ISSN 1870-4069

Optimizing Best Response Dynamics-based Facility Location Games Using Reinforcement Learning

Andrés Burjand Torres Reyes, Rolando Menchaca-Méndez,
Francisco Hiram Calvo-Castro

Instituto Politécnico Nacional,
Centro de Investigación en Computación,
Mexico

atorresr2024@cic.ipn.mx, rmen@cic.ipn.mx, hcalvo@cic.ipn.mx

Abstract. In this article, we propose a model based on Best Response Dynamics (BRD) to examine the behavior of a group of rational agents when an external regulatory entity enforces control policies that influence the agents' dynamics. BRD is valuable for analyzing economic and social phenomena, as it captures the tendency of agents to seek to maximize their individual benefits—a common behavior in these contexts. However, these models frequently converge to a Nash equilibrium, which may not represent a socially optimal outcome. To address this limitation, we suggest introducing an external regulatory agent that employs reinforcement learning to enhance the convergence time to Nash equilibria or, ideally, to guide the system toward socially optimal solutions. We utilize an environment modeled after a Facility Location Game (FLG) to train a reinforcement learning agent and assess the impact of its policies on the FLG's behavior. This methodology presents a novel application of game theory and reinforcement learning for regulating complex systems, with potential implications in economics, social systems, robotics, and engineering. We present preliminary results to support our findings.

Keywords: Reinforcement learning, multi-agent systems, game theory, best response dynamics, facility location games.

1 Introduction

In today's interconnected world, addressing economic, social, and environmental challenges requires a deep understanding of complex systems. These systems often consist of agents (such as individuals, organizations, or entities) that interact and make decisions to optimize their individual benefits. Although such self-interested behaviors frequently result in predictable outcomes, such as Nash equilibria, they may also prevent the system from reaching a social optimum. Even when collective benefit is maximized, an agent can unilaterally adjust its strategy to increase personal gain, disrupting the system's balance.

This paper presents work in progress; it focuses on the initial steps toward addressing the challenge of dynamically regulating self-interested agents in

complex systems to reconcile individual incentives with collective welfare. Specifically, we show preliminary results on improving convergence time to a Nash Equilibrium in a Facility Location Game with players under Best Response Dynamics behavior using a Reinforcement Learning framework.

Game theory provides powerful tools for analyzing agent interactions, with Best Response Dynamics (BRD) being a prominent method for modeling rational decision-making. BRD is a process in which agents use a local search method to achieve outcomes that benefit them individually while driving the system, in general, toward Nash equilibrium [6]. It is particularly effective in modeling economic and social scenarios in which agents aim to maximize individual benefits. This model studies the interaction of rational agents who make decisions to maximize an objective function based on the system's current state, a formulation rooted in the domain of game theory of potential games [17]. However, it suffers from exponential convergence times and an inability to adapt to external interventions or collaborative behaviors.

To overcome these challenges, this work introduces an external regulatory agent equipped with reinforcement learning capabilities. This agent operates independently of the economic agents and intervenes in the system's dynamics by applying incentives or taxes. Unlike traditional models, this regulatory agent dynamically adapts its strategies to optimize regulatory policies, enabling more effective interventions in diverse scenarios. Its primary objectives are as follows:

- Accelerating convergence to Nash Equilibria: Reinforcement learning techniques reduce convergence times from exponential to polynomial, making the process more efficient.
- Guiding the system toward a social optimum: Ensuring outcomes that maximize collective welfare.

The effectiveness of this approach is evaluated using a Facility Location Game, a canonical optimization problem with applications in economics and operations research.

This research contributes an innovative framework for regulating complex systems, bridging game theory and machine learning to address real-world challenges in economic and social contexts.

2 Justification

Efficient regulation of decentralized systems is critical in domains such as energy markets, traffic routing, and public resource allocation, where static, one-size-fits-all policies struggle to address real-time dynamics' inherent volatility and complexity. In energy markets, for instance, the rise of distributed renewable energy sources (e.g., solar panels, wind farms) and fluctuating demand patterns necessitate adaptive pricing and grid-balancing mechanisms to prevent blackouts or the curtailment of renewable generation. Static tariff structures or fixed supply-demand models often fail to account for sudden

weather changes, shifts in consumer behavior, or equipment failures, leading to inefficiencies and instability [9].

