# Decentralized Fictitious Play in Near-Potential Games With Time-Varying Communication Networks

Sarper Aydın[ID], Sina Arefizadeh[ID], and Ceyhun Eksin[ID]

*Abstract*—We study the convergence properties of decentralized fictitious play (DFP) for the class of near-potential games where the incentives of agents are nearly aligned with a potential function. In DFP, agents share information only with their current neighbors in a sequence of time-varying networks, keep estimates of other agents' empirical frequencies, and take actions to maximize their expected utility functions computed with respect to the estimated empirical frequencies. We show that empirical frequencies of actions converge to a set of strategies with potential function values that are larger than the potential function values obtained by approximate Nash equilibria of the potential game. This result establishes that DFP has identical convergence guarantees in near-potential games as the standard fictitious play in which agents observe the past actions of all the other agents.

*Index Terms*—Game theory, autonomous systems, networked control systems, numerical algorithms.

## I. INTRODUCTION

A GAME comprises of multiple agents taking actions to maximize their individual utility functions. Potential games is a special class of game in which there exists a common potential function that captures the incentives of all the agents [1]. Potential games are used to model behavior of agents in multitude of problems including traffic routing problems in transportation systems [2], power allocation in cognitive radio [3], task scheduling in robotics [4]. Many distributed game-theoretic learning dynamics, e.g., best response [1], fictitious play (FP) and its variants, e.g., sampled FP [5], [6], converge to a Nash equilibrium (NE) of any potential game. Nash equilibrium is an action profile in which no agent would benefit by unilaterally deviating from. The convergence of learning dynamics to a NE provides a justification for modeling agents as rational in strategic environments. That is, we observe individuals behaving rationally in reality despite lacking the acuteness, because they learned to act rational over time as they repeatedly interacted with each other in the same environment. Moreover, in technological systems, agents (sensors, robots) can be coded to implement these learning dynamics so that the system performance improves.

However, individual utility functions may deviate from the potential function due to unknown parameters, or the inability of agents to accurately compute the learning rules. Still, we would expect the learning dynamics to reach close to a Nash equilibrium of the potential game when such deviations are small. The class of near-potential games, proposed in [7], [8], quantifies the magnitude of such deviations under which desirable convergence properties continue to hold. Specifically, agents in a near-potential game have utility functions where the difference in utility values between any pair of actions for any agent is close enough to the differences in potential function values of a potential game, so that FP converges to a neighborhood of a NE of the potential game. In [8], the size of the convergence neighborhood is characterized as a function of the closeness of games for common distributed learning algorithms. Here, we characterize the convergence neighborhood based on the closeness of the game to a potential game when agents' actions follow a decentralized version of FP. Recently, FP is also shown to converge in stochastic zero-sum [9], [10] and mean-field games [11].

Agent updates in distributed learning dynamics, e.g., best response, and FP, rely on information from all the agents. In FP, agents assume that other agents pick their actions according to a stationary distribution that is the empirical frequency of their past actions [12]. Given this assumption, agents take actions to maximize their expected utilities computed with respect to these empirical frequencies. Such a response is not possible, when agents do not have access to the past actions of all the other agents, e.g., when agents can only communicate with a subset of the agents at each step. In such a scenario, agents can only keep estimates of the empirical frequencies of the other agents, and take the action that maximizes their expected utility functions computed with respect to the *estimated* empirical frequencies. DFP, developed in [4], is one

such learning algorithm that is fitting to communication over time-varying networks.

Given the DFP dynamics, we show that if the empirical frequencies are far from any equilibrium, i.e., are outside the set of approximate $\epsilon$-Nash equilibria of the near-potential game where the constant $\epsilon$ depends on the closeness of games, then the potential function value of the close potential game increases at every action update (Lemma 2). We note that the set of approximate $\epsilon$-Nash equilibria represents strategies in which an agent's incentive to deviate is not larger than $\epsilon$. Leveraging this result, we show that the empirical frequencies converge to a set of mixed strategies (a distribution over the actions) with potential function values larger than the lowest potential function value obtained by strategies in the set of $\epsilon$-Nash equilibria (Theorem 1 and Corollary 1). The closeness conditions required and the bounds established in these results are identical to the ones for FP dynamics shown in [8]. That is, DFP has the same convergence guarantees as FP in near-potential games, despite the fact that agents communicate only with a subset of the agents at each step. We validate our results in a target assignment problem in which agents are uncertain about the target locations. Numerical experiments show convergence of actions to a NE of the game in which target locations are known, when the near-potential game is close enough to the original game, i.e., when the uncertainty is small.

## II. NEAR-POTENTIAL GAMES

A game consists of $N$ agents represented by the set $\mathcal{N} := \{1, \ldots, N\}$. Each agent $i \in \mathcal{N}$ selects an action $a_i$ over a finite set of actions $\mathcal{A}_i$ to maximize its utility (payoff) $u_i : \mathcal{A}_i \times \mathcal{A}_{-i} \to \mathbb{R}$ where $\mathcal{A}_{-i} := \prod_{j \in \mathcal{N} \setminus \{i\}} \mathcal{A}_j$. We use the standard game-theoretic notation $-i$ to denote the set of all agents except agent $i$. A game $\Gamma$ is a tuple of agents $\mathcal{N}$, action space $\mathcal{A}^N := \prod_{i \in \mathcal{N}} \mathcal{A}_i$, and utility functions $u_i(\cdot)$ for $i \in \mathcal{N}$. Next, we define the class of potential games [1].

