Proportional and Double Imitation Rules for Spectrum Access in Cognitive Radio Networks☆

Stefano Iellamo[a], Lin Chen[b], Marceau Coupechoux[a,*]

[a]*Telecom ParisTech - CNRS LTCI, 46 Rue Barrault, Paris 75013, France.*
Tel: +33 1 45 81 75 88. Fax: +33 1 45 81 31 19
[b]*University Paris-Sud XI - CNRS LRI, 91405 Orsay, France*

Abstract

In this paper, we tackle the problem of opportunistic spectrum access in large-scale cognitive radio networks, where the unlicensed Secondary Users (SU) access the frequency channels partially occupied by the licensed Primary Users (PU). Each channel is characterized by an availability probability unknown to the SUs. We apply population game theory to model the spectrum access problem and develop distributed spectrum access policies based on imitation, a behavior rule widely applied in human societies consisting of imitating successful behaviors. We develop two imitation-based spectrum access policies based on the basic Proportional Imitation (PI) rule and the more advanced Double Imitation (DI) rule given that a SU can only imitate the other SUs operating on the same channel. A systematic theoretical analysis is presented for both policies on the induced imitation dynamics and the convergence properties of the proposed policies to the Nash Equilibrium. Simple and natural, the proposed imitation-based spectrum access policies can be implemented distributedly based on solely local interactions and thus is especially suited in decentralized adaptive learning

*Corresponding author
Email addresses: `stefano.iellamo@telecom-paristech.fr` (Stefano Iellamo), `chen@lri.fr` (Lin Chen), `marceau.coupechoux@telecom-paristech.fr` (Marceau Coupechoux)

environments as cognitive radio networks.

1. Introduction

Cognitive radio [1], with its capability to flexibly configure its transmission parameters, has emerged in recent years as a promising paradigm to enable more efficient spectrum utilization. Spectrum access models in cognitive radio networks can be classified into three categories, namely exclusive use (or operator sharing), commons and shared use of primary licensed spectrum [2]. In the last model, unlicensed secondary users (SU) are allowed to access the spectrum of licensed primary users (PU) in an opportunistic way. In this case, a well-designed spectrum access mechanism is crucial to achieve efficient spectrum usage.

In this paper, we focus on the generic model of cognitive networks consisting of multiple frequency channels, each characterized by a channel availability probability determined by the activity of PUs on it. In such a model, from the SUs perspective, a challenging problem is to coordinate with other SUs in order to opportunistically access the unused spectrum of PUs to maximize its own payoff (e.g., throughput); at the system level, a crucial research issue is to design efficient spectrum access protocols achieving optimal spectrum usage and load balancing on the available channels.

We tackle the spectrum access problem in large-scale cognitive radio networks from an evolutionary game theoretic angle. We formulate the spectrum access problem, show the existence of a Nash Equilibrium (NE) and develop distributed spectrum access policies based on imitation, a behavior rule widely applied in human societies consisting of imitating successful behavior. We study the system dynamics and the convergence of the proposed policies to the NE

when the SU population is large. Simple and natural, the proposed spectrum access policies can be implemented distributedly based on solely local interactions and thus is especially suited in decentralized adaptive learning environments as cognitive radio networks.

In our analysis, we develop imitation-based spectrum access policies where a SU can only imitate the other SUs operating on the same channel. More specifically, we propose two spectrum access policies based on the following two imitation rules: the Proportional Imitation (PI) rule where a SU can sample one other SU; the more advanced adjusted proportional imitation rule with double sampling (Double Imitation, DI) where a SU can sample two other SUs. Under both imitation rules, each SU strives to improve its individual payoff by imitating other SUs with higher payoff. A systematic theoretical analysis is presented for both policies on the induced imitation dynamics and the convergence properties of the proposed policies to the NE.

The key contribution of our work in this paper lies in the systematical application of the natural imitation behavior to address the spectrum access problem in cognitive radio networks, the design of distributed imitation-based channel access policies, and the theoretic analysis on the induced imitation dynamics and the convergence to an efficient and stable system equilibrium. In this paper, we extend the results of [3], where it is assumed that SUs are able to immediately and uniformly imitate any other SU. This assumption makes the theoretical analysis straightforward from the literature on imitation. We assume here that SUs can only imitate SUs on the same channel and obtain a delayed information, as a result of which significant changes should be done in terms of policy design and theoretical analysis.

The rest of the paper is structured as follows. Section 2 discusses related work in the literature. Section 3 presents the system model and Section 4

presents the formulation of the spectrum access game. Section 5 describes the proposed imitation-based spectrum access policies and motivates the choices of proportional and double imitation rules as basis of our policies. In Section 6, we study the system dynamics and the convergence of our algorithms. Section 7 discusses the assumptions of our network model. Section 8 presents simulation based performance evaluation, where our schemes are compared to another decentralized approach called Trial and Error. Section 9 concludes the paper.

2. Related Work

The problem of distributed spectrum access in cognitive radio networks (CRN) has been widely addressed in the literature. A first set of papers assumes that the number of SUs is smaller than the number of channels. In this case, the problem is closely related to the classical Multi-Armed Bandit (MAB) problem [4]. Some recent work has investigated the issue of adapting traditional MAB approaches to the CRN context, among which Anandkumar *et al.* proposed two algorithms with logarithmic regret, where the number of SUs is known or estimated by each SU [5]. Contrary to this literature, we assume in our paper a large population of SUs, able to share the available bandwidth when settling on the same channel.

Another important thrust consists of applying game theory to model the competition and cooperation among SUs and the interactions between SUs and PUs (see [6] for a review). Several papers propose for example algorithms based on no-regret learning (e.g. [7, 8]), which are not guaranteed to converge to the NE. Besides, due to the perceived fairness and allocation efficiency, auction techniques have also attracted considerable research attention and resulted in a number of auction-based spectrum allocation mechanisms (cf. [9] and references therein). The solution proposed in this paper differs from the existing

4

approaches in that it requires only local interactions among SUs and is thus naturally adapted in the distributed environments as CRNs.

Due to the success of applying evolutionary [10] and population game theories [11] in the study of biological and economic problems [11], a handful of recent studies have applied these tools to study resource allocation problems arisen from wired and wireless networks (see e.g. [12, 13]), among which Shakkottai *et al.* addressed the problem of non-cooperative multi-homing of users to WLANs access points by modeling it as a population game [14]. Authors however focus on the system dynamics rather than on the distributed algorithms as we do in this paper. Niyato *et al.* studied the dynamics of network selection in a heterogeneous wireless network using the theory of evolutionary game [15]. The proposed algorithm leading to the replicator dynamics is however based on a centralized controller able to broadcast to all users the average payoff. Our algorithms are on the contrary fully distributed. Coucheney *et al.* studied the user-network association problem in wireless networks with multi-technology and proposed an algorithm based on trial and error mechanisms to achieve the fair and efficient solution [13].

Several theoretical works focus on imitation dynamics. Ackermann *et al.* investigated the concurrent imitation dynamics in the context of finite population symmetric congestion games by focusing on the convergence properties [16]. Berenbrik *et al.* applied the Proportional Imitation Rule to load-balance system resources by focusing on the convergence speed [17]. Ganesh *et al.* applied the Imitate If Better rule[1] (see [19] for a review on imitation rules) in order to load-balance the service rate of parallel server systems [18]. Contrary to our work, it is assumed in [17] and [18] that a user is able to observe the load of

---

[1] Imitate If Better (IIB) is a rule consisting in picking a player and migrating to its strategy if the latter has yielded a higher payoff than the achieved one. IIB is called Random Local Search in [18]

another resource before taking its decision to switch to this resource.

As it is supposed to model human behavior, imitation is mostly studied in economics. In the context of CRN, specific protocol or hardware constraints may however arise so that imitation dynamics are modified, as we show it in this paper. Two very recent works in the context of CRN are [20] and [21], which have the same goals as ours. In [20], authors propose a distributed learning algorithm for spectrum access. User decisions are based on their accumulated experience and they are using a mixed strategy. In [21], imitation is also used for distributed spectrum access. However, the proposed scheme relies on the existence of a common control channel for the sampling procedure. Double imitation is moreover not considered.

3. System Model

In this section, we present the system model of our work with the notations used.

3.1. System Model and PU Operation

We consider the network model shown on Fig. 1 made of a primary network and a secondary network. In the former, a primary transmitter is using on the downlink a set $\mathcal{C}$ of C frequency channels, each with bandwidth B. The primary receivers are operated in a synchronous time-slotted fashion. The secondary network is made of a set $\mathcal{N}$ of N SUs, which try to opportunistically access the channels when they are left free by the PU. We assume that SUs can perform a perfect sensing of PU transmissions, i.e., no collision can occur between the PU and SUs. This assumption is adopted in the literature focusing on resource allocation (see e.g. [22, 23, 5]). The secondary network is also supposed to be sufficiently small so that every SU can receive and decode packets sent on the same channel.
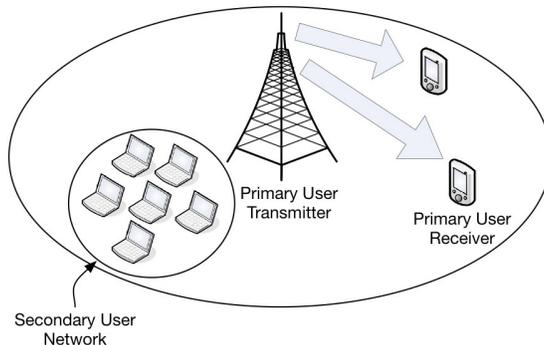
6

Figure 1: Network model: N SUs try to opportunistically access the spectrum left free by a PU. The spectrum is made of C slotted frequency channels.

Let $Z_i(k)$ be the random variable equal to 1 when of channel i is unoccupied by any PU at slot k and 0 otherwise. We assume that the process $\{Z_i(k)\}$ is stationary and independent for each i and k, i.e., the $Z_i(k)$ are i.i.d. random variables for all (i, k). We also assume that at each time slot, channel i is free with probability μ_i, i.e., $\mathbb{E}[Z_i(k)] = \mu_i$. Without loss of generality, we assume $\mu_1 \geq \mu_2 \geq ... \geq \mu_C$. The channel availability probabilities $\mu \triangleq \{\mu_i\}$ are *a priori* not known by SUs.

