary points, which are nonstrict Nash equilibria. However, as the analysis in [25] showed, a vertex cannot be a stationary point of the perturbed dynamics.

Then, the behavior of the recursion (11) nearby stationary points is described by the following.

Proposition 6.3 (Convergence and Nonconvergence): For sufficiently small $\lambda > 0$, let x^* be a stationary point of the ODE (12) corresponding to a strict Nash equilibrium $\alpha^* \in \mathcal{A}$ according to (13). When the reinforcement learning scheme (11) is applied, $\mathbb{P}[\lim_{t\to\infty} x(t) = x^*] > 0$. If, instead, α^* is not a Nash equilibrium, then there exist $\varepsilon > 0$ and $\lambda_0 > 0$ such that $\mathbb{P}[\lim_{t\to\infty} x(t) \in O_\varepsilon(\alpha^*)] = 0$ for all $0 < \lambda < \lambda_0$.

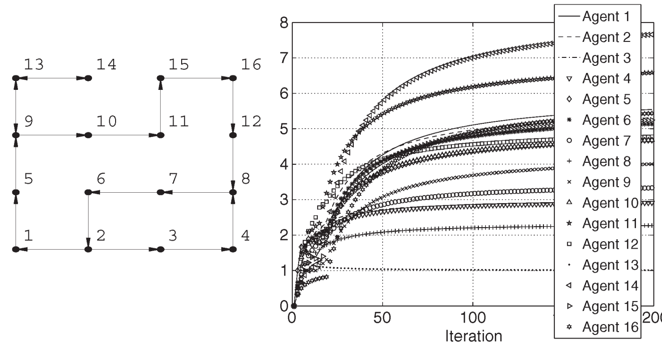
Proof: The proof of the first statement is based on the fact that any stationary point x^*, which corresponds to a strict Nash equilibrium [according to (13)], is a locally asymptotically stable point of the ODE (12). This can be shown by following similar reasoning with [25, Prop. 3.6]. Then, by applying [39, Th. 6.6.1], we conclude that $\mathbb{P}[\lim_{t\to\infty} x(t) = x^*] > 0$ (see also [25, Prop. 3.1]). The proof of the second statement follows from Proposition 6.2 and the fact that the vector field in the vicinity of α^* points toward the interior of $\mathbf{\Delta}$ for any small $\lambda > 0$ (see also [25, Prop. 3.7]). ■

Proposition 6.3 establishes convergence with positive probability of the state-based reinforcement learning to the set of strict Nash equilibria and nonconvergence to action profiles that are not Nash equilibria. Convergence or nonconvergence arguments cannot be established for perturbations of nonstrict Nash equilibria. However, as shown in Section V-C, the "familiarity" weights can be utilized to appropriately shape the set of strict Nash equilibria and eliminate the set of nonstrict Nash equilibria.

In summary, here we showed that 1) reinforcement learning can be modified to incorporate "familiarity" weights in the utility functions and 2) we can establish convergence with positive probability to the set of strict Nash equilibria.

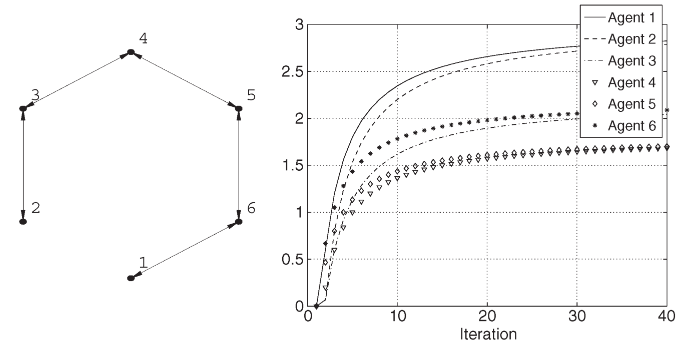
C. Simulations

Here, we illustrate the utility of adaptive play and state-based reinforcement learning on network formation games. To this end, we consider the following two ***examples***: 1) $n = 16$ nodes are placed on the vertices of a rectangular grid, as shown in Fig. 3, such that the neighborhood of each node consists of the two closest nodes along the horizontal and vertical axes, e.g.,



