

Circumventing the Price of Anarchy: Leading Dynamics to Good Behavior

Maria-Florina Balcan¹ Avrim Blum² Yishay Mansour³

¹School of Computer Science, College of Computing, Georgia Institute of Technology

²Department of Computer Science, Carnegie Mellon University

³Blavatnik School of Computer Science, Tel Aviv University
ninamf@cc.gatech.edu avrim@cs.cmu.edu mansour@tau.ac.il

Abstract: Many natural games can have a dramatic difference between the quality of their best and worst Nash equilibria, even in pure strategies. Yet, nearly all work to date on dynamics shows only convergence to *some* equilibrium, especially within a polynomial number of steps. In this work we study how agents with some knowledge of the game might be able to quickly (within a polynomial number of steps) find their way to states of quality close to the *best* equilibrium. We consider two natural learning models in which players choose between greedy behavior and following a proposed good but untrusted strategy and analyze two important classes of games in this context, fair cost-sharing and consensus games. Both games have extremely high Price of Anarchy and yet we show that behavior in these models can efficiently reach low-cost states.

Keywords: Dynamics in Games, Price of Anarchy, Price of Stability, Cost-sharing games, Consensus games, Learning from untrusted experts

1 Introduction

There has been substantial work in the machine learning, game theory, and (more recently) algorithmic game theory communities on understanding the overall behavior of multi-agent systems in which agents follow natural learning dynamics such as (randomized) best/better response and no-regret learning. For example, it is well known that in potential games, best-response dynamics, in which players take turns each making a best-response move to the current state of all the others, is guaranteed to converge to a pure-strategy Nash equilibrium [24, 27]. Significant effort has been spent recently on analyzing various properties of these dynamics and their variations, in particular on their convergence time [1, 26]. No regret dynamics have also been long studied. For example, a well known general result that applies to any finite game is that if players each follow a “no-internal-regret” strategy, then the empirical distribution of play is guaranteed to approach the set of correlated equilibria of the game [17–19].

There has also been a lot of attention recently on fast convergence of both best response and regret minimization dynamics to states with cost comparable to the Price of Anarchy of the game [4, 15, 26, 28], whether or not that state is an equilibrium. This line of work is justified by the realization that in many cases the equilibrium nature of the system itself is less important, and what we care more about is having the

dynamics reach a low-cost state; this is a position we adopt as well.

However, while the above results are quite general, the behavior or equilibrium reached by the given dynamics could be as bad as the *worst* equilibrium in the game (the worst pure-strategy Nash equilibrium in the case of best-response dynamics in potential games, and the worst correlated equilibrium for no-regret algorithms in general games). Even for specific efficient algorithms and even for natural potential games, in general no bounds better than the pure-strategy price of anarchy are known. On the other hand, many important potential games, including fair cost-sharing and consensus games, can have very high-cost Nash equilibria even though they also always have low-cost equilibria as well; that is, their Price of Anarchy is extremely high and yet their Price of Stability is low (we discuss these games more in Section 1.1). Thus, a guarantee comparable with the cost of the worst Nash equilibrium can be quite unsatisfactory.

Unfortunately, in general there have been very few results showing natural dynamics that lead to low cost equilibria or behavior in games of this type, especially within a polynomial number of steps. For potential games, it is known that noisy best-response in which players act in a “simulated-annealing” style manner, also known as log-linear learning, *does* have the property that the states of highest probability in the limit are those minimizing the global poten-

tial [8, 9, 23]. In fact, as temperature approaches 0, these are the only stochastically-stable states. However, as we show in Appendix A.2, reaching such states with non-negligible probability mass may easily require exponential time. For polynomial-time processes, Charikar et al. [11] show natural dynamics rapidly reaching low-cost behavior in a special case of undirected fair cost-sharing games; however, these results require specialized initial conditions and also do not apply to directed graphs.

In this paper we initiate a study of how to aid dynamics, beginning from arbitrary initial conditions, to reach states of cost close to the *best* Nash equilibrium. We propose a novel angle on this problem by considering whether providing more information to simple learning algorithms about the game being played can allow natural dynamics to reach such high quality states. At a high level, there are two barriers to simple dynamics performing well. One is computational: for directed cost-sharing games for instance, we do not even know of efficient *centralized* procedures for finding low-cost states in general. As an optimization problem this is the Directed Steiner Forest problem and the best approximation factor known is $\min(n^{1/2+\epsilon}, N^{4/5+\epsilon}, m^{2/3}N^\epsilon)$ where n is the number of players, N is the number of vertices and m is the number of edges in the graph [12, 16]. The other barrier is incentive-based: even if a low-cost solution were known, there would still be the issue of whether players would individually want to play it, especially without knowing what other players will choose to do. For example, low-cost solutions might be known because people analyzing the specific instance being played might discover and publish low cost global behaviors. In this case, individual players might then occasionally test out their parts of these behaviors, using them as extra inputs to their own learning algorithm or adaptive dynamics to see if they do in fact provide benefit to themselves. The question then is can this process allow low cost states to be reached and in what kinds of games?

Motivated by this question, in this paper we develop techniques for understanding and influencing the behavior of natural dynamics in games with multiple equilibria, some of which may be of much higher social quality than others. In particular we consider a model in which a low cost global behavior is proposed, but individual players do not necessarily trust it (it may not be an equilibrium, and even if it is they do not know if others will follow it). Instead, players use some form of experts learning algorithm where one expert says to play the best response to the current state and another says to follow the proposed strategy.

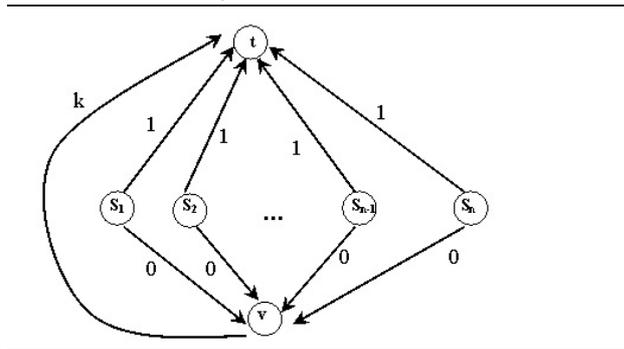
Our model imposes only very mild conditions on the learning algorithm each player uses: different players may use completely different algorithms for deciding among or updating probabilities on their two high-level choices. Assume that the players move in a random order. Will this produce a low-cost solution (even if it doesn't converge to anything)? We consider two variations of this model: a *learn-then-decide* model where players initially follow an "exploration" phase where they put roughly equal probability on each expert, followed by a "commitment" phase where based on their experience they choose one to use from then on, and a *smoothly adaptive* model where they slowly change their probabilities over time. Within each we analyze several important classes of games. In particular, our main results are that for both fair cost-sharing and consensus games, these processes lead to good quality behavior.

Our study is motivated by several lines of work. One is the above-mentioned work on noisy best-response dynamics which reach high-quality states but only after exponentially many steps. Another is work on the *value of altruism* [29] which considers how certain players acting altruistically can help the system reach a better state. The last is the work in [5] which considers a central authority who can *temporarily control* a random subset of the players in order to get the system into a good state. That model is related to the model we consider here, but is much more rigid because it posits two classes of players (one that follows the given instructions and one that doesn't) and does not allow players to adaptively decide for themselves. (See Section 1.2 for more discussion).

