Modeling Social Distancing with Reinforcement Learning

Nejc Ločičnik, Igor Nikolaj Sok, Leon Todorov, Andraž Zrimšek

ABSTRACT: ***This study investigates the emergence of social distancing behaviors in artificial agents using a reinforcement learning (RL) framework. Agents interact in a two-dimensional environment and learn to avoid infected individuals to minimize disease transmission. Drawing inspiration from the adaptive behaviors of ants, agents exchange health information and adjust their behavior accordingly. Initial results demonstrate that agents trained with a basic reward policy show increased separation between healthy and infected individuals, as observed through network metrics such as modularity and clustering. This work highlights the potential of RL in modeling disease dynamics and social distancing strategies.***

I. Introduction

The spread of infectious diseases is a significant challenge in both human and animal populations, prompting natural and artificial systems alike to develop mechanisms for minimizing transmission. Social distancing has emerged as a common adaptive behavior in nature, where organisms avoid close contact with infected individuals to protect themselves and their groups. This phenomenon has been observed across a variety of species and environments, suggesting it provides an evolutionary advantage in mitigating disease transmission risks. In response to the COVID-19 pandemic, social distancing also became a key public health strategy for humans, sparking interest in understanding how such behaviors might emerge and evolve autonomously in artificial agents.

Modeling disease transmission and social distancing behaviors in simulated environments can provide insights into the underlying dynamics of these processes, as well as offer potential applications in fields such as epidemiology, robotics, and swarm intelligence. Traditional approaches often rely on predefined rules to drive agent behaviors, limiting the complexity and adaptability of emergent patterns. By contrast, reinforcement learning (RL) provides a flexible framework where agents learn to navigate environments based on reward structures, allowing for more organic, adaptive behaviors that evolve in response to environmental pressures.

In this study, we aim to model social distancing behaviors using a reinforcement learning approach inspired by natural systems. Agents will learn to minimize disease transmission within a two-dimensional environment by adapting their interactions based on health information they exchange with one another. We will build on existing multi-agent reinforcement learning frameworks, specifically those designed for predator-prey dynamics, to simulate agent behavior under disease-spread conditions. This setup will allow us to explore how reward structures and network adaptations can lead to emergent social distancing behaviors, where agents autonomously avoid infected individuals.

Through our model, we hope to deepen our understanding of how social distancing behaviors emerge and to contribute to the broader field of adaptive multi-agent systems. Ultimately, this research may inform both theoretical models of disease transmission and practical applications in areas requiring coordinated group behavior, such as swarm robotics or public health simulations.

II. Related Work

In the article Predator–prey survival pressure is sufficient to evolve swarming behaviors [1], the authors use a reinforcement learning (RL) approach to model predator and prey behaviors within a cooperative–competitive multi-agent RL framework. Here, predator agents receive rewards for successfully catching prey, while prey agents receive rewards for avoiding capture and staying alive. This approach contrasts significantly with traditional behavior modeling, where predefined rules are often used to drive agents toward expected behaviors. Such rule-based models, however, can fail to capture the complexity and adaptability of real-world dynamics. In contrast, reinforcement learning only presents rewards that encourage or discourage certain actions, allowing for more organic and adaptive behavior development. Through this predator-prey framework, the authors observed emergent behaviors, such as flocking and swarming among prey agents and dispersion tactics among predators. These findings suggest that RL-based approaches can effectively foster diverse and adaptive group behaviors. In our work, we aim to build on this method to model disease spread, adjusting the agent parameters and reward mechanisms to simulate social distancing behaviors.

The complexities of disease spread and natural social distancing behaviors are further explored in Infectious diseases and social distancing in nature [3]. This study examines social distancing as an adaptive response to disease across various animal species, both human and non-human. Social distancing behaviors can emerge as precautionary actions taken by healthy individuals or as physiological responses in infected individuals. The authors analyze the underlying mechanisms driving these behaviors in both infected and non-infected subjects, highlighting how natural populations instinctively modify social interactions to mitigate disease transmission.