3.2. SU Operation

We describe in this section the SU operation and capabilities. As shown in Fig. 2 for a given frequency channel j, time-slots of the primary network are organized into blocks of N_b slots. All SUs are assumed to be synchronized, stay on the same channel during a block and may change their channel at block boundary. Let n_i be the number of SUs operating on channel i.

Assuming perfect sensing of the cognitive users, there is no secondary transmission during slots occupied by the PU (grey slots on the figure). When the PU is idle, SUs share the available bandwidth using a decentralized random access MAC protocol (hatched slots on the figure). The way this MAC protocol is implemented is out of the scope of the paper. Mini-slots can for example be

used at the beginning of each slot in order to perform CSMA, as assumed in [20], or CSMA/CA can be used, as assumed in [24]. For mathematical convenience, we will assume in Sections 5 and 6 that the MAC protocol is perfect and operates like TDMA. Our motivation of such an assumption is to concentrate the analysis on the interactions between SUs and the resulting structural properties of the system equilibria. The results give an upper-bound on the performance of the developed policies. A similar asymptotic analysis has been carried on in [25].

In our work, each SU j is modeled as a rational decision maker, striking to maximize the throughput it can achieve, denoted as T_j^i when j operates on channel i. Assuming a fair MAC protocol and invoking symmetry reasons, all SUs on channel i obtain the same expected throughput, which can be expressed as a function of n_i as $\pi_i(n_i) = \mathbb{E}[T_j^i]$ for all j operating on channel i. It should be noted that $\pi_i(n_i)$ depends on the MAC protocol implemented at the cognitive users. An example is $\pi_i(n_i) = B\mu_i/n_i$ in the case of a perfect MAC protocol operating like TDMA, where B is a constant standing for the channel bandwidth. Generically, $\pi_i(n_i)$ can be rewritten as $\pi_i(n_i) = B\mu_i S(n_i)$ where $S(n_i)$ denotes the throughput of a channel of unit bandwidth without PU. Without loss of generality, we will now assume that $B = 1$. The assumption that SUs on the same channel obtain the same expected throughput can be found in the literature using evolutionary game theory to study spectrum access, see e.g. [15, 26, 20].

Channel availabilities, μ_i, are estimated in the long term by SUs, while the expected throughput π_i and the number of SUs n_i are estimated at the end of each block. In all their transmissions in block b, SUs include in the header of their packets the throughput $\pi_i(n_i)$ obtained in block $b-1$ and the corresponding channel (or strategy) i. We further assume that every SU can overhear at random one or two packets transmitted by SUs on the same channel and decode
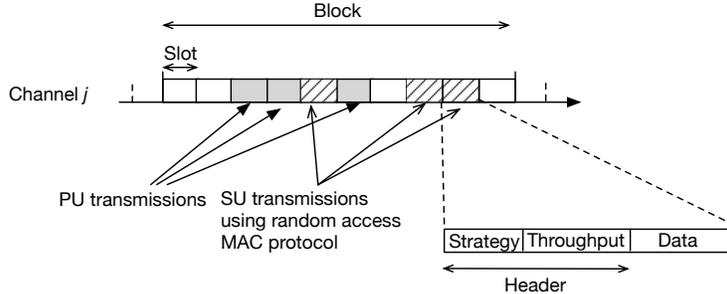
Figure 2: SU operation: N_b slots of a frequency channel form a block, SUs use a MAC protocol for their transmissions and include in the header of their packets a throughput and strategy indication.

the throughput and strategy indications. The overhearing of packets is called a *sampling procedure* and we write $i \rightsquigarrow j$ when SU i samples SU j. Sampling is supposed to be symmetric, i.e., the probabilities $P(i \rightsquigarrow j)$ and $P(j \rightsquigarrow i)$ are identical. After the sampling, SU transmitters communicate to their receiver a channel change order to be executed at the next block boundary.

4. Spectrum Access Game Formulation

To study the interactions among autonomous SUs and to derive distributed channel access policies, we formulate in this section the channel selection problem as a spectrum access game where the players are the SUs and we show the uniqueness of the Nash Equilibrium (NE) when the number of SUs is large. The game is defined formally as follows:

Definition 1. *The spectrum access game G is a 3-tuple $(\mathcal{N}, \mathcal{C}, \{U_j\})$, where $\mathcal{N}$ is the player set, $\mathcal{C}$ is the strategy set of each player. Each player j chooses its strategy $s_j \in \mathcal{C}$ to maximize its payoff function U_j defined as $U_j = \pi_{s_j}(n_{s_j}) = \mathbb{E}[T_j^{s_j}]$.*

The solution of the spectrum access game G is characterized by a Nash Equilibrium [27], a strategy profile from which no player has incentive to deviate

9

unilaterally.

Lemma 1. *For the spectrum access game G, there exists at least one Nash equilibrium.*

Proof. Given the form of the SU payoff function, it follows from [28] that the spectrum access game is a congestion game and a potential game with potential function: $P(n_1, ..., n_C) = \sum_{i \in \mathcal{C}} \sum_{k=1}^{n_i} \pi_i(n_i)$, where n_i is the number of SUs on channel i and $\sum_i n_i = N$. This function takes only a finite set of values and thus achieves a maximum value. $\square$

We now consider the *population* game G, where (1) the number of SUs is large, (2) SUs are small, (3) SUs interact anonymously and (4) payoffs are continuous (see [11] for the discussion on these assumptions). In this model, we focus on the system state $\mathbf{x} \triangleq \{x_i, i \in \mathcal{C}\}$ where x_i denotes the proportion of SUs choosing channel i. In such context, by regarding x_i as a continuous variable, we make the following assumption on the throughput function $S(x_i N)$.

Assumption 1. $S(x_i N)$ *is strictly monotonously decreasing and it holds that* $S(x_i N) \leq 1/(N x_i)$.

We can now establish the uniqueness of the NE in the spectrum access game G for the asymptotic case in the following lemma and theorem.

Lemma 2. *For N sufficiently large, there is no empty channel at NE.*

Proof. Assume, by contradiction, that at a NE, there are no SUs on channel i. Since there are C channels, at a NE, there exists at least one channel where there are at least N/C SUs. Assume that this channel is channel j, i.e., $n_j \geq N/C$. Consider a SU on channel j, its payoff is $\pi_j(n_j) = \mu_j S(N/C)$. From Assumption 1, $\pi_j(n_j) \leq \mu_j C/N$. Now let a SU in channel j switch to channel i, its payoff becomes $\pi_i(1) = \mu_i S(1)$. It holds straightforwardly that $\pi_j(n_j) < \pi_i(1)$ when $N > \frac{\mu_j C}{\mu_i S(1)}$. Hence there is no empty channel at NE $\square$

Theorem 1. *For N sufficiently large, G admits a unique NE, where all SUs get the same payoff. Let y denote the root of $\sum_{i \in \mathcal{C}} S^{-1}\left(\frac{y}{\mu_i}\right) = N$, at the NE, there are $S^{-1}\left(\frac{y}{\mu_i}\right)$ SUs operating on channel i.*

Proof. This theorem is a classical result of population games. See Appendix A for more details. □

We can observe two desirable properties of the unique NE derived in Theorem 1: (1) the NE is optimal from the system perspective as the total throughput of the network achieves its optimum at the NE; (2) at NE, all SUs obtain exactly the same throughput. Note that any state such that $x_i > 0$ for all $i \in \mathcal{C}$ is also system optimal, the NE is one of them. Note also that when N grows indefinitely and as players are symmetric, the NE approaches the Wardrop equilibrium of the system [29].

One critical challenge in the analyzed spectrum access game is the design of distributed spectrum access strategies for rational SUs to converge to the NE. In response to this challenge, we develop in the sequel sections of this paper an efficient spectrum access policy.

5. Imitation-based spectrum access policies

The spectrum access policy we develop is based on *imitation*. As a behavior rule widely observed in human societies, imitation captures the behavior of a bounded rational player that mimics the actions of other players with higher payoff in order to improve its own payoff from one block to the next, while ignoring the effect of its strategy on the future evolutions of the system and forgetting its past experience. The induced imitation dynamics model the spreading of successful strategies under imitation [30]. In this section, we develop two spectrum access policies based on the proportional imitation rule and the double

imitation rule. For tractability reasons, we assume in the next sections that $\pi_i(n_i) = \mu_i/n_i$ on a channel i, i.e., a perfect MAC protocol for SUs.

5.1. Motivation

In this first part, we recall some useful definitions given in [30], we introduce new notations and we provide our motivations.

Definition 2. *A* behavioral rule with single sampling *(resp.* with double sampling*) is a function $F : \mathcal{C}^2 \to \Delta(\mathcal{C})$ (resp. $F : \mathcal{C}^3 \to \Delta(\mathcal{C})$), where $\Delta(\mathcal{C})$ is the set of probability distributions on $\mathcal{C}$ and $F_{i,j}^k$, $\forall i, j, k \in \mathcal{C}$ (resp. $F_{i,j,l}^k$, $\forall i, j, k, l \in \mathcal{C}$) is the probability of choosing channel k in the next iteration (block) after operating on channel i and sampling a SU with strategy j (resp. sampling two SUs with strategies j and l).*

Definition 3. *A behavioral rule with single sampling is* imitating *if $F_{i,j}^k = 0$ when $k \notin \{i, j\}$. A behavioral rule with double sampling is* imitating *if $F_{i,j,l}^k = 0$ when $k \notin \{i, j, l\}$.*

In this paper, we assume that all SUs adopt the same behavioral rule, i.e., the population is *monomorphic* in the sense of Schlag [30] (see e.g. [31, 32, 33] for other papers using this notion).

Schlag has shown in [30] that the proportional imitation rule (PIR) is an *improving* rule, i.e., in any state of the system the expected average payoff is increasing after an iteration of the rule. He has also shown that it is a *dominant* rule, i.e., it always achieves a higher expected payoff improvement than any other improving rule. PIR is moreover the unique dominant rule that never imitates a strategy that achieved a lower payoff and that minimizes the probability of switching among the set of dominant rules.

Schlag has also shown in [34] that the double imitation (DI) rule is the rule that causes less SUs to change their strategy after each iteration among the set

12

of improving behavioral rules with double sampling. As switching may represent a significant cost for today's wireless devices in terms of delay, packet loss and protocol overhead, this property makes PIR and DI particularly attractive. These properties motivate the design of spectrum access policies based on PIR and DI.