*Definition 1 (Potential Games):* A game $\Gamma$ is a potential game, if there exists a function $u : \mathcal{A}^N \to \mathbb{R}$ such that the following relation holds

$$u(a'_i, a_{-i}) - u(a_i, a_{-i}) = u_i(a'_i, a_{-i}) - u_i(a_i, a_{-i}) \quad (1)$$

for any $a'_i \in \mathcal{A}_i$, $a_i \in \mathcal{A}_i$ and $a_{-i} \in \mathcal{A}_{-i}$, and $i \in \mathcal{N}$. The corresponding function $u : \mathcal{A}^N \to \mathbb{R}$ is called the potential function of the game $\Gamma$.

The potential function ($u(\cdot)$) captures the difference in individual payoffs as a result of unilateral deviation in action for every agent. The existence of a potential function assures the existence of a NE in potential games. Furthermore, every maximizer of the potential function is a NE of the potential game [5]. The potential game is *bounded* if there exists a positive finite number $M < \infty$ such that $|u(a)| \le M$ for any $a \in \mathcal{A}^N$.

In order to define the class of near-potential games, we define a notion of closeness between games.

*Definition 2 (Maximum Pairwise Difference):* Let $\Gamma = (\mathcal{N}, \{\mathcal{A}_i, u_i\}_{i \in \mathcal{N}})$ and $\hat{\Gamma} = (\mathcal{N}, \{\mathcal{A}_i, \hat{u}_i\}_{i \in \mathcal{N}})$ be two games with the same set of agents $\mathcal{N}$ and action sets $\{\mathcal{A}_i\}_{i \in \mathcal{N}}$, and

utilities $\{u_i\}_{i \in \mathcal{N}}$ and $\{\hat{u}_i\}_{i \in \mathcal{N}}$. Further, let $d^{\Gamma}_{(a'_i, a)} := u_i(a'_i, a_{-i}) - u_i(a_i, a_{-i})$ be a difference in utility of an agent $i$ by unilateral change to an action $a'_i \in \mathcal{A}_i$, given joint action profile $a = (a_i, a_{-i}) \in \mathcal{A}^N$ in the game $\Gamma$. Then, the maximum pairwise difference $d(\Gamma, \hat{\Gamma})$ between the games $\Gamma$ and $\hat{\Gamma}$ is defined as

$$d(\Gamma, \hat{\Gamma}) := \max_{i \in \mathcal{N}, \, a'_i \in \mathcal{A}_i, \, a \in \mathcal{A}^N} |d^{\Gamma}_{(a'_i, a)} - d^{\hat{\Gamma}}_{(a'_i, a)}|. \quad (2)$$

The maximum pairwise difference (MPD), introduced by [8], defines the distance between two games based on the difference in agent payoffs resulting from unilateral changes to agent actions. A near-potential game is a game that is *close* to a potential game in terms of MPD.

*Definition 3 (Near-Potential Games):* A game $\Gamma$ is a $\delta$ near-potential game if there exists a potential game $\hat{\Gamma}$ within a MPD of $\delta \in \mathbb{R}^+$, i.e., $d(\Gamma, \hat{\Gamma}) \le \delta$.

The class of near potential games relaxes the condition of potential games in (1) similar to ordinal, weighted, or best-response potential games. In this letter, we consider decentralized learning dynamics in $\delta$ near-potential games. The closest potential game to a game $\Gamma$, and the smallest $\delta$ value for $\Gamma$ can be found by solving a convex optimization problem [7]. Here, we assume a potential function $u(\cdot)$ and the associated MPD value $\delta$ are given.

## III. DECENTRALIZED FICTITIOUS PLAY

Fictitious play is a game-theoretic learning mechanism in which each agent repeatedly takes an action to maximize its payoff based on the estimates of other agents' strategies. Agent $i$ forms an estimate of agent $j$'s strategy by keeping an empirical frequency of its past actions.

For simplicity, we assume the action space of agents is common, i.e., $\mathcal{A}_i = \mathcal{A}$ where the number of possible actions is equal to $K \in \mathbb{N}^+$, i.e., $|\mathcal{A}| = K$. We let $\Delta \mathcal{A}$ be the probability space over $\mathcal{A}$. The strategy of agent $i$ is a distribution on the action space denoted as $\sigma_i \in \Delta \mathcal{A}$, where $\sigma_i(a_i) \in [0, 1]$ denotes the probability of selecting action $a_i \in \mathcal{A}$. We define the expected utility of agent $i$ as

$$u_i(\sigma_i, \sigma_{-i}) = \sum_{a \in \mathcal{A}^N} u_i(a_i, a_{-i}) \sigma(a), \quad (3)$$

where $\sigma = (\sigma_i, \sigma_{-i}) \in \Delta \mathcal{A}^N$ is the joint strategy profile.

In standard FP, agents repeatedly take actions in discrete time steps $t = 1, 2, \ldots$. Each agent determines its next action $a_{i,t} \in \mathcal{A}$ to maximize its expected utility assuming other agents play according to a stationary distribution $f_{-i,t} := \{f_{j,t}\}_{j \in \mathcal{N} \setminus \{i\}}$ where $f_{j,t} \in \Delta \mathcal{A}$. The stationary distribution $f_{i,t} \in \Delta \mathcal{A}$ is computed using the past empirical frequency of actions, $f_{i,t} = \frac{1}{t} \sum_{\tau=1}^{t} \sigma_{i,\tau}$, which can equivalently be written as

$$f_{i,t} = \frac{t-1}{t} f_{i,t-1} + \frac{1}{t} \mathbf{1}_{a_{i,t}}. \quad (4)$$

where $\mathbf{1}_{a_{i,t}}$ is a $K \times 1$ vector that is all zeros except a one in the element corresponding to the action taken. The recursive form above allows each agent $i$ to compute the empirical frequencies of other agents $j \in \mathcal{N} \setminus i$ by only keeping the past

**Algorithm 1** DFP for Agent $i$

---

1: **Input:** Local estimates $\upsilon^i_{-i0}$ and networks $\{\mathcal{G}_t\}_{t \geq 1}$.
2: **for** $t = 1, 2, \cdots$ **do**
3:    Share local copies $\{\upsilon^i_{j,t}\}_{j \in \mathcal{N}}$ with $l \in \mathcal{N}_{i,t}$
4:    Update local copies $\upsilon^i_{j,t}$ (5).
5:    Determine action $a_{i,t}$ (6) and update $f_{i,t}$ (4).
6: **end for**

---

empirical frequencies $\{f_{j,t-1}\}_{j \in \mathcal{N} \setminus i}$ in memory and observing their current actions $\{\sigma_{j,t}\}_{j \in \mathcal{N} \setminus i}$.