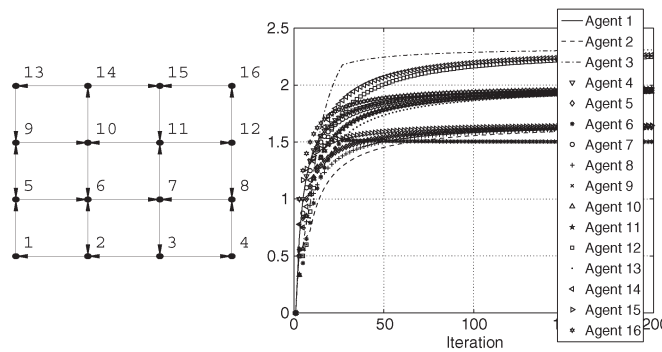
(a) (b)

Fig. 3. Typical response of adaptive play, with $M = 2$ and $p = 0.1$, under $\mathfrak{C}$ with $\kappa_0 = 1/8$ and $\kappa_1 = 0$. (a) Final graph. (b) Running average of mean distance from neighbors with time.



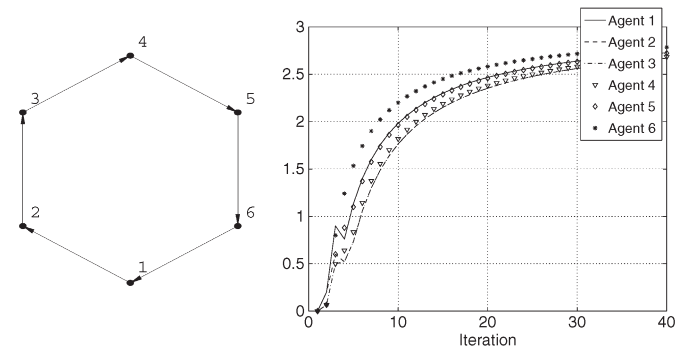
(a) (b)

Fig. 5. Typical response of adaptive play, with $M = 2$ and $p = 0.1$, under $\kappa_0 = 1/4$ and $\kappa_1 = 0$. (a) Final graph. (b) Running average of mean distance from neighbors with time.



(a) (b)

Fig. 4. Typical response of adaptive play, with $M = 2$ and $p = 0.1$, under $\mathfrak{L}$ with $K = 3$, $\kappa_0 = 1/8$, $\kappa_1 = 0$, and $\gamma = 1/2$. (a) Final graph. (b) Running average of mean distance from neighbors with time.



(a) (b)

Fig. 6. Typical response of adaptive play, with $M = 2$ and $p = 0.1$, under $\mathfrak{C}$ with $\kappa_0 = 1/4$ and $\kappa_1 = -1/10$. (a) Final graph. (b) Running average of mean distance from neighbors with time.

$\mathcal{N}_6 = \{2, 5, 7, 10\}$; 2) $n = 6$ nodes are placed on a circle, as shown in Fig. 5, such that the neighborhood of each node consists of the two closest nodes on the circle, e.g., $\mathcal{N}_1 = \{2, 6\}$.

First, let us consider the setup of example 1), where nodes are placed on the vertices of a rectangular grid. A typical response of adaptive play, with $M = 2$ and $p = 0.1$, applied in the connections model $\mathfrak{C}$ with $\kappa_0 = 1/8$ and $\kappa_1 = 0$ is shown in Fig. 3, where we have plotted the final graph and the running average of the mean distance from neighbors. Note that a critically connected network is formed, as expected by Proposition 5.4. Furthermore, the distances among neighboring nodes vary due to the fact that the connections model $\mathfrak{C}$ does not impose any constraint in the internode distances.