Similarly, in traffic routing, rigid signal timings or preprogrammed navigation systems cannot respond to real-time congestion caused by accidents, road closures, or surges in ride-sharing demand. Adaptive traffic management systems, powered by IoT sensors and machine learning, dynamically reroute vehicles and adjust signal cycles to minimize delays and emissions [5].

Public resource allocation—such as distributing emergency aid during disasters or optimizing vaccine delivery during pandemics—also demands real-time adjustments to evolving needs, supply chain disruptions, or demographic inequities. Static policies risk misallocating resources and leaving vulnerable populations underserved, which is a significant concern in fields like public health [2].

These challenges underscore the need for decentralized regulatory frameworks that integrate real-time data, predictive analytics, and feedback loops to balance efficiency, equity, and resilience in dynamic environments.

3 Theoretical Framework

3.1 Game Theory Foundations

As stated by [14], "Game Theory aims to model situations in which multiple participants interact or affect each other's outcomes". These situations are often considered as strategic games, and, according to [7], involve:

- A set of players (the participants) $N = 1, 2, ..., n$,
- Strategy profiles $A = A_1 \times A_2 \times ... \times A_n$, which are the combination of actions chosen by all players in the game, where A_i is the set of actions available to player i
- Utility (or payoff) functions $u_i : A \to \Re$

A *Nash Equilibrium* (NE) is a strategy profile x^* satisfying:

$$x^* \in \mathrm{BR}(x^*)\,,$$

where BR denotes the best-response mapping. In other words, no player can increase their payoff by unilaterally deviating from x^*.

A social optimum is considered a situation that maximizes the total welfare of all players in a game. Mathematically, it is defined as a situation that maximizes a social welfare function, aggregating all players' utilities. The social welfare function $W : A \to \Re$ aggregates individual utilities. One of the most common aggregation methods is the utilitarian social welfare, which sums the utilities of all players:

$$W(a) = \sum_{i \in N} u_i(a), \tag{1}$$

where $a = (a_1, a_2, a_3, ..., a_n)$ is an action profile. An action profile $a^* \in A$ is socially optimal if:

$$a^* \in \arg\max_{a \in A} W(a) = \arg\max_{a \in A} \sum_{i \in N} u_i(a), \tag{2}$$

A very important consideration is the fact that a social optimum is not necessarily a Nash Equilibrium, since certain individual incentives could lead players to deviate from such situations.

According to [13], *potential game* are the ones where a function $\Phi : S \to \mathbb{R}$ exists such that for every player i, any strategy profile $s = (s_i, s_{-i})$, and any alternative strategy s'_i, the change in player i's utility satisfies:

$$u_i(s'_i, s_{-i}) - u_i(s_i, s_{-i}) = \Phi(s'_i, s_{-i}) - \Phi(s_i, s_{-i}), \tag{3}$$

where u_i denotes the utility function of player i.

3.2 Best Response Dynamics (BRD)

Best Response Dynamics (BRD) models rational decision-making in strategic games by assuming agents iteratively update their strategies to maximize individual utilities based on others' actions. This process can be represented as a directed graph where nodes correspond to action profiles, and edges denote transitions via unilateral best-response deviations. At each step, a player switches to a strategy that maximizes their payoff given the current actions of others, driving the system toward equilibrium states. [12]

While BRD converges to Nash equilibria in potential games [17], traditional models face critical limitations:

- They cannot incorporate external interventions,
- They fail to account for learning processes or collaborative behaviors among agents,
- Many existing regulatory frameworks are static and unable to adapt to dynamic systems.

3.3 Reinforcement Learning

Reinforcement Learning (RL) is a machine learning technique where an agent interacts with an environment and tries to obtain the policy that yields the maximum possible reward from said environment, using trial and error as its basis. Using the concept of delayed rewards, RL encapsulates that actions can affect both present and future rewards, improving the decision-making process. RL is also very useful in uncertain environments, because it is designed to keep the focus on proposed objectives, using the Markov Decision Processes to formalize the interaction between an agent, its actions, and its goals. One of the core challenges of RL is strategically balancing exploration (when the agent tries new actions in the hope of better results) and exploitation (when

the agent uses actions that proved helpful in the past), a dilemma absent from supervised and unsupervised learning algorithms. This holistic approach makes RL particularly suitable for real-time decision-making tasks where uncertainty and long-term planning are crucial. [16]