Here, we consider a scenario in which agents communicate over a time-varying network $\mathcal{G}_t = (\mathcal{N}, \mathcal{E}_t)$, where $\mathcal{E}_t$ represents the edge set. In this setting, agent $i$ can only communicate with its neighbors $\mathcal{N}_{i,t} := \{j : (i, j) \in \mathcal{E}_t\}$ at time step $t$. Thus, it cannot compute the empirical frequency of all the other agents using (4). Instead, we replace the actual empirical frequencies $f_{j,t} \in \Delta \mathcal{A}$ with a local copy $\upsilon^i_{j,t} \in \Delta \mathcal{A}$. The local copy $\upsilon^i_{j,t}$ is the estimate of agent $j$'s empirical frequency at time $t$ kept at agent $i$. We let agent $i$'s local copy of its own empirical frequency be equal to its empirical frequency, $\upsilon^i_{i,t} = f_{i,t}$. In each step, agent $i$ updates its local estimate by using the local copies shared by its neighbors,

$$\upsilon^i_{j,t} = \sum_{l \in \mathcal{N}} w^i_{jl,t} \upsilon^l_{j,t}, \tag{5}$$

where $w^i_{jl,t} \geq 0$ is the weight that agent $i$ puts on agent $l$'s estimate of agent $j$'s empirical frequency at time $t$.

In DFP, each agent takes an action that maximizes its expected utility (best-respond) computed using its estimates of others' empirical frequencies $\upsilon^i_{-i,t} := \{\upsilon^i_{j,t-1}\}_{j \in \mathcal{N} \setminus i}$,

$$a_{i,t} \in \arg\max_{a_i \in \mathcal{A}} u_i(a_i, \upsilon^i_{-i,t}). \tag{6}$$

An outline of DFP is given in Algorithm 1.

## IV. CONVERGENCE OF DFP IN NEAR-POTENTIAL GAMES

### A. Game Theoretic Preliminaries

Given a game $\Gamma := \{\mathcal{N}, \mathcal{A}^N, \{u_i\}_{i \in \mathcal{N}}\}$, a NE is a strategy profile such that no agent can individually increase its utility by deviating to another strategy. Given a strategy profile $\sigma^*$, when there exists a profitable deviation for an agent but the increase in utility from this deviation is no more than $\epsilon \geq 0$, the strategy profile $\sigma^*$ is said to be an approximate NE.

*Definition 4 (Approximate NE):* The strategy profile $\sigma^* = (\sigma^*_i, \sigma^*_{-i}) \in \Delta \mathcal{A}^N$ is an $\epsilon$-Nash equilibrium of the game $\Gamma$ for some $\epsilon \geq 0$ if and only if for all $i \in \mathcal{N}$

$$u_i(\sigma^*_i, \sigma^*_{-i}) - u_i(\sigma_i, \sigma^*_{-i}) \geq -\epsilon, \quad \forall \sigma_i \in \Delta \mathcal{A}. \tag{7}$$

The strategy profile $\sigma^*$ is called a pure $\epsilon$-NE if the strategy selects a single action profile $a^* = (a^*_i, a^*_{-i}) \in \mathcal{A}^N$ with probability one. If the $\epsilon$-NE strategy profile is not pure, then it is called an $\epsilon$-mixed NE strategy profile.

We denote the set of $\epsilon$-Nash equilibria using $\Sigma_\epsilon$. We recover the definition of a NE when $\epsilon = 0$ in the above definition. The set of Nash equilibria is denoted with $\Sigma_0$.

We define the set of strategies that achieve potential function values no worse than the potential function value of the worst $\epsilon$-NE as follows:

$$C_\epsilon := \{\sigma \in \Delta \mathcal{A}^N \mid u(\sigma) \geq \min_{y \in \Sigma_\epsilon} u(y)\}. \tag{8}$$

When there is a unique NE, the above set contains the $\Sigma_\epsilon$. When there are multiple equilibria in the game $\Gamma$, the set $C_\epsilon$ contains the equilibria, but it may also contain a continuum of other strategies.

### B. Convergence

We make the following set of assumptions for the time-varying communication network.

*Assumption 1:* The network $\mathcal{G} = (\mathcal{N}, \mathcal{E}_\infty)$ is connected, where $\mathcal{E}_\infty = \{(i, j) | (i, j) \in \mathcal{E}_t$, for infinitely many $t \in \mathbb{N}\}$.
This assumption states that starting from any time $t_0$, there exists a path from agent $j$ to $i$ for any pair of agents $i$ and $j$ when we consider the edges $\bigcup_{t \geq t_0} \mathcal{E}_t$.

*Assumption 2:* There exists a time step $T_B > 0$, such that for any edge $(i, j) \in \mathcal{E}_\infty$ and $t \geq 1$, it holds $(i, j) \in \bigcup_{\tau=0}^{T_B - 1} \mathcal{E}_{t+\tau}$.
This assumption means the edge $(i, j) \in \mathcal{E}_\infty$ also belongs to the edge set $\bigcup_{t_0 + T_B > t \geq t_0} \mathcal{E}_t$ for any time $t_0 > 0$. Assumptions 1 and 2 are referred to as *connectivity* and *bounded communication interval*, respectively in [13].