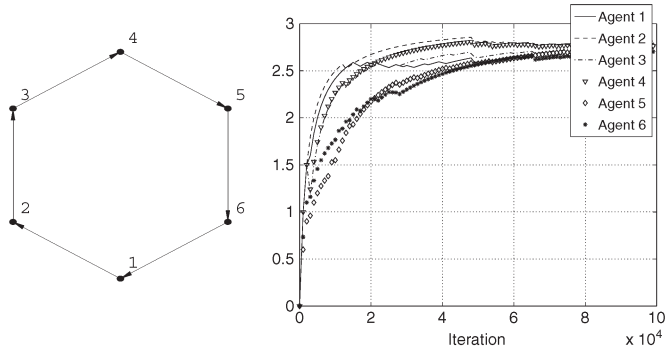
If, instead, the limited connections model $\mathfrak{L}$ is applied with $K = 3$, $\kappa_0 = 1/8$, $\kappa_1 = 0$, and $\gamma = 1/2$, then a typical response is shown in Fig. 4. According to Proposition 5.2, we should expect that adaptive play converges to a connected network such that the internode distance between any two neighboring nodes is no larger than K. Indeed, as we observe in Fig. 4, the running average of the mean distance from neighbors does not exceed K for all agents.

To demonstrate the utility of state-based utility functions in shaping the set of Nash networks, we consider example 2), where nodes are placed on a circle. Under the connections model $\mathfrak{C}$ and the assumed neighborhood layout, there are only two families of critically connected networks, namely, the starlike network in Fig. 5, and the wheel network in Fig. 6. However, the wheel networks are the only strict Nash and efficient networks. The adaptive play and reinforcement learning algorithms introduced here are likely to converge to any Nash equilibrium (starlike or wheel network), although the starlike network is a nonstrict Nash network. Fig. 5 shows a typical response of adaptive play that converges to the starlike Nash network under the connections model $\mathfrak{C}$ with $\kappa_0 = 1/4$ and $\kappa_1 = 0$.

According to Proposition 5.8, it is straightforward to show that, in the connections model $\mathfrak{C}$ with $\kappa_1 \in (-\kappa_0, 0)$, the wheel networks will be the only strict Nash networks. Furthermore, any other (critically connected) network will not be a Nash network. Fig. 6 shows a typical response of adaptive play in the connections model $\mathfrak{C}$ when $\kappa_0 = 1/4$ and $\kappa_1 = -1/10$, where convergence to a wheel network is observed. Under the same framework, Fig. 7 shows a typical response of state-based reinforcement learning (11), where convergence to a wheel network is also observed. Thus, we showed how state-based utility functions can be utilized to exclude convergence from nonefficient Nash networks.

VII. Concluding Remarks and Future Work

We have presented a strategic-form game formulation for the problem of distributed network formation. Some key



(a) (b)

Fig. 7. Typical response of state-based reinforcement learning (11) with $\lambda = 0.01$ and $\nu = 2/3$, under $\mathfrak{C}$ with $\kappa_0 = 1/4$ and $\kappa_1 = -1/10$. (a) Final graph. (b) Running average of mean distance from neighbors with time.

distinguishing features of this paper include the following: 1) directed links and neighborhood constraints; 2) distance-dependent utility functions that guarantee existence of Nash networks; 3) state-based utility functions that can model dynamic phenomena, such as establishment costs, and can shape the set of Nash networks; and 4) conditions that guarantee existence of Nash equilibria for the state-based extension of weakly acyclic games. Although state-based utility functions were not necessarily associated with a specific form of learning dynamics, we showed that, when combined with adaptive play or reinforcement learning, they provide an equilibrium selection approach in network formation games. For example, we showed how efficient graphs can be the only attractors of adaptive play and reinforcement learning when a negative establishment cost is considered. The proposed reinforcement learning scheme also revealed the potential of payoff-based learning approaches (i.e., when nodes only have access to measurements of their utility) for equilibrium selection in network formation.