Building on this, Romano et al. in The trade-off between information and pathogen transmission in animal societies [2] argue that social distancing alone may be insufficient to control disease spread. They note that individuals in a population inherently rely on information exchange, which conveys significant adaptive benefits. This article discusses the balance animals must strike between maintaining necessary social connections and minimizing infection risk, proposing that animals develop “network plasticity” as they weigh the costs and benefits of each social interaction. These trade-offs in social behavior offer insights that are highly relevant for modeling disease spread, as they underscore the complex motivations behind individual actions within a population.

Together, these studies provide essential frameworks and insights into adaptive behavior modeling under environmental pressures. Our work will leverage these principles by employing a reinforcement learning model that integrates disease spread and social distancing, aiming to simulate the interplay of agent interactions and disease transmission dynamics.

III. Methodology

A. Problem Definition

Our objective is to model the spread of infectious diseases within a simulated population of agents that can move freely in a two-dimensional environment and interact with one another. The primary goal is to minimize disease spread by limiting interactions among agents. To achieve this, agents will exchange information about their health status, learning to adjust their behavior to avoid infected peers based on the information they receive.

B. Disease Spread Modeling

The study of Lasius niger ants [4] reveals an intriguing natural strategy for controlling disease spread. When exposed to the fungal pathogen Metarhizium brunneum, these ants dynamically alter their social network structure to reduce transmission risk. Rather than merely avoiding infected individuals, the entire colony adapts its social interactions to limit disease spread.

Both infected and uninfected ants exhibit adaptive behaviors: infected ants spend more time outside the nest, reducing exposure to healthy nest mates, while uninfected ants increase their spatial distance from others, particularly those exposed to the pathogen. These behavioral changes enhance the network's modularity, creating compartments within the social structure that contain the spread of infection.

We will incorporate similar adaptive behavioral adjustments into a reinforcement learning model to study disease transmission dynamics. Agents will be rewarded for exchanging information about their health status and penalized for close contact with infected individuals, thereby promoting social distancing behaviors.

Additionally, the paper suggests that low-level exposure to pathogens may have adaptive benefits. Future model improvements may explore nuanced reward and penalty schemes based on varying exposure levels, as well as the potential for agents to develop immunity through controlled exposure. This would allow for a deeper exploration of the trade-offs between information exchange and disease transmission.

C. Simulation Environment

This study employs a multi-agent reinforcement learning (RL) framework, adapted from the environment developed by Li et al., 2023 [1]. The simulation takes place in a two-dimensional continuous space with periodic boundary conditions, meaning that agents crossing one edge of the square environment reappear on the opposite edge, retaining their velocity.

Agents are modeled to resemble ants (changed from unicycles like in 1), with their body consisting of three connected circles (the back circle being slightly larger) and six legs. Their behavior is driven by a combination of active and passive forces. Active forces, controlled by the agents, include a forward movement force (aF) aligned with their heading direction and a rotational force (aR) enabling changes in heading. Passive forces, inherent to the environment, include drag force (Fd), simulating resistance opposing the agent's velocity, and repulsive force (Fa), which prevents agents from overlapping by pushing them apart. At each simulation step, the agents' positions and velocities are updated by summing all acting forces, with the dynamics governed by:

$$\dot{x} = v, \quad \dot{v} = \frac{ha_F + F_d + F_a}{m}, \quad \dot{\theta} = a_R$$

where x is the agent's position, v its velocity, θ its heading angle, h the unit vector for heading direction, and m the agent's mass. The ant-like design, illustrated in Figure 1, enhances realism while preserving the underlying principles of agent dynamics.

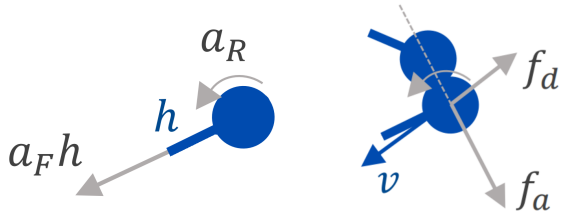

Figure 1. Active (left) and passive (right) agent forces. [1]

To tailor the framework to our objectives, several modifications were implemented. These include:

1) Agent visualization (unicycle to ant) and movement parameters (more "ant-like").
2) Agent perception includes health status of the perceived agents (based on FoV).
3) Keeping track of agent interactions, used for network evaluation.
4) Redefine the reward policy to align with our disease-spread mitigation goals.