1.1 Our Results

As described above we introduce and analyze two models for guiding dynamics to good equilibria, and within these models we prove strong positive results for two classes of games, fair cost sharing and consensus games. In n -player *fair cost sharing* games, players choose routes in a network and split the cost of edges they take with others using the same edge; see Section 3 for a formal definition. These games can model scenarios such as whether to drive one's own car to work or to share public transportation with others (see Figure 1 for a simple example) and can have equilibria as much as a factor of n worse than optimal even though they are always guaranteed to have low-cost equilibria that are only $O(\log n)$ worse than optimal as well. In *consensus games*, players are nodes in a network who each need to choose a color, and they pay a cost for each neighbor of different color than their own (see Section 4). These again can have

Figure 1 A directed cost-sharing game, that models a setting where each player can choose either to drive its own car to work at a cost of 1, or share public transportation with others, splitting an overall cost of k . For any $1 < k < n$, if players arrive one a time and each greedily chooses a path minimizing its cost, then the cost of the equilibrium obtained is n , whereas **OPT** has cost only k .



a wide gap between worst and best Nash equilibrium (in this case, cost $\Omega(n^2)$ for the worst versus 0 for the best).

Our main results for these games are the following. For fair cost-sharing, we show that in the learn-then-decide model, so long as the exploration phase has sufficient (polynomial) length and the proposed strategy is near-optimal, the expected cost of the system reached is $O(\log(n) \log(nm) \mathbf{OPT})$. Thus, this is only slightly larger than that given by the price of stability of the game. For the smoothly-adaptive model, if there are many players of each type (i.e., associated to each (s_i, t_i) pair) we can do even better, with high probability achieving cost $O(\log(nm) \mathbf{OPT})$, or even $O(\mathbf{OPT})$ if the number of players of each type is high enough. Note that with many players of each type the price of anarchy remains $\Omega(n)$ though the price of stability becomes $O(1)$. For consensus games, we show that so long as players place probability $\beta > 1/2$ on the proposed optimal strategy, then with high probability play will reach the exact optimal behavior within a polynomial number of steps. Moreover for certain natural graphs such as the line and the grid, any $\beta > 0$ is sufficient. Note that our results are actually *stronger* than those achievable in the more centralized model of [5], where one needs to make certain minimal degree assumptions on the graph as well.

In both our models it is an easy observation that for any game, if the proposed solution is a good equilibrium, then in the limit the system will *eventually* reach the equilibrium and stay there indefinitely. Our interest, however, is in polynomial-time behavior.

1.2 Related Work

Dynamics and Convergence to Equilibria: It is well known that in potential games, best-response dynamics, in which players take turns each making a best-response move to the current state of all the others, is guaranteed to converge to a pure-strategy Nash equilibrium [24, 27]. Significant effort has been spent recently on the convergence time of these dynamics [1, 26], with both examples of games in which such dynamics can take exponential time to converge, and results on fast convergence to states with cost comparable to the Price of Anarchy of the game [4, 15].

Another well known general result that applies to any finite game is that if players each follow a “no-internal-regret” strategy, then the empirical distribution of play is guaranteed to approach the set of correlated equilibria of the game [17–19]. In particular, for good no-regret algorithms, the empirical distribution will be an ϵ -correlated equilibrium after $O(1/\epsilon^2)$ rounds of play [7]. A recent result of [20] analyzes a version of the weighted-majority algorithm in congestion games, showing that behavior converges to the set of weakly-stable equilibria. This implies performance better than the worst correlated equilibrium, but even in this case the guarantee is no better than the worst pure-strategy Nash equilibrium.

Noisy best-response has been shown in the limit to reach states of minimum global potential [8, 9, 23], and thus provide strong positive results in games with a small gap between potential and cost. However, this convergence may take exponential time, even for fair cost-sharing games. In particular we show in Appendix A.2 that if one makes many copies of the edges of cost 1 in the example of Figure 1, reaching a state with sufficiently many players on the shared path to induce others to follow along would take time exponential in k even if the algorithm can arbitrarily vary its temperature parameter. For details see Appendix A.2.

An alternative line of work assumes that the system starts empty and players join one at a time. Charikar et al. [11] analyze fair cost sharing in this setting on an *undirected* graph where all players have a common sink. Their model has two phases: in the first, players enter one at a time and use a greedy algorithm to connect to the current tree, and in the second phase the players undergo best-response dynamics. They show that in this case a good equilibrium (one that is within only a $\text{polylog}(n)$ factor of optimal) is reached. We discuss this result further in Appendix A. We remark here that for directed graphs, it is easy to construct simple examples where this process reaches an equilibrium that is $\Omega(n)$ from optimal (see Fig-

ure 1). For scheduling on unrelated machines with makespan cost function, the natural greedy algorithm to assign incoming jobs is an $O(m)$ approximation and a more sophisticated online algorithm guarantees an $O(\log m)$ approximation [3]. This implies that in a two phase model, where in the first phase the jobs join one at a time and use the greedy algorithm (or that of [3]), and in the second phase they perform a best response dynamics, achieves cost within a factor $O(m)$ (or $O(\log m)$) of optimal. This is in contrast to the unbounded Price of Anarchy in general.

Taxation: There has also been work on using taxes to improve the quality of behavior [13, 14]. Here the aim is via taxes to *adjust* the utilities of each player, such that the only Nash equilibria in the new game correspond to optimal or near-optimal behavior in the original game. In contrast, our focus is to identify how dynamics can be made to reach a good result without changing the game or adjusting utilities, but rather by injecting more information into the system.

Public service advertising: Finally, the *public service advertising* model of [5] also uses the idea of a proposed strategy in order to move players into a good equilibrium. In the model of [5], the players only once select (randomly) between following the proposed strategy or not. Players then stick with their decision while those that decided not to follow the proposed strategy settle on some equilibrium for themselves (given the other players actions are fixed). Then, in the last phase all players perform a best response dynamics to converge to a final equilibrium (the convergence is guaranteed since the discussion is limited to potential games). In contrast, in our models the players continuously randomly re-select between following the proposed strategy or performing a best response. This continuous randomization makes the analysis of our dynamics technically more challenging. Conceptually, the benefit of our model is that the players are “symmetric” and can continuously switch between the two alternatives, which better models selfish behavior. This continuous process is what enables the players to both explore and exploit the two alternative actions.

2 A Formal Framework

2.1 Notation and Definitions

We start by providing general notations and definitions. A game is denoted by a tuple

$$\mathcal{G} = \langle N, (\mathcal{S}_i), (\text{cost}_i) \rangle$$

where N is a set of n players, \mathcal{S}_i is the finite action space of player $i \in N$, and cost_i is the cost func-

tion of player i . The joint action space of the players is $\mathcal{S} = \mathcal{S}_1 \times \dots \times \mathcal{S}_n$. For a joint action $s \in \mathcal{S}$ we denote by s_{-i} the actions of players $j \neq i$, i.e., $s_{-i} = (s_1, \dots, s_{i-1}, s_{i+1}, \dots, s_n)$. The cost function of player i maps a joint action $s \in \mathcal{S}$ to a real non-negative number, i.e., $\text{cost}_i : \mathcal{S} \rightarrow \mathbb{R}^+$. Every game has associated a social cost function $\text{cost} : \mathcal{S} \rightarrow \mathbb{R}$ that maps a joint action to a real value. In the cases discussed in this paper the social cost is simply the sum of players’ costs, i.e.,

$$\text{cost}(s) = \sum_{i=1}^n \text{cost}_i(s).$$

The optimal social cost is

$$\mathbf{OPT}(\mathcal{G}) = \min_{s \in \mathcal{S}} \text{cost}(s).$$

We sometimes overload notation and use \mathbf{OPT} for a joint action s that achieves cost $\mathbf{OPT}(\mathcal{G})$.