A few directions in which this paper could be extended include the following: 1) designing alternative utility functions; 2) reducing communication complexity; and 3) designing alternative distributed learning processes. In particular, although the networks emerging through the proposed scheme exhibit desirable properties, e.g., connectivity, bounded-hop diameter, and small number of links, different scenarios may require alternative properties. For example, minimal number of links may not be desirable due to issues related to sensitivity to failures. Furthermore, although we analytically showed that the proposed reinforcement learning scheme locally converges to the strict Nash equilibria, it would be desirable to establish global convergence arguments, which is currently an open research problem, which is not necessarily restricted to network formation games.

Appendix

Proof of Proposition 5.1

Part 1: Suppose a network, i.e., α, is not a Nash equilibrium. Suppose further that it is not connected. Then, there exist $i, j \in \mathcal{I}$ such that $j \in \mathcal{N}_i$ and $(j \to i) \subseteq \alpha$. (Recall our assumption that connectivity is feasible with the underlying neighborhood structure.) In the $\mathfrak{C}$ framework, setting $\alpha_i' = \alpha_i \cup j$ increases the utility of agent i by $1 - (\kappa_0 + \kappa_1) > 0$ without decreasing the utility of any other agent. Furthermore, since $\kappa_1 \geq 0$

$$v_i\left(\left(\alpha_i', \alpha_{-i}\right), \alpha_i'\right) \geq v_i\left(\left(\alpha_i', \alpha_{-i}\right), \alpha_i\right) > v_i(\alpha, \alpha_i). \quad (14)$$

Therefore, $\alpha_i' \in \mathrm{IS}_i(\alpha)$ and, if we define the coordination function $\phi : \mathcal{A} \to \mathbb{R}$ such that

$$\phi(\alpha) \triangleq \sum_{i \in \mathcal{I}} v_i(\alpha, \alpha_i) \quad (15)$$

(i.e., ϕ is the value of the graph), then $\phi(\alpha_i', \alpha_{-i}) > \phi(\alpha_i, \alpha_{-i})$. Therefore, the coordination property is satisfied.

Now, suppose that α is not a Nash equilibrium but is connected, and let $\alpha_i' \in \mathrm{BR}_i(\alpha)$. We can assume that agent i maintains connectivity to all of $\mathcal{N}_i$. Otherwise, by the preceding arguments, we can replace α_i' with another $\alpha_i'' \in \mathrm{BR}_i(\alpha)$ by adding links that maintain connectivity. As a result, the key difference between the new network (α_i', α_{-i}) and the old network (α_i, α_{-i}) is that agent i maintained connectivity with fewer links. This does not reduce the utility of other agents. Again, since $\kappa_1 \geq 0$, the action α_i' satisfies $\alpha_i' \in \mathrm{IS}_i(\alpha)$. Therefore, the coordination property is satisfied when we define the coordination function ϕ as in (15).

Part 2: In moving from the connections model $\mathfrak{C}$ to the limited connections model $\mathfrak{L}$, simple connectivity to neighbors is insufficient. Rather, neighbors must be within a radius of K to contribute to benefits. Now, suppose that a network, i.e., α, is not a Nash equilibrium. Furthermore, assume that there exist $i, j \in \mathcal{I}$ such that $j \in \mathcal{N}_i$ and $\mathrm{dist}_\alpha(j, i) > K$, i.e., neighbor j is outside of the benefit radius K. In the $\mathfrak{L}$ framework, setting $\alpha_i' = \alpha_i \cup j$ changes the utility of agent i by

$$\beta_i(\alpha')\left(1-\gamma\frac{d_i(\alpha')}{1+d_i(\alpha')}\right)-\beta_i(\alpha)\left(1-\gamma\frac{d_i(\alpha)}{1+d_i(\alpha)}\right)-(\kappa_0+\kappa_1).$$