*Assumption 3:* There exists a scalar $0 < \eta < 1$, such that the following statements hold for all $i \in \mathcal{N}$, $j \in \mathcal{N}$ and $t = 1, 2, \ldots$,
 (i) If $l \in \mathcal{N}_{i,t} \cup \{i\}$, then $w^i_{jl,t} \geq \eta$. Otherwise, $w^i_{jl,t} = 0$,
 (ii) $w^i_{ii,t} = 1$,
(iii) $\sum_{l \in \mathcal{N}_{i,t} \cup \{i\}} w^i_{jl,t} = 1$.
Assumption 3(i) makes sure that agents only put positive weight on their current neighbors' estimates in (5). Assumption 3(ii) ensures $\upsilon^i_{i,t} = f_{i,t}$ for all $t > 0$. Assumption 3(iii) means that if we construct an $N \times N$ weights matrix $W_{j,t}$ associated with the updates of local estimates on agent $j$'s empirical frequecies where $[W_{j,t}]_{i,l} = w^i_{jl,t}$, then the matrix $W_{j,t}$ has to be row stochastic for all times.

Assumptions 1-3 ensure that local information stored by agent $i \in \mathcal{N}$ reaches an agent $j \in \mathcal{N} \setminus \{i\}$ in finite time. Next result provides a rate for the convergence of the local copies of empirical frequencies $\upsilon^i_{j,t}$ to the actual empirical frequencies $f_{j,t}$—see [14] for the proof.

*Lemma 1 (Proposition 1, [14]):* Suppose Assumptions 1-3 hold. If $f_{j0} = \upsilon^i_{j,0}$ holds for all pairs of agents $j \in \mathcal{N}$ and $i \in \mathcal{N}$, then the local copies $\{\upsilon^i_t\}^{i \in \mathcal{N}}_{t \geq 0}$ converge to the empirical frequencies $\{f_t\}_{t \geq 0}$ with rate $O(\log t/t)$, i.e., $||\upsilon^i_{j,t} - f_{j,t}|| = O(\log t/t)$ for all $j \in \mathcal{N}$ and $i \in \mathcal{N}$.
The proof relies on the properties of row stochastic matrices formed by the weights $\{w^i_{jl,t}\}_{l \in \mathcal{N}}$. Note that we do not assume the weights form a doubly stochastic matrix, i.e., $W_{j,t}$ does not have to be column stochastic. The column stochasticity would require agents to coordinate the weights they use to update their local estimates of agent $j$'s empirical frequency.

Next, we assume the potential function is bounded.

*Assumption 4:* The potential function $u : \Delta \mathcal{A}^N \to \mathbb{R}$ of the potential game $\hat{\Gamma}$ is bounded.

Using the above assumptions, we provide a lower bound for the change in the expected potential function values between consecutive time steps.

*Lemma 2:* Suppose Assumptions 1-4 hold. Let $\Gamma$ be a $\delta$ near-potential game for some $\delta \geq 0$. The potential function is given by $u(\cdot)$. We denote the empirical frequency sequence generated by the DFP algorithm as $\{f_t\}_{t \geq 1}$. If the empirical frequency $f_t$ is outside the $\epsilon$-NE set for $\epsilon \geq 0$, then given a large enough $T > 0$ we have

$$u(f_{t+1}) - u(f_t) \geq \frac{\epsilon - N\delta}{t+1} - O\left(\frac{\log t}{t^2}\right) \quad \text{for all } t \geq T. \quad (9)$$

*Proof:* Taylor's expansion of the expected utility yields

$$u(f_{t+1}) - u(f_t) = \sum_{i=1}^{N} \sum_{a_i \in \mathcal{A}} u(a_i, f_{-i,t})\big(f_{i,t+1}(a_i) - f_{i,t}(a_i)\big) \\ + O(\|f_{i,t+1} - f_{i,t}\|^2), \quad (10)$$

where $f_{i,t+1}(a_i)$ denotes the probability of selecting action $a_i \in \mathcal{A}$ according to the empirical frequency $f_{i,t+1}$. Using the empirical frequency updates (4), we have

$$u(f_{t+1}) - u(f_t) \\ = \sum_{i=1}^{N} \sum_{a_i \in \mathcal{A}} \frac{1}{t+1} u(a_i, f_{-i,t})\big(\mathbf{1}_{(a_{i,t}=a_i)} - f_{i,t}(a_i)\big) + O\left(\frac{1}{t^2}\right) \quad (11)$$

where $\mathbf{1}_{(a_{i,t}=a_i)}$ is a $K \times 1$ vector with all elements equal to zero except the element that corresponds to $a_i$. Since $u(f_{i,t}, f_{-i,t}) = \sum_{a_i \in \mathcal{A}_i} u(a_i, f_{-i,t}) f_{i,t}(a_i)$, we have,

$$u(f_{t+1}) - u(f_t) \\ = \frac{1}{t+1} \sum_{i=1}^{N} \big(u(a_{i,t}, f_{-i,t}) - u(f_{i,t}, f_{-i,t})\big) + O\left(\frac{1}{t^2}\right). \quad (12)$$

Using $d(\Gamma, \hat{\Gamma}) \leq \delta$, the above equality and Definition 2 imply

$$u(f_{t+1}) - u(f_t) \geq \frac{1}{t+1} \sum_{i=1}^{N} (u_i(a_{i,t}, f_{-i,t}) - u_i(f_{i,t}, f_{-i,t}) - \delta) \\ + O\left(\frac{1}{t^2}\right). \quad (13)$$