Given a joint action s , the *Best Response (BR)* of player i is the set of actions $BR_i(s)$ that minimizes its cost, given the other players actions s_{-i} , i.e., $BR_i(s_{-i}) = \arg \min_{a \in \mathcal{S}_i} \text{cost}_i(a, s_{-i})$.

A joint action $s \in \mathcal{S}$ is a *pure Nash Equilibrium (NE)* if no player $i \in N$ can benefit from unilaterally deviating to another action, namely, every player is playing a best response action in s , i.e., $s_i \in BR_i(s_{-i})$ for every $i \in N$. A best response dynamics is a process in which at each time step, some player which is not playing a best response switches its action to a best response action, given the current joint action. In this paper we focus on potential games, which have the property that any best response dynamics converges to a pure Nash equilibrium [24].

Let $\mathcal{N}(\mathcal{G})$ be the set of Nash equilibria of the game \mathcal{G} . The *Price of Anarchy (PoA)* is defined as the ratio between the maximum cost of a Nash equilibrium and the social optimum, i.e.,

$$\max_{s \in \mathcal{N}(\mathcal{G})} \text{cost}(s) / \mathbf{OPT}(\mathcal{G}).$$

The *Price of Stability (PoS)* is the ratio between the minimum cost of a Nash equilibrium and the social optimum, i.e.,

$$\min_{s \in \mathcal{N}(\mathcal{G})} \text{cost}(s) / \mathbf{OPT}(\mathcal{G}).$$

For a class of games, the PoA and PoS are the maximum over all games in the class

2.2 The Model

We now describe the formal model that we consider. Initially, players begin in some arbitrary state,

which could be a high-cost equilibrium or even a state that is not an equilibrium at all. Next an entity, perhaps the designer of the system or a player who has studied the system well, proposes some better global behavior \mathcal{B} . \mathcal{B} may or may not be an equilibrium behavior, our only assumption is that it have low overall cost. Now, players move one at a time in a random order. Each player, when it is their turn to move, chooses among two options. The first is to simply behave greedily and to make a best-response move to the current configuration. The second option is to follow their part of the given behavior \mathcal{B} . These two high-level strategies (best-response or follow \mathcal{B}) are viewed as two “experts” and the player then runs some learning algorithm aiming to learn which of these is most suitable for himself. Note that best-response is an *abstract* option — the specific action it corresponds to may change over time.

Because moves occur in an asynchronous manner, there are multiple reasonable ways to model the feedback each player gives to its learning algorithm: for instance, does it consider the average cost since the player’s previous move or just the cost when it is the player’s turn to go, or something in between? In addition, does it get to observe the cost of the action it did not take (the full information model) or only the action it chose (the bandit model)? To abstract away these issues, and even allow different players to address them differently, we consider here two models that make only very mild assumptions on the kind of learning and adaptation made by players.

Learn then Decide model: In this model, players follow an “exploration” phase where each time it is their turn to move, they flip a coin to decide whether to follow the proposed behavior \mathcal{B} or to do a best-response move to the current configuration. We assume that the coin gives probability at least β to \mathcal{B} , for some constant $\beta > 0$. Finally, after some common time T^* , all players switch to an “exploitation” phase where they each commit in an arbitrary way based on their past experience to follow \mathcal{B} or perform best response from then on. (The time T^* is selected in advance.)

The above model assumes some degree of coordination: a fixed time T^* after which all players make their decisions. One could imagine instead each player i having its own time T_i^* at which it commits to one of the experts, perhaps with the time itself depending on the player’s experience. In the Smoothly-Adaptive model below we even more generally allow players to smoothly adjust their probabilities over time as they like, subject only to a constraint on the amount by which probabilities may change between time steps.

Smoothly Adaptive model: In this model, there are no separate exploration and exploitation phases. Instead, each player i maintains and adjusts a value p_i over time. When player i is chosen to move, it flips a coin of bias p_i to select between the proposed behavior or a best-response move, choosing \mathcal{B} with probability p_i . We allow the players to use arbitrary adaptive learning algorithms to adjust these probabilities with the sole requirement that learning proceed slowly. Specifically, using p_i^t to denote the value of p_i at time t , we require that

$$|p_i^t - p_i^{t+1}| \leq \Delta$$

for a sufficiently (polynomially) small quantity Δ , and furthermore that for all i , the initial probability $p_i^0 \geq p^0$ for some overall constant $0 < p^0 < 1$. Note that the algorithm may update p_i even in time steps at which it does not move. The learning algorithm may use any kind of feedback or weight-updating strategy it wants to (e.g., gradient descent [30, 31], multiplicative updating [10, 21, 22]) subject to this bounded step-size requirement.

We say that the probabilities are (T, β) -good if for any time $t \leq T$ we have for all i , $p_i^t > \beta$. (Note that if $\Delta < (p^0 - \beta)/T$ then clearly the probabilities are (T, β) -good.)

We point out that while one might at first think that any natural adaptive algorithm would learn to favor best-response (always decreasing p_i), this depends on the kind of feedback it uses. For instance, if the algorithm considers only its cost immediately after it moves, then indeed by definition best-response will appear better. However, if it considers its cost immediately *before* it moves (comparing that to what its cost would have been had it chosen the other alternative) or even the sum total cost since its previous move, then \mathcal{B} might appear better. Our model allows users to update in any way they wish, so long as the updates are sufficiently gradual.

Finally, as mentioned in the introduction, in both our models it is an easy observation that for any game, if the proposed solution is a good equilibrium, then in the limit (as $T^* \rightarrow \infty$ in the Learn-then-Decide model or as $\Delta \rightarrow 0$ in the Smoothly Adaptive model) the system will *eventually* reach the equilibrium and stay there indefinitely. Our interest, however, is in polynomial-time behavior.

3 Fair Cost Sharing

The first class of games we study in this paper, because of its rich structure and wide gap between price

of anarchy and price of stability, is that of *fair cost sharing games*. These games are defined as follows. We are given a graph $G = (V, E)$, which can be directed or undirected, where each edge $e \in E$ has a nonnegative cost $c_e \geq 0$. There is a set $N = \{1, \dots, n\}$ of n players, where player i is associated with a source s_i and a sink t_i . The strategy set of player i is the set \mathcal{S}_i of $s_i - t_i$ paths. In an outcome of the game, each player i chooses a single path $P_i \in \mathcal{S}_i$. Given a vector of players' strategies $s = (P_1, \dots, P_n)$, let x_e be the number of agents whose strategy contains edge e . In the fair cost sharing game the cost to agent i is

$$\text{cost}_i(s) = \sum_{e \in P_i} \frac{c_e}{x_e}$$

and the goal of each agent is to connect its terminals with minimum total cost. The social cost of an outcome $s = (P_1, \dots, P_n)$ is defined to be

$$\text{cost}(P_1, \dots, P_n) = \sum_{e \in \cup_i P_i} c_e.$$

It is well known that fair cost sharing games are potential games [2, 24] and the price of anarchy in these games is $\Theta(n)$ while the price of stability is $H(n)$ [2], where $H(n) = \sum_{i=1}^n 1/i = \Theta(\log n)$. In particular, the potential function for these games is

$$\Phi(s) = \sum_{e \in E} \sum_{x=1}^{x_e} c_e/x,$$

which satisfies the following inequality:

Fact 1 *In fair cost sharing, for any $s \in \mathcal{S}$ we have: $\text{cost}(s) \leq \Phi(s) \leq H(n) \cdot \text{cost}(s)$.*

For ease of notation, we assume in this section that the proposed strategy \mathcal{B} is the socially optimal behavior **OPT**, so we can identify $P_i^{\text{OPT}} = P_i^{\mathcal{B}}$ as the behavior proposed by \mathcal{B} to player i . If \mathcal{B} is different from **OPT**, then we simply lose the corresponding approximation factor.