5.2. Spectrum Access Policy Based on Proportional Imitation

Algorithm 1 presents our proposed spectrum access policy based on the proportional imitation rule, termed as PISAP. The core idea is: At each iteration t, each SU (say j) randomly selects another SU (say j') on the same channel; if the payoff at $t-1$ of the selected SU (denoted $U_{j'}(t-1)$) is higher than its own payoff at $t-1$ (denoted $U_j(t-1)$), the SU imitates the strategy of the selected SU at the next iteration with a probability proportional to the payoff difference, with coefficient the imitation factor σ^2. The payoff and the strategy at $t-1$ of the sampled SU are read from the packet header.

Algorithm 1 PISAP: Executed at each SU j

1: **Initialization**: Set the imitation factor σ
2: At $t = 0$, randomly choose a channel to stay and store the payoff $U_j(0)$.
3: **while** at each iteration $t \geq 1$ **do**
4: Randomly select a SU j'
5: **if** $U_j(t-1) < U_{j'}(t-1)$ **then**
6:   Migrate to the channel $s_{j'}(t-1)$ with probability $p = \sigma(U_{j'}(t-1) - U_j(t-1))$
7: **end if**
8: **end while**

5.3. Spectrum Access Policy Based on Double Imitation

In this subsection, we turn to a more advanced imitation rule, the double imitation rule [34] and propose the DI-based spectrum access policy, termed

---

[2] One way of setting σ is to set $\sigma = 1/(\omega - \alpha)$, where ω and α are two exogenous parameters such that $U_j \in [\alpha, \omega], \forall j \in \mathcal{C}$. In our case, $\omega = 1$ and $\alpha = 0$ can be chosen.

as DISAP. Under DISAP, each SU randomly samples two SUs on the same channel by decoding two packet headers. It then imitates them with a certain probability determined by the payoff differences. The spectrum access policy based on the double imitation is detailed in Algorithm 2, in which each SU j (with payoff U_j and strategy i at $t-1$) randomly samples two other SUs j_1 and j_2 (operating at $t-1$ on channel i_1 and i_2 respectively and, without loss of generality, with utilities $U_{j_1} \leq U_{j_2}$) and updates the probabilities of switching to channels i_1 and i_2, denoted as p_{j_1} and p_{j_2} respectively.

5.4. Discussion

As pointed in [34], double imitation may seem complicated compared to the proportional imitation rule. We can however extract the following properties [34]: DI is an imitating rule, i.e., a SU never chooses a channel that is not in his sample; switching probabilities are continuous in the sampled payoffs and increase with payoff differences; for a joint sample with three different channels, the most successful channel is chosen more likely; a SU never imitates another SU that obtains a lower payoff.

Note that PISAP is clearly different from the proportional imitation rule presented in [30] in the sense that a SU is not able to uniformly sample another SU across the network. The radio constraint indeed forces it to sample a SU on the same channel. As in this case, current strategies are identical, imitation is based on the previous iteration.

If every SU is able to uniformly sample another SU in the network and to imitate the current strategy, the system dynamics is straightforward to obtain.It is indeed shown e.g. in [35] that in the asymptotic case (assuming continuous time for simplicity), the proportional imitation rule generates a population dynamics

14

Algorithm 2 DISAP: Executed at each SU j.

1: **Initialization**: Set the parameters ω, α and $\sigma = 1/(\omega - \alpha)$.
    Define $[A]^+ \triangleq \max\{0, A\}$ and $Q(U) \triangleq 2 - \frac{U-\alpha}{\omega-\alpha}$.

2: At $t = 0$ and $t = 1$, randomly choose a channel and store the payoff $U_j(0)$.

3: **while** at each iteration $t \geq 2$ **do**

4:    Let i and U_j be resp. the channel and the payoff of j at $t - 1$.

5:    Randomly sample two SUs j_1 and j_2 (with channels i_1 and i_2 and with payoffs U_{j_1} and U_{j_2} resp. at $t-1$). Suppose w.l.o.g. that $U_{j_1} \leq U_{j_2}$.

6:    **if** $|\{i, i_1, i_2\}| = 1$, i.e., $i = i_1 = i_2$ **then**

7:        Go to channel i.

8:    **else if** $|\{i, i_1, i_2\}| = 2$ **then**

9:        **if** $i = i_1$, $i \neq i_2$ and $U_j \leq U_{j_2}$ **then**

10:            $p_{j_2} = \frac{\sigma}{2} Q(U_j)(U_{j_2} - U_j)$.
            Switch to channel i_2 w.p. p_{j_2} and go to channel i w.p. $1 - p_{j_2}$.

11:        **else if** $i_1 = i_2$, $i \neq i_1$ and $U_j \leq U_{j_1} = U_{j_2}$ **then**

12:            $p_{j_1} = \frac{\sigma}{2}(Q(U_{j_1}) + Q(U_j))(U_{j_1} - U_j)$.
            Switch to channel i_1 w.p. p_{j_1} and go to channel i w.p. $1 - p_{j_1}$.

13:        **end if**

14:    **else if** $|\{i, i_1, i_2\}| = 3$ **then**

15:        **if** $U_j \leq U_{j_1} \leq U_{j_2}$ **then**

16:            $p_{j_1} = \frac{\sigma}{2} \left[Q(U_j)(U_{j_1} - U_{j_2}) + Q(U_{j_2})(U_{j_1} - U_j) \right]^+$.
            $p_{j_2} = \frac{\sigma}{2} \left[Q(U_{j_1})(U_{j_2} - U_j) + Q(U_{j_2})(U_{j_1} - U_j) \right] - p_{j_1}$.
            Switch to channel i_1 w.p. p_{j_1}, to channel i_2 w.p. p_{j_2} and go to channel i w.p. $1 - p_{j_1} - p_{j_2}$.

17:        **else if** $U_{j_1} \leq U_j \leq U_{j_2}$ **then**

18:            $p_{j_2} = \frac{\sigma}{2} \left[Q(U_{j_1})(U_{j_2} - U_j) + Q(U_{j_2})(U_{j_1} - U_j) \right]^+$.
            Switch to channel i_2 w.p. p_{j_2} and go to channel i w.p. $1 - p_{j_2}$.

19:        **end if**

20:    **else**

21:        Go to channel i.

22:    **end if**

23: **end while**

---

described by the following set of differential equations:

$$\dot{x}_i(t) = \sigma x_i(t)[\pi_i(t) - \bar{\pi}(t)], \quad \forall i \in \mathcal{C}, \tag{1}$$

where $\bar{\pi} \triangleq \sum_{i \in \mathcal{C}} x_i \pi_i$ denotes the expected payoff of all SUs in the network. This equation can be easily solved as:

$$x_i(t) = \left(x_i(0) - \frac{\mu_i}{\sum_{l \in \mathcal{C}} \mu_l} \right) e^{-\left(\sum_{l \in \mathcal{C}} \mu_l \right)\sigma t} + \frac{\mu_i}{\sum_{l \in \mathcal{C}} \mu_l}, \quad \forall i \in \mathcal{C}. \tag{2}$$

The imitation dynamics induced by PIR thus converges exponentially in time to an evolutionary equilibrium, which is also the NE of G.

With the same assumption, the double imitation rule generates in the asymptotic case an aggregate monotone dynamics [34, 36], which is defined as follows:

$$\dot{x}_i = \frac{x_i}{\omega - \alpha} \left[1 + \frac{\omega - \bar{\pi}}{\omega - \alpha} \right] (\pi_i - \bar{\pi}), \quad \forall i \in \mathcal{C}, \tag{3}$$

whose solution is

$$x_i(t) = \left(x_i(0) - \frac{\mu_i}{\sum_{l \in \mathcal{C}} \mu_l} \right) e^{-\frac{\bar{\pi}}{\omega - \alpha}\left(1 + \frac{\omega - \bar{\pi}}{\omega - \alpha} \right)t} + \frac{\mu_i}{\sum_{l \in \mathcal{C}} \mu_l}, \quad \forall i \in \mathcal{C}. \tag{4}$$

As a consequence, DI converges exponentially in time to the NE of the spectrum access game G, however at a higher rate than PIR because by definition $\sigma = \frac{1}{\omega - \alpha}$ and ω and α being upper and lower bounds on payoffs, $\frac{\omega - \bar{\pi}}{\omega - \alpha} \geq 0$. From (2) and (4), it turns out that the aggregate monotone dynamics is a time-rescaled version of the replicator dynamics, as pointed in [10]. We will see in the next section that both dynamics continue to play an important role in our model.

As desirable properties, the proposed imitation-based spectrum access policies (both PISAP and DISAP) are stateless, incentive-compatible for selfish autonomous SUs and requires no central computational unit. The spectrum

assignment is achieved by local interactions among autonomous SUs. The autonomous behavior and decentralized implementation make the proposed policies especially suitable for large scale cognitive radio networks. The imitation factor σ controls the tradeoff between the convergence speed and the channel switching frequency in that larger σ represents more aggressiveness in imitation and thus leads to fast convergence, at the price of more frequent channel switching for the SUs.

6. Imitation Dynamics and Convergence

We have seen that proportional imitation and double imitation rules generate a replicator dynamics and an aggregate monotone dynamics. In the sequel analysis, we study the induced imitation dynamics and the convergence of the proposed spectrum access policies PISAP and DISAP, which take into account the constraint imposed by SU radios.

6.1. System Dynamics

In this subsection, we first derive in Theorem 2 the dynamics for a generic imitation rule F with large population. We then derive in Lemma 3, Theorem 3 and Theorem 4 the dynamics of the proposed proportional imitation policy PISAP and study its convergence. The counterpart analysis for the double imitation policy DISAP is explored in Lemma 4, Theorem 5 and Theorem 6.

We start by introducing the notation used in our analysis. At an iteration, we label all SUs performing strategy i (channel i in our case) as SUs of type i and we refer to the SUs on s_j as neighbors of SU j. We denote $n_i^l(t)$ the number of SUs on channel i at iteration t and operating on channel l at $t-1$. It holds that $\sum_{l \in \mathcal{C}} n_i^l(t) = n_i(t)$ and $\sum_{i \in \mathcal{C}} n_i^l(t) = n_l(t-1)$. For a given state $s(t) \triangleq \{s_j(t), j \in \mathcal{N}\}$ of the system at iteration t and for a finite population of size N, we denote $p_i(t) \triangleq n_i(t)/N$ the proportion of SUs of type i and

17

$p_i^l(t) \triangleq n_i^l(t)/N$ the proportion of SUs migrating from channel l to i. We use x instead of p to denote these proportions when N is large. It holds that $p \to x$ when $N \to +\infty$.