D. Basic Reward Policy

Our initial reward policy aimed to produce social distancing patterns in agent behavior is based on direct collisions between agents as the primary form of information exchange. Collisions between agents of the same status (healthy-healthy or infected-infected) were rewarded to encourage grouping behavior. In contrast, collisions between agents of different statuses (healthy-infected or vice versa) were penalized to discourage close contact to limit disease spread.

The above mention basic reward policy is demonstrated in isolation with all healthy agents in figure 2. On the left figure we penalized each agent upon collision with a -1 reward, while on the right figure we rewarded each colliding agent with a $+1$ reward. This demonstrates how a very simple change in the reward policy effects the learned behavior.

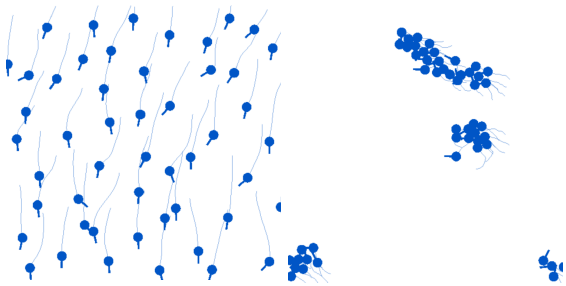

Figure 2. Demonstration of simple reward policies - punish (right) or reward (left) collisions.

Building on this foundation, we introduced a **diminishing reward system** to refine agent behavior further. This system rewarded agents for diverse interactions while penalizing risky

collisions. To prevent excessive rewards from repeated interactions between the same agents, we incorporated a diminishing factor based on recent interactions:

$$\text{reward}(a,b) = \begin{cases} -\lambda & \text{if } sick(a) \neq sick(b) \\ +\sigma * \gamma(1 - recent(a,b)) & \text{otherwise.} \end{cases}$$

Here, `recent(a,b)` is initialized to 1 upon interaction and decreases by a factor of 0.9 at each step. This mechanism ensures diminishing rewards for repeated collisions, encouraging diverse and meaningful interactions that better reflect real-world dynamics.

To further enhance agent behavior, we introduced optional reward components that address specific aspects of agent-environment dynamics. These additions allow for greater adaptability to different scenarios:

1) **Wall Collision Penalty** → In non-periodic environments, where agents encounter boundaries, a wall collision penalty discourages agents from colliding with walls:

$$\text{reward}(a) = \begin{cases} -\lambda & \text{if a collides with any wall} \\ 0 & \text{otherwise.} \end{cases}$$

2) **Control Penalty** → To mimic energy consumption during movement, a control penalty was introduced. This reward is proportional to the magnitude of the agent's control inputs (aF for forward force and aR for rotational force), encouraging agents to exhibit conservative movement:

$$\text{reward}(aF, aR) = -(\alpha|aF| + \beta|aR|)$$

These optional components provide additional flexibility to tailor agent behavior for specific objectives while preserving the simplicity and general applicability of the core reward system. Together, the diminishing reward mechanism and optional components have demonstrated their effectiveness in producing well-performing agents, with interaction networks resembling real-world examples. Ant behavior continues to serve as a baseline for comparison, validating the model's adaptability across species and scenarios.

NOTE: this stuff gets moved to discussion?: We plan to start with this simple reward policy before scaling up simulation complexity with more nuanced agent interactions. For example, we may introduce cooperative behavior by allowing agents to exchange resources or other benefits, which could model cooperative dynamics where agents work together to achieve common goals.

