
INSTITUTO POLITÉCNICO NACIONAL

CENTRO DE INVESTIGACIÓN EN COMPUTACIÓN


# T E S I S

## Modelado de movimiento colectivo en enjambres usando aprendizaje por refuerzo

PARA OBTENER EL GRADO DE:

**MAESTRÍA EN CIENCIAS DE LA COMPUTACIÓN**

P R E S E N T A:


**Ing. Marco Antonio Jiménez Limas**

DIRECTORES DE TESIS:

Dr. Juan Carlos Chimal Eguía

Dr. Ponciano Jorge Escamilla Ambrosio



Ciudad de México

Junio 2022

SIP-13
REP 2017

# INSTITUTO POLITÉCNICO NACIONAL
## SECRETARIA DE INVESTIGACIÓN Y POSGRADO

### *ACTA DE REGISTRO DE TEMA DE TESIS Y DESIGNACIÓN DE DIRECTOR DE TESIS*

Ciudad de México, a 03 de mayo del 2022

El Colegio de Profesores de Posgrado del **Centro de Investigación en Computación** en su Sesión
(Unidad Académica)

Extraordinaria No 06 celebrada el día 28 del mes marzo de 2022, conoció la solicitud presentada por el (la) alumno (a):

| Apellido Paterno: | JIMÉNEZ | Apellido Materno: | LIMAS | Nombre (s): | MARCO ANTONIO |
|---|---|---|---|---|---|

Número de registro: A 2 0 0 3 8 2

del Programa Académico de Posgrado: **Maestría en Ciencias de la Computación**

Referente al registro de su tema de tesis; acordando lo siguiente:

1.- Se designa al aspirante el tema de tesis titulado:

***"Modelado de movimiento colectivo en enjambres usando aprendizaje por refuerzo"***

Objetivo general del trabajo de tesis:

Desarrollar una simulación computacional de un enjambre de agentes capaces de presentar comportamientos colectivos emergentes a partir de su interacción con el entorno, implementando aprendizaje por refuerzo a nivel individual.

2.- Se designa como Directores de Tesis a los profesores:

Director: **Dr. Juan Carlos Chimal Eguía** 2° Director: **Dr. Ponciano Jorge Escamilla Ambrosio**

No aplica: ☐

3.- El Trabajo de investigación base para el desarrollo de la tesis será elaborado por el alumno en:

**Centro de Investigación en Computación**

que cuenta con los recursos e infraestructura necesarios.

4.- El interesado deberá asistir a los seminarios desarrollados en el área de adscripción del trabajo desde la fecha en que se suscribe la presente, hasta la aprobación de la versión completa de la tesis por parte de la Comisión Revisora correspondiente.

Director(a) de Tesis
**Dr. Juan Carlos Chimal Eguía**

2° Director de Tesis
**Dr. Ponciano Jorge Escamilla Ambrosio**

Aspirante
**C. Marco Antonio Jiménez Limas**

**Presidente del Colegio**
**Dr. Francisco Hiram Calvo Castro**

SIP-14
REP 2017

# INSTITUTO POLITÉCNICO NACIONAL

## SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

*ACTA DE REVISIÓN DE TESIS*

En la Ciudad de México siendo las 12:00 horas del día 20 del mes de junio del 2022 se reunieron los miembros de la Comisión Revisora de la Tesis, designada por el Colegio de Profesores de Posgrado de: Centro de Investigación en Computación para examinar la tesis titulada:

**"Modelado de movimiento colectivo en enjambres usando aprendizaje por refuerzo"**

del (la) alumno (a):

| Apellido Paterno: | JIMÉNEZ | Apellido Materno: | LIMAS | Nombre (s): | MARCO ANTONIO |
|---|---|---|---|---|---|

Número de registro: A 2 0 0 3 8 2

Aspirante del Programa Académico de Posgrado: **Maestría en Ciencias de la Computación**

Una vez que se realizó un análisis de similitud de texto, utilizando el software antiplagio, se encontró que el trabajo de tesis tiene 08 % de similitud. **Se adjunta reporte de software utilizado.**

Después que esta Comisión revisó exhaustivamente el contenido, estructura, intención y ubicación de los textos de la tesis identificados como coincidentes con otros documentos, concluyó que en el presente trabajo **SI** ☐ **NO** ☒ **SE CONSTITUYE UN POSIBLE PLAGIO.**

**JUSTIFICACIÓN DE LA CONCLUSIÓN:** *(Por ejemplo, el % de similitud se localiza en metodologías adecuadamente referidas a fuente original)*

Las referencias son las mismas que otros documentos y por ello se repiten.

**Es responsabilidad del alumno como autor de la tesis la verificación antiplagio, y del Director o Directores de tesis el análisis del % de similitud para establecer el riesgo o la existencia de un posible plagio.

Finalmente y posterior a la lectura, revisión individual, así como el análisis e intercambio de opiniones, los miembros de la Comisión manifestaron **APROBAR** ☒ **SUSPENDER** ☐ **NO APROBAR** ☐ la tesis por **UNANIMIDAD** ☒ o **MAYORÍA** ☐ en virtud de los motivos siguientes:

Cumple con los términos para su examen final.

**COMISIÓN REVISORA DE TESIS**

Dr. Juan Carlos Chimal Eguía
Director de Tesis

Dra. Elsa Rubio Espino

M. en C. Germán Téllez Castillo

Dr. Perciano Jorge Escamilla Ambrosio
2° Director de Tesis

Dr. Mario Eduardo Rivero Ángeles

Dra. Sandra Dinora Orantes Jiménez

INSTITUTO POLITÉCNICO NACIONAL
CENTRO DE INVESTIGACIÓN EN COMPUTACIÓN

Dr. Francisco Hiram Calvo Castro
PRESIDENTE DEL COLEGIO DE PROFESORES

# INSTITUTO POLITÉCNICO NACIONAL

## SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

*CARTA DE AUTORIZACIÓN DE USO DE OBRA PARA DIFUSIÓN*

En la Ciudad de México el día 7 del mes de junio del año 2022, el que suscribe Marco Antonio Jiménez Limas alumno del programa Maestría en Ciencias de la Computación con número de registro A200382, adscrito a Centro de Investigación en Computación, manifiesta que es autor intelectual del presente trabajo de tesis bajo la dirección del Dr. Juan Carlos Chimal Eguía, y el Dr. Ponciano Jorge Escamilla Ambrosio y cede los derechos del trabajo intitulado Modelado de movimiento colectivo en enjambres usando aprendizaje por refuerzo, al Instituto Politécnico Nacional, para su difusión con fines académicos y de investigación.

Los usuarios de la información no deben reproducir el contenido textual, gráficas o datos del trabajo sin el permiso expresado del autor y director(es). Este puede ser obtenido escribiendo a la siguiente dirección de correo marcojimenez@ciencias.unam.mx. Si el permiso se otorga, al usuario deberá dar agradecimiento correspondiente y citar la fuente de este.

Marco Antonio Jiménez Limas

# RESUMEN

El comportamiento colectivo de poblaciones de animales es un fenómeno que ha resultado de gran interés por la singularidad de los movimientos exhibidos y por ser un proceso resultante de la autoorganización. Se denomina a este tipo de movimiento agregado *comportamiento de swarm* o de *enjambre.* Diversos enfoques se han propuesto para modelar este tipo de comportamiento, dentro de los que destacan los modelos basados en agentes (MBA), los cuales se utilizan ampliamente para modelar sistemas complejos que contienen agentes espontáneos e interactivos, y es una poderosa herramienta para analizar comportamientos globales de sistemas complejos. Sin embargo, la mayoría de los modelos de movimiento de enjambre utilizan enfoques en el que las reglas son implementadas y la autoorganización resulta de dichas reglas preestablecidas. En el presente trabajo, se propone un modelo de movimiento colectivo de un enjambre de agentes utilizando el enfoque de MBA, pero en el que los agentes no tienen reglas predefinidas a seguir; en su lugar, los agentes tienen implementado un modelo de aprendizaje por refuerzo a nivel individual (RL por sus siglas en inglés), para que de esta forma ellos aprendan qué comportamientos son mejores para evitar la pérdida de vecinos, y así evaluar la emergencia de comportamiento colectivo a partir de la interacción usando las acciones aprendidas que resultan más benéficas. Se realizaron 30 simulaciones de tres modelos distintos: i) modelo de Vicsek para todos los agentes; ii) modelo de RL para un agente, y modelo de Vicsek para el resto de los agentes; y iii) modelo de RL para todos los agentes. Esto se hizo para comparar el tiempo de simulación que los agentes tardaban en orientarse y agruparse exhibiendo comportamiento colectivo al usar reglas predefinidas y al no usarlas; la lógica utilizada para el modelo de RL fue evitar la pérdida de vecinos imitando la necesidad de cohesión en grupos de animales en la naturaleza. Así, los agentes del modelo de RL fueron aprendiendo qué acciones les permitían perder menos vecinos. Se compararon los tres modelos utilizados, para evaluar la política de acción de los agentes con reglas preestablecidas y con el modelo de RL implementado a nivel individual para un agente y para todos los agentes. Los resultados obtenidos mostraron que es posible que en un sistema de agentes emerja

comportamiento colectivo derivado de la interacción de cada agente con su entorno, implementando RL a nivel individual. El tiempo de simulación que tardó un agente en alinearse con el resto fue más alto al implementar el aprendizaje por refuerzo que en el modelo de Vicsek ($9.814 \times 10^6$ vs. 142.72 ticks respectivamente), y fue considerablemente más alto en el modelo de aprendizaje por refuerzo para todos los agentes respecot al modelo de Vicsek ($11.749 \times 10^6$ ticks vs. 142.72 ticks respectivamente). Resulta ampliamente interesante que la lógica utilizada para el modelaje de RL haya sido similar a los modelos que utilizan enfoques más mecánicos, sobre todo para el estudio de las propiedades y mecanismos que rigen y subyacen a los fenómenos colectivos.

Con este trabajo se logró ir un paso más a profundidad en la forma de modelar a los enjambres, no solo siguiendo reglas, sino dándoles a los agentes un objetivo individual que les llevó a aprender comportamientos individuales, de los cuales emergió el comportamiento colectivo.

# ABSTRACT

The collective behavior of animal populations is a phenomenon that has been of great interest due to the uniqueness of the movements exhibited and because it is a process resulting from self-organization. This type of aggregate movement is called *swarm behavior*. Several approaches have been proposed to model this type of behavior, among which agent-based models (ABM) stand out, which are widely used to model complex systems that contain spontaneous and interactive components. ABMs are a powerful tool for analyzing the global behavior of complex systems. However, most swarm movement approaches use models in which rules are implemented and self-organization results from these pre-established rules. Here, a collective movement model of a swarm of agents is proposed using the ABM approach, but in which the agents do not have predefined rules to follow; instead, the agents have implemented a reinforcement learning model at the individual level (RL), so that in this way they learn what behaviors are better to avoid the loss of neighbors, and thus evaluate the emergence of collective behavior from interaction using the learned actions that are most beneficial. 30 simulations of three different models were carried out: i) Vicsek model for all agents; ii) RL model for an agent, and Vicsek model for the rest of the agents; and iii) RL model for all agents. This was done to compare the simulation time it took for agents to orient themselves and group together exhibiting collective behavior when using and not using predefined rules; the logic used for the RL model was to avoid neighbor loss by mimicking the need for cohesion in groups of animals in nature. Thus, the RL model agents learned which actions allowed them to lose fewer neighbors.

The three models used were compared to evaluate the action policy of the agents with pre-established rules and with the RL model implemented at the individual level for an agent and for all agents. The results obtained showed that it is possible that collective behavior emerges in a system of agents derived from the interaction of each agent with its environment, implementing RL at the individual level. The simulation time it took for an agent to align with the rest was higher when implementing RL than in the Vicsek model ($9.814 \times 10^6$ ticks vs. 142.72 ticks respectively), and it was considerably higher in the RL model for all agents with respect to the Vicsek model ($11.749 \times 10^6$ ticks vs. 142.72

ticks respectively). It is widely interesting that the logic used for RL modeling has been similar to models that use more mechanical approaches, especially for the study of the properties and mechanisms that govern and underlie collective phenomena.

With this work, it was possible to go a step further in the way of modeling swarms, not only following rules, but also giving agents an individual objective that led them to learn individual behaviors, from which collective behavior emerged.

# AGRADECIMIENTOS

Agradezco a mi novia, Fernanda Borjas, por ayudarme y apoyarme siempre y en todo momento, por ayudarme a redactar y dar estructura al escrito de esta tesis.

Agradezco a mi familia, en especial a mi mamá Lupita Limas, por siempre apoyarme y por emocionarse más de mis logros que lo que yo me emociono.

Agradezco a CONACyT, por el apoyo económico brindado durante la realización de este proyecto.

Agradezco al Instituto Politécnico Nacional, al Centro de Investigación en Computación, por el apoyo que me brindaron durante mi estancia.

Agradezco a mis directores de tesis, el Dr. Juan Carlos Chimal Eguía y el Dr. Ponciano Jorge Escamilla Ambrosio, por su orientación, apoyo y por la confianza que tuvieron en mí.