3.4 Facility Location Games

The Facility Location Game is an optimization problem where the goal is to determine which facilities to open and how to assign customers cost-effectively. Given a set of facilities F and a set of customers U, each facility $i \in F$ has a fixed, non-negative opening cost f_i. Additionally, serving a customer $j \in U$ from a facility $i \in F$ incurs a non-negative service cost c_{ij}, which depends on the specific facility–customer pair. The objective is to minimize the total cost, which consists of the sum of the opening costs of the selected facilities and the service costs of assigning each customer to an open facility. This requires making two key decisions: selecting the facilities to open and determining the optimal assignment of customers to these facilities while ensuring every customer is served [8, 11].

Facility Location Games, as potential games, share the property of guaranteed convergence to a Nash equilibrium, although the upper bound is considered exponential (i.e., $O(n^m)$). Another important characteristic of potential games is that they have a central function, called the potential function, which is optimized by the actions of all the players [18].

4 Related Work

A state-of-the-art search reveals several articles that analyze and apply Best Response Dynamics as a modeling technique. For example, [3] studied public goods games played on networks with possibly non-reciprocal relationships between players, where they explained how and why a Nash equilibrium is not always achieved in games on directed networks (which implies unequal relationships); this paper poses an interesting background since taking into consideration the nature of the relation between the agents in the model can improve its plausibility, and it could be interesting to study how the regulating agent could help achieve a Nash equilibrium.

Some researchers have utilized evolutionary game theory to understand spatial collective decision-making behaviors, such as [22]. In this case, they developed incentive mechanisms (reward and punishment) to investigate asynchronous BRD of anti-coordinating agents. This approach could be fruitful compared to the one proposed for this thesis. It is also interesting to point out the distinction between coordinating and anti-coordinating agent actions—former when, if one strategy prevails, agents in the system will be favored to follow it; latter when individuals take the opposite action if most game partners make the same choice [19].

A recent study by [4] explores discrete opinion dynamics in social networks with stubborn agents, where conformists adopt the most common opinions

from their neighbors, but stubborn agents remain unaffected by others. This research transforms the opinion dynamics into an n-strategy evolutionary game model with best-response updating, shedding light on how information influences strategy evolution. When agents have complete information about all their neighbors' opinions, the game becomes a potential game, guaranteeing the existence of at least one pure-strategy Nash equilibrium (PNE) and ensuring convergence to a PNE through asynchronous BRD. However, multiple PNEs often arise, complicating predictions of the evolutionary outcome. An interesting extension of this work is provided when information is limited, as agents can observe their neighbors with a probability of less than one. In this case, the game results in a unique stationary strategy distribution if stubborn agents are present, and the corresponding PNE becomes globally stable under both synchronous and asynchronous dynamics. This finding introduces the idea that a combination of stubborn agents and limited information can function as an equilibrium-selection mechanism, making the evolutionary outcome more predictable. The authors use numerical simulations on various network types, demonstrating how the distribution of stubborn agents and the available information level can significantly influence the final evolutionary outcome, showing how agents with different opinions converge to the PNE.

Reinforcement Learning is a well-respected machine learning paradigm in the literature. It is common to find uses in control theory, for instance, [10], where they used RL for heading control for unmanned sailboats using a backstepping sliding mode approach; here they propose an RL-based controller that enhances tracking performance and robustness against disturbances by integrating adaptive compensation mechanisms. The simulations show it outperforms existing methods, demonstrating RL's effectiveness in improving control strategies for uncertain and dynamic systems. There are a variety of uses as well in financial applications, such as [21], where they present a novel approach to equity portfolio optimization by integrating spectral analysis, portfolio theory, and deep reinforcement learning. In [15], the researchers show a predictive-based reinforcement learning (PRL) model to improve credit assessment for manufacturers and importers; by integrating predictive analytics and reinforcement learning, PRL enhances credit-scoring accuracy, decision-making, and financial stability.

It is also possible to find multiple articles that use both game theory and reinforcement learning to model complex problems. For instance, [1] explores the intersection of these two fields to model cyber-physical human systems by "proposing a computationally feasible approach to simultaneously model multiple humans as decision-makers, instead of determining the decision dynamics of the intelligent agent of interest and forcing the others to obey certain kinematic and dynamic constraints imposed by the environment." This multi-agent method could have certain advantages over the use of a regulatory agent, but it is also more complex; there is also an opportunity to find the intersection between both ideas, because many social and political scenarios include groups of

decision-makers capable of adaptation while having a regulatory agent with a different nature.