Denote by F a generic imitation rule under the channel constraint. In the case of a simple imitation rule (e.g. PISAP), F is characterized by the probability set $\{F_{j,k}^i\}$, where $F_{j,k}^i$ denotes the probability that a SU choosing strategy j at the precedent iteration imitates another SU choosing strategy k at the precedent iteration and then switches to channel i at next iteration after imitation. Instead, by applying a double imitation rule (e.g. DISAP), we can characterize F by the probability set $\{F_{j,k,l}^i\}$, where $F_{j,k,l}^i$ denotes the probability that a SU choosing strategy j at the precedent iteration imitates two neighbors choosing respectively strategy k and strategy l at the precedent iteration and then switches to channel i at the next iteration after imitation. In both cases the only way to switch to a channel i is to imitate a SU that was on channel i. That means $F_{j,k}^i = 0$, $\forall k \neq i$ (PISAP) and $F_{j,k,l}^i = 0$, $\forall k, l \neq i$ (DISAP).

At the initialization phase (iterations 0 and 1), each SU randomly chooses its strategy with uniform distribution. In the asymptotic case, we thus have $\forall i \in \mathcal{C}$, $x_i(0) > 0$ and $x_i(1) > 0$ a.s. After that, the system state at iteration $t + 1$, denoted as $\mathbf{p(t+1)}$ ($\mathbf{x(t+1)}$ in the asymptotic case), depends on the states at iteration t and $t - 1$.

We have now the following theorem that relates the finite and asymptotic cases.

Theorem 2. *For any imitation rule F, if the imitation among SUs of the same type occurs randomly and independently, then $\forall \delta > 0$, $\epsilon > 0$ and any initial state $\{\tilde{x}_i(0)\}$, $\{\tilde{x}_i(1)\}$ such that $\forall i \in \mathcal{C}$, $\tilde{x}_i(0) > 0$ and $\tilde{x}_i(1) > 0$, there exists $N_0 \in \mathbb{N}$ such that if $N > N_0$, $\forall i \in \mathcal{C}$, the event $|p_i(t) - x_i(t)| > \delta$ occurs with probability less than ϵ for all t, where $p_i(0) = x_i(0) = \tilde{x}_i(0)$, $p_i(1) = x_i(1) = \tilde{x}_i(1)$. In the*

case of a simple imitation policy it holds that

$$x_i(t+1) = \sum_{j,l,k \in \mathcal{C}} \frac{x_j^l(t) x_j^k(t)}{x_j(t)} F_{l,k}^i \quad \forall i \in \mathcal{C}$$

Differently, a double imitation policy yields:

$$x_i(t+1) = \sum_{j,l,k,z \in \mathcal{C}} \frac{x_j^l(t) x_j^k(t) x_j^z(t)}{[x_j(t)]^2} F_{l,\{k,z\}}^i \quad \forall i \in \mathcal{C}$$

Proof. The proof consists of first showing the theorem holds for iteration $t = 2$ and then proving the case $t \geq 3$ by induction. The detail is in Appendix B. $\square$

Theorem 2 is a result on the short run adjustments of large populations under any generic imitation rule F: the probability that the behavior of a large population differs from the one of an infinite population is arbitrarily small when N is sufficiently large. In what follows, we study the convergence of PISAP and DISAP specifically.

6.2. PISAP Dynamics and Convergence

In this section, we now focus on PISAP and derive the induced imitation dynamics in the following analysis.

Lemma 3. *On the proportional imitation policy PISAP under channel constraint, it holds that*

$$x_i^j(t+1) = \sum_{l,k \in \mathcal{C}} \frac{x_j^l(t) x_j^k(t)}{x_j(t)} F_{l,k}^i \quad \forall i, j \in \mathcal{C}. \tag{5}$$

Proof. The proof is straightforward from the analysis in the proof of Theorem 2.

$\square$

Theorem 3. *The proportional imitation policy PISAP under channel constraint generates the following dynamics in the asymptotic case:*

$$x_i(t+1) = x_i(t-1) + \sigma \pi_i(t-1) x_i(t-1) - \sigma \sum_{j,l \in \mathcal{C}} \pi_l(t-1) \frac{x_j^i(t) x_j^l(t)}{x_j(t)} \quad (6)$$

where $\pi_i(t)$ denotes the expected payoff of an individual SU on channel i at iteration t.

Proof. See Appendix C. $\qquad\square$

Although we are not able to prove it theoretically, we observe via extensive numerical experiments that (6) converges to the NE. The formal proof is left for future work. To get more in-depth insight on the dynamics (6), we notice that under the following approximation:

$$\sum_{l \in \mathcal{C}} \pi_l(t-1) \frac{x_j^l(t)}{x_j(t)} \approx \bar{\pi}(t-1), \quad (7)$$

where $\bar{\pi}(t-1)$ is the average individual payoff for the whole system at iteration $t-1$, noticing $\sum_j x_j^i(t) = x_i(t-1)$, (6) can be written as:

$$x_i(t+1) = x_i(t-1) + \sigma x_i(t-1)[\pi_i(t-1) - \bar{\pi}(t-1)]. \quad (8)$$

Note that the approximation (7) states that in any channel j at iteration t, the proportions of SUs coming from any channel l are representative of the whole population.

Under the approximation (7), given the initial state $\{x_i(0)\}$, $\{x_i(1)\}$, we can decompose (8) into the following two independent discrete-time replicator
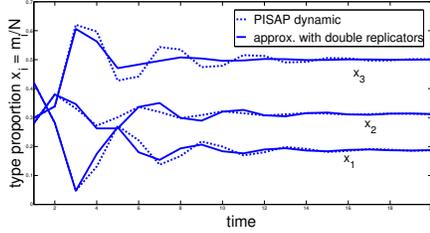
20

Figure 3: PISAP dynamics and its approximation by double replicator dynamics.
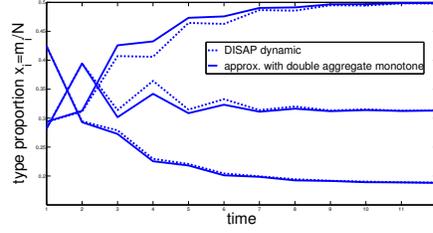

Figure 4: DISAP dynamics and its approximation by double aggregate monotone dynamics.

dynamics:

$$\begin{cases} x_i(u) = x_i(u-1) + \sigma x_i(u-1)[\pi_i(u-1) - \bar{\pi}(u-1)] \\ x_i(v) = x_i(v-1) + \sigma x_i(v-1)[\pi_i(v-1) - \bar{\pi}(v-1)], \end{cases} \quad (9)$$

where $u = 2t$, $v = 2t + 1$. The two equations in (9) illustrate the underlying system dynamics hinged behind PISAP under the approximation (7): it can be decomposed into two independent delayed replicator dynamics that alternatively occur at the odd and even iterations, respectively. The following theorem establishes the convergence of (9) to a unique fixed point, which is also the NE of the spectrum access game G.

Theorem 4. *Starting from any initial point, the system described by* (9) *converges to a unique fixed point which is also the NE of the spectrum access game* G.

Proof. The proof, of which the detail is provided in Appendix D, consists of showing that the mapping described by (9) is a contraction mapping. $\square$

As an illustrative example, Fig. 3 shows that the double replicator dynamics provides an accurate approximation of the system dynamics induced by PISAP.

6.3. DISAP Dynamics and Convergence

We now focus on DISAP and derive the induced imitation dynamics.

Lemma 4. *On the double imitation policy DISAP under channel constraint, it holds that*

$$x_i^j(t+1) = \sum_{l,k,z \in \mathcal{C}} \frac{x_j^l(t) x_j^k(t) x_j^z(t)}{[x_j(t)]^2} F_{l,k,z}^i \quad \forall i,j \in \mathcal{C}. \qquad (10)$$

Proof. The proof is straightforward from the analysis in the proof of Theorem 2.

$\square$

Theorem 5. *The double imitation policy DISAP under channel constraint generates the following dynamics in the asymptotic case:*

$$x_i(t+1) = x_i(t-1) + \sum_j x_j^i(t) Q(\bar{\pi}_j(t-1))(\pi_i(t-1) - \bar{\pi}_j(t-1)), \qquad (11)$$

where $\bar{\pi}_j(t-1) = \sum_k \frac{x_j^k(t)}{x_j(t)} \pi_k(t-1)$ and $Q(U) \triangleq 2 - \frac{U-\alpha}{\omega-\alpha}$.

Proof. See Appendix E. $\square$

Again, we are not able to prove it analytically and leave the formal proof as future work. However, we observe via extensive numerical experiments that (11) converges to the NE and, as shown in Fig. 5, is also characterized by a smoother and faster convergence with respect to the proportional imitation dynamics (equation (6)).

By performing the approximation $\bar{\pi}_j(t-1) \approx \bar{\pi}(t-1)$ for all j, (11) can be written as:

$$x_i(t+1) \quad = \quad x_i(t-1) + x_i(t-1) Q(\bar{\pi}(t-1))(\pi_i(t-1) - \bar{\pi}(t-1)). \quad (12)$$

Given the initial state $\{x_i(0)\}$, $\{x_i(1)\}$, we can now decompose (12) into the following two independent discrete-time aggregate monotone dynamics:

$$\begin{cases} x_i(u) = x_i(u-1) + x_i(u-1)[2 - \bar{\pi}(u-1)][\pi_i(u-1) - \bar{\pi}(u-1)] \\ x_i(v) = x_i(v-1) + x_i(v-1)[2 - \bar{\pi}(v-1)][\pi_i(v-1) - \bar{\pi}(v-1)], \end{cases} \tag{13}$$

where $u = 2t$, $v = 2t + 1$. The underlying system dynamics can thus be decomposed into two independent delayed aggregate monotone dynamics that alternatively occur at the odd and even iterations, respectively. The following theorem establishes the convergence of (13) to a unique fixed point which is also the NE of the spectrum access game G. The proof follows exactly the same analysis as that of Theorem 4.

Theorem 6. *Starting from any initial point, the system described by (13) converges to a unique fixed point which is also the NE of the spectrum access game G.*

As an illustrative example, Fig. 4 shows that the double aggregate dynamics provides an accurate approximation of the system dynamics induced by DISAP.

7. Discussion

This paper is a first step to systematically apply imitation rules to cognitive radio networks. There are several points to be tackled in order to make the model more realistic.