Since agent i added a link, $\beta_i(\alpha') \geq \beta_i(\alpha) + 1$ and $d_i(\alpha') \leq d_i(\alpha)$. These inequalities imply that the change in utility is at least

$$\left(1 - \gamma\frac{d_i(\alpha)}{1 + d_i(\alpha)}\right) - (\kappa_0 + \kappa_1) > 1 - \gamma - (\kappa_0 + \kappa_1)$$

which is positive by assumption. Therefore, $\alpha_i' \in \mathrm{BR}_i(\alpha)$. Furthermore, since $\kappa_1 \geq 0$, we have $1 - \gamma - \kappa_0 > 0$, and therefore, the condition (14) is also satisfied, i.e., $\alpha_i' \in \mathrm{IS}_i(\alpha)$. Thus, if we define ϕ as in (15), then the game satisfies the coordination property.

Now, suppose that α is not a Nash equilibrium but satisfies $\mathrm{dist}_\alpha(j, i) \leq K$ for all i, j such that $j \in \mathcal{N}_i$. Let $\alpha_i' \in \mathrm{BR}_i(\alpha)$. Again, we can assume that agent i maintains connectivity (within radius K) to all of $\mathcal{N}_i$. However, *unlike* the $\mathfrak{C}$ framework, maintaining this connectivity to neighbors does not imply that other nodes have maintained connectivity to their neighbors within a radius K.

Let us decompose the two actions α_i and α_i' in terms of links that were 1) kept, 2) added, and 3) dropped. Specifically, define the disjoint sets

$$N_{\text{keep}} = \alpha_i \cap \alpha_i' \tag{16a}$$

$$N_{\text{add}} = \alpha_i' \backslash N_{\text{keep}} \tag{16b}$$

$$N_{\text{drop}} = \alpha_i \backslash N_{\text{keep}}. \tag{16c}$$

Each of these sets is a subset of $\mathcal{N}_i$. Since α_i' is a better reply, then

$$|\mathcal{N}_i|\left(1-\gamma\frac{d_i(\alpha')}{1+d_i(\alpha')}\right)-\kappa_0|N_{\text{add}}\cup N_{\text{keep}}|-\kappa_1|N_{\text{add}}| > |\mathcal{N}_i|\left(1-\gamma\frac{d_i(\alpha)}{1+d_i(\alpha)}\right)-\kappa_0|N_{\text{drop}}\cup N_{\text{keep}}|.$$

Since α started with no deficient agents, $d_i(\alpha) = 0$, and thus

$$\kappa_0\left(|N_{\text{drop}}|-|N_{\text{add}}|\right)-\kappa_1|N_{\text{add}}| > |\mathcal{N}_i|\gamma\frac{d_i(\alpha')}{1+d_i(\alpha')}. \tag{17}$$

The left-hand side in (17) is bounded from above by $(|\mathcal{N}_i|-1)\kappa_0$. Assume that the network α' has deficient agents in $B_i(\alpha'; K)$. Then, the right-hand side of (17) is bounded from below by $|\mathcal{N}_i|/2$, and thus

$$\kappa_0 > \frac{|N_i|}{|N_i|-1}\cdot\frac{\gamma}{2}.$$

This contradicts the assumed condition on κ_0. Accordingly, α' must not have deficient agents in $B_i(\alpha_i'; K)$. Intuitively, the assumed bound on κ_0 implies that an agent will not sacrifice downstream deficiency just to reduce its number of links. Since $\kappa_1 \geq 0$, $\alpha_i' \in \text{IS}_i(\alpha)$. In addition, since the network α' has no deficient agents, none of the utilities of agents other than i has been reduced. Therefore, by defining ϕ as in (15), the game satisfies the coordination property. ■