By the contrapositive statement of Lemma 3 and via Lemma 1, if $f_t \notin \Sigma_\epsilon$, then there exists at least one agent $i^o \in \mathcal{N}^o \subseteq \mathcal{N}$ where $|\mathcal{N}^o| \geq 1$, whose local estimates $v_{-i_o,t}^{i_o}$ are outside $(\epsilon - O(\log t/t))$-NE region for large enough $t$, i.e., $v_{-i_o,t}^{i_o} \notin \Sigma_{\epsilon - O(\log t/t)}$. Hence, agent $i_0$'s utility value changes by at least $\epsilon - O(\log t/t)$,

$$u_{i_o}(a_{i_o,t}, v_{-i_o,t}^{i_o}) - u_{i_o}(v_{i_o,t}, v_{-i_o,t}^{i_o}) \geq \epsilon - O\left(\frac{\log t}{t}\right). \quad (14)$$

Since each agent $i \in \mathcal{N} \setminus \mathcal{N}^o$ best responds to local copies $v_{-i,t}^i$,

$$u_i(a_{i,t}, v_{-i,t}^i) - u_i(v_{i,t}, v_{-i,t}^i) \geq 0. \quad (15)$$

Using Lipschitz continuity of the mixed extension of the utility function and Lemma 1, it holds again for all $i \in \mathcal{N}$,

$$u_i(a_{i,t}, f_{-i,t}) - u_i(a_{i,t}, v_{-i,t}^i) \geq -O\left(\frac{\log t}{t}\right), \quad (16a)$$

$$u_i(v_{i,t}, v_{-i,t}^i) - u_i(f_{i,t}, f_{-i,t}) \geq -O\left(\frac{\log t}{t}\right). \quad (16b)$$

Therefore, summing the left-hand sides of (16a), (16b), with (14) or (15) for all agents $i \in \mathcal{N}$ yields,

$$\sum_{i=1}^{N}(u_i(a_{i,t}, f_{-i,t}) - u_i(f_{i,t}, f_{-i,t}) - \delta) \geq \epsilon - O\left(\frac{\log t}{t}\right). \quad (17)$$

Thus, when (17) is substituted into (13), we obtain the desired lower bound in (9). ∎

The result above implies that the potential function values will improve with each update of the algorithm in the near-potential game $\Gamma$ if the empirical frequencies are outside an $\epsilon$-equilibrium for some $\epsilon > N\delta$ and time $t$ is large enough. For the standard FP, the same improvement relation holds when we replace $O(\log(t)/t^2)$ term with $O(1/t^2)$—see [8, Lemma 5.3]. That is, the rate loss $O(\log(t))$ due to the local estimates trailing behind the actual empirical frequencies (Lemma 1) appears as a slow down in the improvement of potential function values in (9).

Since the potential function is bounded within the space of actions, it cannot increase indefinitely which implies that the sequence will enter a set of strategies with potential function values comparable to that of approximate Nash equilibria. We formalize this intuition in the following result.

*Theorem 1:* Suppose Assumptions 1-4 hold. Let $\{f_t\}_{t \geq 1}$ be the sequence of empirical frequencies generated by Algorithm 1. Then, for any $\epsilon > 0$, there exists a time $T_\epsilon$ and $\delta > 0$, such that for all $t > T_\epsilon$, $f_t \in C_{N\delta + \epsilon}$ where $C_{N\delta + \epsilon}$ is defined in (8).

*Proof:* The proof has two steps. First, we show that the region $\Sigma_{N\delta + \epsilon'}$ for some $\epsilon > \epsilon' > 0$ is visited infinitely often. Assume that there exists a long enough time $\hat{T}$ such that for $t > \hat{T}$, it holds that $f_t \notin \Sigma_{N\delta + \epsilon'}$. The difference in potential function values between consecutive time steps after long enough time $t > \hat{T}$ becomes,

$$u(f_{t+1}) - u(f_t) \geq \frac{1}{t+1}\left(N\delta + \epsilon' - N\delta - O\left(\frac{\log t}{t}\right)\right). \quad (18)$$

$$> \frac{\epsilon'}{2(t+1)} > 0. \quad (19)$$

Further, summing over the consecutive time steps for all $t > \hat{T}$ provides

$$\limsup_{t \to \infty} u(f_{t+1}) - u(f_{\hat{T}}) \geq \sum_{t=\hat{T}+1}^{\infty} \frac{\epsilon'}{2(t+1)}. \quad (20)$$

Since $u(\cdot)$ is bounded (Assumption 4), the left-hand side of the inequality above has to be finite, while the right-hand side is not. This contradicts our assumption yielding $\Sigma_{N\delta + \epsilon'}$ is infinitely visited by $f_t$ for all $t > \hat{T}$.