A promising direction for advancing this field is integrating evolutionary game theory with reinforcement learning, leveraging non-cooperative game theory to address the dynamic and complex nature of the agents' interactions. The research developed by [20] proposes a hybrid framework where a non-cooperative competition dynamically selects policy update modes using Nash equilibrium, ensuring diversity in agent strategies, while a cooperative collaboration balances exploration and convergence. The system can overcome local optima and adapt more effectively to dynamic conditions by allowing evolutionary algorithms to drive environment-independent exploration. This approach highlights the potential of combining game-theoretic principles with reinforcement learning. It suggests pathways for enhancing models of regulatory and adaptive agents in socio-political and multi-agent systems, aligning well with the challenges and opportunities identified for this work.

5 Proposed Solution and Methodology

5.1 Description of the Simulated Environment

This simulated environment will be a graph-based representation of a facility location game. The base will consist of a weighted, undirected, connected tree graph $G = (V, D, E, W)$, where:

- V (nodes):
  - Each with an associated client demand.
  - Potential facility locations $F \subseteq V$.
- D (node weights): node weights represent the demand at each node, such that $\forall d \in D,\ d \in \mathbb{N}$.
- E (edges): connections between nodes (e.g., roads, transit links).
- W (edge weights): edge weights represent serving costs (e.g., distance, congestion, transportation fees), such that $\forall w \in W,\ w \in \mathbb{N}$.

The decision to use a tree was made to ensure convergence via BRD. There are three main reasons: the potential is bounded ($\phi \geq 0$); each best response reduces ϕ or leaves it unchanged if equilibrium is reached; and the absence of cycles prevents infinite loops.

The group of players acting under Best Response Dynamics can be defined as a set $N = \{player_1, player_2, \ldots, player_n\}$, where each player i chooses a location $f_i \in F$ to build its unique uncapacitated facility with no associated building cost, and this location becomes exclusive to i while i decides to keep it. We also define a distance function $d_G(x, y)$ giving the shortest-path distance between any two vertices $x, y \in V$ and a profit function $U_i(f_i, f_{-i})$ depending on the locations of all players, defined in equation (4):

$$U_i(f_i, f_{-i}) = \sum_{\substack{c \in V \\ f_i = \text{nearest}(c)}} D(c) \cdot \big(1 - d_G(c, f_i)\big). \tag{4}$$

Since this FLG is considered under the potential games framework, we use a global potential function ϕ that reflects system-wide efficiency:

$$\phi(f) = \sum_{c \in V} D(c) \cdot d_G\big(c, \text{nearest}(f)\big). \tag{5}$$

Here, ϕ represents the total weighted distance from all clients to their nearest facility. Thus, players' strategies directly impact ϕ.

The best-response update for player i at time step $t+1$ is described in equation (6):

$$f_i^{(t+1)} = \arg \max_{f \in F \setminus f_{-i}^{(t)}} U_i\big(f, f_{-i}^{(t)}\big). \tag{6}$$

where $f_{-i}^{(t)}$ are the locations of the other players at time t. When the distance calculation between customers and facilities results in a tie, it is broken by random assignment.

Given these definitions, the BRD process proceeds as follows:

1. Start with an initial random configuration $(x_1^{(0)}, x_2^{(0)}, \ldots, x_n^{(0)})$ with no overlap.
2. At each time step t, select a player at random to update their position to their best response given the current positions of others.
3. Halt when no player can improve their utility (PNE).

Under this design, a Nash equilibrium arises when no player can unilaterally move to capture more clients (i.e., $\nexists f_i' : U_i(f_i', f_{-i}) > U_i(f_i, f_{-i})$).

After success with this simple simulation, certain changes could improve the model's real-life applicability. For example, we could simulate a more dynamic graph where road conditions change over time, include traffic congestion, model evolving client demand with stochastic variations, etc.