- It is assumed in this paper that a SU can capture another SU packet for sampling with probability 1. Assuming a capture probability less than 1 would have the same effect as decreasing the value of σ in PISAP, i.e., it would slow down the convergence speed of the proposed algorithms.

23

- It is assumed that all SUs can all hear each other on the same channel. A more realistic setting would consider a graph of possible communications between SUs. In this case, our algorithms are not any more ensured to converge. This point is left for further work. A promising approach is to use the results of the literature on 'learning from neighbors', which studies the conditions under which efficient actions are adopted by a population if agents receive information only from their neighbors (see e.g. [37]).

- In this paper, SUs are supposed to provide in their packet header the exact average throughput that can be obtained on a given channel. We have investigated in [24] the effect of providing only an estimate of the average throughput. Assuming the use of CSMA/CA as SU MAC protocol, we have shown by simulations that our algorithms continue to converge in this more realistic context.

- For mathematical convenience, we have assumed in this paper that the SU MAC protocol was perfect and could act as TDMA. Although unrealistic, this approach gives an upper bound on the performance of our policies. Also, the analysis can be extended with other more realistic MAC protocols by adapting the utility functions. Particularly, we have investigated in [24] the use of CSMA/CA and shown by simulations the convergence of our policies.

- It is assumed that a generic PU transmits with a certain probability in TDMA-like mode. If there are multiple PU transmitters it is possible to distinguish two cases:

  1. The transmission of each PU covers the totality of SU receivers. This scenario boils down to the case of a unique generic PU transmitter.

  2. The transmission of one or more PUs covers a sub-set of the SU re-

ceivers. In this case, different SUs may have different perceptions of the environment and a further analysis, based on the fact that the channel availability probabilities are now dependent on both channel i and SU j, should be carried on. This point is left for future work.

8. Performance Evaluation

In this section, we conduct simulations to evaluate the performance of the proposed imitation-based channel access policies (PISAP and DISAP) and demonstrate some intrinsic properties of the policies, which are not explicitly addressed in the analytical part of the paper.

For performance comparison, we also show the results obtained by simulating Trial and Error [38] (shortened into $T\&E$ in the following). The latter has been chosen as it is, to the best of our knowledge, one of the best existing mechanisms that 1) applies to our model and 2) is guaranteed to converge to a NE. In T&E, players locally implement a state machine, so that at each iteration each player is characterized by a state, which is defined by the triplet $\{current\_mood, benchmark\_mood, benchmark\_strategy\}$. Players current mood (the four possible moods are: *content*, *watchful*, *hopeful* and *discontent*) reflects the machine reaction to its experience of the environment. A NE is reached when everybody is in state *content*.

8.1. Simulation Settings

We simulate two cognitive radio networks, termed Network 1 and Network 2. We study the performance of our algorithms on Network 1, and compare their convergence behaviors and fairness to the ones obtained by T&E on Network 2.

- **Network 1:** We consider $N = 50$ SUs, $C = 3$ channels characterized by the availability probabilities $\mu = [0.3, 0.5, 0.8]$.

- **Network 2:** We set $N = 10$, $C = 2$ and $\mu = [0.2, 0.8]$.

Note that the introduction of Network 2 has been necessary as the dynamics induced by T&E turns out to be very slow to converge on the bigger Network 1 (after 10^5 iterations convergence is still not achieved).

We assume that the block duration is long enough, so that the SUs, regardless of the occupied channel, can evaluate their payoff without errors. T&E learning parameters (i.e., experimentation probability and benchmark mood acceptance ratio) are set at each iteration according to [39].

8.2. System Dynamics

In Fig. 5, the trajectories described by (6) and (11) are compared. The first part of the curves is characterized by important variations. This can be interpreted by the overlap of two replicator/aggregate monotone dynamics at odd and even instants, as explained in Section 6. We observe that, in the asymptotic case, DISAP outperforms PISAP as it is characterized by less pronounced wavelets and a faster convergence. However, both dynamics correctly converge to an evolutionary equilibrium. It is easy to check that the converged equilibrium is also the NE of G and the system optimum, which confirms our theoretic analysis. The dynamics presented in Fig. 5 are valid in an asymptotic case, when the number of SUs is large. We now turn our attention to small size scenarios.

8.3. Convergence with finite number of SUs

We study in this section the convergence of PISAP and DISAP on Network 1 ($N = 50, C = 3$). Fig. 6 and Fig. 7 show a realization of our algorithms. We notice that an imitation-stable equilibrium is achieved progressively following the dynamics characterized by (6) and (11). The equilibrium is furthermore very close to the system optimum: we can in fact check that, according to

26

Theorem 1, the proportion of SUs choosing channels 1, 2 and 3 at the system optimum is 0.1875, 0.3125 and 0.5 respectively; in the simulation results we observe that there are 9, 16 and 25 SUs settling on channels 1, 2 and 3 respectively. We also notice on this example that DISAP convergence is faster than PISAP convergence.

We now focus on Network 2 ($N = 10, C = 2$) and compare T&E convergence behavior (Fig. 8) to the trends of PISAP (Fig. 9) and DISAP (Fig. 10). It is easy to notice that T&E converges in a much slower and more chaotic way with respect to PISAP and DISAP. With T&E, the search of a NE may turn out to be extremely long (in the realization depicted in Fig. 8, e.g., convergence is achieved within $3.5 \cdot 10^3$ iterations). On the contrary, PISAP and DISAP converge within 75 and 32 iterations respectively.

8.4. System Fairness

We now turn to the analysis of the fairness of the proposed spectrum access policies. To this end, we adopt the Jain's fairness index [40], which varies in $[0, 1]$ and reaches its maximum, when the resources are equally shared amongst users. Fig. 11 and Fig. 12, whose curves represent an average over 10^3 independent realizations on Network 2 of our algorithms and of T&E respectively, show that PISAP and DISAP clearly outperform T&E in terms both of fairness and convergence speed. In fact, while our system turns out to be very fair from the early iterations, T&E needs $6 \cdot 10^3$ iterations to get its system to reach a fairness value of 0.85. From Fig. 11, one can further infer that indeed DISAP converges more rapidly than PISAP: for example, a fairness index of 0.982 is reached at $t = 100$ by DISAP and at $t = 200$ by PISAP.

8.5. Switching Cost

At last, we concentrate on the switching frequency of the three algorithms because switching may represent a significant cost for today's wireless devices

in terms of delay, packet loss and protocol overhead. In Fig. 13, we define the switching cost at iteration t as the number of strategy switches between 0 and t. After 200 iterations, the switching cost of DISAP and PISAP has stabilized because convergence has been reached. On the contrary, T&E exhibits a fast growing cost.

8.6. Imperfect observations of PU activity

We assumed so far that cognitive radio users observe the channel activity of the primary user without errors. In this section, we investigate the performance of the proposed algorithms when the PU activity is *imperfectly* observed by the SUs. We denote by P_e the SU probability of error in detecting the PU activity and by Q_e the miss detection probability of an idle PU (probability of false alarm). The expected value of the payoff experienced by SUs on channel i can be written as follows:

$$\mathbb{E}[\pi_i(n_i)] = \sum_{m=0}^{n_i-1} \binom{n_i - 1}{m}(1 - Q_e)^m Q_e^{n_i-1-m}(1 - Q_e)\frac{\mu_i}{m+1}, \qquad (14)$$

which does not depend on P_e because a miss detection of the PU activity does not affect the throughput of any SU.

We now want to evaluate the impact of Q_e on the expected throughput estimates. To this end, we calculate the values taken by (14) for different values of Q_e and for different numbers of SUs. Results are shown in Fig. 14. Surprisingly, we see that the estimates under sensing errors rapidly converge to the values calculated for the ideal case with no miss detections (i.e., $Q_e = 0$). This is due to the fact that a trade-off arises. On the one hand, a SU, which is unable to detect a free slot experiences a penalty in its throughput. On the other hand, there are less SUs in average accessing free slots, which results in a higher throughput. As shown in Fig. 14, the two effects counterweight when
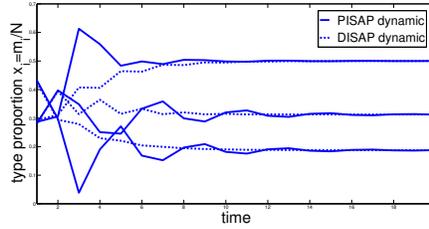
28

Figure 5: PISAP and DISAP system dynamics in the asymptotic case.
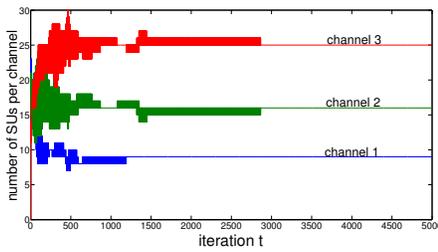

Figure 6: PISAP on Network 1: number of SUs per channel as a function of the number of iterations.
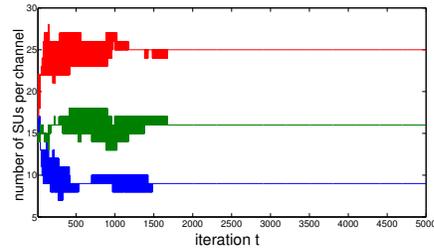
Figure 7: DISAP on Network 1: number of SUs per channel as a function of the number of iterations.

the number of SUs gets larger. Hence, one can infer that in practice the impact of miss detections of the PU activity/inactivity is limited for a number of SUs on the same channel greater than 5.
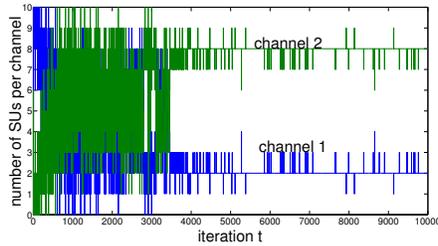

Figure 8: T&E on Network 2: number of SUs per channel as a function of the number of iterations.
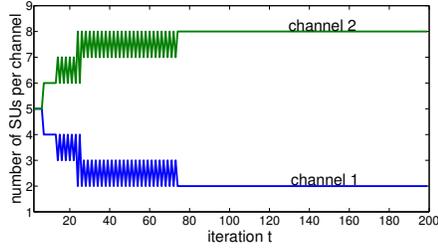
Figure 9: PISAP on Network 2: number of SUs per channel as a function of the number of iterations.
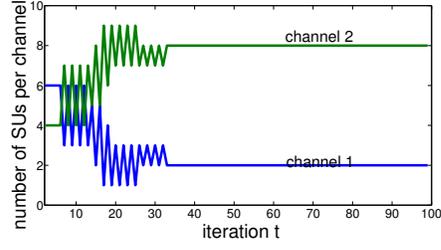