In the second part, we show that if $f_t \in C_{N\delta + \epsilon}$ at a large enough $t$, then $f_t$ will remain in $C_{N\delta + \epsilon}$. Consider the case that $f_t \in \Sigma_{N\delta + \epsilon'}$. Observe that $\|f_{t+1} - f_t\| = O(\frac{1}{t})$ as per (4). Thus, there exists some $T'' > \hat{T}$ and for all $t > T'' > \hat{T}$ such that we have $f_{t+1} \in \Sigma_{N\delta + \epsilon}$ by Lemma 1 and the Lipschitz continuity of the mixed extension of the potential function. Next, we
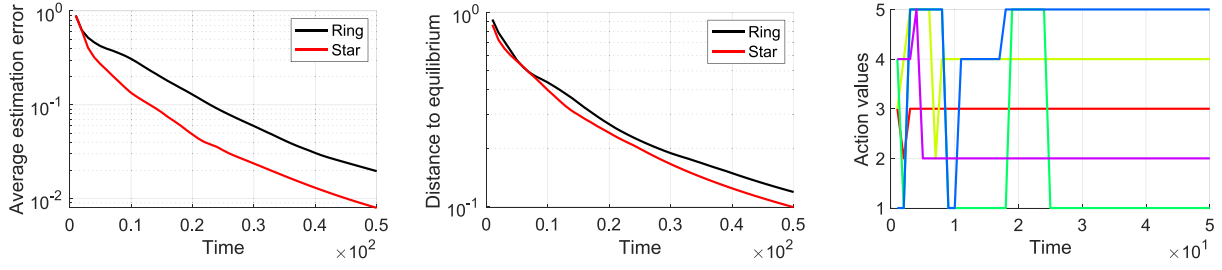
Fig. 1.  DFP in target assignment game with unknown target locations. In the ring network, each agent $i \neq N$ shares its information only with agent $i+1$, and $i = N$ sends to $i = 1$. In the star network, there exists a central agent connected to all the other agents. Network weights are set as $w_{il,t}^j = 0.75$ and $w_{jl,t}^j = 0.25/|\mathcal{N}_{i,t}|, \forall j \in \mathcal{N}_{i,t}$. We generate 20 replications with random initial positions and signals for each network type. Agents' positions are sampled from identical and independent normal distributions with mean 0 and variance 0.1 for each axis of the positions. Targets are positioned around a circle centered at origin with radius 1 with equal distances to each other. (Left) Average estimation error $\frac{1}{N(N-1)} \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N} \setminus \{i\}} ||f_{it} - v_{it}^j||$ over 20 runs. (Middle) Average distance to Nash equilibrium $\frac{1}{N} \sum_{i \in \mathcal{N}} ||f_{it} - a_i^*||$ over 20 runs. (Right) Action selections from a single run.

consider the case that $f_t \in C_{N\delta + \epsilon} \setminus \Sigma_{N\delta + \epsilon'}$. Using (18) and the definition of $C_{N\delta + \epsilon}$, it holds that

$$u(f_{t+1}) > u(f_t) \geq \min_{y \in \Sigma_{N\delta + \epsilon}} u(y). \tag{21}$$

Thus there exists some $T'' > \hat{T}$ such that if $f_t \in C_{N\delta + \epsilon}$ for $t \geq T''$, then $f_{t+1} \in C_{N\delta + \epsilon}$. ∎

The following result shows that in the limit empirical frequencies converge to $C_{N\delta}$.

*Corollary 1:* Suppose Assumptions 1-4 hold. Let $\{f_t\}_{t \geq 1}$ be generated by Algorithm 1. Then, the empirical frequencies $\{f_t\}_{t \geq 0}$, converge to the set $C_{N\delta}$.

*Proof:* Consider the set $C_{N\delta + 1/q}$ for $q \in \mathbb{Z}^+$. By Theorem 1, there exists a time $T_q$ such that for all $t > T_q, f_t \in C_{N\delta + 1/q}$ for any $q \in \mathbb{Z}^+$. Lemma 5 states that for any $\xi > 0, C_{N\delta + 1/q} \subset S_\xi$ for some $q \in \mathbb{Z}^+$, where $S_\xi$ is the $\xi > 0$ neighborhood of $C_{N\delta}$ (see (26) for a formal definition). Thus, there exists a time $T_\xi$ such that for all $t > T_\xi, f_t \in S_\xi$ for any $\xi > 0$ and some $q \in \mathbb{Z}^+$. From the definition of $S_\xi$, the following holds,

$$\limsup_{t \to \infty} \min_{z \in C_{N\delta}} ||z - f_t|| < \xi. \tag{22}$$

Since, $\xi > 0$ can be arbitrarily small and $||z - f_t|| \geq 0$, we can conclude that $\lim_{t \to \infty} \min_{z \in C_{N\delta}} ||z - f_t|| = 0$. ∎

The above corollary states that the joint empirical frequencies converge to a set of strategies that have potential values larger than the worst potential value obtained by an approximate NE strategy. Note that it also implies convergence of DFP to a NE in exact potential games, i.e., when $\delta = 0$, capturing the convergence results in [4], [15].

*Remark 1:* While our results show that the empirical frequencies $\{f_t\}_{t \geq 1}$ generated by DFP will be on par with potential function values of $N\delta$-NE strategies of the potential game ($C_{N\delta}$), it is not clear whether the sequence $\{f_t\}_{t \geq 1}$ would remain within the neighborhood of a single Nash equilibrium or visit neighborhoods of multiple Nash equilibria that belong to the set $C_{N\delta}$. For standard FP, the sequence $\{f_t\}_{t \geq 1}$ converges to a neighborhood of a single NE given that any two Nash equilibria of the game are far enough from each other— see [8, Th. 5.2]. We conjecture that $\{f_t\}_{t \geq 1}$ in DFP will also converge to a neighborhood of a single NE under the same assumption.

## V. NUMERICAL EXPERIMENTS

We consider a target assignment problem with $N = 5$ autonomous agents with the objective to cover all $K = 5$ targets using minimum effort. We represent this objective with the following utility function

$$u_i(a_i = k, a_{-i}) = \frac{\sum_{a_i \in \mathcal{A}} a_i \mathbf{1}_{a_{-ik} = 0}}{\sum_{a_i \in \mathcal{A}} a_i d_{ik}}, \tag{23}$$

where $k \in \mathcal{A}$, and $\mathbf{1}_{a_{-ik} = 0} \in \{0, 1\}$ is a binary value that it is equal to 1 if none of the other agents $j \in \mathcal{N} \setminus \{i\}$ select target $k$, and otherwise it is equal to 0. The distance vector of agent $i$ is $d_i = [d_{i1}, \ldots, d_{ik}, \ldots, d_{iK}] \in \mathbb{R}_+^K$, where $d_{ik} = ||\theta_i - \theta_k||$ is the distance between position vectors of agent $i, \theta_i \in \mathbb{R}^2$ and target $k, \theta_k \in \mathbb{R}^2$ in a 2D plane. According to the utility function above, agent $i$ only receives a positive payoff by selecting target $k$, if no other agent selects the target $k$. In this case, the payoff agent $i$ receives is inversely proportional to the distance of the agent to the target selected. Hence, no agent has more utility by changing its target, if other agents cover remaining targets. The payoff function ensures that any action profile that is one-to-one assignment between agents and targets is a NE.