Overview of the Results and Analysis: Our main results for fair cost sharing are the following. If we have many players of each type (the type of a player is determined by its source s_i and destination t_i) then in both the learn-then-decide and smoothly adaptive models we can show that with high probability, behavior will reach a state of cost within a logarithmic factor of **OPT** within a polynomial number of steps. Moreover, for the learn-then-decide model, even if we do not have many players of each type we can show that the *expected* cost at the end of the process will be low.

The high level idea of the analysis is that we first prove that so long as each player randomizes with probability near to 50/50, with high probability the overall cost of the system will drop to within a logarithmic factor of **OPT** in a polynomial number of steps; moreover, at that point both the best response and the proposed actions are pretty good strategies from the individual players point of view. To finish the analysis in the ‘‘Learn then Decide’’ model, we show that in the remaining steps of the exploration phase the *expected* cost does not increase by much; using properties of the potential function, we then show that in the final ‘‘decision’’ round T^* , the overall potential cannot increase substantially either, which in turn implies a bound on the increase in overall cost.

For the adaptive model, one key difficulty in the analysis is to show that if the system reaches a state where the social cost is low and both abstract actions are pretty good for most players, the cost never goes high again. We are able to show that this indeed is the case as long as there are many players of each type, no matter how the players adjust their probabilities p_i^t or make their choice between best-response and following the proposed behavior.

3.1 The Main Arguments

We begin with the following key lemma that is useful in the analysis of both learn-then-decide and smoothly adaptive models.

Lemma 2 *Consider a fair cost sharing game. There exists a $T = \text{poly}(n)$, such that if the probabilities are (T, β) -good for constant $\beta > 0$, then with high probability the cost at time T will be at most $O(\text{OPT} \log(mn))$.*

Moreover, if we have at least $c \log(nm)$ players of each (s_i, t_i) pair for sufficiently large constant c , then with high probability the cost at time T will be $O(\text{OPT})$.

Proof: We begin with the general case. Let n_e^{opt} denote the number of players who use edge e in **OPT**. We partition edges into two classes. We say an edge is a ‘‘high traffic’’ edge if $n_e^{\text{opt}} > c \log(nm)$ where c is a sufficiently large constant ($c = 32/\beta$ suffices for the argument below). We say it is a ‘‘low traffic’’ edge otherwise.

Define $T_0 = 2n \log n$. With high probability, by time T_0 each player has had a chance to move at least once. We assume in the following that this indeed is the case. Note that as a crude bound, at this point the cost of the system is at most $n^2 \cdot \text{OPT}$ (each player will move to a path of cost-share at most **OPT** and therefore of actual cost at most $n \cdot \text{OPT}$). Next, by

Chernoff bounds and the union bound, our choice of c implies that with high probability each high-traffic edge e has at least $\beta n_e^{opt}/2$ players on it at all time steps $T \in [T_0, T_0 + n^3]$; in particular, Chernoff bounds imply that each of the at most mn^3 events has probability at least $1 - e^{-\beta n_e^{opt}/8} \geq 1 - 1/(mn)^4$. In the remaining analysis, we assume this indeed is the case as well.

Let \mathbf{OPT}_i denote the cost of player i in \mathbf{OPT} , so that $\mathbf{OPT} = \sum_i \mathbf{OPT}_i$. Our assumption above implies that for any time step T under consideration, if player i follows the proposed strategy P_i^{OPT} , its cost will be at most $c \log(nm) \mathbf{OPT}_i$. In particular, its cost on the low-traffic edges in P_i^{OPT} can be at most a factor $c \log(nm)$ larger than its cost on those edges under \mathbf{OPT} , and its cost on high-traffic edges is at most a $2/\beta$ factor larger than its cost on those edges under \mathbf{OPT} .

We now argue as follows. Let \mathbf{cost}_T denote the cost of the system at time T . If

$$\mathbf{cost}_T \geq 2c \log(nm) \mathbf{OPT},$$

then the expected cost of a random player is at least

$$2c \log(nm) \mathbf{OPT}/n.$$

On the other hand, if player i is chosen to move at time T , from the above analysis its cost after the move (whether it chooses \mathcal{B} or best response) will be at most $c \log(nm) \mathbf{OPT}_i$. The expected value of this quantity over players i chosen at random is at most $c \log(nm) \mathbf{OPT}/n$. Therefore, if

$$\mathbf{cost}_T \geq 2c \log(nm) \mathbf{OPT},$$

the expected drop in potential at time T is at least

$$c \log(nm) \mathbf{OPT}/n.$$

Finally, since the cost at time T_0 was at most $n^2 \cdot \mathbf{OPT}$, which implies by Fact 1 the value of the potential was at most $n^2(1 + \log(n)) \mathbf{OPT}$, with high probability this cannot continue for more than $O(n^3)$ steps. Formally, we can apply Hoeffding-Azuma bounds for supermartingales as follows: let us define $Q = c \log(nm) \mathbf{OPT}$ and

$$\Delta_T = \max(\Phi_T - \Phi_{T-1} + Q/n, -2Q)$$

and consider running this process stopping when $\mathbf{cost}_T < 2Q$. Let

$$X_T = \Phi_0 + \Delta_1 + \dots + \Delta_T.$$

Then throughout the process we have

$$\mathbf{E}[X_T | X_1, \dots, X_{T-1}] \leq X_{T-1}$$

and

$$|X_T - X_{T-1}| \leq 2Q,$$

where the first inequality holds because our analysis showing an expected decrease in potential of at least Q/n is true even if we cap all decreases to a maximum of $2Q$ as in the definition of Δ_T . So, by Hoeffding-Azuma (see Theorem 10 in Appendix B), after n^3 steps in the non-stopped process with high probability we would have

$$X_T - X_0 \leq \frac{1}{2} n^2 Q,$$

which is not possible since by definition of X_T we have

$$\Phi_T \leq \Phi_0 + (X_T - X_0) - TQ/n$$

which would be negative. Therefore, with high probability stopping must occur before this time as desired.

Finally, if we have at least $c \log(nm)$ players of each type, then there are no low-traffic edges and so we do not need to lose the $c \log(nm)$ factor in the argument.

We now present a second lemma which will be used to analyze the ‘‘Learn then Decide’’ model.

Lemma 3 *Consider a fair cost sharing game in the Learn-then-Decide model. If the cost of the system at time T_1 is $O(\mathbf{OPT} \log(mn))$, and $T = T_1 + \text{poly}(n) < T^*$, then the expected value of the potential at time T is $O(\mathbf{OPT} \log(mn) \log(n))$.*

Proof: First, as argued in the proof of Lemma 2, with high probability for any player i and any time $t \in [T_1, T]$, the cost for player i to follow the proposed strategy at time t is at most $c \log(nm) \mathbf{OPT}_i$ for some constant c . Let us assume below that this is indeed the case.