Figure 10: DISAP on Network 2: number of SUs per channel as a function of the number of iterations.
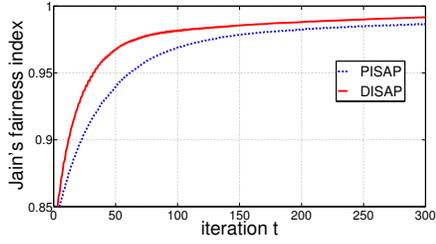

Figure 11: PISAP and DISAP on Network 1: Jain's fairness index as a function of the number of iterations (average over 10^3 realizations).
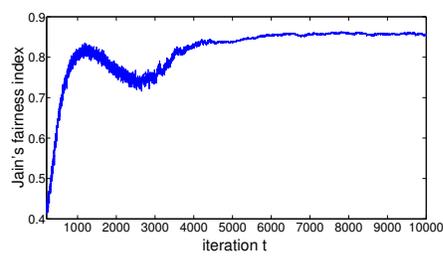

Figure 12: T&E on Network 2: Jain's fairness index as a function of the number of iterations (average over 10^3 realizations).
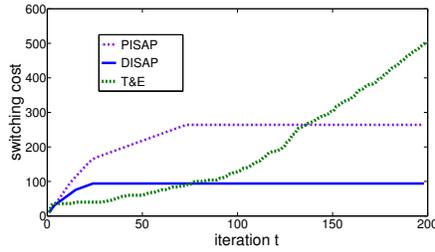

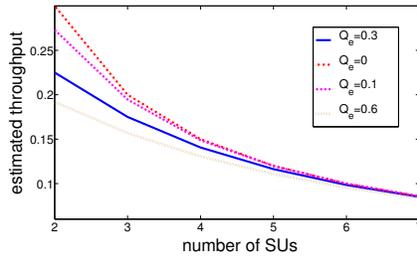Figure 13: Switching cost of PISAP, DISAP and T&E in Network 2.

Figure 14: Expected throughput deviation under sensing errors ($\mu_i = 0.6$).

9. Conclusion and Further Work

In this paper, we address the spectrum access problem in cognitive radio networks by applying population game theory and develop two imitation-based spectrum access policies. In our model, a SU can only imitate the other SUs operating on the same channel. This constraint makes the basic proportional imitation and double imitation rules irrelevant in our context. These two imitation rules are thus adapted to propose PISAP, a proportional imitation spectrum access policy, and DISAP, a double imitation spectrum access policy. A systematic theoretical analysis is presented on the induced imitation dynamics and the convergence properties of the proposed policies to the Nash equilibrium. Simulation results show the efficiency of our algorithms even for small size scenarios. It is also shown that PISAP and DISAP outperform Trial and Error in terms of convergence speed and fairness. As an important direction of the future work, we plan to investigate the imitation-based channel access problem in the more generic multi-hop scenario where SUs can imitate their neighbors and derive the relevant channel access policies.

[1] S. Haykin, Cognitive Radio: Brain-Empowered Wireless Communications, IEEE Journal on Selected Areas in Communications 23 (2) (2005) 201–220.

[2] M. Buddhikot, Understanding Dynamic Spectrum Access: Models, Taxonomy and Challenges, in: Proc. IEEE Dynamic Spectrum Access Networks (DySPAN), Dublin, Ireland, 2007.

[3] S. Iellamo, L. Chen, M. Coupechoux, A. V. Vasilakos, Imitation-based Spectrum Access Policy for Cognitive Radio Networks, in: Proc. International Symposium on Wireless Communication Systems (ISWCS), Paris, France, 2012.

[4] A. Mahajan, D. Teneketzis, Foundations and Applications of Sensor Management, Springer-Verlag, 2007, Ch. Multi-armed Bandit Problems, pp. 121–151.

[5] A. Anandkumar, N. Michael, A. Tang, Opportunistic Spectrum Access with Multiple Users: Learning under Competition, in: Proc. IEEE International Conference on Computer Communication (INFOCOM), San Diego, CA USA, 2010.

[6] J. Neel, J. Reed, R. Gilles, Convergence of Cognitive Radio Networks, in: Proc. IEEE Wireless Communications and Networking Conference (WCNC), Atlanta, Georgia USA, 2004.

[7] N. Nie, C. Comaniciu, Adaptive Channel Allocation Spectrum Etiquette for Cognitive Radio Networks, ACM Mobile Networks and Applications (MONET) 11 (6) (2006) 779–797.

[8] Z. Han, C. Pandana, K. Liu, Distributive Opportunistic Spectrum Access for Cognitive Radio Using Correlated Equilibrium and No-Regret Learning, in: Proc. IEEE Wireless Communications and Networking Conference (WCNC), Hong Kong, China, 2007.

[9] L. Chen, S. Iellamo, M. Coupechoux, P. Godlewski, An Auction Framework for Spectrum Allocation with Interference Constraint in Cognitive Radio Networks, in: Proc. IEEE International Conference on Computer Communication (INFOCOM), San Diego, California USA, 2010.

[10] J. Hofbauer, K. Sigmund, Evolutionary Game Dynamics, Bull. Amer. Math. Soc. 40 (4) (2003) 479–519.

[11] W. H. Sandholm, Population Games and Evolutionary Dynamics, The MIT Press, 2010.

[12] H. Tembine, E. Altman, R. El-Azouzi, Y. Hayel, Bio-inspired Delayed Evolutionary Game Dynamics with Networking Applications, Telecommunication Systems 47 (1) (2011) 137–152.

[13] P. Coucheney, C. Toutati, B. Gaujal, Fair and Efficient User-Network Association Algorithm for Multi-Technology Wireless Networks, in: Proc. IEEE International Conference on Computer Communication (INFOCOM), Rio de Janeiro, Brazil, 2009.

[14] S. Shakkottai, E. Altman, A. Kumar, Multihoming of Users to Access Points in WLANs: A Population Game Perspective, IEEE Journal on Selected Areas in Communications 25 (6) (2007) 1207–1215.

[15] D. Niyato, E. Hossain, Dynamics of Network Selection in Heterogeneous Wireless Networks: An Evolutionary Game Approach, IEEE Transactions on Vehicular Technology 58 (4) (2009) 2008–2017.

[16] H. Ackermann, P. Berenbrink, S. Fischer, M. Hoefer, Concurrent Imitation Dynamics in Congestion Games, in: Proc. ACM Symposium on Principles of Distributed Computing (PODC), Calgary, Canada, 2009.

[17] P. Berenbrik, T. Friedetzky, L. Goldberg, P. Goldberg, Distributed Selfish Load Balancing, in: Proc. ACM-SIAM Symposium on Discrete Algorithms (SODA), San Francisco, CA, 2011.

[18] A. Ganesh, S. Lilienthal, D. Manjunath, A. Proutiere, F. Simatos, Load Balancing via Random Local Search in Closed and Open Systems, in: Proc. ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems, New York, NY, 2010.

[19] C. Alos-Ferrer, K. H. Schlag, The Handbook of Rational and Social Choice, Oxford University Press, 2009, Ch. Imitation and Learning.

[20] X. Chen, J. Huang, Evolutionarily Stable Spectrum Access, to appear in IEEE Transactions on Mobile Computing.

[21] X. Chen, J. Huang, Imitative Spectrum Access, in: Proc. International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), Paderborn, Germany, 2012.

[22] S. Huang, X. Liu, Z. Ding, Opportunistic Spectrum Access in Cognitive Radio Networks, in: Proc. IEEE International Conference on Computer Communication (INFOCOM), Phoenix, Arizona USA, 2008.

[23] M. Maskery, V. Krishnamurthy, Q. Zhao, Decentralized Dynamic Spectrum Access for Cognitive Radios: Cooperative Design of a Non-Cooperative Game, IEEE Transactions on Communications 57 (2) (2009) 459–469.

[24] S. Iellamo, L. Chen, M. Coupechoux, Imitation-based Spectrum Access Policy for CSMA/CA-based Cognitive Radio Networks, in: Proc. IEEE Wireless Communications and Networking Conference (WCNC), Paris, France, 2012.

[25] X. Chen, J. Huang, Spatial Spectrum Access Game: Nash Equilibria and Distributed Learning, in: Proc. ACM SIGMOBILE International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc), Hilton Head Island, South Carolina USA, 2012.

[26] L. Lai, H. E. Gamal, H. Jiang, H. V. Poor, Cognitive Medium Access: Exploration, Exploitation, and Competition, IEEE Transactions on Mobile Computing 10 (2) (2011) 239–253.

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

[28] I. Milchtaich, Congestion Games with Player-Specific Payoff Functions, Games and Economic Behavior 13 (1) (1996) 111–124.

[29] A. Haurie, P. Marcotte, On the Relationship between Nash, Cournot and Wardrop Equilibria, Networks 15 (3) (1985) 295–308.

[30] K. H. Schlag, Why Imitate, and if so, How ? A Boundedly Rational Approach to Multi-Armed Bandits, Journal of Economic Theory 78 (1) (1998) 130–156.

[31] C. Alos-Ferrer, F. Shi, Imitation with Asymmetric Memory, Economic Theory 49 (1) (2012) 193–215.

[32] B. Pradelski, H. Young, Learning Efficient Nash Equilibria in Distributed Systems, Games and Economic behavior 75 (2) (2012) 882–897.

[33] J. Elias, F. Martignon, E. Altman, Joint Pricing and Cognitive Radio Network Selection: a Game Theoretical Approach, in: Proc. International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), Paderborn, Germany, 2012.

[34] K. H. Schlag, Which One Should I Imitate ?, Journal of Mathematical Economics 31 (4) (1999) 493–522.

[35] W. H. Sandholm, Local Stability under Evolutionary Game Dynamics, Theoretical Economics 5 (1) (2010) 27–50.

[36] L. Sammuelson, J. Zhang, Evolutionary Stability in Asymmetric Games, Journal of Economic Theory 57 (2) (1992) 363–391.

[37] C. Alós-Ferrer, S. Weidenholzer, Contagion and Efficiency, Journal of Economic Theory 143 (1) (2008) 251–274.

[38] H. P. Young, Learning by Trial and Error, Games and Economic Behavior 65 (2) (2009) 626–643.