Agradezco a los miembros de mi jurado, la Dra. Elsa Rubio Espino, el M. en C. Germán Téllez Castillo, la Dra. Sandra Dinora Orantes Jiménez, y el Dr. Mario Eduardo Rivero Ángeles, por sus valiosas contribuciones que hicieron al trabajo, por tomarse el tiempo de revisar la tesis y hacer correcione

# ÍNDICE DE CONTENIDO

RESUMEN ..... I

ABSTRACT ..... III

AGRADECIMIENTOS ..... V

1. INTRODUCCIÓN ..... 3

1.1. Marco teórico ..... 3

1.1.1. Sistemas complejos y sus propiedades ..... 3

1.1.2. Autoorganización en enjambres ..... 10

1.1.3. Modelación de sistemas complejos ..... 12

1.1.4. Aprendizaje por refuerzo ..... 15

1.2. Justificación y aplicaciones ..... 17

1.3. Hipótesis ..... 21

1.4. Objetivos ..... 22

1.4.1. General ..... 22

1.4.2. Particulares ..... 22

2. MÉTODOS ..... 23

2.1. Modelo de movimiento colectivo (modelo de Vicsek) ..... 23

2.2. Aprendizaje de un agente (modelo de QL) ..... 27

2.3. Aprendizaje de enjambre (modelo de QL) ..... 31

3. RESULTADOS ..... 32

3.1. Modelo de movimiento colectivo (modelo de Vicsek) ..... 32

3.2. Aprendizaje de un agente (modelo de QL) ..... 33

3.3. Aprendizaje de enjambre (modelo de QL) ..... 36

4. DISCUSIÓN ..... 39

5. CONCLUSIONES Y TRABAJO A FUTURO ..... 42

6. REFERENCIAS BIBLIOGRÁFICAS ..... 44

# ÍNDICE DE FIGURAS

Figura 1.1.....4
Figura 1.2.....5
Figura 1.3.....7
Figura 1.4.....8
Figura 1.5.....10
Figura 1.6.....12
Figura 1.7.....14
Figura 2.1.....24
Figura 2.2.....24
Figura 2.3.....25
Figura 3.1.....32
Figura 3.2.....33
Figura 3.3.....33
Figura 3.4.....34
Figura 3.5.....35
Figura 3.6.....35
Figura 3.7.....36
Figura 3.8.....37
Figura 3.9.....38

# 1. INTRODUCCIÓN

## 1.1. Marco teórico

### 1.1.1. Sistemas complejos y sus propiedades

La ciencia de la complejidad ha surgido como un novedoso enfoque de investigación para estudiar los sistemas complejos (Phelan, 2001). Estos sistemas se caracterizan por ser dinámicos, no lineales, caóticos y multi-dimensionales, y por estar formados por una gran cantidad de componentes e interacciones (Waldrop, 1993).

A diferencia de otros enfoques de estudio, una de las particularidades de la complejidad es que engloba diferentes disciplinas; en pocas palabras, es una colección de teorías y herramientas conceptuales de distintos bagajes. Por ello, los sistemas que estudia pueden ser de una gran variedad de naturalezas: físicos, biológicos, químicos o incluso sociales. Dada esta diversidad, puede parecer extraño estudiarlos bajo un mismo enfoque; sin embargo, a diferencia de otras disciplinas científicas que tienden a enfocarse en los componentes mismos de cada sistema, la ciencia de la complejidad se enfoca en cómo los componentes dentro de un sistema se relacionan entre sí (Siegenfeld & Bar-Yam, 2020).

Como se mencionó anteriormente, los sistemas complejos tienen propiedades que los hacen sistemas de estudio muy interesantes; una de las más relevantes es que el comportamiento del sistema no puede ser entendido o predecido estudiando sus componentes de manera individual (Goldstein, 2011). Sin embargo, los sistemas complejos tienen muchas otras propiedades interesantes para su estudio, las cuales se listan a continuación.

#### a. Interacciones

Los sistemas complejos están compuestos por una gran cantidad de elementos que interactúan entre ellos y con el ambiente que les rodea. Estos componentes forman redes de interacción, las cuales pueden proveer información del sistema que dificulta el

estudio de los componentes de forma aislada o la predicción completa de su comportamiento futuro. Además, los componentes de un sistema también pueden ser sistemas completamente nuevos, lo que lleva a sistemas de sistemas, siendo interdependientes entre sí (Mitchell, 2009; Rivkin & Siggelkow, 2007).

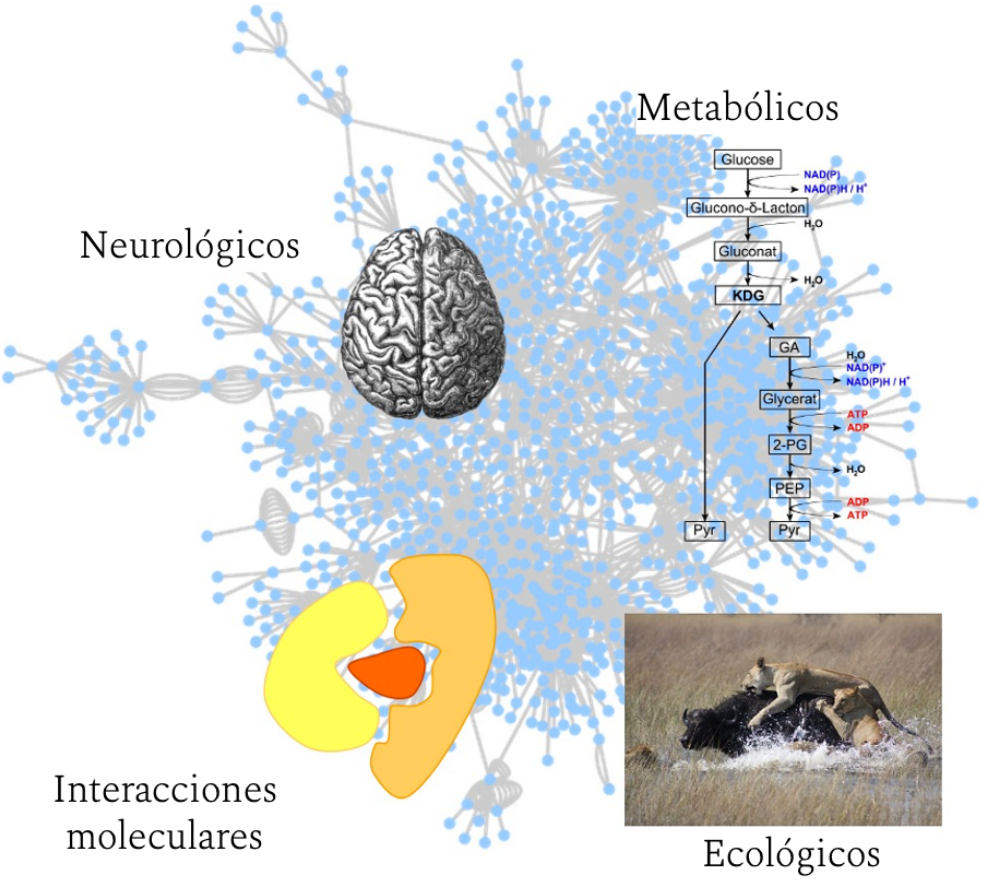


**Figura 1.1.** Redes de interacciones en fenómenos biológicos. Diversos fenómenos tienen una vasta cantidad de interacciones a diferentes niveles de organización; desde las interacciones celulares a nivel molecular hasta las interacciones ecológicas entre especies a nivel ecosistémico, los sistemas complejos se caracterizan por las interacciones entre sus componentes.

Uno de los desafíos más importantes del estudio de los sistemas complejos es no solo visualizar los componentes y sus interacciones, sino también comprender cómo estas interacciones dan lugar a las diferentes propiedades del sistema. Algunos ejemplos de estas interacciones en sistemas complejos son las miles de millones de neuronas que interactúan en el cerebro humano, las numerosas moléculas que interactúan en el metabolismo, las interacciones entre componentes celulares e incluso las interacciones

entre especies y factores abióticos en un ecosistema (Figura 1.1). Es importante notar que las interacciones en los sistemas complejos ocurren a muchas escalas distintas de organización, por lo que pueden encontrarse desde nivel molecular hasta nivel ecosistémico o espacial (Wimsatt, 1972).

## b. Emergencia

La emergencia se refiere a la existencia de propiedades colectivas, es decir, propiedades o características que no poseen los componentes de manera individual, sino que *emergen* a partir de las interacciones con otros componentes (Lichtenstein & Plowman, 2009). En sistemas simples, las propiedades del sistema pueden ser entendidas o descritas al estudiar las propiedades de sus componentes individuales. Esto no ocurre en sistemas complejos, en los que las propiedades del sistema a menudo no pueden entenderse o predecir a partir del estudio de sus componentes debido a que la interacción entre ellos genera información, estructuras y comportamientos colectivos no triviales a escalas más grandes (Domenico et al., 2019). Algunos ejemplos de emergencia en sistemas complejos son las múltiples células que forman un organismo vivo; desde la formación de células, tejidos, órganos, sistemas hasta un individuo (Cohen & Harel, 2007). Otro ejemplo está dado por los miles de millones de neuronas en un cerebro que dan lugar a la conciencia e inteligencia (Figura 1.2).

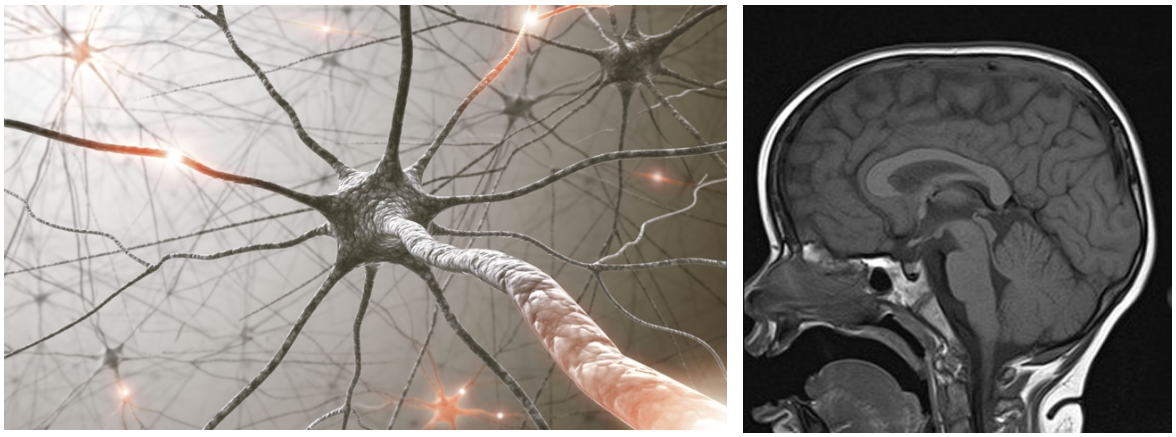

**Figura 1.2.** Las neuronas en nuestro cerebro constantemente se comunican a través de señales químicas que transmiten estímulos; su actividad individual puede ser estudiada a nivel fisiológico, sin embargo, la cognición es una propiedad que surge a raíz de las interacciones de miles de neuronas, es decir, es una propiedad que emerge a partir de las acciones e interacciones de las células cerebrales.

### c. Dinámica y caos

Los sistemas pueden ser estudiados a partir del cambio en su estado a través del tiempo; su estado puede ser descrito en función de las variables que mejor caracterizan al sistema. A medida que un sistema cambia de estado, también lo hacen así las variables que lo describen, a menudo en respuesta a su ambiente. Este cambio se denomina *lineal* si es proporcional al tiempo, al estado del sistema o a los cambios en el ambiente, y se denomina *no lineal* si no es proporcional a cualquiera de estas variables (Baker & Gollub, 1996).

Los sistemas complejos son en su mayoría *no lineales*, sus cambios de estado ocurren a diferentes ritmos dependiendo de su estado actual y de las condiciones del ambiente. Asimismo, también pueden llegar a tener estados estables en los que se mantienen igual a pesar de sufrir perturbaciones, o estados inestables en los que los que se ven seriamente afectados por una ligera perturbación (Rasband, 2015). Se considera un estado estable a los atractores del sistema, es decir estados a los cuales el sistema dinámico tiende conforme el tiempo avanza y de los cuales una vez que se llega a ellos, el sistema regresa a ese estado aun cuando hay perturbaciones, o en todo caso pasa a otro estado estable si es que el sistema presenta múltiples estados estables (May, 1977). En algunos casos, ligeros cambios en el ambiente pueden alterar completamente el comportamiento del sistema, lo que se conoce como bifurcaciones, transiciones de fase o puntos de inflexión (Devaney, 2018).

Algunos sistemas son caóticos, extremadamente sensibles a condiciones iniciales e impredecibles a largo plazo, mostrando el llamado *efecto mariposa* (Strogatz, 1994). Un sistema complejo también puede depender de la trayectoria, es decir, su estado futuro depende no solo de su estado presente, sino también del conjunto de estados que ha tenido en el pasado (Domenico et al., 2019). Un sistema complejo que presenta este tipo de dinámica es la volatilidad financiera en el mercado de valores (Litimi et al., 2018). Otro ejemplo es el clima, el cual cambia de manera impredecible en el tiempo. El sistema de ecuaciones es un modelo matemático simplificado para la convección atmosférica; este modelo es no lineal, no periódico, tridimensional y determinista (Broer & Takens,

2011). Las ecuaciones de Lorenz han sido objeto de cientos de artículos de investigación sobre sistemas no lineales y dinámica caótica (Figura 1.3).

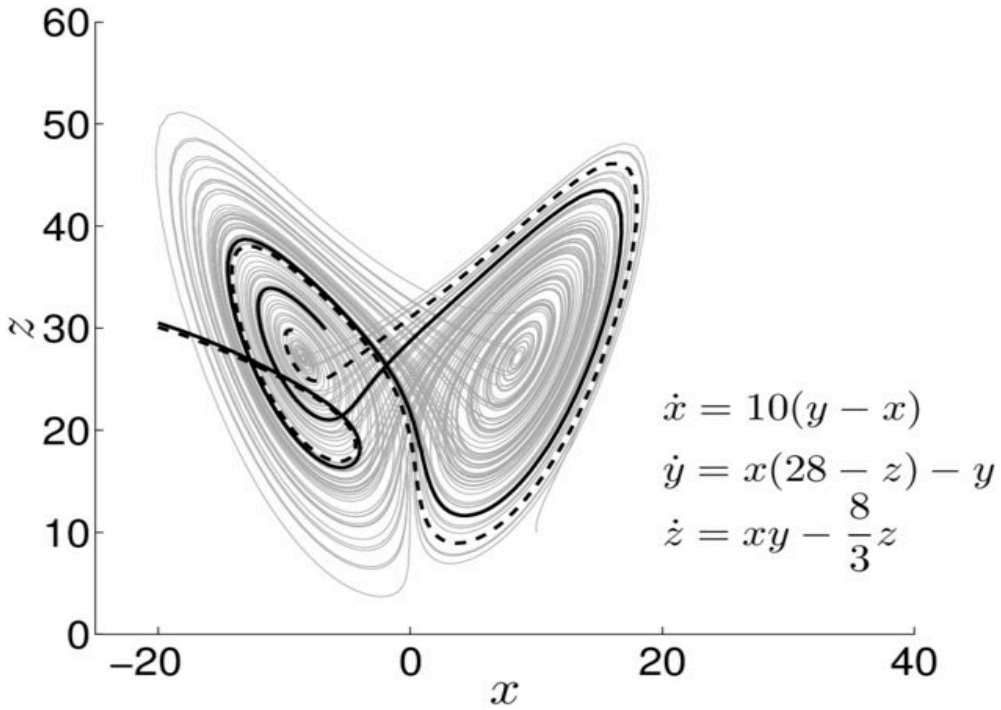


Figura 1.3. Uno de los sistemas caóticos más estudiados es el sistema de ecuaciones de Lorenz, cuyo resultado es un atractor caótico. Las líneas negras continuas y la línea discontinua son dos trayectorias que comienzan como trayectorias vecinas. Debido a la naturaleza caótica del atractor de Lorenz, la separación entre las trayectorias aumenta a medida que evoluciona sobre el atractor. Tomada de Kabiraj (2012).