Next, the above bound implies that if the cost at time $t \in [T_1, T]$ is \mathbf{cost}_t , then the expected decrease in potential caused by a random player moving (whether following the proposed strategy or performing best response) is at least

$$(\mathbf{cost}_t - c \log(nm) \mathbf{OPT})/n;$$

in particular, \mathbf{cost}_t/n is the expected cost of a random player before its move, and $c \log(nm) \mathbf{OPT}/n$ is an upper bound on the expected cost of a random player after its move. Note that this is an expectation over randomness in the choice of player at time t , conditioned on the value of \mathbf{cost}_t . In particular, since this holds true for any value of \mathbf{cost}_t , we can take expectation over the entire process from time T_1 up to t , and we have that if

$$\mathbf{E}[\mathbf{cost}_t] \geq c \log(nm) \mathbf{OPT}$$

then

$$\mathbf{E}[\Phi_{t+1}] \leq \mathbf{E}[\Phi_t],$$

where Φ_t is the value of the potential at time step t .

Finally, since $\Phi_t \leq \log(n) \mathbf{cost}_t$, this implies that if

$$\mathbf{E}[\Phi_t] \geq c \log(nm) \log(n) \mathbf{OPT}$$

then

$$\mathbf{E}[\Phi_{t+1}] \leq \mathbf{E}[\Phi_t].$$

Since for any value of \mathbf{cost}_t we always have

$$\mathbf{E}[\Phi_{t+1}] \leq \mathbf{E}[\Phi_t] + c \log(nm) \mathbf{OPT}/n,$$

this in turn implies by induction on t that

$$\begin{aligned} \mathbf{E}[\Phi_t] &\leq \\ &c \log(nm) \log(n) \mathbf{OPT} + c \log(nm) \mathbf{OPT}/n \end{aligned}$$

for all $t \in [T_1, T]$, as desired.

Our main result in the ‘‘Learn then Decide’’ model is the following:

Theorem 4 *For fair cost sharing games in the Learn then Decide model, a polynomial number of exploration steps T^* is sufficient so that the expected cost at any time $T' \geq T^*$ is $O(\log(n) \log(nm) \mathbf{OPT})$.*

Proof: From Lemma 2, there exists $T = \text{poly}(n)$ such that with high probability the cost of the system will be at most $O(\mathbf{OPT} \log(mn))$ at some time $T_1 \in [T^* - T, T^*]$. From Lemma 3, this implies the expected value of the potential at time T^* right before the final exploitation phase is $O(\mathbf{OPT} \log(mn) \log(n))$. Finally, we consider the decisions made at time T^* . When a player chooses to make a best-response move, this can only decrease potential. When a player chooses \mathcal{B} , this could *increase* potential. However, for any edge e in the proposed solution \mathcal{B} , the total increase in potential caused by edge e over *all* players who have e in their proposed solution is at most

$$c_e \cdot H(n_e^{\text{opt}}) = O(c_e \log n).$$

This is because whenever a new player makes a decision to commit to following the proposed strategy and using edge e , all previous players who made that commitment after time T and whose proposed strategy uses edge e are still there. Thus, the total increase in potential after time T^* is at most $O(\mathbf{OPT} \log n)$.

Since after all players have committed, potential can only decrease, this implies that the expected value of the potential at any time $T' \geq T^*$ is $O(\mathbf{OPT} \log(mn) \log(n))$. Therefore, the expected cost at time T' is at most this much as well, as desired.

We now use Lemma 2 to analyze fair cost sharing games in the adaptive learning model when the number of players n_i of each type (i.e., associated to each (s_i, t_i) pair) is large.

Theorem 5 *Consider a fair cost sharing game in the adaptive learning model satisfying $n_i = \Omega(m)$ for all i . There exists a $T_1 = \text{poly}(n)$ such that if the probabilities are (T_1, β) -good for constant $\beta > 0$, then with high probability, for all $T \geq T_1$ the cost at time T is $O(\log(nm) \mathbf{OPT})$.*

Moreover, there exists constant c such that if $n_i \geq \max[m, c \log(mn)]$ then with high probability, for all $T \geq T_1$ the cost at time T is $O(\mathbf{OPT})$.

Proof: First, by Lemma 2 we have that with high probability at some time $T_0 \leq T_1$, the cost of the system reaches $O(\mathbf{OPT} \log(mn))$. The key to the argument now is to prove that once the cost becomes low, it will *never* become high again. To do this we use the fact that n_i is large for all i . Our argument which follows the proof of Theorem 4.3 in [6] is as follows.

Let U be the set of all edges in use at time T_0 along with all edges used in \mathcal{B} . In general, we will insert an edge into U if it is ever used from then on in the process, and we never remove any edge from U , even if it later becomes unused. Let c^* be the total cost of all edges in U . So, c^* is an upper bound on the cost of the current state and the only way c^* can increase is by some player choosing a best response path that includes some edge not in U . Now, notice that any time a best-response path for some (s_i, t_i) player uses such an edge, the total cost of all edges inserted into U is at most c^*/n_i , because the current player can always choose to take the path used by the least-cost player of his type and those n_i players are currently sharing a total cost of at most c^* . Thus, any time new edges are added into U , c^* increases by at most a multiplicative $(1 + 1/n_i)$ factor. We can insert edges into U at most m times, so the final cost is at most

$$\mathbf{cost}(T_0)(1 + 1/n_{i^*})^m,$$

where $n_{i^*} = \min_i n_i$. This implies that as long as $n_{i^*} = \Omega(m)$ we have $\mathbf{cost}(T) = O(\mathbf{cost}(T_0))$ for all $T \geq T_0$. Thus, overall the total cost remains $O(\mathbf{OPT} \log(mn))$.

If $n_i \geq \max[m, c \log(mn)]$, we simply use the improved guarantee provided by Lemma 2 to say that with high probability the cost of the system reaches $O(\mathbf{OPT})$ within $T_0 \leq T_1$ time steps, and then apply the charging argument as above in order to get the desired result.

Variations: One could imagine a variation on our framework where rather than proposing a fixed behavior \mathcal{B} , one can propose a more complex strategy such as “follow \mathcal{B} until time step T^* and then follow \mathcal{B}' ” or “follow \mathcal{B} until time step T^* and then perform best-response”. In the latter case, the Smoothly-Adaptive model with (T^*, β) -good probabilities becomes essentially a special case of the Learn-then-Decide model and all results above for the Learn-then-Decide model go through immediately. On the other hand, this type of strategy requires a form of global coordination that one would prefer to avoid.

4 Consensus Games

Another interesting class of games we consider is that of consensus games. Here, players are vertices in an n -vertex graph G , and each have two choices of action, *red* or *blue*. Players incur a cost equal to the number of neighbors they have of different color, and the overall social cost is the sum of costs of each player. The potential function for consensus games is simply half the social cost function, or equivalently the number of edges having endpoints of different color. While in these games optimal behavior is trivial to describe (all red or all blue for a total cost of 0) and is an equilibrium, they also have extremely poor equilibria as well. For example consider two cliques of $n/2$ vertices each, with each node also having $\alpha n/2$ neighbors in the other clique for some $0 < \alpha < 1$. In this case, there is a Nash equilibrium with one clique red and one clique blue, for an overall cost of $\Omega(n^2)$.¹ This is substantially worse than optimal (either by an $\Omega(n^2)$ or infinite factor, depending on whether one allows an additive offset or not).

Unlike in the case of fair cost-sharing games, for consensus games there is no hope of quickly achieving near-optimal behavior for *all* $\beta > 0$. In particular, for any $\beta < 1/2$ there exists $\alpha > 0$ in the example above such that if players choose the proposed strategy \mathcal{B} with probability β , with high probability *all* nodes have more neighbors performing best response than following \mathcal{B} . Thus, it is easy to see by induction that no matter what strategy \mathcal{B} is, all best-response players will remain with their initial color. Moreover, this remains true for exponentially in n many steps.² On

¹Intuitively one can think of this example as two countries, one using English units and one using the metric system, neither wanting to switch.