5.2 Preliminary description of the proposed solution

As we discussed earlier, there are two main problems with agents that act selfishly in any given game: the time complexity to achieve Nash Equilibrium is usually exponential, and many states are not socially optimal since the algorithms they follow usually lead to local optimums; this is the case as well for BRD [4]. We decided to address both of these issues by implementing an agent with a different nature, one that can alter certain game conditions by modifying the rules, applying incentives, and learning how to obtain the optimal values for the variables it can control using reinforcement learning. Fig 1 illustrates this idea. Fig 1 represents the process through which the agent obtains information of the environment given by the Facility Location Game with BRD to decide how to act through regulations and incentives (with policy π), and also takes rewards that help it improve its policy. Reinforcement learning depends on the agent-environment framework since it pays attention to the current state of the

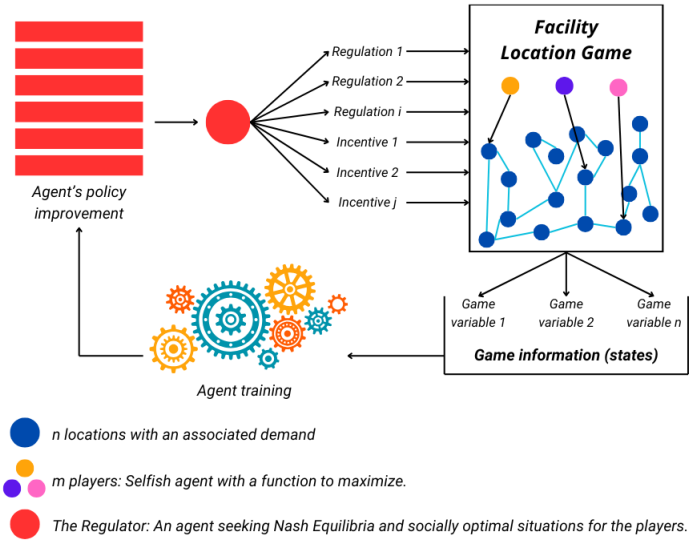

Fig. 1. Proposed architecture.

situation and considers its space of action. The first step is creating a simulation of the Facility Location Game, described in the previous section. Once the simulation is operating, we can extract the set of states corresponding to the environment. These states will be encoded to capture spatial relationships with a set of variables such as the positions of facilities and clients, the distribution of demand, the number of players, the players' decisions, costs, etc, as well as the potential function, which is the function to maximize.

This paper is centered on the objective of reducing convergence time. The process of balancing both objectives, given that we proposed a Multi-Objective Reinforcement Learning framework, is still being developed.

The regulatory agent (RA) will take actions a_t from a predefined action space A ($a_t \in A = \{\text{tax on locations}, \text{incentives for players}, \dots\}$) depending on the state S_t; from this space, it will select a policy to pursue its two main objectives:

1. Accelerate convergence to NE, which we formally consider as the reduction to polynomial time ($O(n)$).
2. Obtaining socially optimal situations, even if it means escaping Nash Equilibria

This policy will be iteratively improved with the reward function assigned to the environment, which will give the necessary incentives to the RA to achieve an optimal policy, which is the policy that secures the best positions for the players, and, therefore, the best rewards for the RA. A policy π can be defined as:

$$\pi : S_t \rightarrow A.$$

Each state S_t at time t consists of:

- Graph structure G_t: The adjacency matrix of the current network.
- Facility locations (Both potential and taken) F_t: A binary vector indicating which nodes have facilities.
- Demand distribution D_t: A vector assigning demand values to client nodes.
- Cost matrix C_t: Transportation costs from each client to its assigned facility.

The final reward signal will most likely be based on arguments like: cost minimization (e.g., transportation or setup costs), Stagnation (when there are no changes in players' positions or utilities), and/or convergence incentives (e.g., penalizing large deviations from equilibrium solutions). A proposed reward function $R(S_t)$ is:

$$R(S_t) = \Delta\phi_t - I. \tag{7}$$

Where I is the number of iterations, this way, the reward is proportional to how much the general utility is improved and is reduced by the iterations it takes to achieve NE.

The final implementation of the RA will make use of different variables to achieve its goal, which will be its *action space*, like:

- Conveniently changing the player selection order (instead of it being random) to prioritize players with higher potential to improve system efficiency.
- Utility function weighting: Dynamically adjusting α and β weights corresponding to the *sum_ demands* and *sum_ costs* variables in the utility function.
- Altering the tie-breaking rules
- Facility activation/deactivation

It will use a model-free approach called Proximal Policy Optimization, which is suitable for discrete action spaces.