[39] B. S. Pradelski, H. P. Young, Efficiency and Equilibrium in Trial and Error Learning, Discussion paper, Department of Economics, University of Oxford (2010).

[40] R. Jain, D. Chiu, W. Hawe, A Quantitative Measure of Fairness and Discrimination for Resource Allocation in Shared Computer Systems, Research Report TR-301, DEC (1984).

[41] R. Abraham, J. Marsden, T. Ratiu, Manifolds, Tensor Analysis, and Applications, Springer-Verlag, 1988.

Appendix A. Proof of theorem 1

We first show that any point $\mathbf{x} = (x_i)_{i \in \mathcal{C}}$ cannot be a NE if there exists i_1 and i_2 such that $\pi_{i_1}(N x_{i_1}) < \pi_{i_2}(N x_{i_2})$. Otherwise, consider the strategy profile $\mathbf{x}'$ where ϵN SUs move from channel i_1 to i_2. For N large and with sufficient small ϵ, it follows from the continuity of $\pi_i(x_i)$ that $\pi_{i_1}(N(x_{i_1} - \epsilon)) < \pi_{i_2}(N(x_{i_2} + \epsilon))$, which indicates that by switching from i_1 to i_2, one can increase its payoff. We then proceed to show the second part of the theorem. To this end, let y denote the payoff of any SU at the NE, we have: $\mu_i S(x_i N) = y$, $\forall i \in \mathcal{C}$. It follows that $x_i N = S^{-1}\left(\frac{y}{\mu_i}\right)$. Noticing that $\sum_i x_i = 1$, at the NE, we have:

$$\sum_{i \in \mathcal{C}} S^{-1}\left(\frac{y}{\mu_i}\right) = N. \tag{A.1}$$

Since a NE is ensured to exist, (A.1) admits at least a solution y. Moreover, it follows from the strict monotonicity of S in Assumption 1 that its inverse

function S^{-1} is also strictly monotonous. Hence (A.1) admits a unique solution. We thus complete the proof.

Appendix B. Proof of theorem 2

We prove the statement for $t = 2$. The case for $t \geq 3$ is analogous to [30], which can be shown by induction and is therefore omitted.

Define the random variable $w_i^j(c)$ such that

$$
w_i^j(c) = \begin{cases} 1 & \text{if SU } c \text{ is on channel } j \text{ at iteration } t = 1 \\ & \text{and migrates to channel } i \text{ at } t = 2 \\ 0 & \text{otherwise} \end{cases} . \tag{B.1}
$$

We now distinguish two cases: proportional and double imitation.

Appendix B.1. Proportional imitation

By definition, if $j \neq s_c(1)$, it holds that $w_i^j(c) = 0$. Otherwise, c imitates with probability $\frac{n_{s_c(1)}^k}{n_{s_c(1)}}$ a SU that was using channel k at $t = 0$ and that is currently $(t = 1)$ on the same channel as c $(s_c(1))$, and then migrates to channel i with probability $F_{s_c(0),k}^i$. Note that we allow for self-imitation in our algorithm. At initial states, all strategies are supposed to be chosen by at least one SU (N is large), so that $n_{s_c(1)} \neq 0$. We thus have:

$$
\mathbb{P}[w_i^j(c) = 1] = \begin{cases} 0 & \text{if } j \neq s_c(1) \\ \sum_{k \in \mathcal{C}} \dfrac{n_{s_c(1)}^k}{n_{s_c(1)}} F_{s_c(0),k}^i & \text{otherwise} \end{cases} . \tag{B.2}
$$

We can now derive the population proportions at iteration $t = 2$ as:

$$
p_i^j(2) = \frac{1}{N} \sum_{c \in \mathcal{N}} w_i^j(c) \quad \forall i, j \in \mathcal{C}. \tag{B.3}
$$

The expectations of these proportions can now be written as (using the Kronecker delta $\delta_{i,j}$):

$$\mathbb{E}[p_i^j(2)] \;=\; \frac{1}{N}\sum_{c\in\mathcal{N}}\mathbb{P}[w_i^j(c)=1] \tag{B.4}$$

$$=\; \frac{1}{N}\sum_{c\in\mathcal{N},k\in\mathcal{C}}\frac{n_{s_c(1)}^k(1)F_{s_c(0),k}^i\delta_{j,s_c(1)}}{n_{s_c(1)}(1)} \tag{B.5}$$

$$=\; \frac{1}{N}\sum_{h,l,k\in\mathcal{C}}\frac{n_h^l(1)n_h^k(1)F_{l,k}^i\delta_{j,h}}{n_h(1)} \tag{B.6}$$

$$=\; \frac{1}{N}\sum_{l,k\in\mathcal{C}}\frac{n_j^l(1)n_j^k(1)F_{l,k}^i}{n_j(1)} \tag{B.7}$$

$$=\; \sum_{l,k\in\mathcal{C}}\frac{\tilde{x}_j^l(1)\tilde{x}_j^k(1)}{\tilde{x}_j(1)}F_{l,k}^i. \tag{B.8}$$

It follows that

$$\mathbb{E}[p_i(2)] \;=\; \sum_{j\in\mathcal{C}}\mathbb{E}[p_i^j(2)] = \sum_{j,l,k\in\mathcal{C}}\frac{\tilde{x}_j^l(1)\tilde{x}_j^k(1)}{\tilde{x}_j(1)}F_{l,k}^i. \tag{B.9}$$

As $w_i^j(c)$ and $w_i^j(d)$ are independent random variables for $c\neq d$ and since the variance of $w_i^j(c)$ is less than 1, the variance of $p_i^j(2)$ and $p_i(2)$ for any $i,j\in\mathcal{C}$ are less than $1/N$ and C/N, respectively. It then follows the Bienaymé-Chebychev inequality that

$$\forall i\in\mathcal{C},\mathbb{P}[\{|p_i(2)-\mathbb{E}[p_i(2)]|>\delta\}]<\frac{C}{(N\delta)^2}. \tag{B.10}$$

Choosing N_0 such that $\frac{C}{(N_0\delta)^2}<\epsilon$ concludes the proof for $t=2$. The proof can then be induced to any t as in [30].

38

Appendix B.2. Double imitation

If $j \neq s_c(1)$, it holds that $w_i^j(c) = 0$. Otherwise, c imitates with probability $\frac{n_{s_c(1)}^k}{N_{s_c(1)}} \frac{n_{s_c(1)}^z}{n_{s_c(1)}}$ two SUs that were using respectively channel k and channel z at $t = 0$ and that are currently ($t = 1$) on the same channel as c ($s_c(1)$), and then migrates to channel i with probability $F_{s_c(0),k,z}^i$.

The proof follows in the steps of the proportional imitation and only the main passages will be sketched out. We allow a SU to sample twice the same SU on the channel, so that:

$$\mathbb{P}[w_i^j(c) = 1] = \begin{cases} 0 \text{ if } j \neq s_c(1) \\ \displaystyle\sum_{k,z \in \mathcal{C}} \frac{n_{s_c(1)}^k}{n_{s_c(1)}} \frac{n_{s_c(1)}^z}{n_{s_c(1)}} F_{s_c(0),k,z}^i \quad \text{otherwise.} \end{cases} \tag{B.11}$$

We then derive the proportions expectations:

$$\mathbb{E}[p_i^j(2)] = \frac{1}{N} \sum_{c \in \mathcal{N}} \mathbb{P}[w_i^j(c) = 1] \tag{B.12}$$

$$= \frac{1}{N} \sum_{c \in \mathcal{N}, k, z \in \mathcal{C}} \frac{n_{s_c(1)}^k(1)}{n_{s_c(1)}(1)} \frac{n_{s_c(1)}^z(1)}{n_{s_c(1)}(1)} F_{s_c(0),k,z}^i \delta_{j,s_c(1)} \tag{B.13}$$

$$= \sum_{l,k,z \in \mathcal{C}} \frac{\tilde{x}_j^l(1) \tilde{x}_j^k(1) \tilde{x}_j^z(1)}{[\tilde{x}_j(1)]^2} F_{l,k,z}^i. \tag{B.14}$$

It follows that:

$$\mathbb{E}[p_i(2)] = \sum_{j \in \mathcal{C}} \mathbb{E}[p_i^j(2)] \tag{B.15}$$

$$= \sum_{j,l,k,z \in \mathcal{C}} \frac{\tilde{x}_j^l(1) \tilde{x}_j^k(1) \tilde{x}_j^z(1)}{[\tilde{x}_j(1)]^2} F_{l,k,z}^i. \tag{B.16}$$

The rest of the proof for the double imitation follows the same way as that of proportional imitation.

Appendix C. Proof of theorem 3

Recall the analysis in [30]. In this reference, equation (10) states that $F_{i,j}^j = F_{j,i}^i + \sigma[\pi_j - \pi_i]$. We can now characterize $\{F_{l,k}^i\}$ for PISAP as:

$$
F_{l,k}^i = \begin{cases}
0 & \text{if } l, k \neq i \\
F_{i,l}^l + \sigma[\pi_i(t-1) - \pi_l(t-1)] & \text{if } k = i \text{ and } l \neq i \\
1 - F_{k,i}^i - \sigma[\pi_k(t-1) - \pi_i(t-1)] & \text{if } l = i \text{ and } k \neq i \\
1 & \text{if } l = k = i
\end{cases}.
$$

The above four equations state that: (1) If none of the involved channels is i then the probability to switch to channel i is null (F is imitating); (2) The switching probability is proportional to the payoff difference; (3) If a SU does not imitate, it stays on the same channel; (4) If a SU imitates another SU with the same strategy, its strategy is not modified. Equation (5) can now be written as follows:

$$
\begin{aligned}
x_i^j(t+1) &= \sum_{l \neq i} \frac{x_j^l(t) x_j^i(t)}{x_j(t)} \left(F_{i,l}^l + \sigma[\pi_i(t-1) - \pi_l(t-1)] \right) \\
&\quad + \sum_{k \neq i} \frac{x_j^i(t) x_j^k(t)}{x_j(t)} \left(1 - F_{i,k}^i - \sigma[\pi_k(t-1) - \pi_i(t-1)] \right) + \frac{x_j^{i2}}{x_j} \\
&= \sum_{l \neq i} \frac{x_j^l(t) x_j^i(t)}{x_j(t)} \left(1 + \sigma[\pi_i(t-1) - \pi_l(t-1)] \right) + \frac{x_j^{i2}}{x_j} \\
&= \sum_{l \neq i} \frac{x_j^l(t) x_j^i(t)}{x_j(t)} \sigma[\pi_i(t-1) - \pi_l(t-1)] + \sum_{l \in \mathcal{C}} \frac{x_j^l(t) x_j^i(t)}{x_j(t)} \\
&= x_j^i(t) + \sum_{l \in \mathcal{C}} \frac{x_j^l(t) x_j^i(t)}{x_j(t)} \sigma[\pi_i(t-1) - \pi_l(t-1)].
\end{aligned}
$$

This concludes the proof.