²Of course, in the limit as the number of time steps $T \rightarrow \infty$ eventually we will observe a sequence in which all players select \mathcal{B} , and if \mathcal{B} is an equilibrium the system will then remain there forever. Thus if \mathcal{B} is “all blue” the system will in the limit converge to optimal behavior. However, our interest is in polynomial time convergence.

the other hand, this argument breaks down for $\beta > 1/2$. In fact, we will show that for $\beta > 1/2$, for any graph G , the system will in polynomial time reach optimal behavior (if $\mathcal{B} = \mathbf{OPT}$).

It is interesting to compare this with the centralized model of [5] in which a central authority takes control of a random constant fraction of the players and aims to use them to guide the selfish behavior of the others. In that setting, even simple low-degree graphs can cause a problem. For instance, consider a graph G consisting of $n/4$ 4-cycles each in the initial equilibrium configuration *red, red, blue, blue*. If the authority controls a random β fraction of players (for any constant $\beta < 1$), with high probability a constant fraction of the 4-cycles contain no players under the authority’s control and will remain in a high-cost state. On the other hand, it is easy to see that this specific example will perform well in both our learn-then-decide or smoothly-adaptive models. We show that in fact *all* graphs G perform well, though this requires care in the argument due to correlations that may arise.

We assume that the proposed behavior \mathcal{B} is the optimal behavior “all blue” and prove that with high probability the configuration will reach this state after $O(n \log^2 n)$ steps. We begin with a simple preliminary lemma.

Lemma 6 *If X_1, \dots, X_d are $\{0, 1\}$ -valued random variables with $\Pr(X_i = 1) \geq p$, then no matter how they are correlated,*

$$\Pr(\text{MAJORITY}(X_1, \dots, X_d) = 1) \geq 2p - 1.$$

Proof: The expected number of 1’s is at least pd . It is also at most

$$d \cdot p_{MAJ} + (d/2) \cdot (1 - p_{MAJ})$$

where

$$p_{MAJ} = \Pr(\text{MAJORITY}(X_1, \dots, X_d) = 1).$$

Solving, we get $p_{MAJ} \geq 2p - 1$.

We now present our main result of this section.

Theorem 7 *In both Learn-then-Decide and Smoothly-Adaptive models, for any constant $\beta > 1/2$, if the proposed behavior \mathcal{B} is optimal then with high probability play will become optimal in $O(n \log^2 n)$ steps. For the Smoothly-Adaptive model we assume behavior is $(cn \log^2 n, \beta)$ -good for sufficiently large constant c .*

Proof: In both models the dynamics contain two random processes: the random order in which players

move and the random coins flipped by the players to determine how they want to move. In order to prove the theorem it will be helpful to separate these two processes and consider them each in turn.

First, consider and fix a random sequence S of players to move. Let T_1 denote the time by which all players have moved at least once according to S , and more generally let T_{t+1} denote the time by which all players have moved at least once since time T_t . Since players move i.i.d. in a random order, with high probability $T_{t+1} \leq T_t + 3n \log(n)$ for all $t = 1, \dots, n$. We assume in the following that this indeed is the case; in fact, we can allow S to be an arbitrary, adversarially-chosen sequence subject to this constraint.

Fixing S , we now consider the coin flips of the individual players. We will prove by induction that each player has probability at least q_t of being blue at time T_t (not necessarily independently) for

$$q_t = (1 - \gamma)q_{t-1} + \gamma \quad \text{for } \gamma = 2\beta - 1.$$

Equivalently, we can write this as

$$1 - q_t = (1 - q_{t-1})(1 - \gamma).$$

Since $\gamma > 0$ (because $\beta > 1/2$), this in turn implies that $t = O(\log n)$ is sufficient to reach $1 - q_t \leq 1/n^2$, meaning that with high probability all nodes are blue as desired.

We prove this bound on q_t as follows. Consider the nodes who move at times $T_{t-1} + 1, T_{t-1} + 2, \dots, T_t$ in order. When some node v moves, by induction each neighbor w of v has probability at least q_t of being blue (though these may not be independent). By assumption, v chooses \mathcal{B} with some probability $\beta' \geq \beta > 1/2$ and chooses best-response with probability $1 - \beta'$. Therefore, the probability v becomes blue is at least:

$$\begin{aligned} & \beta' + (1 - \beta')\Pr(\text{majority of nbrs of } v \text{ are blue}) \\ & \geq \beta' + (1 - \beta')(2q_{t-1} - 1) \quad (\text{by Lemma 6}) \\ & = (1 - (2\beta' - 1))q_{t-1} + (2\beta' - 1) \\ & \geq (1 - \gamma)q_{t-1} + \gamma \quad (\text{for } \gamma = 2\beta - 1) \end{aligned}$$

as desired. Thus, with high probability all nodes are blue by time T_t for $t = O(\log n)$, and by our assumption on S this occurs within $O(n \log^2 n)$ steps.

The general result above requires $\beta > 1/2$, and as noted earlier there exists graphs G and initial configurations such that the process will fail for any $\beta < 1/2$. On the other hand, for several “nice” graphs such as the line or grid, *any* constant $\beta > 0$ is sufficient. For this we assume that best-response will only ask to switch color if the new color is *strictly* better than the current color.

Theorem 8 *For the line and d -dimensional grid graphs (constant d), for any $\beta > 0$, if the proposed action \mathcal{B} is optimal then with high probability play will reach optimal in $\text{poly}(n)$ steps. For the case of the Smoothly-Adaptive model we assume behavior is (T, β) -good for T a sufficiently large polynomial in n .*

Proof: Assume \mathcal{B} is “all blue”. On the line, if any two neighbors become blue, they will remain blue indefinitely. Similarly in the grid, if any d -dimensional cube becomes blue, the nodes in the cube will also remain blue indefinitely. On the line, any neighbors, in any two consecutive steps, have probability at least β^2/n^2 of becoming blue, and on the d -dimensional grid, any cube, in any 2^d consecutive steps, has probability at least $(\beta/n)^{2^d}$ of becoming all blue. Therefore with high probability all nodes become blue in a polynomial number of steps.

5 Conclusions and Future Directions

In this paper we initiate a study of how to aid dynamics, beginning from arbitrary initial conditions, to reach states of cost close to the *best* Nash equilibrium. We propose a novel angle on this problem by considering how providing more information to simple learning algorithms about the game being played can allow the dynamics to reach such low-cost states. We show that for fair cost-sharing and consensus games, proposing a good solution and allowing users to adaptively decide for themselves between that solution and best-response behavior will efficiently lead to near-optimal configurations, so long as users adapt sufficiently slowly. Both of these games have the property that all by itself, random best response may easily end up in a high-cost state: cost $\Omega(n \cdot \mathbf{OPT})$ for fair cost-sharing and cost $\Omega(n^2)$ for consensus, whereas with this additional aid the states reached have cost only $\text{polylog}(n) \cdot \mathbf{OPT}$ for cost-sharing and 0 for consensus.

Open questions and future directions

Our results for fair cost sharing games in the adaptive learning model hold only for the case when the number of players n_i of each type is large. One natural open question is whether similar positive results can be shown for the general case, i.e., when the number of players n_i of each type is arbitrary. It would also be interesting to broaden the types of learning algorithms the players can use. In particular, for the Smoothly Adaptive model, how large a learning rate Δ (or for the Learn-then-Decide model, how small a cutoff time T^*) can one allow and still produce comparable positive results on the quality of the final outcome?