In the Preliminary Results section, we implement a lightweight regulatory agent based on a tabular reinforcement-learning scheme with an ε-greedy Monte Carlo policy as a proof of concept. At each time step, the agent "intervenes" by selecting exactly one of the currently non-converged players—dynamically prioritizing those whose move is estimated to yield the greatest improvement in global efficiency. Concretely, the action space at state S_t has size equal to the number of active players n. To keep the state representation compact, we discretize each player's individual utility $u_i^{(t)}$ and the overall potential function $\phi^{(t)}$ into a small number of bins (from "very low" to "very high"), which were estimated using statistical data from the simulations run using only FLG + BRD; the resulting tuple

$$s_t = \big(\mathrm{bin}(u_1^{(t)}), \ldots, \mathrm{bin}(u_n^{(t)}), \mathrm{bin}(\phi^{(t)})\big). \tag{8}$$

serves as the index into a Q-table $Q(s, a)$, which is initialized with a small positive constant (10^{-5}) to encourage early exploration. At each decision point, with

probability ε the agent picks a random valid player (exploration), and otherwise it exploits by choosing

$$a_t \;=\; \arg\max_{a \in \mathcal{A}(s_t)} Q(s_t, a)\,. \tag{9}$$

Immediately after the chosen player best-responds and the game state advances, we compute a scalar reward

$$R(S_t) \;=\; \phi^{(t)} \;-\; \phi^{(t-1)} \;-\; I \;\times pw, \tag{10}$$

where I is the current iteration count (to penalize long trajectories) and pw is a tunable penalty weight. During learning, we perform the incremental Bellman update

$$Q(s_t, a_t) \;\leftarrow\; Q(s_t, a_t) \;+\; \alpha\Big[R(S_t) \;+\; \gamma \max_{a'} Q(s_{t+1}, a') \;-\; Q(s_t, a_t)\Big], \tag{11}$$

with learning rate $\alpha = 0.1$ and discount factor $\gamma = 0.99$. Once the Facility Location Game converges, a Monte Carlo end-of-episode pass backpropagates the final reward through the entire episode history: computing the return G backward and adjusting each visited Q-entry by

$$Q(s_t, a_t) \;\leftarrow\; Q(s_t, a_t) \;+\; \alpha\,\big(G - Q(s_t, a_t)\big). \tag{12}$$

To balance exploration and exploitation over successive episodes, the exploration rate is decayed multiplicatively:

$$\varepsilon \;\leftarrow\; 0.995\,\varepsilon. \tag{13}$$

Preliminary experiments show that this simple regulatory intervention significantly accelerates convergence of best-response dynamics, reducing both the number of iterations and the variance of the potential-function trajectory.

6 Preliminary Results

In fig 2 we show how the previously defined simulation results look using a specific seed (66). The simulation was implemented in Python, and the source code (including the Facility Location Game environment, Best Response Dynamics logic, and visualization tools) is publicly available on GitHub under an open-source license. The repository can be accessed at: https://github.com/Burjand/facility_location_game.git. These were the used hyperparameters:

- Number of nodes: 100
- Number of potential facilities: 80
- Number of BRD players: 10
- Seed: 66