## d. Autoorganización

Otra propiedad sumamente interesante de los sistemas complejos es la autoorganización. La autoorganización se define como la emergencia espontánea de orden en sistemas físicos y naturales (Kauffman, 1993). También se puede definir como el proceso mediante el cual los elementos que componen a los sistemas complejos se organizan para llegar a un estado estable (Yackinous, 2015). Los sistemas complejos pueden autoorganizarse para producir patrones no triviales de forma espontánea sin un modelo, a partir de las interacciones no lineales que se dan entre ellos (Ball, 1999). Este fenómeno puede producir estructuras físicas o funcionales, o comportamientos dinámicos colectivos.

Existen variados ejemplos de procesos de autoorganización en la naturaleza, como las poblaciones de estorninos que muestran patrones complejos de bandadas, o los peces en grandes cardúmenes que exhiben complejos comportamientos colectivos, o algunos grupos de insectos sociales en el forrajeo y realización de tareas complejas (Couzin & Krause, 2003) (Figura 1.4). Además, otros procesos como el desarrollo embrionario también presentan este fenómeno, en el que un cigoto se divide y finalmente se autoorganiza en la forma compleja de un organismo (Domenico et al., 2019).

El estudio de la autoorganización se ha vinculado estrechamente con nuevos campos de estudio, como la inteligencia artificial y la ciberseguridad (Dobson et al., 2019; Ranganathan & Kira, 2003). La simulación, especialmente usando modelos de agentes, se ha convertido en una de las herramientas más comunes para investigar los mecanismos que dan lugar a este tipo de organización en sistemas (Green et al., 2008; Sayama, 2015).

**Figura 1.4.** La autoorganización es un fenómeno recurrente en colonias de hormigas: los comportamientos colectivos complejos surgen como producto de las interacciones entre muchos individuos, cada uno siguiendo un conjunto simple de reglas, sin ningún tipo de control centralizado. Los individuos no poseen conocimiento global de las necesidades de la colonia, únicamente reaccionan sólo a su entorno local.

**e. Adaptación**

Los sistemas complejos pueden adaptarse y evolucionar; están constantemente cambiando y respondiendo a los estímulos del ambiente en el que se encuentran (Stonier & Yu, 1994). Esta adaptación puede ocurrir a numerosas escalas: cognitiva, a través del aprendizaje y adquisición de herramientas; social, mediante el intercambio de información a través de interacciones; o incluso evolutiva, a través de la variación genética y la selección natural (Domenico et al., 2019).

Estos sistemas tienen diferentes características que les permiten evolucionar, como la *robustez*, que es la capacidad para soportar perturbaciones; la *resiliencia*, que es la capacidad de volver al estado original después de una gran perturbación; o la *adaptación* misma, que es la capacidad de alterar el propio sistema para seguir siendo funcional y sobrevivir. Los sistemas complejos con estas propiedades se conocen como sistemas adaptativos complejos (Holland, 1992).

Un fenómeno presente en sistemas complejos adaptativos es la *histéresis* o *estados estables alternativos*; esto implica que el sistema puede existir en múltiples estados (conjuntos de condiciones únicas). Estos estados alternativos no son transitorios y, por lo tanto, se consideran estables en escalas de tiempo relevantes. Los sistemas pueden pasar de un estado estable a otro, en lo que se conoce como *cambio de régimen* o *de estado*. Debido a las retroalimentaciones que existen en estos sistemas complejos, éstos muestran resistencia a los cambios de régimen y, por lo tanto, tienden a permanecer en un estado a menos que las perturbaciones sean lo suficientemente grandes (Krasnosel'skii & Pokrovskii, 2012).

Múltiples estados pueden persistir bajo condiciones ambientales iguales, a lo que llamamos *histéresis*. Diversos estudios (Faassen et al., 2015; Scheffer, 1989; Scheffer & Carpenter, 2003) sugieren que los estados discretos están separados por umbrales, en contraste con los sistemas que cambian suave y continuamente a lo largo de un gradiente de una condición del sistema determinada. Un ejemplo de este comportamiento son los ecosistemas lacustres que sufren eutrofización (Figura 1.5), la cual depende de la

cantidad de nutrientes presentes en el sistema (variable independiente) y la concentración de biomasa de algas (variable de respuesta).

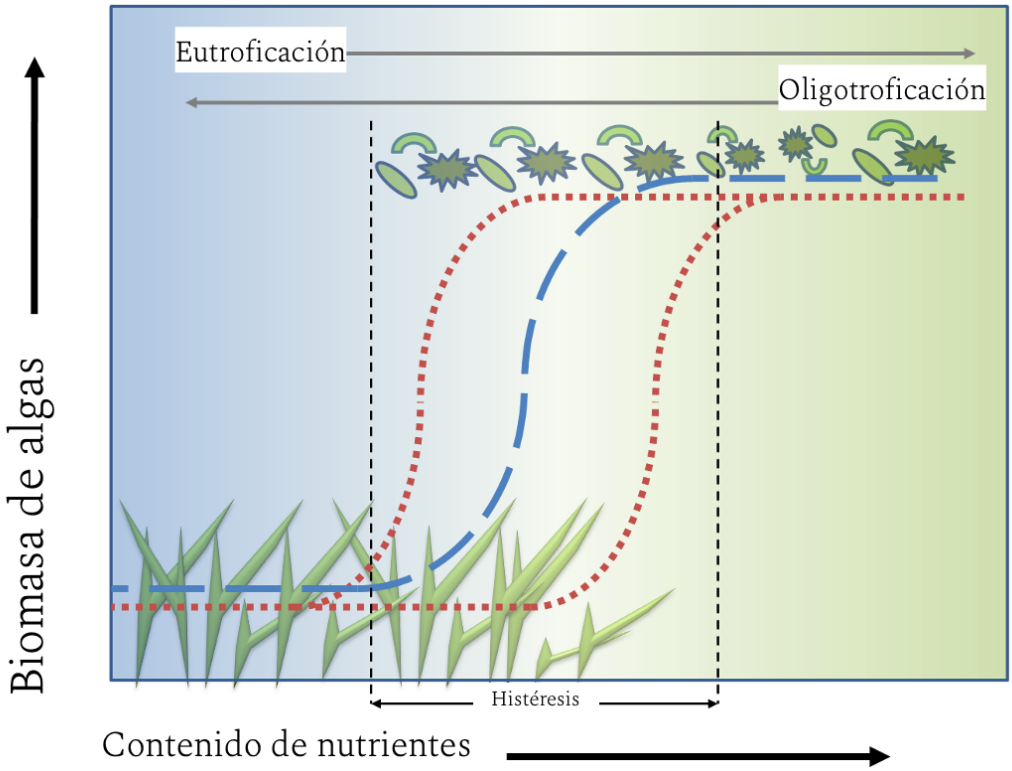


**Figura 1.5.** A menudo se muestra que los lagos responden de forma no lineal a la eutrofización (puntos azules) y en muchos casos se han asumido histéresis (puntos naranjas). La histéresis es un comportamiento no lineal en el que el estado de un lago no solo depende de su entrada actual sino también del estado anterior. Como resultado, podrían ocurrir dos estados estables dadas las mismas condiciones (zona de histéresis). Tomada de Janssen et al., (2015).

### 1.1.2. Autoorganización en enjambres

Uno de los fenómenos que ha captado el interés de científicos y estudiosos desde hace mucho tiempo es el comportamiento de poblaciones de diferentes organismos que se organizan en peculiares formaciones agregadas (Bak, 2013). Bancos de peces, bandadas de pájaros, enjambres de abejas y hormigas se encuentran entre los ejemplos más reconocibles (Camazine et al., 2020) (Figura 1.6). Se denomina a este tipo de movimiento agregado *comportamiento de swarm* o de *enjambre* (Parrish & Edelstein-Keshet, 1999).

El comportamiento colectivo complejo surge de las interacciones de los individuos que exhiben comportamientos simples por sí mismos (Bak, 2013). Las colonias de hormigas

y las colmenas, por ejemplo, tienen la interesante propiedad de que un gran número de ellas llevan a cabo procesos de manera muy organizada con un comportamiento aparentemente decidido que mejora su supervivencia colectiva (Beni, 2004; Bonabeau et al., 1997). Sorprendente y paradójicamente, estos insectos parecen utilizar reglas de interacción muy simples. ¿Cómo es posible que los *enjambres* de criaturas con un poder cerebral y unas capacidades de comunicación relativamente bajos puedan organizarse sin ningún control central en comportamientos que parecen exhibir un propósito global? (Bonabeau et al., 1999).

La autoorganización es un proceso en el que surge alguna forma de orden o coordinación global a partir de las interacciones locales entre los componentes de un sistema inicialmente desordenado. Este proceso es espontáneo, es decir, no está dirigido ni controlado por ningún agente o subsistema interno (Zhang, 2015). La autoorganización generalmente se basa en tres condiciones (Bonabeau et al., 1999): (*i*) fuertes interacciones dinámicas no lineales, que pueden incluir retroalimentaciones positivas o negativas; (*ii*) una amplificación de las fluctuaciones, es decir, la necesidad de eventos aleatorios; y (*iii*) múltiples y complejas interacciones: pueden ser directas (visuales, físicas o químicas) o indirectas (estigmergia o a través del ambiente) (Young & Crawford, 2004).

A pesar de estas propiedades, la definición de autoorganización permanece elusiva. Serra & Zanarini (2013) describen el concepto de autoorganización generalmente como "un comportamiento altamente organizado incluso en ausencia de un diseño predeterminado". Una definición novedosa de autoorganización presentada aquí es el comportamiento del sistema que mantiene sus puntos operativos en o cerca de una frontera óptima de Pareto. Esta noción de eficiencia constituye una característica central de los principios fundamentales de la inteligencia colectiva, la cual se ha descrito como la inteligencia emergente de grupos de agentes autónomos simples (Fleischer, 2005).

**Figura 1.6.** El comportamiento de enjambre es un comportamiento colectivo exhibido por entidades, particularmente animales, que se agregan de manera autoorganizada. Desde la perspectiva de la modelación, es un comportamiento emergente que surge de reglas simples que son seguidas por individuos y que no involucra ningún tipo de control central.

### 1.1.3. Modelación de sistemas complejos

El interés por entender la inteligencia colectiva ha fomentado la formulación de modelos confiables y herramientas de simulación con el objetivo de desentrañar los mecanismos que subyacen a dichos fenómenos. Una de las disciplimas en las que la inteligencia colectiva ha tenido mayor incidencia es el área de la vida artificial, la cual es la disciplina que intenta entender las propiedades principales de sistemas vivos, mediante el uso de modelos y simulaciones computacionales (Bedau, 2003). Dentro de esta área, diferentes enfoques se han utilizado para el modelaje de estos comportamientos, ya que la inteligencia colectiva resultante de estos fenómenos resulta de gran interés para aplicaciones como la robótica, la ciberseguridad, la optimización de procesos o patrones de tráfico en sistemas de movilidad y transporte (Liu & Passino, 2000).

Los modelos propuestos han variado en grados de idealización y complejidad de la descripción de los enjambres, sus entornos y las interacciones entre ellos (Touma et al., 2010). Un método de modelado matemático para la autoorganización es el uso de ecuaciones diferenciales y otro método son los autómatas celulares (Wolfram, 2002). Uno de los enfoques más utilizados para el estudio de fenómenos de autoorganización y en general para el modelaje de los sistemas complejos es el Modelado Basados en Agentes (MBA). El MBA tiene sus raíces en el modelado de sistemas adaptativos complejos (SAC). Un SAC puede autoorganizarse espontáneamente y reconstruir sus componentes dinámicamente para sobrevivir en el medio ambiente.

El MBA es un método de modelado ascendente (o bottom-up). Se utiliza ampliamente para modelar sistemas complejos que contienen agentes espontáneos e interactivos, y es una poderosa herramienta para analizar comportamientos globales de sistemas complejos (Zhang, 2015). Los MBA modelan la dinámica de los sistemas adaptativos basados en el mecanismo de adaptación de los individuos. Es completamente diferente a los modelos de ecuaciones diferenciales, los cuales son modelos descendentes (o top-down) que se enfocan en el fenómeno general y no en los componentes individuales del sistema (Helbing, 2012).

El MBA se ha utilizado como un enfoque de estudio de comportamientos y mecanismos a nivel sistema, ya que permite capturar las características esenciales de las entidades del sistema y sus interacciones (Manson et al., 2012). Este enfoque se ha utilizado con éxito en la simulación de distintos tipos de procesos, como la autoorganización molecular, la dinámica de difusión, el tráfico y la gestión de flujo, e inclusive en simulaciones de mercado (bolsa de valores), entre muchos otros. Generalmente, el MBA se puede utilizar en los siguientes casos (Bonabeau, 2002): (i) las interacciones entre agentes son no-lineales y discretas; (ii) las interacciones son complejas y heterogéneas; (iii) los factores espaciales son muy importantes y las ubicaciones de los agentes no son fijas; y (iv) los agentes muestran comportamientos complejos y diversos, incluido el aprendizaje y la adaptación. El MBA se puede usar junto con otras técnicas de modelado como dinámica sistemática o ser complementado con estadística (Zhang, 2015).