E. Model Performance Measures

To evaluate whether social distancing patterns emerge in agent behavior, we transform agent interactions into a network and analyze its structure, following approaches demonstrated in [4]. Throughout each evaluation step, interactions between agents are tracked and recorded in an $n \times n$ interaction matrix, where n is the total number of agents. Two agents are considered to be interacting upon collision. At the end of an evaluation run, this interaction matrix is normalized to construct the network. Nodes in the network represent agents, and edges are created between nodes if their interaction value exceeds a specified threshold (0.01). The actual values from the interaction matrix are used as edge weights, and each node is annotated with the health status of the corresponding agent.

This network is then used to calculate appropriate network measures such as clustering, modularity (between infected and non-infected), network density and more. This allows us to get a better understanding of how agents are interacting with one another. We expect all of the mentioned metrics to increase as same health status agents interact more between themselves.

Emergent social distancing behavior should also be clearly observable in the simulation visualization. We expect healthy agents to avoid infected ones, and vice versa. This behavior will be especially evident if we increase the agent density within the environment. In such cases, we should observe infected agents becoming isolated, forming empty regions around them, while healthy agents fill the remaining space in the simulation.

IV. Results

To test the performance of our initial reward policy, we ran an episode consisting of 5000 steps with both a random untrained network and a trained network. For each case, we constructed a network of all the interactions throughout the episode, which is displayed in Figure 3. In the network visualizations, healthy agents are colored blue, and infected agents are colored orange.

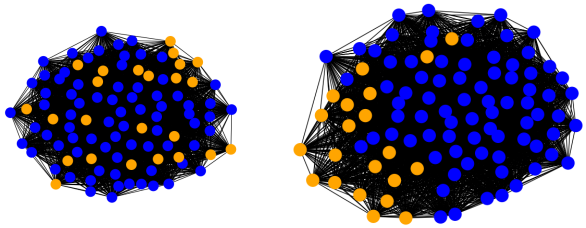

Figure 3. Random interaction network (left), learned interaction network (right).

In the random network, no clear structure is apparent. However, in the learned network, a separation begins to emerge, with infected agents interacting less with healthy ones, but still interacting with other infected agents. This behavior aligns with our expectations. The structure becomes more pronounced when we filter the network by keeping only edges with weights greater than 0.2, representing significant interactions. The filtered network is shown in Figure 4.

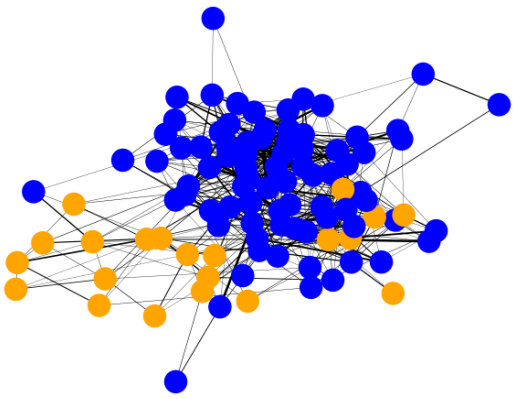

Figure 4. Filtered learned interaction network.

In the filtered network, it is clearly observed that infected agents interact much less with healthy agents. To gain a deeper understanding of the networks and to assess metrics that are harder to discern from visual inspection alone, we calculated

several network metrics, which are presented in Table IV. These metrics are relevant to pathogen transmission.

Net. Metric	Random	Trained
Clustering	0.041	0.064
Modularity	-0.016	0.050
Density	0.728	0.779
Efficiency	0.864	0.889

An increase in modularity suggests greater separation between the two groups, which leads to denser clusters and higher clustering values. These changes indicate that basic social distancing behavior has begun to emerge, with infected and healthy agents interacting less with each other. However, to gain a more comprehensive understanding of the agents' behavior and its impact on disease transmission, further testing is needed.

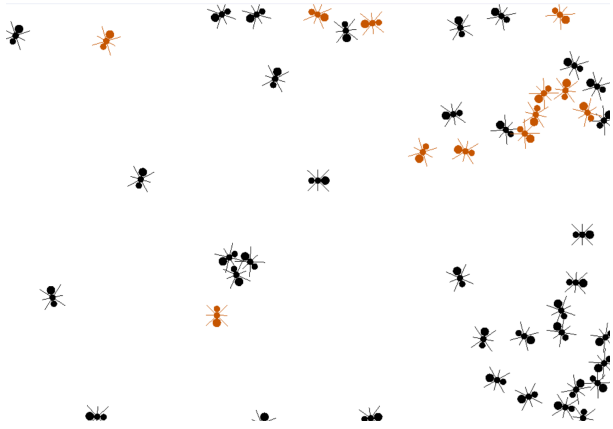

Figure 5. Visualization of our best performing policy on 75 agents (20% infected).