Proof of Proposition 5.4 (Critically connected → Nash): Let α^* correspond to a critically connected network. Suppose for some agent $i \in \mathcal{I}$ and some action α_i', $\alpha_i' \neq \alpha_i^*$, then

$$v_i\left(\left(\alpha_i', \alpha_{-i}^*\right), \alpha_i^*\right) > v_i\left(\left(\alpha_i^*, \alpha_{-i}^*\right), \alpha_i^*\right) \tag{18}$$

i.e., agent i's utility of α_i' is greater than that of α_i^*. From the proof of Proposition 5.1, we can assume that α' is also connected. As in (16), we can write $\alpha_i^* = N_{\text{keep}} \cup N_{\text{drop}}$ and $\alpha_i' = N_{\text{keep}} \cup N_{\text{add}}$. Clearly, if $N_{\text{drop}} = \varnothing$, then (18) cannot hold since α^* is connected. Assume that $N_{\text{drop}} \neq \varnothing$. The utility of agent i in case of α^* is equal to

$$v_i\left(\left(\alpha_i^*, \alpha_{-i}^*\right), \alpha_i^*\right) = (n-1) - \kappa_0\left(|N_{\text{keep}}| + |N_{\text{drop}}|\right).$$

In case of α', the utility of i is

$$v_i\left(\left(\alpha_i', \alpha_{-i}^*\right), \alpha_i^*\right) = (n-1) - \kappa_0\left(|N_{\text{keep}}| + |N_{\text{add}}|\right).$$

Thus

$$v_i\left(\left(\alpha_i', \alpha_{-i}^*\right), \alpha_i^*\right) - v_i\left(\left(\alpha_i^*, \alpha_{-i}^*\right), \alpha_i^*\right) = \kappa_0\left(|N_{\text{drop}}| - |N_{\text{add}}|\right) > 0.$$

The only possibility for (18) to hold is if $|N_{\text{drop}}| > |N_{\text{add}}|$.

We now show that $|N_{\text{drop}}| > |N_{\text{add}}|$ contradicts α^* being a critically connected network.

- For each element of N_{add}, construct a path in α^* to i. These paths must pass through $N_{\text{keep}} \cup N_{\text{drop}}$.
- Since $|N_{\text{drop}}| > |N_{\text{add}}|$, there exists a $k^* \in N_{\text{drop}}$ that is not part of any of these paths.
- Construct a path in α' from k^* to i. This path, prior to reaching i, must pass through $(N_{\text{keep}} \cup N_{\text{add}})$. Prior to hitting a node in $(N_{\text{keep}} \cup N_{\text{add}})$, this path lies in α^*.
- The conclusion is a path from k^* to an element of N_{add} or an element of N_{keep}. In either case, the path can be continued in α^* to i without passing through k^*. This contradicts the critically connected assumption on α^*.

As a result, α^* cannot be a Nash equilibrium.

(Nash → Critically connected) Suppose a Nash network is not critically connected. Then, there exists an agent i that can drop a direct link to an agent $j \in \mathcal{N}_i$ but still maintain connectivity to j and, hence, receive the benefits of j without incurring the maintenance cost of j. Therefore, the original network cannot be a Nash network. ■