Respecto a los fenómenos de autoorganización, el MBA resulta un enfoque sumamente útil para la simulación del comportamiento colectivo. En 1987, Reynolds creó un modelo de agentes llamado *boids*, que es un modelo de comportamiento colectivo para simular el movimiento de una bandada de pájaros (Reynolds, 1987). Cada agente o boid se implementa como un ente independiente que navega de acuerdo con su propia percepción del entorno dinámico, y sigue las siguientes reglas: (i) la regla de la separación, que dicta que un boid debe alejarse de otros agentes cercanos a él, con la idea de simular la evitación de colisiones en el aire; (ii) la regla de la alineación, que dicta que un boid debe moverse en la dirección general en la que se mueve la parvada, promediando las velocidades y direcciones de los otros agentes; y (iii) la regla de la cohesión, que indica que un agente debe minimizar la exposición al exterior de la parvada moviéndose hacia el centro percibido de la parvada (Figura 1.7). La interacción de los agentes siguiendo estas reglas simples da lugar a un comportamiento colectivo complejo muy similar al observado en parvadas de poblaciones de estorninos en la naturaleza.

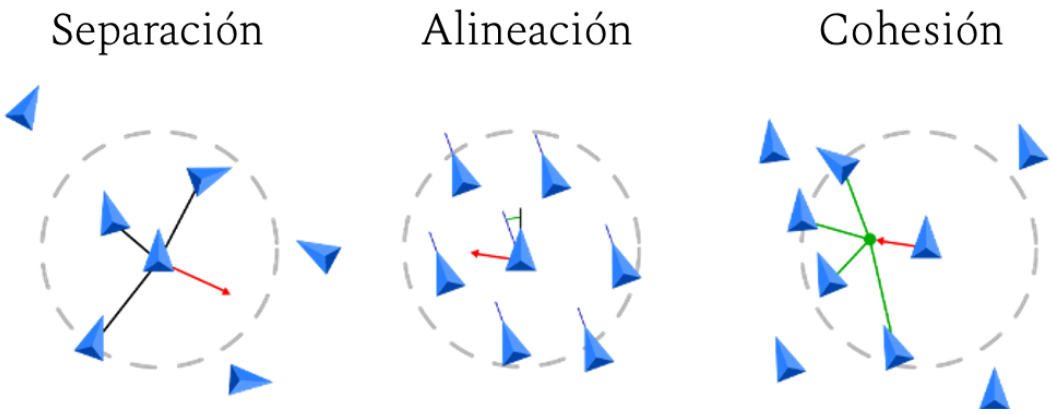


**Figura 1.7.** Reglas de colectividad del MBA "boids" (Reynolds, 1987): Separación, Alineación y Cohesión.

El modelo de *boids* ha sido ampliamente utilizado para diversos fines y como punto de partida para el entendimiento de los mecanismos subyacentes de la inteligencia colectiva. Es por ello que el MBA resulta un enfoque de gran interés y utilidad para estudios relacionados con fenómenos de autoorganización en sistemas complejos, en particular del comportamiento colectivo, foco central de la presente tesis.

Sobre esta base, han surgido un gran número de modelos de movimiento colectivo (Cavagna et al., 2015; Olfati-Saber, 2006). Sin embargo, la mayoría de los modelos de movimiento de enjambre existentes solo se enfocan en el comportamiento individual o de corto plazo del grupo sin considerar el comportamiento de aprendizaje inteligente. Por ejemplo, observando el comportamiento de los enjambres de entes biológicos para evitar a los depredadores de manera efectiva, es muy importante que los individuos tomen decisiones acertadas, además de seguir a sus vecinos en movimientos cooperativos. De hecho, el mecanismo de retroalimentación formado por el comportamiento de aprendizaje cooperativo inteligente es la clave para hacer que el comportamiento de enjambre sea adaptativo y acumulativo, y también la clave para comprender los sistemas de enjambre. La capacidad de evitar a los depredadores es una de las manifestaciones típicas de la adaptabilidad ambiental del sistema de enjambre biológico, que puede estudiarse como el problema de control cooperativo para los sistemas de enjambre que enfrentan amenazas dinámicas en un entorno desconocido (Lan et al., 2020).

### 1.1.4. Aprendizaje por refuerzo

El aprendizaje por refuerzo (RL por sus siglas en inglés) se define como el aprendizaje por prueba-error a partir del desempeño de un agente y su retroalimentación del entorno o un evaluador externo (Sutton & Barto, 2018). En otras palabras, es un tipo de aprendizaje en el que un agente debe aprender un comportamiento para resolver un problema a través de interacciones de prueba y error con un ambiente dinámico (Kaelbling et al., 1996).

El RL implica que los agentes aprenden de sus interacciones, entre ellos y con el ambiente; además, este tipo de aprendizaje está orientado a objetivos. Los problemas de RL implican aprender qué hacer, cómo asignar situaciones a acciones, para maximizar una señal de recompensa numérica (Sutton & Barto, 2018). Existen tres características distintivas del RL: (*i*) son problemas de circuito cerrado, ya que las acciones del sistema de aprendizaje influyen en sus entradas posteriores; (*ii*) los agentes no tienen instrucciones o reglas sobre qué acciones tomar, a diferencia de otros tipos de aprendizaje automático, sino que debe descubrir qué acciones producen la mayor

recompensa al probarlas; y (*iii*) en algunos casos las acciones pueden afectar no solo la recompensa inmediata sino también la siguiente situación y, a través de ella, todas las recompensas subsecuentes (Sutton & Barto, 2018).

Los modelos de RL generalmente comienzan con agentes completos, interactivos y que buscan completar un objetivo explícito; pueden detectar aspectos de sus entornos y pueden elegir acciones para influir en sus entornos (Kuremoto et al., 2008). Estos agentes también pueden ser componentes de un sistema de comportamiento más grande (como de un grupo de entes sociales que exhiben patrones complejos de comportamiento). En este caso, el agente interactúa directamente con el resto de los componentes del grupo e interactúa indirectamente con el entorno o ambiente del sistema.

Uno de los desafíos que surgen en el aprendizaje por refuerzo, y no en otros tipos de aprendizaje, es la disyuntiva existente entre la *exploración* y la *explotación* (Sutton & Barto, 2018). Con la finalidad de obtener la mayor cantidad de recompensa, un agente de aprendizaje por refuerzo debe realizar acciones que ha probado en el pasado y que le han brindado recompensas. Sin embargo, para descubrir tales acciones, tiene que probar acciones que no ha realizado antes. El agente tiene que explotar lo que ya sabe para obtener una recompensa, pero también tiene que explorar para tomar mejores decisiones en el futuro. En una tarea estocástica, cada acción debe intentarse muchas veces para obtener una estimación fiable de su recompensa esperada, y es a esto a lo que se conoce como la disyuntiva exploración-explotación, la cual solo surge en el aprendizaje por refuerzo.

El RL se considera una de las tecnologías centrales en el diseño de sistemas inteligentes (Lan et al., 2020). Sus características han sido ampliamente estudiadas y aplicadas en los campos de la inteligencia artificial, el aprendizaje automático y el control automático. Asimismo, el RL se ha utilizado para el estudio de inteligencia colectiva, como el comportamiento de enjambre o de swarm (Hüttenrauch et al., 2019; Kuremoto et al., 2008; Lan et al., 2020). Aunque el RL está diseñado a partir del punto de vista de un agente individual, también puede resultar de utilidad en el aprendizaje óptimo de comportamientos colectivos (Kuremoto et al., 2008).

Uno de los avances más importantes en el RL fue el desarrollo de un algoritmo de control conocido como Q-learning (Watkins, 1989). Su forma más simple, Q-learning de un paso, se define por:

$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha[R_{t+1} + \gamma \min_{a} Q(S_{t+1}, a) - Q(S_t, A_t)] \quad (1)$$

La idea fundamental de este algoritmo es que la función de valor de acción aprendida ($\boldsymbol{Q}$) depende de dos cosas: de su estado actual ($\boldsymbol{S_t}$) y de la acción más óptima que puede tomar el agente dado su estado actual ($\boldsymbol{A_t}$). El objetivo es aprender las acciones más óptimas que un agente puede tomar bajo diferentes circunstancias, dependiendo de la recompensa que cada par de acción-estado le provee (Sutton & Barto, 2018).

Este método brinda a los agentes la capacidad de aprender a actuar de manera óptima en los dominios markovianos experimentando las consecuencias de las acciones, sin necesidad de que construyan mapas de los dominios (Watkins & Dayan, 1992).

En el presente trabajo, se propone un modelo de movimiento colectivo de un enjambre de agentes utilizando el enfoque de MBA, pero en el cual los agentes no tengan reglas predefinidas a seguir, en su lugar los agentes tendrán implementado un modelo de aprendizaje por refuerzo a nivel individual, para que de esta forma ellos aprendan qué comportamientos tomar a nivel individual y de esta forma llegar a un comportamiento colectivo emergente. Esto con la finalidad de ir un paso atrás en el proceso de modelado, en lugar de modelar al agente de forma individual con reglas predefinidas, poder modelar el aprendizaje que tiene el agente y que da lugar a sus comportamientos individuales y al comportamiento del sistema. De esta manera puede ser más entendible el por que o como es que se dan estos comportamientos en la naturaleza, e incluso puede ser una forma de modelar o tratar de replicar estos comportamientos en sistemas artificiales (como sistemas de robots).

## 1.2. Antecedentes

Como se mencionó con anterioridad, el movimiento colectivo es un fenómeno que ha sido ampliamente estudiado durante mucho tiempo, especialmente en las áreas de la dinámica no lineal, ecología, psicología social e incluso en la robótica. Desde un punto

de vista evolutivo, se ha propuesto que el comportamiento colectivo prevalece en muchas especies ya que representa una ventaja adaptativa, como la reducción del riesgo a ser depredado, o la provisión de eficiencia en el apareamiento, o incluso facilitar la búsqueda de alimento (Werner & Dyer, 1992). Sin embargo, el estudio y entendimiento del movimiento colectivo y de los enjambres resulta una tarea muy compleja de ejecutar de manera experimental; para ello, el modelado y las simulaciones computacionales han resultado una forma alternativa de estudiar estos fenómenos. Estas herramientas han proporcionado una forma concreta de probar y derivar nuevas teorías sobre el comportamiento colectivo (Ward et al., 2001). Por ejemplo, el modelo *boids* de Reynolds (1987) representa uno de los primeros trabajos que buscan entender los mecanismos subyacentes al movimiento colectivo utilizando herramientas de simulación computacional. Mataric (1992) desarrolló con éxito un sistema robots para producir un comportamiento colectivo; en su trabajo, afirma que éste último es la combinación ponderada de una serie de interacciones básicas: evitación de colisiones, seguimiento, dispersión, agregación y orientación. Al programar cada uno de estos comportamientos en varios robots y luego establecer un peso que determinara cuál tenía más probabilidades de ejecutarse, Mataric pudo producir un comportamiento colectivo bastante sofisticado.

Yendo un paso más allá en la simulación, y considerando la fisiología y el ambiente de las especies, Ward y colaboradores (2001) investigaron el uso de controladores sensoriales evolucionados para producir conducta colectiva en peces artificiales. En su estudio, crearon un conjunto de agentes artificiales en un mundo artificial con "peligros" y "alimento"; en cada agente, implementaron un cerebro de red neuronal artificial simple que controlaba su movimiento bajo diferentes circunstancias. Su trabajo destaca el papel de la fisiología de las especies en la comprensión del comportamiento y el papel del ambiente en el desarrollo de sistemas sensoriales.

Otros trabajos, como el de Theraulaz y Bonabeau (1995), se enfocan más en la cooperación orientada a objetivos; en su trabajo, desarrollaron una población de agentes que construyeron colectivamente una estructura de nido depositando ladrillos de acuerdo con la percepción del entorno local y un conjunto de reglas de comportamiento.

En cuanto a implementaciones en el área de la robótica, Martinoli (1999) realizó un trabajo pionero con robots, quien utilizó la evolución artificial para sintetizar un sistema de control neuronal a un grupo de robots, a los que se les pidió que encontraran "artículos de comida" distribuidos aleatoriamente en un área. En algunos casos, los individuos evolucionados mostraron comportamientos colectivos interesantes, como explorar el área en parejas. Por otro lado, Baldassare y colaboradores (2003) desarrollaron un conjunto de experimentos en los que se simularon robots para que tuvieran la capacidad de agregarse y moverse juntos hacia un objetivo; los autores mostraron que los individuos evolucionados muestran patrones de comportamiento interesantes en los que los grupos de robots actúan como una sola unidad.

Otros enfoques han sido utilizados en el modelado del comportamiento colectivo, como el descrito por Ecoffet y colaboradores (2020), quienes abordan el problema de aprendizaje de estrategias cooperativas en enjambres de robots. En su trabajo, se preguntan cómo permitir que cada robot en un enjambre aprenda el comportamiento socialmente óptimo cuando este comportamiento es individualmente subóptimo, e incorporan teoría de juegos para estudiar dicha disyuntiva.

En general, los estudios realizados en materia de comportamiento colectivo utilizando herramientas como el modelado y la simulación computacional demuestran que estas técnicas, al ser implementadas tomando en consideración la fenomenolofía y comportamiento de los individios modelados, permiten obtener como resultado fenómenos de autoorganización que surgen de las interacciones entre los individuos (agentes o robots), y entre los individuos y el entorno. Esto representa un poderoso método para estudiar y recrear el comportamiento colectivo.

## 1.3. Justificación y aplicaciones

Como se mencionó anteriormente, los enjambres o *swarms* están compuestos de agentes autoorganizados que llevan a cabo tareas en conjunto. Se ha observado que la eficiencia del tiempo y la eficiencia energética se pueden mejorar considerablemente si los individuos se organizan en grupos para realizar sus funciones (Sumpter, 2010). Por ejemplo, las aves son capaces de generar formaciones durante la migración para volar

con mayor eficiencia energética, los peces en bandadas son más efectivos para defenderse de los depredadores que los individuales, las hormigas siempre cooperan en grupos para cazar y migrar (Couzin, 2009; Giardina, 2008; Zhu et al., 2017).

La mayoría de las aplicaciones del estudio de los sistemas de enjambre se encuentran dentro de la teoría del control y la teoría de grafos (Ren & Cao, 2011). En particular, uno de los mayores intereses que se tiene para aplicar el modelaje de sistemas de comportamiento colectivo en la realidad es en experimentos con vehículos aéreos no tripulados (Zhu et al., 2017). Se han realizado diversos experimentos usando sistemas de enjambre (principalmente mediante el uso de múltiples vehículos terrestres autónomos). Azuma y colaboradores (2013) diseñaron un marco de control de transmisión para la coordinación de un sistema de enjambre, utilizando un grupo de vehículos terrestres autónomos para probar el método propuesto.