Figure 5 visualizes a simulation of our best-performing policy. The orange ants represent infected ants, while the black ants represent healthy ones. Although this figure appears to indicate the behavior described by the network statistics (with same-health-status agents tending to group), it should be noted that the separation between healthy and infected agents is not as pronounced as hoped. This is primarily due to the fact that interactions are based on collisions, and agents only perceive the 6-8 nearest agents in their field of view (FoV). This limitation causes agents of different statuses to often move much closer to each other than intended.

We also conducted some preliminary experiments with alternative methods of information exchange to produce clearer visualizations. Inspired by ant behavior, we implemented a simplified pheromone system, where each agent leaves a mark in the surrounding area that accumulates into a concentration heatmap that decays over time. These pheromones could be either positive (indicating "safety" areas from healthy agents) or negative (indicating "danger" from infected agents). If the concentration of pheromones at an agent's location reaches a certain threshold, the agent perceives it. This would theoretically provide agents with an indication of safety or danger ahead, helping them maintain greater distance from infected agents.

Unfortunately, this approach did not work as expected. We believe the issue lies in the fact that there isn't a clear correlation between the observations (pheromone concentrations) and the actions that would lead to better rewards. With the collision-based interaction system, agents simply adjust their movement to either avoid or approach other agents based on their health status, which leads to higher cumulative rewards. In contrast, with the pheromone system, agents detect a concentration but cannot determine the exact actions needed to improve their situation. Although theoretically, actions should be to stop (set force to 0) for positive pheromones and flee for negative ones, the lack of a direct feedback loop between pheromone detection and reward-driven actions made it challenging for the agents to effectively use the pheromone system.

V. Discussion and Future Work

As this is a preliminary report, we will use this discussion section to outline our future work.

Our primary goal is to replicate the experiment described in [4]. To achieve this, we plan to perform multiple runs for three scenarios: random networks, learned networks without infected agents, and learned networks with infected agents. The random network will serve as a baseline to normalize the metrics from our learned networks. We will then compare changes in network structure before and after the introduction of a pathogen. Throughout these experiments, we will track the same agents in pairs of pre- and post-introduction conditions. Additionally, we may continue refining our reward policies or explore alternative information exchange mechanisms beyond simple collisions.

As the final presentation requires a presentation video, we are thinking of implementing dynamic disease spread, where healthy agents can become infected, which would include infected agent mortality after a set number of time steps. This would allow us to directly visualize the effectiveness of social distancing by observing whether the disease dies off (due to infected agents not spreading the infection and eventually dying in isolation) or if all agents succumb to the pathogen.

CONTRIBUTIONS: **LT** prepared/fixed the environment setup and did the basic avoid/touch experiments, **INS** and **AZ** did the reward policy experiments and the network statistics, **NL** did the alternative interactions experiment and organized/polished the report. Each member wrote their own parts of the report.

References

[1] Jianan Li, Liang Li, and Shiyu Zhao. "Predator–prey survival pressure is sufficient to evolve swarming behaviors". In: *New Journal of Physics* 25.9 (2023), p. 092001.

[2] Valéria Romano, Cédric Sueur, and Andrew JJ MacIntosh. "The tradeoff between information and pathogen transmission in animal societies". In: *Oikos* 2022.10 (2022), e08290.

[3] Sebastian Stockmaier et al. "Infectious diseases and social distancing in nature". In: *Science* 371.6533 (2021), eabc8881.

[4] Nathalie Stroeymeyt et al. "Social network plasticity decreases disease transmission in a eusocial insect". In: *Science* 362.6417 (2018), pp. 941–945.