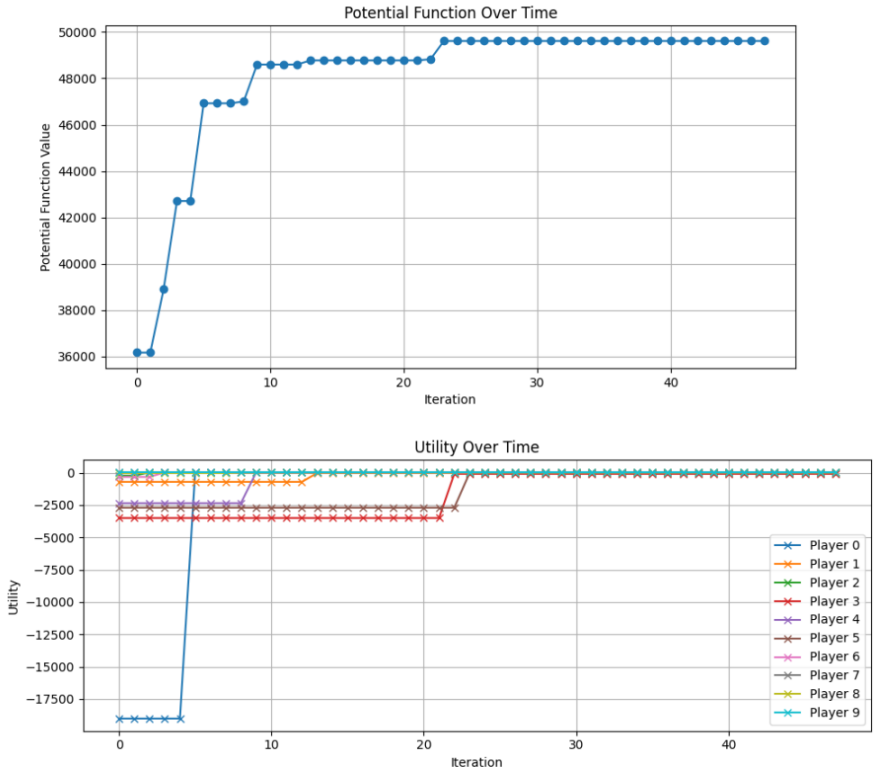

Fig. 2. FLG with BRD simulation results.

Running 1000 simulations with different seeds chosen randomly, 100 nodes, 80 potential facilities, and 10 players, the average number of iterations to achieve Nash Equilibrium was 58.515, and the average potential function value at the end was 60597.483. After implementing the regulatory agent with the conditions stated before, we ran again 1000 simulations with different seeds chosen randomly and the same parameters, and obtained that the average number of iterations to achieve Nash Equilibrium was 43.817, and the average potential function value at the end was 50949.478.

7 Scientific Novelty

The majority of studies involving multi-agent modeling have two main approaches regarding their adaptability to achieve objectives: The first is that agents act selfishly and non-cooperatively, maximizing their own utility function; the second is that each agent can learn, allowing them to adapt to situations in a more "intelligent" manner. The novelty of this work lies in a variation of the first modeling approach, introducing an agent capable of

modifying the system's rules so that agents can reach Nash Equilibria more quickly despite lacking the adaptability provided by learning and cooperating in the second approach.

8 Limitations and Future Work

Despite these promising benefits, the modeling technique has notable limitations. One of the most significant is its inability to incorporate the agents' capacity for cooperative behavior, long-term adaptation, and learning from past actions. This limitation arises because the framework assumes purely rational and selfish players. A potential area for future improvement —either in this work or in subsequent studies— would be enhancing the model to simulate agents' learning abilities and collaborative behavior, which can turn this into a Multi-agent Reinforcement Learning problem. Some other future works that could be developed based on this one could be developing computational methods to scale the model for larger, more complex systems and testing the framework in real-world scenarios to validate its assumptions and refine its applicability.

Acknowledgments. This research was funded in part by the "Secretaría de Ciencia, Humanidades, Tecnología e Innovación" (Secihti) of Mexico and the "Instituto Politécnico Nacional" under grant SIP 20253428.

References

1. Albaba, B. M., Yildiz, Y.: Modeling cyber-physical human systems via an interplay between reinforcement learning and game theory. Annual Reviews in Control, vol. 48, pp. 1–21 (1 2019) doi: 10.1016/j.arcontrol.2019.10.002
2. Barret, H., Ortmann, W., Dawson, L., Saenz, C., Reis, A.: Resource allocation and priority setting, vol. 3. Springer (2016), https://www.ncbi.nlm.nih.gov/books/NBK435776/
3. Bayer, P., Kozics, G., Szőke, N. G.: Best-response dynamics in directed network games. Journal of Economic Theory, vol. 213, pp. 105720 (8 2023) doi: 10.1016/j.jet.2023.105720
4. Cao, W., Zhang, H., Kou, G., Zhang, B.: Discrete opinion dynamics in social networks with stubborn agents and limited information. Information Fusion, vol. 109, pp. 102410 (4 2024) doi: 10.1016/j.inffus.2024.102410
5. Eremina, L., Mamoiko, A., Aohua, G.: Application of distributed and decentralized technologies in the management of intelligent transport systems. Intelligence Robotics, vol. 3, no. 2, pp. 149–61 (1 2023) doi: 10.20517/ir.2023.09
6. Feldman, M., Snappir, Y., Tamir, T.: The Efficiency of Best-Response Dynamics. Springer (1 2017), https://doi.org/10.1007/978-3-319-66700-3_15
7. Hara, K.: Coalitional strategic games. Journal of Economic Theory, vol. 204, pp. 105512 (7 2022) doi: 10.1016/j.jet.2022.105512
8. Iloglu, S., Albert, L. A., Michini, C.: Facility location and restoration games. Computers Operations Research, vol. 174, pp. 106896 (2025) doi: https://doi.org/10.1016/j.cor.2024.106896