Otra de las áreas en las que el modelado de comportamiento colectivo ha sido ampliamente utilizado es en el área de la optimización (Lim & Jain, 2009). Especialmente, el algoritmo ACO (*Ant Colony Optimization*) y sus variantes han sido aplicados en diversos problemas de optimización, como el problema del vendedor viajero (*TSP* por sus siglas en inglés), coloración de grafos, planificación de rutas, entre muchos otros (Bonabeau et al., 2000). Además, en el área de sistemas de potencia también se ha aplicado este enfoque para problemas relacionados con sistemas de energía, flujos de potencia óptimos, confiabilidad y seguridad del sistema de potencia, entre otros (Del Valle et al., 2008).

Asimismo, el modelado del comportamiento colectivo ha comenzado a incidir en el área de la bioinformática (Lim & Jain, 2009). Las diferentes técnicas de optimización derivadas de este enfoque han sido utilizadas para la agrupación de datos de expresión génica, construcción de árboles filogenéticos, predicción estructural de macromoléculas orgánicas como los ácidos nucleicos y las proteínas, y problemas de agrupamiento molecular (Das et al., 2008).

Finalmente, se han reportado estudios en los que se utilizan herramientas derivadas del modelado de comportamiento colectivo en la realización de hardware (Lim & Jain, 2009).

Entre ellos se incluyen el generador de números aleatorios basado en la optimización de colonias, matrices de puertas programables en campo, circuitos digitales, así como filtros de respuesta de impulso infinito (Duan & Yu, 2007). Por otro lado, el hardware basado en la optimización de enjambres de partículas para el entrenamiento de redes neuronales, el diseño y ajuste de controladores, robots móviles, sistemas de sensores de tolerancia a fallas, así como el hardware basado en la optimización de colonias de hormigas para redes de sensores inalámbricos (Johnson et al., 2008).

Los resultados que se han obtenido de la aplicación del modelado de comportamiento colectivo han generado un creciente interés en esta área, haciendo esta investigación muy atractiva para aplicaciones de optimización y control, especialmente en vista de la capacidad de los sistemas inteligentes de enjambre para hacer frente a fallas y entornos cambiantes (Bonabeau et al., 2000). Actualmente, se percibe a la inteligencia colectiva como un nuevo paradigma importante en optimización y control, por lo que es muy factible pensar que seguirán surgiendo nuevas aplicaciones prácticas del modelado de la inteligencia colectiva. El presente trabajo contribuye a la comprensión de cómo funcionan estos sistemas en la naturaleza, para así poder utilizar dicho conocimiento en diferentes contextos prácticos.

## 1.4. Motivación

Hay varias especies de insectos, aves, peces, bacterias en la naturaleza que presentan comportamientos colectivos durante su movimiento. Este tipo de sistemas son muy interesantes y sorprendentes ya que tienen la capicidad de realizar tareas sin que haya un líder dentro o fuera del grupo guiándolos. El tipo de movimiento colectivo que realizan estos enjambres es muy fluido, e incluso viéndolo a otra escala, pareciera que es un organismo el que se está moviendo, pero en realidad es un conjunto de agentes que están interactuando entre ellos, y dan lugar a este tipo de movimientos fluidos. Hay varios modelos que logran replicar este tipo de movimiento colectivo, pero aún no hay una respuesta de porqué es que cierto grupo de animales realizan estos movimientos, o cómo es que aprendieron qué comportamientos indivduales seguir para poder realizar esto. Los MBA parten de reglas de interaccion agente-agente y agente-ambiente, y con esto

pueden replicar ciertos comportamientos observados en la naturaleza, pero en la naturaleza no hay reglas que los individuos sigan: ellos tienen comportamientos que han aprendido durante su vida. Poder realizar un modelo en el cual los agentes aprendan comportamientos en lugar de seguir reglas puede ayudar a entender mejor cómo es que los agentes naturales funcionan, y dar una mejor idea de por qué es que hay tantos grupos de animales que tienen estos comportamientos. También puede ayudar a tener una metodología o enfoque para aplicarlo en casos prácticos, como en la robótica de enjambre y que así los robots tengan un funcionamiento más parecido al que tienen los animales que forman enjambres.

## 1.5. Hipótesis

Es posible que en un sistema de agentes emerja comportamiento colectivo derivado de la interacción de cada agente con su entorno, implementando aprendizaje reforzado a nivel individual.

## 1.6. Objetivos

### 1.6.1. General

Desarrollar una simulación de un enjambre de agentes capaz de presentar comportamiento colectivo emergente a partir de la interacción de cada agente con su entorno, implementando aprendizaje por refuerzo a nivel individual.

### 1.6.2. Particulares

- Identificar características sensoriales que posibilitan la emergencia de comportamiento colectivo en enjambres biológicos.
- Seleccionar un MBA en el cual se de un movimiento colectivo en el enjambre para usarlo como modelo de referencia
- Realizar un MBA en el que se le aplique RL a cada agente a nivel individual
- Seleccionar formas de medir o cuantificar si el modelo de RL es capaz de presentar comportamientos similares al MBA de referencia
- Realizar experimentos para evaluar el modelo.

# 2. MÉTODOS

Para modelar el movimiento colectivo de un enjambre de agentes, de tal forma que éste resultara del comportamiento individual aprendido por cada agente, el trabajo se dividió en tres partes:

a) En la primera parte se modeló el movimiento colectivo del enjambre en dos dimensiones, para lo que se utilizó como base el Modelo de Vicsek (Vicsek et al., 1995).
b) Una vez obtenido el enjambre con movimiento colectivo, en la segunda parte se agregó un agente nuevo al cual no se le implementó el modelo de Vicsek, sino un algoritmo de aprendizaje por refuerzo mediante el cual aprendió a coordinar su movimiento con el del enjambre. De este paso se observó que el agente aprendió comportamientos similares a los de los demás agentes.
c) Finalmente se hizo un modelo en el cual no se implementó el modelo de Vicsek a ninguno de los agentes; en su lugar, se implementó un algoritmo de aprendizaje por refuerzo a nivel individual en todos los agentes.

## 2.1. Modelo de movimiento colectivo (modelo de Vicsek)

Dado que los sistemas de enjambres son sistemas complejos, se utilizó el enfoque de MBA para realizar esta simulación. En específico, se utilizó el modelo de Vicsek para describir las interacciones entre agentes y obtener movimiento colectivo. El modelo se implementó en NetLogo (Wilensky, 1999), el cual es un software que permite realizar modelos y simulaciones computacionales de sistemas multi-agente en 2D y 3D. En este software se puede desarrollar la simulación de un modelo escribiendo código en una pestaña del IDE (acrónimo de Integrated Development Environment, en inglés), el cual tiene un lenguaje propio que facilita el modelado usando agentes (Figura 2.1a). Además, tiene una interfaz gráfica en la cual se pueden variar parámetros del modelo de manera dinámica durante la simulación, misma que es muy sencilla de utilizar para el usuario y

que permite controlar la ejecución de la simulación sin tener que recurrir al código (Figura 2.1b).

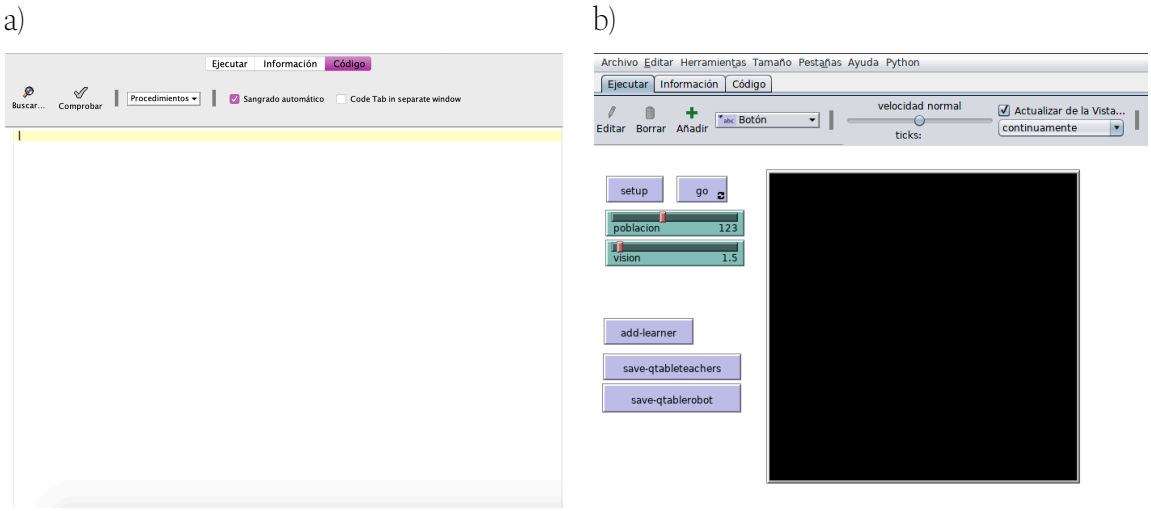


**Figura 2.1.** Componentes de NetLogo. **(a)** Entorno de código de Netlogo, en el que se codifica el modelo a implementar en la simulación. **(b)** Interfaz gráfica de NetLogo, en la que pueden controlarse los valores de los parámetros, así como diversos aspectos de la ejecución de la simulación.

El modelo consistió de los siguientes componentes: un conjunto de $n$ agentes idénticos donde cada agente tenía una posición ($\boldsymbol{x}$), una velocidad ($\boldsymbol{v}$), un rango de visión ($\boldsymbol{r}$) y un ángulo de orientación ($\boldsymbol{\theta}$) (Figura 2.2).

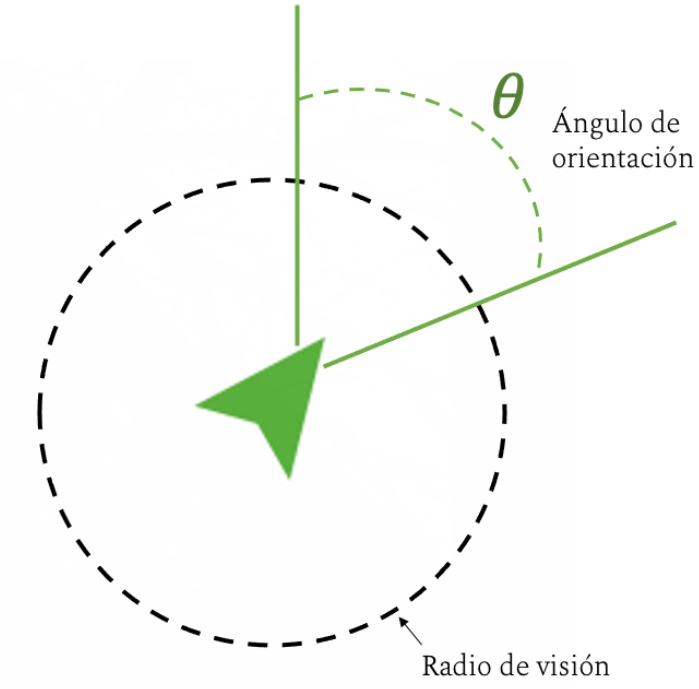


**Figura 2.2.** Radio de visión y ángulo de orientación de los agentes.

Los agentes se encontraban situados en un espacio cuadrado de dos dimensiones con fronteras periódicas. Una frontera periódica es una condición de frontera espacial que se le pone a un modelo, así el espacio es un toroide. Es decir, el extremo derecho del espacio está conectado con el extremo izquierdo del espacio, y la parte superior del espacio está conectada con la parte inferior del espacio. De esta manera, si un agente se mueve y llega al extremo derecho del espacio, el agente al seguir moviéndose aparecerá en el extremo izquierdo del espacio. Inicialmente los agentes se posicionaron de manera aleatoria en el espacio con una orientación también aleatoria (Figura 2.3).

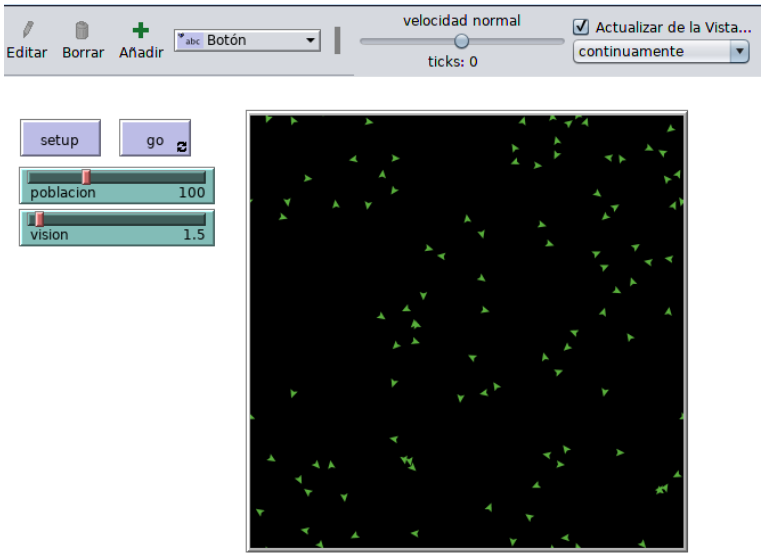


**Figura 2.3.** Posición y orientación inicial de los agentes.

La rapidez de movimiento de los agentes fue constante, así que la velocidad solo dependía del cambio de orientación. Cabe destacar que en las simulaciones en NetLogo el tiempo es discreto, y va avanzando en pasos discretos de una unidad generalmente llamada *tick.* Así, en cada paso de tiempo, cada uno de los agentes actualizó su posición bajo la siguiente ecuación:

$$x_i^{t+1} = x_i^t + v_i^t \Delta t \qquad (2)$$

La ecuación 1 describe cómo el agente $i$ actualizó su posición. Donde $x_i^t$ y $v_i^t$ son la posición y velocidad del agente $i$ al tiempo $t$ respectivamente, y $\Delta t = 1$. En las simulaciones desarrolladas se tomó la rapidez como constante con un valor de 1 unidad de espacio por unidad de tiempo, es decir $\| v_i^t \| = 1 \forall\, t \in i$.