Appendix D. Proof of theorem 4

We prove the convergence of (9) by showing that the mapping described by (9) is a contraction. A contraction mapping is defined [41] as follows: let (X, d) be a metric space, $f \colon X \to X$ is a contraction if there exists a constant $k \in [0, 1)$ such that $\forall x, y \in X$, $d(f(x), f(y)) \leq kd(x, y)$, where $d(x, y) = ||x - y|| = \max_i |x_i - y_i|$. Such an f is called a contraction and admits a unique fixed point, to which the mapping described by f converges.

Noticing that

$$d(f(x), f(y)) = ||f(x) - f(y)|| \leq \left|\left|\frac{\partial f}{\partial x}\right|\right| d(x, y), \tag{D.1}$$

it suffices to show that the Jacobian $\left|\left|\dfrac{\partial f}{\partial x}\right|\right| \leq k$. In our case, it suffices to show that $||J||_\infty \leq k$, where $J = (J_{ij})_{i,j \in \mathcal{C}}$ is the Jacobian of the mapping described by one of the equation in (9), defined by $J_{ij} = \dfrac{\partial x_i(u)}{\partial x_j(u-1)}$.

Recall that $\pi_i = \frac{\mu_i}{N x_i}$ and $\bar\pi = \sum_l \frac{\mu_l}{N}$, (9) can be rewritten as:

$$x_i(u) = x_i(u-1) + \sigma \left[\frac{\mu_i}{N} - x_i(u-1) \sum_l \frac{\mu_l}{N}\right]. \tag{D.2}$$

It follows that

$$J_{ij} = \begin{cases} 1 - \sum_l \dfrac{\mu_l}{N} & \text{if } j = i \\[2mm] 0 & \text{otherwise} \end{cases}. \tag{D.3}$$

Hence

$$||J||_\infty = \max_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}} |J_{ij}| = 1 - \sum_l \frac{\mu_l}{N} < 1, \tag{D.4}$$

which shows that the mapping described by (9) is a contraction. It is further

41

easy to check that the fixed point of (9) is $x^* = \frac{\mu_i}{\sum_{l \in \mathcal{N}} \mu_l}$, which is also the unique NE of G.

Appendix E. Proof of theorem 5

We start from the following equation (we skip the reference to time on the right hand side after the first line for the sake of clarity):

$$x_i^j(t+1) = \sum_{l,k,z \in \mathcal{C}} \frac{x_j^l(t) x_j^k(t) x_j^z(t)}{[x_j(t)]^2} F_{l,k,z}^i \qquad (\text{E.1})$$

$$= \frac{x_j^i}{x_j^2} \sum_{k \neq i} x_j^{k2} [F_{i,k,k}^i + F_{k,i,k}^i + F_{k,k,i}^i] +$$

$$\frac{x_j^{i2}}{x_j^2} \sum_{k \neq i} x_j^k [F_{k,i,i}^i + F_{i,k,i}^i + F_{i,i,k}^i] +$$

$$\frac{x_j^i}{x_j^2} \sum_{k \neq i} \sum_{l \notin \{k,i\}} x_j^l x_j^k [F_{i,k,l}^i + F_{k,i,l}^i + F_{k,l,i}^i] +$$

$$\frac{x_j^{i3}}{x_j^2}. \qquad (\text{E.2})$$

The second equality can be understood as follows. $F_{l,k,z}^i \neq 0$ only if at least one of the indices $l, k,$ or z is equal to i. The first sum of the right hand side (RHS) is obtained when two indices are equal and different from i, the third one is equal to i. The second sum is obtained when one index is different from i and the two others are equal to i. The third sum is obtained when one index is equal to i and the two others are different and different from i. The last term corresponds to the case where all indices are equal to i (in this case, obviously, $F_{i,i,i}^i = 1$). Now we have:

$$F_{i,k,k}^i + F_{k,i,k}^i + F_{k,k,i}^i = 2F_{k,i,k}^i + 1 - F_{i,k,k}^k, \qquad (\text{E.3})$$

$$F_{k,i,i}^i + F_{i,k,i}^i + F_{i,i,k}^i = F_{k,i,i}^i + 2(1 - F_{i,i,k}^k), \qquad (\text{E.4})$$

$$F_{i,k,l}^i + F_{k,i,l}^i + F_{k,l,i}^i = 1 - F_{i,k,l}^{k,l} + F_{l,i,k}^i + F_{k,i,l}^i, \qquad (\text{E.5})$$

42

where $F_{i,k,l}^{k,l} = F_{i,k,l}^{k} + F_{i,k,l}^{l}$. Above, we used the fact that $\forall (i, j, k, l),\; F_{i,j,k}^{l} = F_{i,k,j}^{l}$ (i.e., there is no order in the sampling of two individuals) and $F_{i,k,l}^{i} + F_{i,k,l}^{k} + F_{i,k,l}^{l} = 1$ (i.e., with probability one, the SU goes to channel i, j or k at the next iteration).

Moreover, we note that:

$$
\frac{x_j^i}{x_j^2}\left[\sum_{k\neq i} x_j^{k2} + 2x_j^i \sum_{k\neq i} x_j^k + \sum_{k\neq i}\sum_{l\notin\{k,i\}} x_j^l x_j^k + x_j^{i2}\right]
$$

$$
= \frac{x_j^i}{x_j^2}\left[\sum_{k} x_j^{k2} + 2x_j^i \sum_{k\neq i} x_j^k + \sum_{k\neq l} x_j^l x_j^k - x_j^i \sum_{l\neq i} x_j^l - x_j^i \sum_{k\neq i} x_j^k\right]
$$

$$
= \frac{x_j^i}{x_j^2}\sum_{k,l} x_j^l x_j^k
$$

$$
= x_j^i \tag{E.6}
$$

We used here the fact that $\sum_k x_j^k = x_j$.

Equation (E.2) can now be written (we skip the reference to time on the RHS, all x_j^i are functions of t):

$$
\begin{aligned}
x_i^j(t+1) &= x_j^i + \frac{x_j^i}{x_j^2}\sum_{k\neq i} x_j^{k2}[2F_{k,i,k}^i - F_{i,k,k}^k] + \\
&\quad \frac{x_j^{i2}}{x_j^2}\sum_{k\neq i} x_j^k[F_{k,i,i}^i - 2F_{i,i,k}^k] + \\
&\quad \frac{x_j^i}{x_j^2}\sum_{k\neq i}\sum_{l\notin\{k,i\}} x_j^l x_j^k[F_{l,i,k}^i + F_{k,i,l}^i - F_{i,k,l}^{k,l}].
\end{aligned} \tag{E.7}
$$

We now use the following property of the double imitation [34][3] for $i \notin \{j, k\}$:

$$
F_{i,j,k}^{j,k} - F_{j,i,k}^i - F_{k,i,j}^i = \frac{1}{2}Q(\pi_k(t-1))(\pi_j(t-1) - \pi_i(t-1)) +
$$

---

[3]In this reference, the equation (3) of Theorem 1 is wrong. The correct formula is however given in the proof of the theorem in Appendix.

43

$$\frac{1}{2}Q(\pi_j)(\pi_k(t-1) - \pi_i(t-1)). \qquad \text{(E.8)}$$

Payoffs π are functions of $t-1$ because imitation is based on the payoff obtained at the previous iteration). In particular, for $j = k$, we obtain:

$$F_{i,k,k}^k - 2F_{k,i,k}^i \quad = \quad Q(\pi_k(t-1))(\pi_k(t-1) - \pi_i(t-1)). \qquad \text{(E.9)}$$

From these equations, we can simplify (E.7) into (skipping again reference to time on the RHS):

$$\begin{aligned}
x_i^j(t+1) \quad = \quad & x_j^i + \frac{x_j^i}{x_j^2}\sum_k x_j^{k2}Q(\pi_k)(\pi_i - \pi_k) + \\
& \frac{x_j^{i2}}{x_j^2}\sum_k x_j^k Q(\pi_i)(\pi_i - \pi_k) + \\
& \frac{x_j^i}{x_j^2}\sum_k \sum_{l\notin\{k,i\}} x_j^l x_j^k Q(\pi_l)(\pi_i - \pi_k). \qquad \text{(E.10)}
\end{aligned}$$

The term in the last double summation has been obtained by using (E.8), separating the expression in two double sums and interchanging indices j and k in the first double sum. Note also that all terms of the involved sums are null for $k = i$.

We now obtain:

$$\begin{aligned}
x_i^j(t+1) \quad = \quad & x_j^i + \frac{x_j^i}{x_j^2}\sum_k x_j^k(\pi_i - \pi_k)\left[x_j^k Q(\pi_k) + x_j^i Q(\pi_i) + \sum_{l\notin\{k,i\}} x_j^l Q(\pi_l)\right], \\
= \quad & x_j^i + x_j^i \sum_k \frac{x_j^k}{x_j}(\pi_i - \pi_k)\sum_l \frac{x_j^l}{x_j}Q(\pi_l) \\
= \quad & x_j^i + x_j^i Q(\bar\pi_j)(\pi_i - \bar\pi_j), \qquad \text{(E.11)}
\end{aligned}$$

where $\bar\pi_j(t-1) = \sum_k \frac{x_j^k(t)}{x_j(t)}\pi_k(t-1)$ can be interpreted as the average payoff at

the previous iteration of SUs settling now on channel j. We now have:

$$
\begin{aligned}
x_i(t+1) &= \sum_j x_i^j(t+1) \\
&= \sum_j \left[x_j^i(t) + x_j^i(t)Q(\bar{\pi}_j(t-1))(\pi_i(t-1) - \bar{\pi}_j(t-1)) \right] \\
&= x_i(t-1) + \sum_j x_j^i(t)Q(\bar{\pi}_j(t-1))(\pi_i(t-1) - \bar{\pi}_j(t-1)).
\end{aligned}
$$

$$(\text{E.12})$$

We used the fact that $\sum_j x_j^i(t+1) = x_i(t)$. This concludes the proof.