References

[1] G. Chasparis and J. S. Shamma, "Efficient network formation by distributed reinforcement," in *Proc. IEEE 47th Conf. Decision Control*, Cancun, Mexico, Dec. 2008, pp. 1690–1695.
[2] R. Cohen and D. Raz, "Cost effective resource allocation of overlay routing relay nodes," in *Proc. IEEE INFOCOM*, Apr. 2011, pp. 3236–3244.
[3] Y. Wu, K. Yang, and J. Shen, "Source selection routing algorithms in integrated cellular networks," *IET Commun.*, vol. 2, no. 1, pp. 98–106, Jan. 2008.
[4] Z. Li and P. Mohapatra, "QRON: QoS-aware routing in overlay networks," *IEEE J. Sel. Areas Commun.*, vol. 22, no. 1, pp. 29–40, Jan. 2004.
[5] S.-J. Lee, S. Banerjee, P. Sharma, P. Yalagandula, and S. Basu, "Bandwidth-aware routing in overlay networks," in *Proc. IEEE INFOCOM*, 2008, pp. 1732–1740.
[6] M. Khan, H. Tembine, and A. Vasilakos, "Evolutionary coalitional games: Design and challenges in wireless networks," *IEEE Wireless Commun.*, vol. 19, no. 2, pp. 50–56, Apr. 2012.
[7] G. Smaragdakis, N. Laoutaris, V. Lekakis, A. Bestavros, J. Byers, and M. Roussopoulos, "Selfish overlay network creation and maintenance," *IEEE/ACM Trans. Netw.*, vol. 19, no. 6, pp. 1624–1637, Dec. 2011.
[8] M. Jackson and A. Wolinsky, "A strategic model of social and economic networks," *J. Econ. Theory*, vol. 71, no. 1, pp. 44–74, Oct. 1996.
[9] V. Bala and S. Goyal, "A noncooperative model of network formation," *Econometrica*, vol. 68, no. 5, pp. 1181–1229, Sep. 2000.
[10] K. Sohrabi, J. Gao, V. Ailawadhi, and G. J. Pottie, "Protocols for self-organization of a wireless sensor network," *IEEE Pers. Commun.*, vol. 7, no. 5, pp. 16–27, Oct. 2000.
[11] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Commun. Mag.*, vol. 40, no. 8, pp. 102–114, Aug. 2002.
[12] R. Aumann and R. Myerson, *Endogenous Formation of Links Between Players and Coalitions: An Application of the Shapley Value*, A Roth, Ed. Cambridge, U.K.: Univ. Press, 1988, pp. 175–191.
[13] R. Myerson, *Game Theory: Analysis of Conflict*. Cambridge, MA: Harvard Univ. Press, 1991.
[14] B. Dutta and S. Mutuswami, "Stable networks," *J. Econ. Theory*, vol. 76, no. 2, pp. 322–344, Oct. 1997.
[15] R. Myerson, "Graphs and cooperation in games," *Math. Oper. Res.*, vol. 2, no. 3, pp. 225–229, Aug. 1977.
[16] B. Dutta and M. Jackson, "The stability and efficiency of directed communication networks," *Rev. Econ. Des.*, vol. 5, no. 3, pp. 251–272, Sep. 2000.
[17] A. Watts, "A dynamic model of network formation," *Games Econ. Behav.*, vol. 34, no. 2, pp. 331–341, Feb. 2001.
[18] M. Jackson and A. Watts, "The evolution of social and economic networks," *J. Econ. Theory*, vol. 106, no. 2, pp. 265–295, Oct. 2002.

[19] N. Olaizola and F. Valenciano, "One-way flow network formation under constraints," Univ. Basque Country, Biscay, Basque Country, 2012.

[20] E. Arcaute, R. Johari, and S. Mannor, "Network formation: Bilateral contracting and myopic dynamics," *IEEE Trans. Autom. Control*, vol. 54, no. 8, pp. 1765–1778, Aug. 2009.

[21] B. Skyrms and R. Pemantle, "A dynamic model of social network formation," *Proc. Nat. Acad. Sci. USA*, vol. 9716, pp. 9340–9346, Aug. 2000.

[22] P. Bonacich and T. Liggett, "Asymptotics of a matrix-valued markov chain arising in sociology," *Stochastic Process. Appl.*, vol. 104, no. 1, pp. 155–171, Mar. 2003.

[23] G. Koloniari and E. Pitoura, "A game-theoretic approach to the formation of clusterd overlay networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 23, no. 4, pp. 589–597, Apr. 2012.

[24] S. Goyal, *Connections: An Introduction to the Economics of Networks*. Princeton, NJ: Princeton Univ. Press, 2007.

[25] G. Chasparis and J. S. Shamma, "Distributed dynamic reinforcement of efficient outcomes in multiagent coordination and network formation," *Dyn. Games Appl.*, vol. 2, no. 1, pp. 18–50, 2012.