It would also be interesting to extend this model to the case of multiple proposed solutions, some of potentially much higher quality than others, and still show that these dynamics reach good states. In this case, one would need to make additional assumptions on the learning algorithms the players are using. In particular, if players are trying to decide between various learning dynamics (best response, regret minimization, noisy best response, etc.), and if they use an experts learning algorithm to decide between these actions, can we show that if following the the best of the dynamics is good for all players and if the experts learning algorithms have appropriate properties, then the players will do nearly as well as if they had used the best dynamics?

Finally, in both of the games we consider players with the same profile (the same source and destination in the fair cost sharing case) receive the same global advice, which can also be easily communicated as a single message. It would be interesting to also analyze games where players with the same profile might receive different global advice.

One way to view our model is as follows. Starting from the original game, we create a meta-game in which each player is playing one of the two abstract actions, best response and the proposed strategy. We then show that if the players learn in the new meta-game, with restrictions only on the learning rate, then this results in good behavior in the original game. More generally, it would be interesting to further explore the idea of designing “meta games” on the top of the actual games and showing that natural behavior in these meta games induce good behavior in the original game of interest.

Acknowledgments

This work was supported in part by ONR grant N00014-09-1-0751, by AFOSR grant FA9550-09-1-0538, by NSF grant CCF-0830540, by the IST Programme of the European Community under the PASCAL2 Network of Excellence IST-2007-216886, by a grant from the Israel Science Foundation (grant No. 709/09) and by grant No. 2008-321 from the United States-Israel Binational Science Foundation (BSF). This publication reflects the authors’ views only.

References

- [1] H. Ackermann, H. Röglin, and B. Vöcking. On the impact of combinatorial structure on congestion games. In *Proc. 47th FOCS*, 2006.
- [2] E. Anshelevich, A. Dasgupta, J. Kleinberg, E. Tardos, T. Wexler, and T. Roughgarden. The Price of Stability for Network Design with Fair Cost Allocation. In *FOCS*, 2004.
- [3] J. Aspnes, Y. Azar, A. Fiat, S. A. Plotkin, and O. Waarts. On-line routing of virtual circuits with applications to load balancing and machine scheduling. *J. ACM (JACM)*, 44(3):486–504, 1997.
- [4] B. Awerbuch, Y. Azar, A. Epstein, and A. Skopalik. Fast convergence to nearly optimal solutions in potential games. In *ACM Conference on Electronic Commerce*, 2008.
- [5] M.-F. Balcan, A. Blum, and Y. Mansour. Improved Equilibria via Public Service Advertising. In *ACM-SIAM Symposium on Discrete Algorithms*, 2009.
- [6] M.-F. Balcan, A. Blum, and Y. Mansour. The Price of Uncertainty. In *Tenth ACM Conference on Electronic Commerce*, 2009.
- [7] A. Blum and Y. Mansour. Chapter 4: Learning, Regret-Minimization, and Equilibria. In N. Nisan, T. Roughgarden, E. Tardos, and V. V. Vazirani, editors, *Algorithmic Game Theory*. Cambridge University Press, 2007.
- [8] L. Blume. The statistical mechanics of strategic interaction. *Games and Economic Behavior*, 5:387–424, 1993.
- [9] L. Blume. How noise matters. *Games and Economic Behavior*, 44:251–271, 2003.
- [10] N. Cesa-Bianchi, Y. Freund, D. Helmbold, D. Hausler, R. Schapire, and M. Warmuth. How to use expert advice. *Journal of the ACM*, 44(3):427–485, 1997.
- [11] M. Charikar, H. Karloff, C. Mathieu, J. Naor, and M. Saks. Online multicast with egalitarian cost sharing. In *Proc. 20th SPAA*, 2008.
- [12] C. Chekuri, G. Even, A. Gupta, and D. Segev. Set Connectivity Problems in Undirected Graphs and the Directed Steiner Network Problem. In *ACM-SIAM Symposium on Discrete Algorithms*, 2008.
- [13] R. Cole, Y. Dodis, and T. Roughgarden. How much can taxes help selfish routing? In *ACM-EC*, 2003.
- [14] R. Cole, Y. Dodis, and T. Roughgarden. Pricing network edges for heterogeneous selfish users. In *STOC*, 2003.
- [15] A. Fanelli, M. Flamminia, and L. Moscardelli. Speed of convergence in congestion games under best response dynamics. In *Proc. 35th ICALP*, 2008.
- [16] M. Feldman, G. Kortsarz, and Z. Nutov. Improved approximation algorithms for directed steiner forest. In *ACM-SIAM Symposium on Discrete Algorithms*, 2009.
- [17] D. Foster and R. Vohra. Calibrated learning and correlated equilibrium. *Games and Economic Behavior*, 21:40–55, 1997.
- [18] D. Foster and R. Vohra. Regret in the on-line decision problem. *Games and Economic Behavior*, 29:7–36, 1999.
- [19] S. Hart and A. Mas-Colell. A simple adaptive procedure leading to correlated equilibrium. *Econometrica*, 68:1127–1150, 2000.

- [20] R. Kleinberg, G. Piliouras, and E. Tardos. Multiplicative updates outperform generic no-regret learning in congestion games. In *Proc. 41th STOC*, 2009.
- [21] N. Littlestone, P. M. Long, and M. K. Warmuth. Online learning of linear functions. In *Proc. of the 23rd Symposium on Theory of Computing*, pages 465–475. ACM Press, New York, NY, 1991. See also UCSC-CRL-91-29.
- [22] N. Littlestone and M. K. Warmuth. The weighted majority algorithm. *Information and Computation*, 108(2):212–261, 1994.
- [23] J. Marden and J. Shamma. Revisiting log-linear learning: asynchrony, completeness, and payoff-based implementation. Submitted, September 2008.
- [24] D. Monderer and L. Shapley. Potential games. *Games and Economic Behavior*, 14:124–143, 1996.
- [25] R. Motwani and P. Raghavan. *Randomized Algorithms*. Cambridge University Press, 1997.
- [26] N. Nisan, T. Roughgarden, E. Tardos, and V. Vazirani, editors. *Algorithmic Game Theory*. Cambridge, 2007.
- [27] R. Rosenthal. A class of games possessing pure-strategy nash equilibria. *Int. Journal Game Theory*, 2(1), 1973.
- [28] T. Roughgarden. Intrinsic Robustness of the Price of Anarchy. In *STOC*, 2009.
- [29] Y. Sharma and D. P. Williamson. Stackelberg thresholds in network routing games or the value of altruism. In *ACM Conference on Electronic Commerce*, 2007.
- [30] B. Widrow and M. E. Hoff. Adaptive switching circuits. *1960 IRE WESCON Convention Record*, pages 96–104, 1960. Reprinted in *Neurocomputing* (MIT Press, 1988).
- [31] M. Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the 20th International Conference on Machine Learning*, 2003.

A Further Discussion of Other Dynamics

A.1 Players Entering One at a Time

As mentioned in Section 1.2, one form of natural dynamics that have been studied in potential games is where the system starts empty and players join one at a time. Charikar et al. [11] analyze this setting for fair cost sharing on an *undirected* graph where all players have a common sink. They consider a two-phase process. In Phase 1, the players arrive one by one and each connects to the root by greedily choosing a path minimizing its cost, i.e., each selects a greedy (best response) path relative to the selection of paths by the previous players. In Phase 2, players are allowed to change their paths in order to decrease their costs, namely, in the second step players play best response

dynamics. Charikar et al. [11] show that interestingly the sum of the players’ costs at the end of the first step will be within an $O(\log^2 n)$ factor of the cost of a socially optimal solution (which in this case is defined to be a minimum Steiner tree connecting the players to the root). This then implies that the cost of the Nash equilibrium achieved in the second step, as well as all states reached along the way, are $O(\log^3 n)$ close to **OPT**.