The target positions are unknown. Each agent receives private signals $\vartheta_{kt}^i$ at each time step $t$ about the position of each target $k$. The private signals $\vartheta_t^i = [\vartheta_1^i, \ldots, \vartheta_K^i]^T$ for each agent $i$ come from a multivariate normal distribution with mean $\theta = [\theta_1, \ldots, \theta_K]^T$ and covariance matrix $\sigma I$, where $\sigma = 0.5$ and $I \in \mathbb{R}^{K \times K}$ is the identity matrix. We assume agents receive signals up until time $\tau = 10$. At time $\tau$, agent $i$'s point estimate of target $k$'s position is given by $\hat{\theta}_k = (1/\tau) \sum_{t=1}^\tau \vartheta_t^i$. Final time $T_f$ is set to 50. If the beliefs and distances were identical, the game with payoffs in (23) is a potential game. The existence of different beliefs and different distances to targets creates a near-potential game.

The final joint action profile $a_{T_f}$ is equal to a pure NE $a^*$ in all of the replications for both network types validating Theorem 1 and Corollary 1. Fig. 1(Left) verifies the convergence rate $O(\log t/t)$ of estimation error as it goes to 0. Fig. 1(Middle) shows the average rate of convergence of the empirical frequencies to a NE of the target assignment game. Agents learn each others' empirical frequencies faster in the

star network. Fig. 1(Right) shows that around $T_f/2 = 25$ agents reach a NE.

## VI. CONCLUSION

We studied the convergence properties of DFP in near-potential games when the communication network is time-varying. We showed empirical frequencies of actions converge to a set of strategies with potential function values that are large enough compared to approximate Nash equilibria. These results established that convergence properties of DFP are identical to that of FP in near-potential games.

## APPENDIX

*Lemma 3:* Let $(\sigma', \sigma'') \in \Delta\mathcal{A}^N \times \Delta\mathcal{A}^N$ be strategies such that $||\sigma' - \sigma''|| \leq \xi$, for a small enough $\xi > 0$. If $\sigma' \in \Sigma_\alpha$, then $\sigma'' \in \Sigma_{\alpha+\beta}$, where $\Sigma_\alpha$ and $\Sigma_{\alpha+\beta}$ are $\alpha$−NE and $\alpha + \beta$−NE sets with respect to the order given $\alpha \geq 0$ and $\beta \geq 0$.

*Proof:* To define an approximate $\epsilon$−NE set, let $\psi : \Delta\mathcal{A}^N \to \mathbb{R}$ be a function as follows,

$$\psi(\sigma) = -\max_{i\in\mathcal{N}, a_i\in\mathcal{A}} (u_i(a_i, \sigma_{-i}) - u_i(\sigma_i, \sigma_{-i})). \quad (24)$$

Then, it holds that a strategy profile $\sigma = (\sigma_i, \sigma_{-i})$ is an $\epsilon$−NE, if and only if $\psi(\sigma) \geq -\epsilon$. By Lipschitz continuity of expected the utility function (3), the difference $u_i(a_i, \sigma_{-i}) - u_i(\sigma_i, \sigma_{-i})$ is Lipschitz continuous. Using the fact that $\psi(\sigma)$ is defined as maximum over finite set of such differences, $\psi(\sigma)$ is also Lipschitz continuous. Hence, there exists a Lipschitz constant $L_\psi \in \mathbb{R}^+$ such that for any $||\sigma' - \sigma''|| \leq \xi \leq \beta/L_\psi$, it holds, $|\psi(\sigma') - \psi(\sigma'')| \leq \beta$. Thus, by the definitions of $\psi : \Delta\mathcal{A}^N \to \mathbb{R}$ and $\epsilon$−NE in (7), if $\sigma' \in \Sigma_\alpha$, then $\sigma'' \in \Sigma_{\alpha+\beta}$. ∎

We use the notion of an upper semi-continuous correspondence in the next set of results.

*Definition 5 (Upper Semi-continuous Correspondence):* A correspondence $h : X \Rightarrow Y$ is upper semi-continuous, if one of the following statements hold,

- For any $\bar{x} \in X$ and any open neighborhood $V$ of $h(\bar{x})$, there exists a neighborhood $U$ of $\bar{x}$, such that $h(x) \subset V$, and $h(x)$ is a compact set for all $x \in U$.
- $Y$ is compact, and the set, i.e., its graph, $\{(x, y)|x \in X, y \in h(x)\}$ is closed.

*Lemma 4:* Let $h : \mathbb{R} \Rightarrow \Delta\mathcal{A}^N$ be the correspondence representing the set of $\alpha$-NE strategies, i.e.,

$$h(\alpha) = \Sigma_\alpha = \{\sigma \in \Delta\mathcal{A}^N | \psi(\sigma) \geq -\alpha\} \quad (25)$$

where $\psi$ is defined in (24). Then, the correspondence $h : \mathbb{R} \Rightarrow \Delta\mathcal{A}^N$ is upper semi-continuous.