Así, en cada paso de tiempo, el agente $i$ cambiaba su velocidad mediante el cambio de su orientación $\theta_i$. La actualización del ángulo de orientación dependía de la percepción sensorial del entorno del agente, en la cual dentro de su rango de visión revisa el promedio de las orientaciones de los otros agentes y calcula la diferencia de ángulos entre su orientación y la orientación promedio de los agentes vecinos. Esto está descrito en la siguiente ecuación:

$$p_i^t = \frac{1}{n_i} \sum_{|x_j - x_i| \leq r} v_j^t \tag{3}$$

donde $p_i^t$ es el promedio de orientación de los vecinos del agente $i$ en el instante $t$, $n_i$ es el total de agentes que están dentro del rango de visión del agente $i$, y $v_j^t$ es la velocidad del agente $j$ en el instante $t$ en donde al ser la rapidez constante la velocidad depende completamente de la orientación del agente en el instante $t$. Para calcular la diferencia angular entre el agente $i$ y el promedio de orientaciones de los vecinos, se hizo uso de la siguiente ecuación:

$$S_i^t = arccos\left(\frac{p_i^t \cdot v_i^t}{\| p_i^t \|}\right) \tag{4}$$

donde $S_i^t$ es la diferencia angular entre el agente $i$ y el promedio de orientaciones de sus vecinos en el instante $t$. A esta variable se le llamará de ahora en adelante el estado del agente $i$ en el instante $t$. Una vez que el estado del agente era obtenido, el agente cambiaba su orientación angular dependiendo del estado en el que se encontraba, es decir, tomaba una acción basándose en su estado actual:

$$v_i^{t+1} = a_i^t(s_i^t) \tag{5}$$

Así se puede ver que la velocidad del agente $i$ en el instante $t + 1$ dependía de la acción tomada en el instante $t$, que a su vez dependía del estado del agente en el tiempo $t$.

Dado que la acción del agente era rotar cierto ángulo, se definió un ángulo de rotación máximo $\theta_{max}$, ya que en cada paso de tiempo el agente estaba limitado en el ángulo que podía girar. De esta forma, en cada instante de tiempo el agente podía girar a lo más $\theta_{max}$

grados. De esta manera, se tiene que la acción tomada por el agente se puede escribir como:

$$a_i^t(s_i^t) = \begin{cases} s_i^t & si \quad |s_i^t| \leq \theta_{max} \\ \theta_{max} & si \quad s_i^t > \theta_{max} \\ -\theta_{max} & si \quad s_i^t < -\theta_{max} \end{cases} \tag{6}$$

Siguiendo estas ecuaciones, cada agente irá actualizando su posición espacial dependiendo de la orientación de sus vecinos.

Una vez implementada la simulación, con el objetivo de obtener el tiempo promedio de ticks que le toma al sistema obtener una orientación predeterminada utilizando el modelo de Vicsek, se realizaron 30 corridas de la simulación, y se obtuvo la media y la desviación estándar de los ticks.

## 2.2. Aprendizaje de un agente (modelo de QL)

Una vez que se obtuvo el movimiento colectivo en el enjambre siguiendo la metodología descrita en la sección 2.1, se procedió a agregar un agente nuevo al sistema. El agente nuevo no tenía implementado el mismo modelo que los demás agentes del enjambre; en su lugar, a este agente se le implementó un modelo de aprendizaje por refuerzo, cuyo objetivo fue maximizar la cantidad de vecinos que tenía el agente. Nótese que a diferencia del modelo de Vicsek, el objetivo de este modelo de aprendizaje no es alinearse con el resto de los agentes, sino, perder la menor cantidad de vecinos posible. Lo anterior se realizó con la finalidad de comprobar si el agente era capaz de aprender comportamientos que le permitieran tener un dinámica de movimiento similar a la del enjambre. Con este modelo de aprendizaje su buscó utilizar un esquema más realista en cuanto al movimiento colectivo observado en dinámicas naturales, ya que se han planteado muchos beneficios asociados al movimiento agrupado, como la evitación de depredadores (Milinski & Heller, 1978) y la alimentación colectiva (Pitcher et al., 1982).

Dado este esquema, el agente tenía las mismas variables de estado y la misma regla de actualización de posición que los agentes descritos en la sección anterior, sin embargo, las decisiones de rotación las tomaba usando el algoritmo de Q-learning (Watkins, 1989).

El agente tuvo que aprender una política que minimizara una recompensa negativa (se le conoce como política al acto de tomar una acción dependiendo del estado actual en el que se encuentre el agente). Para ello, se discretizó el espacio de estados en los que se podía encontrar el agente y el espacio de acciones que el agente podía realizar. El espacio de estados resultante se describe de la siguiente manera:

$$K_s = [-150, -120, -90, -80, -70, -60, -50, -40, -30, -20, -10, -3, -1, 0, 1, 3, 10, 20, 30, 40, 50, 60, 70, 80, 90, 120, 150] \quad (7)$$

donde:

$$K_s = \begin{cases} -150 & si \quad \Delta\theta \leq -150 \\ -120 & si \quad -150 < \Delta\theta \leq -120 \\ -90 & si \quad -120 < \Delta\theta \leq -90 \\ \vdots & \\ 150 & si \quad 120 < \Delta\theta \leq 150 \end{cases} \quad (8)$$

De esta forma, el agente solo se puede encontrar en 27 posibles estados dependiendo de la diferencia de orientación entre él y el promedio de orientaciones de sus vecinos. El espacio de acciones resultante se describe de la siguiente manera:

$$K_a = [30, 20, 10, 3, 1, 0, -1, -3, -10, -20, -30] \quad (9)$$

Así, el agente puede realizar 11 acciones posibles, es decir, puede decidir girar entre 11 posibles ángulos. De esta manera se puede obtener una matriz de estado-acción $K_a \times K_s$, la cual el agente utilizó para darle un valor a las acciones que tomará dependiendo del estado en el que se encuentre. A esta matriz se le conoce como tabla Q, y ayuda al agente a tomar acciones que minimicen una recompensa negativa. Todos los valores de la tabla Q se inicializan con un valor de cero, el cual se va actualizando utilizando la ecuación (1).

La forma en que se evalúa la política que toma el agente es mediante una penalización, en la que si pierde vecinos después de tomar cierta acción, obtiene una penalización de 1, y si mantiene la misma cantidad de vecinos o adquiere más, obtiene un valor de 0, es decir:

$$c_i^{t+1} = \begin{cases} 1 & si \quad n_i^{t+1} < n_i^t \\ 0, & para\ cualquier\ otro\ caso \end{cases} \tag{10}$$

donde $c^{t+1}$ es la penalización o recompensa al tiempo $t+1$, $n^{t+1}$ y $n^t$ son las cantidades de vecinos al tiempo $t+1$ y al tiempo $t$ respectivamente. De esta forma, una vez que el agente tomaba una política (estado $i$, acción $j$), se evaluaba que tan buena fue la elección mediante esta recompensa, y con eso se actualiza la entrada de la tabla Q utilizando la regla de aprendizaje de Q-learning (Watkins, 1989).

Para la implementación del modelo, a los parámetros de la ecuación (1) se les dieron los siguientes valores: $\alpha$=0.005, $\gamma$ =1 , donde $\alpha$, $\gamma$ son el índice de aprendizaje y el factor de descuento respectivamente.

Siguiendo esta regla de actualización para los valores de la tabla Q, el agente puede ir encontrando una política que minimiza el costo con el paso de las iteraciones. El uso de la tabla Q le permite al agente ver qué acción puede ser la mejor dependiendo del estado en el que se encuentre. La acción que toma el agente en cada paso de tiempo se seleccionó usando el método de epsilón-greedy:

$$a_i^t = \begin{cases} argmin\ Q_i(s_i^t, a'), & con\ probabilidad\ 1-\epsilon \\ una\ acción\ aleatoria, & con\ probabilidad\ \epsilon \end{cases} \tag{11}$$

El método epsilón-greedy consiste en seleccionar un valor de épsilon ($\epsilon$), en el cual se toma una acción al azar con probabilidad $\epsilon$, y con probabilidad de $1-\epsilon$ se toma la acción que se considera como la mejor. Este es un método de optimización que ayuda a no caer en mínimos o máximos locales, por lo que se favorece la exploración del espacio de acciones y se evita tomar siempre la misma acción. Para este caso se seleccionó un valor de $\epsilon$ igual a 0.01.

Siguiendo esta metodología, se modeló al nuevo agente para que interactuara con el resto de los agentes que tenían implementado el modelo de Vicsek, con el objetivo de vislumbrar si era capaz de aprender a moverse de manera colectiva con ellos.

Se realizaron 30 corridas de la simulación con el objetivo de obtener el tiempo promedio de ticks que le toma al agente con el modelo de QL alinearse con el resto de los agentes, y se obtuvo la media y la desviación estándar de los ticks.

Además, para cuantificar el aprendizaje de comportamientos el agente, se midió la tasa de pérdida de vecinos en cada paso de tiempo. La tasa de pérdida de vecinos al tiempo $t(p(t))$ se calcula como sigue:

$$p(t) = \begin{cases} 0 & si \quad n^t \geq n^{t-1} \\ \frac{n^{t-1} - n^t}{n^{t-1}} & si \quad n^t < n^{t-1} \end{cases} \tag{12}$$

Nótese que si $n^t \geq n^{t-1}$ entonces $p(t) = 0$, y si $n^t < n^{t-1}$, entonces:

$$p(t) = \frac{n^{t-1} - n^t}{n^{t-1}} = \frac{n^{t-1}}{n^{t-1}} - \frac{n^t}{n^{t-1}} = 1 - \frac{n^t}{n^{t-1}} \tag{13}$$

donde $n^t$ es la cantidad de vecinos al tiempo $t$. De aquí podemos ver que $n^{t-1} > 0 \; \forall \, t$ , ya que $n^t < n^{t-1}$ y $n^t \geq 0 \; \forall \, t$, pues no se puede tener una cantidad negativa de vecinos. Así que la ecuación (12) está bien definida.

Nótese que de $n^t < n^{t-1}$ se observa que $\frac{n^t}{n^{t-1}} < 1$, así que $0 < 1 - \frac{n^t}{n^{t-1}}$. Así, si $n^t < n^{t-1}$, entonces $p(t) > 0$. Se sabe que el valor mínimo que puede tomar $n^t$ es 0, por lo que si $n^t = 0$ y $n^{t-1} \neq 0$, $p(t) = 1$.

De esta forma, se observa que $0 \leq p(t) \leq 1$, el cual se puede interpretar como el porcentaje de vecinos perdidos en cada paso de tiempo. Para cada una de las 30 simulaciones, se midió la tasa de pérdida de vecinos del agente $(p(t))$.

Finalmente, una vez que el agente aprendió a moverse con el enjambre, se obtuvo la tabla Q con los valores de cada acción para cada estado en cada simulación, y se hizo un promedio de las tablas de cada una de las 30 simulaciones para ver qué política habían aprendido los agentes y compararla con la política del modelo de Vicsek.

## 2.3. Aprendizaje de enjambre (modelo de QL)

Dado que el objetivo del presente trabajo fue modelar el comportamiento colectivo de un enjambre utilizando aprendizaje por refuerzo a nivel individual en cada agente, se procedió a aplicar la metodología descrita en la sección 2.2 a cada agente. De esta forma, ningún agente en la simulación tenía implementado el modelo de Vicsek. Así, se tuvieron *n* agentes que interactuaban entre ellos e iban aprendiendo comportamientos de movimiento utilizando Q-Learning para encontrar la política que mejor funcionara, es decir, aquella que minimizara el costo. De esta forma se obtuvo un modelo en el cual emergió comportamiento colectivo a partir de interacciones de los agentes con ambiente, en donde cada agente tenía implementado el modelo de aprendizaje reforzado a nivel individual descrito en la sección 2.2.

Al igual que en las secciones anteriores, se realizaron 30 simulaciones para este caso, se midió el tiempo promedio que tardaron los agentes en obtener la misma orientación y así lograr un movimiento colectivo.

# 3. RESULTADOS

## 3.1. Modelo de movimiento colectivo (modelo de Vicsek)

A partir de la implementación del modelo de Vicsek en NetLogo, se obtuvo una simulación del movimiento colectivo de los agentes. Se pudo observar que todos los agentes logran obtener la misma orientación (Figura 3.1a), muy similar a lo que se puede observar en las parvadas de pájaros (Figura 3.1b).

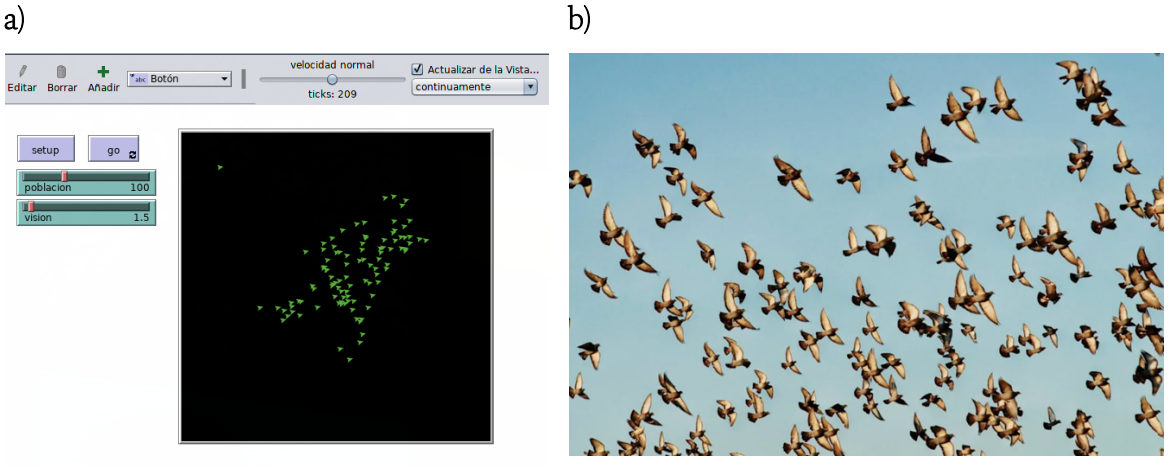


**Figura 3.1**. **(a)** Orientación de los agentes después de la simulación del modelo de Vicsek en NetLogo. **(b)** Imagen de la orientación de los individuos de una parvada real.