Note that in the directed case the result above does not hold, in fact such dynamics can lead to very poor equilibria. Figure 1 shows an example where if the players arrive one by one and each connects to the root by greedily choosing a path minimizing its cost, then the cost of the equilibrium obtained can be much worse than the cost of **OPT**. The optimal solution which is also a Nash equilibrium in this example is (P_1, \dots, P_n) where $P_i = s_i \rightarrow v \rightarrow t$ for each i ; however the solution obtained if the players arrive one at a time and each connects to the root by greedily choosing a path minimizing its cost is (P'_1, \dots, P'_n) where $P'_i = s_i \rightarrow t$ for each player i . Clearly, $\text{cost}(P'_1, \dots, P'_n) = n$ which is much worse than $\text{cost}(\mathbf{OPT}) = k$. Moreover, if one modifies the example by making many copies of the edge of cost k , then even if behavior begins in a *random* initial configuration, with high probability each edge of cost k will have few players on it and so best-response behavior will lead to the equilibrium of cost n .

A.2 A Lower Bound for Noisy Best-Response

In noisy best-response dynamics (also called log-linear learning [23]), when it is player i ’s turn to move, it probabilistically chooses an action with a probability that decreases exponentially with the gap between the cost of that action and the cost of the best-response action. The rate of decrease is controlled by a temperature term τ , much like in simulated annealing. In fact, the dynamics can be viewed as a form of simulated annealing with the global potential as the objective function. At zero temperature, the dynamics is equivalent to standard (non-noisy) best response, and at infinite temperature the dynamics is completely random. While it is known that with appropriate temperature control this process in the limit will stabilize at states of optimum global potential, we show here there exist cost sharing instances such that no dynamics of this form can achieve expected cost $o(n \cdot \mathbf{OPT} / \log n)$ within a polynomial number of time steps.

We begin with a definition capturing a broad range of dynamics of this form.

Definition 1 *Let generalized noisy best response dynamics be any dynamics where players move one at a time in a random order, and when it is a given player i 's turn to move, it probabilistically selects among its available actions. The sole requirement is that if action a is a worse response for player i than action b to the current state S , and furthermore a has been a worse response than b to all past states S' , then the probability of choosing a should be at most the probability of choosing b .*

The above definition captures almost any natural individual learning-based dynamics. We now show that no dynamics of this form can achieve expected cost $o(n \cdot \mathbf{OPT} / \log n)$ within a polynomial number of time steps for fair cost-sharing.

Theorem 9 *For the fair cost sharing game, no generalized noisy best response dynamics can achieve expected cost $o(n \cdot \mathbf{OPT} / \log n)$ within a polynomial number of time steps.*

Proof: We consider a version of the ‘‘cars or public transit’’ example of Figure 1, but where each player has n cars (options of cost 1 that cannot be shared by others). For this problem, we can describe the evolution of the system as a random walk on a line, where the current position t indicates the number of players currently using the public transit (the shared edge of cost k). The exact probabilities in this walk depend on specifics of the dynamics and may even change over time, but one immediate fact is that so long as the walk has never reached $t \geq k$, the shared edge of cost k is a worse response to any player than its edges of cost 1. Therefore, by definition of generalized noisy best response, each player has at most a $1/n$ chance of choosing the shared edge, and at least a $1 - 1/n$ chance of choosing a private edge, when it is that player’s turn to move. Since position t corresponds to a t/n fraction of players on the shared edge and a $1 - t/n$ fraction on private edges, this in turn implies that given that the random walk is in position $1 \leq t \leq k - 1$,

1. the probability $p_{t,t+1}$ of moving to position $t + 1$ is at most $\frac{1}{n}(1 - \frac{t}{n})$, and
2. the probability $p_{t,t-1}$ of moving to position $t - 1$ is at least $(1 - \frac{1}{n})\frac{t}{n}$,

(with remaining probability $1 - p_{t,t+1} - p_{t,t-1}$ the walk remains in position t). In particular,

$$p_{t,t+1}/p_{t,t-1} \leq \frac{n-t}{t(n-1)} \leq 1/t.$$

We now argue that the expected time for this walk to reach position k is superpolynomial in n for

$k = \log n$. In particular, consider a simplified version of the above Markov chain where $p_{t,t+1}/p_{t,t-1} = 1/t$ (rather than $\leq 1/t$) and we delete all self-loops except at the origin (so $p_{t,t+1} = 1/(t+1)$ and $p_{t,t-1} = t/(t+1)$ for $1 \leq t \leq k-1$). Deleting self-loops can only decrease the expected time to reach position k since it corresponds to simply ignoring time spent in self-loops, and the same for setting $p_{t,t+1}/p_{t,t-1} = 1/t$. So, it suffices to show the expected time for this simplified walk to reach position k is superpolynomial in n .

For convenience, set $p_{k,k-1} = 1$. We can now solve for the stationary distribution π of this chain. In particular, the simplified Markov chain is now equivalent to a random walk on an undirected multigraph with vertices v_0, v_1, \dots, v_k having one edge between v_k and v_{k-1} , $(k-1)$ edges between v_{k-1} and v_{k-2} , $(k-1)(k-2)$ edges between v_{k-2} and v_{k-3} , and in general $(k-1)(k-2) \dots t$ edges between v_t and v_{t-1} for $1 \leq t \leq k-1$. In addition, node v_0 has $(n-1) \cdot (k-1)!$ edges in self-loops since the probability $p_{0,0}$ is at least $(n-1)/n$. Therefore, since the stationary distribution of an undirected random walk is proportional to the degree of each node [25], we have that

$$\pi_k \leq 1/(n \cdot (k-1)!) < 1/k!.$$

Lastly, since the expected time h_{kk} between consecutive visits to node v_k satisfies $h_{kk} = 1/\pi_k$ by the Fundamental Theorem of Markov Chains, the expected time h_{0k} to reach v_k from v_0 is at least $1/\pi_k$ as well. So, the expected time to reach v_k is at least $k!$ which is superpolynomial in n for $k = \log(n)$ (or even $k = \omega(\log n / \log \log n)$).

Finally, the fact that the expected time to reach position k is superpolynomial in n implies that the probability of reaching position k within a polynomial number of time steps is less than $1/\text{poly}(n)$: specifically, if the walk has probability p of reaching position k in T time steps starting from position 0, then by the Markov property the expected time to reach position k is at most T/p . Moreover, so long as the walk has not yet reached position k , the cost of the system is $\Theta(n) = \Theta(n \cdot \mathbf{OPT}/k)$. Thus, the expected cost of the system within a polynomial number of time steps is $\Omega(n \cdot \mathbf{OPT} / \log n)$ as desired.

B Hoeffding-Azuma

For completeness we present here the Hoeffding-Azuma concentration bound for supermartingales.

Theorem 10 *Suppose X_0, X_1, X_2, \dots is a supermartingale, namely that*

$$\mathbf{E}[X_k | X_1, \dots, X_{k-1}] \leq X_{k-1},$$

for all k . Suppose also that for all k we have

$$|X_k - X_{k-1}| \leq C.$$

Then, for any $\lambda > 0$ and any $n \geq 1$,

$$\mathbf{Pr}[X_n \geq X_0 + \lambda] \leq e^{-\lambda^2/(2C^2n)}.$$