*Proof:* Since $\psi$ is a (Lipschitz) continuous function from Lemma 3, the set $h(\alpha)$ is closed by [8]. As a result, the graph of the correspondence $h$ is closed. Hence, $\Sigma_\alpha$ is compact as $\Sigma_\alpha \subseteq \Delta\mathcal{A}^N$ is bounded, which satisfies Definition 5. ∎

*Lemma 5:* Let $C_{N\delta}$ and $C_{N\delta+1/q}$ for $q \in \mathbb{Z}^+$ be the closed sets defined as in (8). Further, we define the set $S_\xi$ for $\xi > 0$ as follows,

$$S_\xi := \{\sigma \in \Delta\mathcal{A}^N \mid \min_{y\in C_{N\delta}} ||\sigma - y|| < \xi\}. \quad (26)$$

Then, for any $\xi > 0$ and for some $q \in \mathbb{Z}^+$, it holds that $C_{N\delta+1/q} \subset S_\xi$.

*Proof:* Using the properties of upper semi-continuity (Definition 5), there exists $\zeta > 0$ such that the set $\{\sigma \in \Delta\mathcal{A}^N | u(\sigma) \geq \min_{y\in\Sigma_{N\delta}} u(y) - \zeta\}$ is a subset of $\xi$-neighborhood of the set $C_{N\delta}$, i.e., $\{\sigma \in \Delta\mathcal{A}^N \mid u(\sigma) \geq \min_{y\in\Sigma_{N\delta}} u(y) - \zeta\} \subseteq S_\xi$. Next, since the correspondence $h$ (25) is upper semi-continuous by Lemma 4, for any $\xi' > 0$ there exists a large enough $q$ such that $\Sigma_{N\delta+1/q}$ is contained in $\xi'$ neighborhood of $\Sigma_{N\delta}$. That is, for any $\xi' > 0$, for any point $z \in \Sigma_{N\delta+1/q}$, there exists $y \in \Sigma_{N\delta}$ such that $||z - y|| \leq \xi'$. Given that potential function $u$ is Lipschitz continuous with Lipschitz constant $L$, and defining $\zeta/L \geq \xi' > 0$, the following holds for large enough $q$,

$$\min_{z\in\Sigma_{N\delta+1/q}} u(z) \geq \min_{y\in\Sigma_{N\delta}} u(y) - \zeta. \quad (27)$$

Hence, for any $z \in C_{N\delta+1/q}$, the inequality (27) is satisfied. Thus, for large enough $q$ and any $z \in C_{N\delta+1/q}$, it also holds $z \in S_\xi$ and $C_{N\delta+1/q} \subset S_\xi$. ∎

## REFERENCES

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

[2] J. R. Marden, G. Arslan, and J. S. Shamma, "Joint strategy fictitious play with inertia for potential games," *IEEE Trans. Autom. Control*, vol. 54, no. 2, pp. 208–220, Feb. 2009.

[3] S. M. Perlaza, S. Lasaulce, H. Tembine, and M. Debbah, "Learning to use the spectrum in self-configuring heterogenous networks: A logit equilibrium approach," in *Proc. 5th Int. ICST Conf. Perform. Eval. Methodol. Tools*, 2011, pp. 565–571.

[4] C. Eksin and A. Ribeiro, "Distributed fictitious play for multiagent systems in uncertain environments," *IEEE Trans. Autom. Control*, vol. 63, no. 4, pp. 1177–1184, Apr. 2017.

[5] D. Monderer and L. S. Shapley, "Fictitious play property for games with identical interests," *J. Econ. Theory*, vol. 68, no. 1, pp. 258–265, 1996.

[6] B. Swenson, S. Kar, and J. Xavier, "Single sample fictitious play," *IEEE Trans. Autom. Control*, vol. 62, no. 11, pp. 6026–6031, Nov. 2017.

[7] O. Candogan, I. Menache, A. Ozdaglar, and P. A. Parrilo, "Flows and decompositions of games: Harmonic and potential games," *Math. Oper. Res.*, vol. 36, no. 3, pp. 474–503, 2011.

[8] O. Candogan, A. Ozdaglar, and P. A. Parrilo, "Dynamics in near-potential games," *Games Econ. Behav.*, vol. 82, pp. 66–90, Nov. 2013.

[9] M. O. Sayin, F. Parise, and A. Ozdaglar, "Fictitious play in zero-sum stochastic games," 2021. [Online]. Available: arXiv:2010.04223.

[10] D. S. Leslie, S. Perkins, and Z. Xu, "Best-response dynamics in zero-sum stochastic games," *J. Econ. Theory*, vol. 189, Sep. 2020, Art. no. 105095.

[11] S. Perrin, J. Perolat, M. Laurière, M. Geist, R. Elie, and O. Pietquin, "Fictitious play for mean field games: Continuous time analysis and applications," in *Proc. NeurIPS*, 2020, pp. 1–6. [Online]. Available: https://arxiv.org/abs/2007.03458

[12] H. P. Young, *Strategic Learning and Its Limits*. Oxford, U.K.: Oxford Univ. Press, 2004.

[13] A. Nedic and A. Ozdaglar, "Distributed subgradient methods for multiagent optimization," *IEEE Trans. Autom. Control*, vol. 54, no. 1, pp. 48–61, Jan. 2009.

[14] S. Arefizadeh and C. Eksin, "Distributed fictitious play in potential games with time varying communication networks," in *Proc. IEEE 53rd Asilomar Conf. Signals Syst. Comput.*, 2019, pp. 1755–1759.

[15] B. Swenson, S. Kar, and J. Xavier, "Empirical centroid fictitious play: An approach for distributed learning in multi-agent games," *IEEE Trans. Signal Process.*, vol. 63, no. 15, pp. 3888–3901, Aug. 2015.