Como se mencionó anteriormente, para poder medir el tiempo promedio de ticks que le toma al sistema obtener una orientación predeterminada, se realizaron 30 corridas de la simulación.

El tiempo promedio que le tomó al enjambre alinearse con la misma orientación fue relativamente bajo (promedio = 142.72 ticks, $\sigma$ = 40.25 ticks). Esto se debe a que todos los agentes del sistema saben de antemano las reglas que deben seguirse para llegar a este tipo de comportamiento.

En la Figura 3.2 podemos visualizar la política que siguen los agentes cuando se utiliza el modelo de Vicsek; estos agentes ya tienen implementadas las reglas a seguir, por lo que ya conocen qué acción deben tomar dependiendo del estado en el que se encuentren. De esta forma se puede observar que si un agente tiene una diferencia de 120 grados con el resto de sus vecinos, entonces deberá girar 120 grados. Cuando cada

agente del sistema sigue estas reglas, se obtiene el comportamiento colectivo (Figura 3.3).

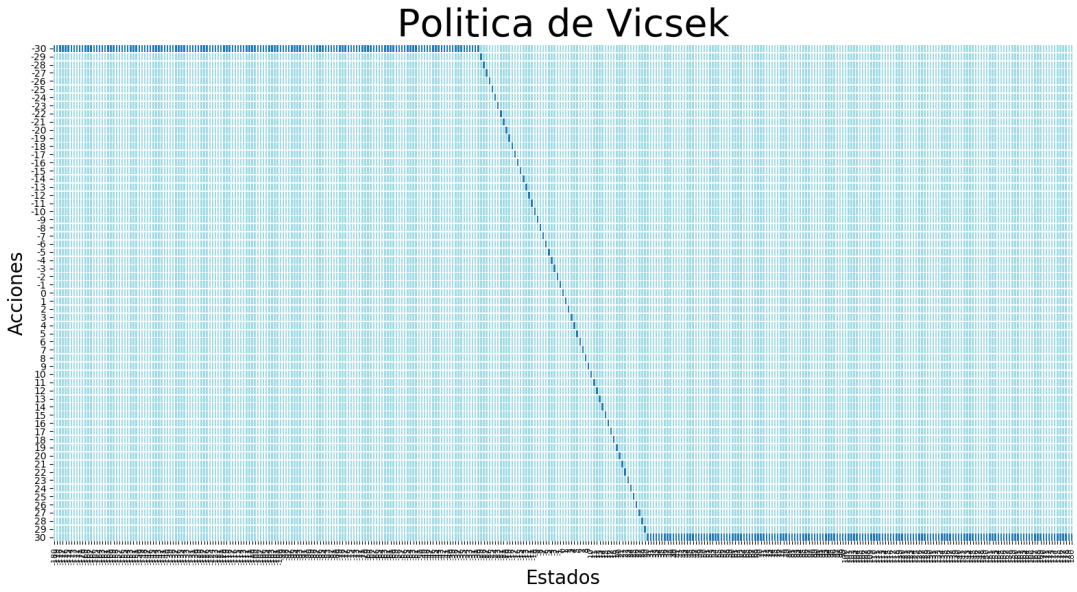


**Figura 3.2.** Tabla de estados y acciones que siguen los agentes utilizando el modelo de Vicsek.

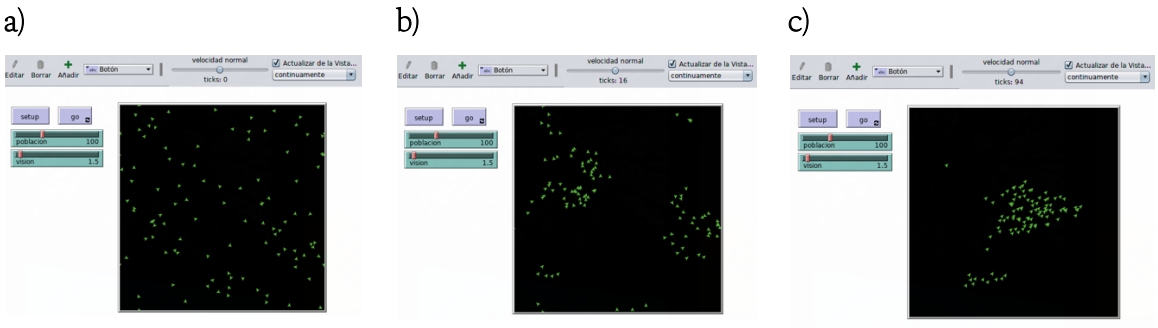


**Figura 3.3.** Comportamiento del enjambre en 3 tiempos distintos. **(a)** Tiempo inicial: se observa que los agentes están posicionados y orientados de manera aleatoria. **(b)** Tiempo medio: se observa que los agentes empiezan a formar cúmulos agregados orientados entre ellos. **(c)** Tiempo final: se observa que los agentes se organizaron con una orientación similar en cúmulos agrupados.

## 3.2. Aprendizaje de un agente (modelo de QL)

Una vez que el enjambre se orientó en la misma dirección, se introdujo a un agente nuevo que tenía el modelo de aprendizaje por refuerzo. Como se mencionó anteriormente, este agente contaba con las mismas variables de estado y la misma regla de actualización de posición que los agentes que tenían el modelo de Vicsek, sin embargo, las decisiones de rotación las tomaba usando el algoritmo de Q-learning (Watkins, 1989).

A partir de las simulaciones realizadas, se observó que el agente fue aprendiendo a orientarse igual que sus vecinos, y que exhibía un comportamiento similar al de los demás agentes (Figura 3.4).

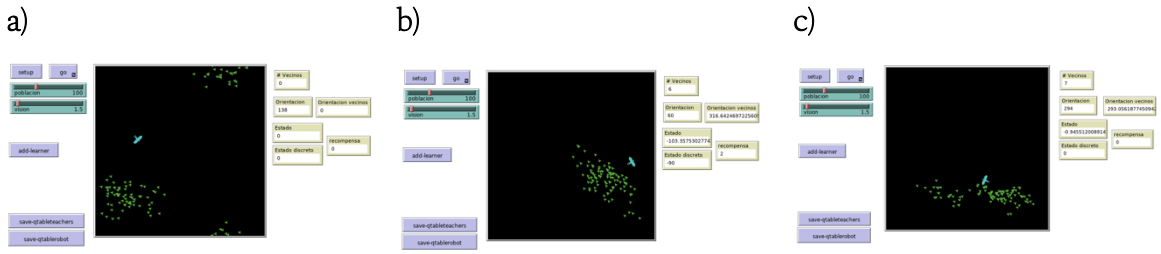


**Figura 3.4.** Comportamiento del agente con el modelo de QL en 3 tiempos distintos. **(a)** Tiempo inicial: se observa que el agentes está posicionado y orientado de manera aleatoria. **(b)** Tiempo medio: se observa que el agente empieza a interactuar con el enjambre, pero aún no ha aprendido comportamiento que le permitan orientarse con el grupo. **(c)** Tiempo final: se observa que el agente tiene una dinámica de movimiento muy similar a la del enjambre.

Asimismo, se midió el tiempo promedio que el agente tardaba en aprender una política que le permitiera tener un movimiento similar al del enjambre. Para este caso, el tiempo promedio fue considerablemente más alto que en el modelo de Vicsek, ya que fue de $9.814 \times 10^6$ ticks, y la desviación estándar fue de $1.103 \times 10^6$ ticks.

Como se mencionó en la sección de métodos, se midió la tasa de pérdida de vecinos $(p(t))$ del agente para cuantificar el aprendizaje de comportamientos del mismo. Esta tasa se fue midiendo en cada simulación, para poder visualizar su variación conforme el agente iba aprendiendo qué políticas eran mejores (es decir, qué políticas minimizaban su costo). La Figura 3.5 muestra el cambio en la tasa de pérdida de vecinos promedio a lo largo del tiempo. Puede notarse como al transcurrir el tiempo, el agente va disminuyendo su pérdida de vecinos, hasta quedarse estable en un valor de $p(t)$ que oscila alrededor de 0.1.

Una vez que el agente aprendió a moverse con el enjambre, se obtuvo la tabla Q con los valores de cada acción para cada estado en cada simulación, y se hizo un promedio de las tablas de cada una de las 30 simulaciones para ver qué política habían aprendido los agentes y compararla con la política del modelo de Vicsek. Puede observarse que el agente aprendió qué acciones tomar dependiendo del estado en el que se encuentra muy similares a los del modelo de Vicsek; en resumen, se puede observar que aprendió el

comportamiento general del modelo de Vicsek, pero hay algunos estados para los cuales aún no logró encontrar la mejor acción, como en el caso en el que se encuentra a –80 grados, para el cual la acción aprendida fue girar 20 grados, no obstante, la acción óptima sería girar –30 grados (Figura 3.6).

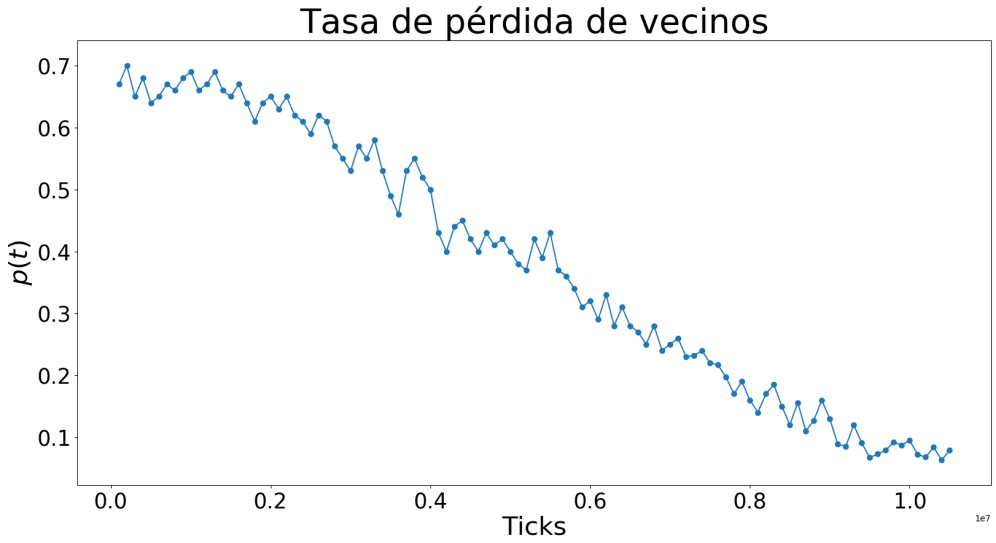


**Figura 3.5.** Cambio en la tasa de pérdida de vecinos con el paso de tiempo (ticks) a nivel agente. Para obtener la tasa de pérdida de vecinos promedio para cada tick, se promedió el valor de ésta en cada tick de las 30 simulaciones realizadas.

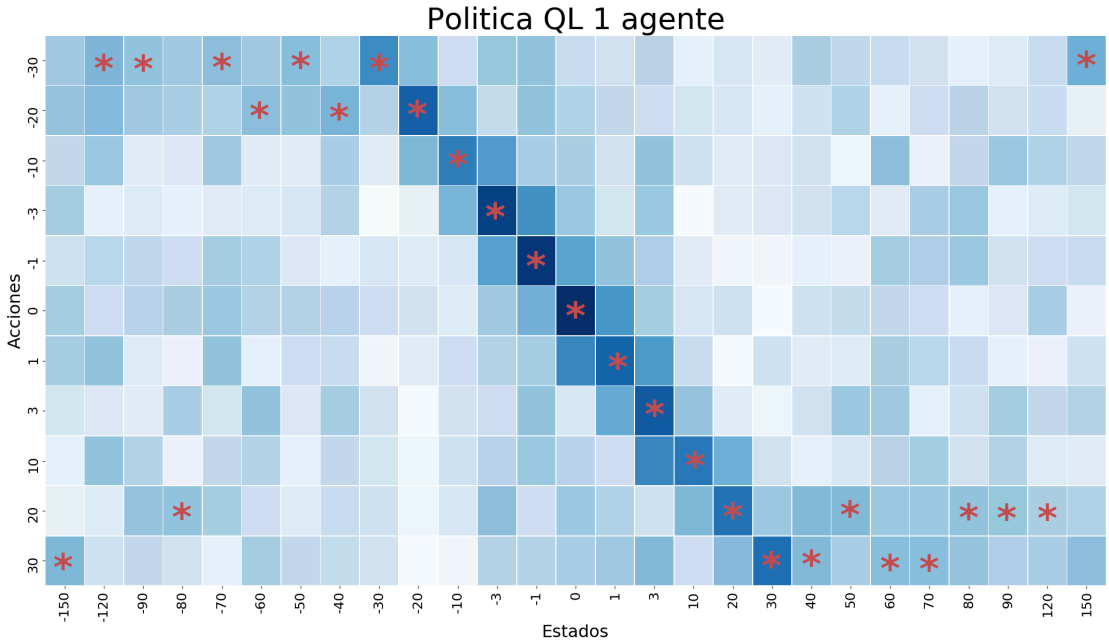


**Figura 3.6.** Tabla de estados y acciones que sigue el agente con el modelo de QL. Se puede observar que el agente aprendió qué acciones tomar dependiendo del estado en el que se encuentra muy similares a los del modelo de Vicsek. Los asteriscos indican cuál fue la acción aprendida por el agente como "mejor" opción posible dependiendo de su estado actual.

## 3.3. Aprendizaje de enjambre (modelo de QL)

Para la última parte, en la que ningún agente tenía el modelo de Vicsek implementado y en su lugar los 100 agentes tenían implementado el modelo de aprendizaje por refuerzo a nivel individual, se obtuvieron los siguientes resultados.

A partir del modelo basado en agentes implementado en NetLogo, se observó que el enjambre de agentes al que se le implementó el modelo de aprendizaje por refuerzo tuvo un comportamiento colectivo similar al del enjambre del modelo de Vicsek, una vez aprendidas las políticas que les resultaban más beneficiosas (Figura 3.7).