9. Lammers, I., Diestelmeier, L.: Experimenting with Law and Governance for Decentralized Electricity Systems: Adjusting Regulation to Reality? Sustainability, vol. 9, no. 2, pp. 212 (2 2017) doi: 10.3390/su9020212
10. Li, C.-M., Zhang, B.-L., Cao, Y.-L., Yin, B.: Reinforcement learning-based backstepping sliding mode heading control for unmanned sailboats. Ocean Engineering, vol. 327, pp. 120936 (2025) doi: https://doi.org/10.1016/j.oceaneng.2025.120936
11. Li, X., Lu, X.: An approximate cost recovery scheme for the k-product facility location game with penalties. Theoretical Computer Science, vol. 1021, pp. 114933 (2024) doi: https://doi.org/10.1016/j.tcs.2024.114933
12. Mimun, H. A., Quattropani, M., Scarsini, M.: Best-response dynamics in two-person random games with correlated payoffs. Games and Economic Behavior, vol. 145, pp. 239–262 (3 2024) doi: 10.1016/j.geb.2024.03.011
13. Monderer, D., Shapley, L. S.: Potential games. Games and Economic Behavior, vol. 14, no. 1, pp. 124–143 (5 1996) doi: 10.1006/game.1996.0044
14. Nisan, N., Roughgarden, T., Tardos, E., Vazirani, V. V.: Algorithmic Game Theory: Computing in Games. Cambridge University Press (1 2007), http://ebooks.cambridge.org/chapter.jsf?bid= CBO9780511800481cid= CBO9780511800481A011
15. Razaque, A., Beishenaly, A., Kalpeyeva, Z., Uskenbayeva, R., Nikolaevna, M. A.: A reinforcement learning and predictive analytics approach for enhancing credit assessment in manufacturing. Decision Analytics Journal, vol. 15, pp. 100560 (2025) doi: https://doi.org/10.1016/j.dajour.2025.100560
16. Sutton, R. S., Barto, A. G.: Reinforcement learning: An introduction. The MIT Press (2020)
17. Swenson, B., Murray, R., Kar, S.: On Best-Response Dynamics in Potential Games. SIAM Journal on Control and Optimization, vol. 56, no. 4, pp. 2734–2767 (1 2018) doi: 10.1137/17m1139461
18. Swenson, B., Murray, R., Kar, S., Poor, H. V.: Best-response dynamics in continuous potential games: Non-convergence to saddle points. 2018 52nd Asilomar Conference on Signals, Systems, and Computers, pp. 310–315 (Oct 2018) doi: 10.1109/acssc.2018.8645541
19. Yang, K., Huang, C., Dai, Q., Yang, J.: The effects of attribute persistence on cooperation in evolutionary games. Chaos Solitons and Fractals, vol. 115, pp. 23–28 (8 2018) doi: 10.1016/j.chaos.2018.08.018
20. Yu, J., Zhang, Y., Sun, C.: Balance of exploration and exploitation: Non-cooperative game-driven evolutionary reinforcement learning. Swarm and Evolutionary Computation, vol. 91, pp. 101759 (11 2024) doi: 10.1016/j.swevo.2024.101759
21. Yu, P., Liu, S., Jin, C., Gu, R., Gong, X.: Optimization-based spectral end-to-end deep reinforcement learning for equity portfolio management. Pacific-Basin Finance Journal, vol. 91, pp. 102746 (2025) doi: https://doi.org/10.1016/j.pacfin.2025.102746
22. Zhu, Y., Xia, C.: Asynchronous best-response dynamics of networked anti-coordination game with payoff incentives. Chaos Solitons and Fractals, vol. 172, pp. 113503 (5 2023) doi: 10.1016/j.chaos.2023.113503