[26] J. Marden and A. Wierman, "Overcoming limitation of game-theoretic distributed control," in *Proc. 48th IEEE Conf. Decision Control*, Dec. 2009, pp. 6466–6471.

[27] J. Marden, G. Arslan, and J. Shamma, "Cooperative control and potential games," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 39, no. 6, pp. 1393–1407, Dec. 2009.

[28] D. Monderer and L. Shapley, "Potential games," *Games Econ. Behav.*, vol. 14, no. 1, pp. 124–143, May 1996.

[29] D. Lewis, *Convention: A Philosophical Study*. Oxford, U.K.: Blackwell, 2002.

[30] P. Vanderschraaf, *Learning and Coordination*. New York: Routledge, 2001.

[31] H. P. Young, *Individual Strategy and Social Structure*. Princeton, NJ: Princeton Univ. Press, 1998.

[32] R. Komali, A. B. MacKenzie, and R. P. Gilles, "Effect of selfish node behavior on efficient topology design," *IEEE Trans. Mobile Comput.*, vol. 7, no. 9, pp. 1057–1070, Sep. 2008.

[33] H. P. Young, "The evolution of conventions," *Econometrica*, vol. 61, no. 1, pp. 57–84, Jan. 1993.

[34] L. Breiman, *Probability*. Philadelphia, PA: SIAM, 1992.

[35] M. Kandori, G. Mailath, and R. Rob, "Learning, mutation, and long-run equilibria in games," *Econometrica*, vol. 61, no. 1, pp. 29–56, Jan. 1993.

[36] W. B. Arthur, "On designing economic agents that behave like human agents," *J. Evol. Econ.*, vol. 3, no. 1, pp. 1–22, 1993.

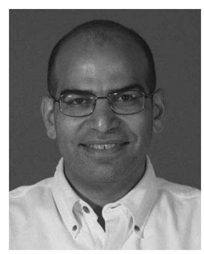
[37] T. Börgers and R. Sarin, "Learning through reinforcement and replicator dynamics," *J. Econ. Theory*, vol. 77, no. 1, pp. 1–14, Nov. 1997.

[38] I. Erev and A. Roth, "Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria," *Amer. Econ. Rev.*, vol. 88, no. 4, pp. 848–881, Sep. 1998.

[39] H. J. Kushner and G. G. Yin, *Stochastic Approximation Algorithms and Applications*. New York: Springer-Verlag, 2003.

Georgios C. Chasparis (S'04–M'08) was born in Athens in 1978. He received the B.S. degree in mechanical engineering from National Technical University of Athens, Athens, Greece, in 2001 and the M.Sc. and Ph.D. degrees from the University of California, Los Angeles, in 2004 and 2008, respectively.

From 2008 to 2010, he was a Postdoctoral Fellow with the Georgia Institute of Technology. Since 2010, he has been with the Department of Automatic Control, Lund University, Lund, Sweden, where he is currently a Postdoctoral Fellow. His research interests are learning in games, distributed control and optimization, and operations research.



Jeff S. Shamma received the B.S. degree in mechanical engineering from Georgia Institute of Technology, Atlanta, in 1983 and the Ph.D. degree in systems science and engineering from the Massachusetts Institute of Technology (MIT), Cambridge, in 1988.

He has held faculty positions with the University of Minnesota, Minneapolis; the University of Texas, Austin; and the University of California, Los Angeles; and visiting positions with Caltech and MIT. In 2007, he returned to Georgia Institute of Technology, where he is currently a Professor of electrical and computer engineering and the Julian T. Hightower Chair in Systems and Control. His research interest is feedback control and systems theory, with his recent area of interest being game theory and decision making for multiagent models in engineered and societal networked systems.

Dr. Shamma is currently an Associate Editor of the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B (2009, present) and *Games* (2012–present). He previously served on the Air Force Scientific Advisory Board (2008, 2011). He was a recipient of the National Science Foundation Young Investigator Award (1992) and the American Automatic Control Council Donald P. Eckman Award (1996).