Para este caso, el tiempo promedio fue considerablemente más alto que en el modelo de Vicsek, ya que fue de $11.749 \times 10^6$ ticks, y la desviación estándar fue de $1.453 \times 10^6$ ticks.

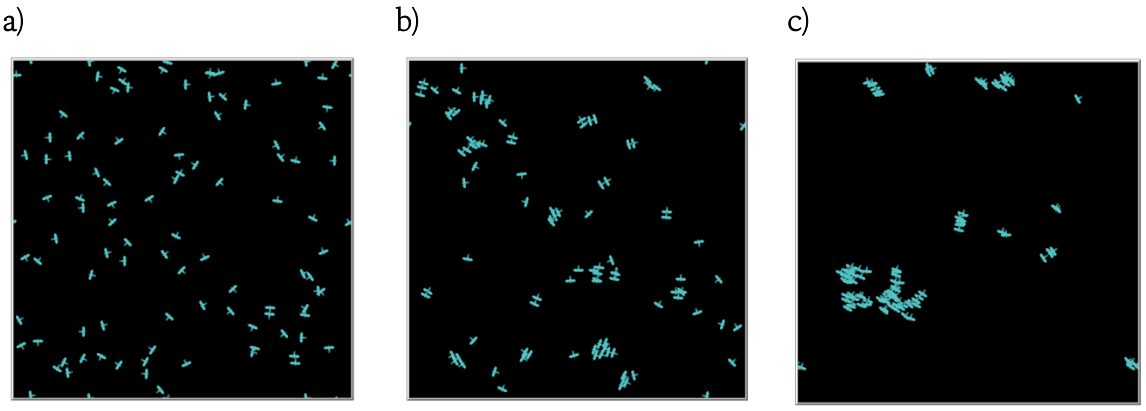


**Figura 3.7.** Comportamiento del enjambre con el modelo de QL en 3 tiempos distintos. (a) Tiempo inicial: se observa que los agentess están posicionados y orientados de manera aleatoria. (b) Tiempo medio: se observa que los agentes empiezan a interactuar entre sí; algunos agentes empiezan a aprender comportamientos que les permiten orientarse y formar cúmulos, mientras que otro siguen dispersos sin aprender comportamientos que les permitan orientarse con el grupo. (c) Tiempo final: se observa un comportamiento similar al movimiento colectivo a través de la formación de cúmulis de agentes con orientaciones iguales; no obstante, todavía hay agentes que no han aprendido comportamientos que les permitan formar parte del enjambre.

De igual manera, se utilizó la métrica de la tasa de pérdida de vecinos para ver cómo se comportaba el sistema con el paso del tiempo y evaluar si los agentes iban aprendiendo una política que les permitiera tener movimiento colectivo. En la Figura 3.8 se observa la variación de la tasa de pérdida de vecinos en cada tick. Puede notarse como al transcurrir el tiempo, los agentes en promedio van disminuyendo su tasa de pérdida de vecinos, hasta quedarse estable en un valor de $p(t)$ que oscila alrededor de 0.2. A diferencia del modelo de 1 agente con QL (Figura 3.5), en este caso se observa que los

agentes tardan más tiempo en llegar a un valor de $p(t)$ estable, y que éste valor es mayor que en el caso de 1 agente.

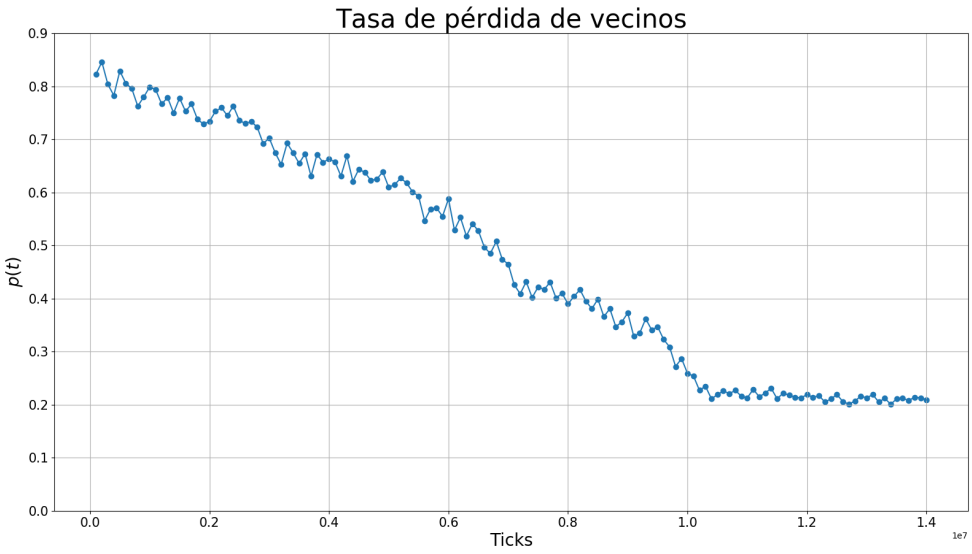


**Figura 3.8.** Cambio en la tasa de pérdida de vecinos con el paso de tiempo (ticks) a nivel sistema. Para obtener la tasa de pérdida de vecinos promedio para cada tick, se promedió el valor de ésta en cada tick de las 30 simulaciones realizadas.

Finalmente, una vez que los agentes desarrollaron movimiento colectivo, se promedió el valor de la tabla Q de todos los agentes y de todas las corridas de las simulaciones. Esto permitió visualizar qué acciones fueron las que los agentes consideraron mejores para cada estado en el que se encontraban. Los resultados se pueden observar en la Figura 3.9. A diferencia de la política de 1 agente con el modelo de QL (Figura 3.6), puede observarse que al promediar el aprendizaje de todos los agentes, el comportamiento es mucho más similar al del modelo de Vicsek. En este escenario, la acción idónea aprendida por los agentes para cada estado es la misma que en el modelo de Vicsek. Sin embargo, puede observarse que pese a que hay una acción predilecta evidente para cada estado, hay otras acciones que resultan también bastante buenas para tomar en cada estado, es decir, hay agentes que no aprendieron la mejor acción, pero aprendieron otras acciones que también pueden funcionarles para evitar la pérdida de vecinos (como en el caso del estado –1 grado, se observa que muchos agentes aprendieron que la mejor opción era girar –1 grado, pero otros aprendieron que girar –3 o 0 grados también resultan acciones adecuadas para no perder vecinos).

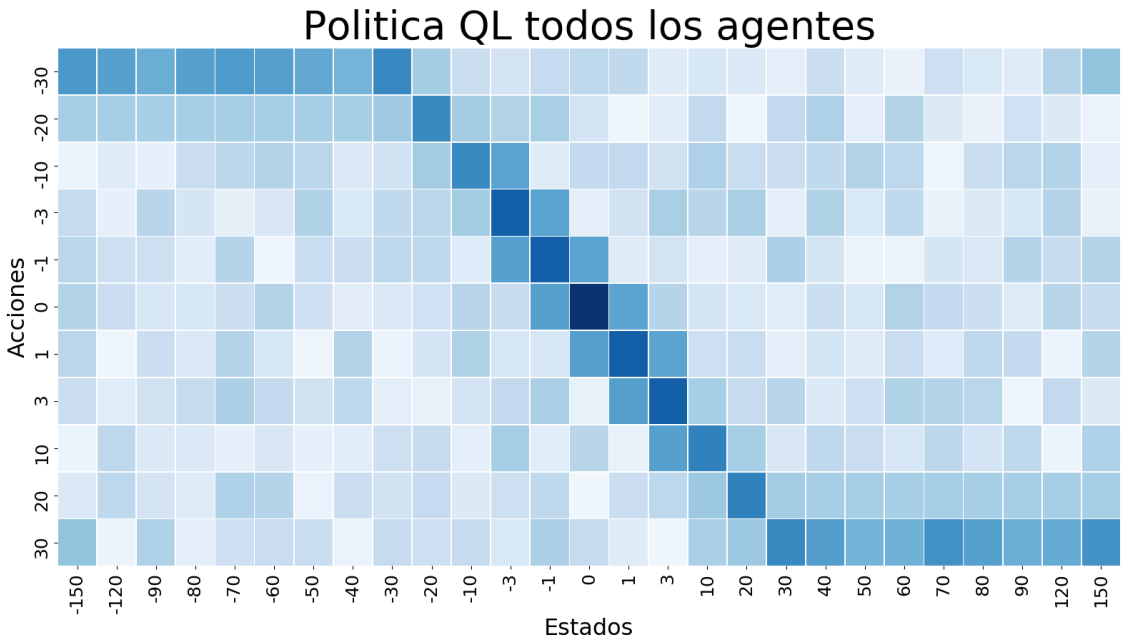


**Figura 3.9.** Promedio de las tablas Q de estados y acciones de los agentes con el modelo de QL. Se puede observar que en promedio los agentes aprendieron qué acciones tomar dependiendo del estado en el que se encuentran muy similares a los del modelo de Vicsek.

# 4. DISCUSIÓN

Los resultados obtenidos muestran que es posible que en un sistema de agentes emerja comportamiento colectivo derivado de la interacción de cada agente con su entorno, implementando aprendizaje reforzado a nivel individual.

Puede observarse que utilizar el enfoque que permite minimizar la pérdida de vecinos en el modelo de aprendizaje por refuerzo (Figura 3.9) resulta adecuado para llegar a un comportamiento bastante similar al que tiene el sistema de agentes al que se le implementó el modelo Vicsek (Figura 3.2). Es decir, minimizar la pérdida de vecinos como parámetro de toma de decisiones para los agentes permite la emergencia de comportamiento colectivo. Esto resulta en una estrategia para mantener la cohesión del grupo cuando los agentes solo pueden tener conocimiento de su entorno a nivel local. En ese caso, si los agentes quieren permanecer juntos y formar un enjambre, deben enfocarse en mantener la misma orientación que sus vecinos.

Los resultados obtenidos de los tres tipos de simulaciones (modelo de Vicsek, aprendizaje de un agente con QL, aprendizaje de todos los agentes con QL) muestran cómo es que el tiempo promedio que el enjambre tarda en obtener la misma orientación va aumentando conforme hay más agentes que no saben las reglas y deben ir aprendiéndolas.

Cuando todos los agentes saben qué comportamiento deben seguir (modelo de Vicsek) tardan 142.72 ticks en promedio en obtener la misma orientación y tener un movimiento colectivo. Sin embargo, cuando se agrega un agente que no conoce la política de Vicsek y que tiene implementado el modelo de QL, el tiempo que tarda en lograr orientarse en la misma dirección que los demás agentes es 100,000 veces mayor que en el modelo de Vicsek, ya que en este escenario, el agente debe ir aprendiendo las acciones que le permiten aumentar su recompensa (mantener el mayor número de vecinos en su rango de visión).

Para el último modelo, en el que todos los agentes tenían que aprender simultáneamente mediante el algoritmo de QL, el tiempo aumenta considerablemente respecto al modelo de Vicsek (142.72 ticks vs. $11.749 \times 10^6$ ticks) y respecto al modelo de un agente con el

modelo de QL y el resto con el modelo de Vicsek ($9.814 \times 10^6$ ticks vs. 142.72 ticks respectivamente). Estos resultados resultan congruentes con lo esperado, ya que en el primer caso solo se está midiendo el tiempo que tardan los agentes en orientarse, y en el tercer caso se mide el tiempo que tardan los agentes en aprender adicional al tiempo que tardan en orientarse, por lo cual se está comparando el tiempo de procesos distintos, lo cual no es lo más óptimo para evaluar el desempeño del modelo. Para poder tener una mejor comparación de estos dos casos, se guardaron los valores de la tabla Q una vez que los agentes habían aprendido a orientarse, y se corrieron 30 simulaciones con esos valores de la tabla Q. Se puede notar que el tiempo promedio fue más parecido al tiempo del modelo de Vicsek, ya que fue de 387.24 ticks, y la desviación estándar fue de 76.496 ticks.

Puede verse que el tiempo que tardan los agentes del modelo de QL en lograr el comportamiento colectivo es similar al tiempo que les toma a los agentes del modelo de Vicsek. Cabe señalar que sí hay un incremento en el tiempo que tardan los agentes en orientarse, debido a que no hay una política fija y a que los agentes no siempre toman la mejor acción ya que están usando el método épsilon-greedy. Esto genera que ocasionalmente los agentes tomen acciones al azar con una probabilidad dada, por lo que es natural que tarden más en orientarse en la misma dirección.

Este resultado es importante ya que los agentes con el modelo de QL logran llegar a un movimiento colectivo similar al observado en distintos enjambres, a pesar de que tarden un poco más de tiempo en lograrlo que con el modelo de Vicsek. Estos tiempos se podrían ir minimizando al explorar otros algoritmos de aprendizaje por refuerzo u otra métrica a optimizar, no obstante, el resultado es interesante en el sentido de que la lógica utilizada para la simulación tiene más sentido en un contexto natural. Es decir, es normal pensar que los individuos de un enjambre buscan agregarse en grupos y no perder a sus vecinos, ya que esto les provee de diversos beneficios, entre los que destacan la evitación de depredadores, la alimentación colectiva, el ahorro de energía y la eficiencia de viaje (Milinski & Heller, 1978; Pitcher et al., 1982).

Por ello, que el resultado de utilizar esta lógica para el modelaje haya sido similar a los modelos que utilizan enfoques más mecánicos se torna interesante para el estudio de las propiedades y mecanismos que rigen y subyacen a los fenómenos colectivos.

El enfoque y los resultados de este trabajo están siendo aplicados a un problema de conexión y comunicación en drones, el cual consiste en lo siguiente: se tiene una cantidad *n* de drones que deben cubir cierta área; en esa área, los drones tienen un rango de señal, el objetivo es que ellos doten de señal al area, pero cuando no haya suficiente abasto entonces se puedan mover los demas drones de forma que se pueda aumentar el campo de cobertura. Utilizando un enfoque de MBA y RL se podría modelar este problema de forma que los drones sean los agentes y tengan que aprender comportamientos que les permitan maximizar el rango de cobertura que tienen. Como se mencionó, este trabajo aún está en proceso y no ha concluido, pero muestra el potencial que tiene utilizar este enfoque y una de las muchas aplicaciones que puede tener este trabajo.

Otro trabajo que se tiene en puerta y que utiliza los resultados de este proyecto es el de aplicar este enfoque en robots y ya no en simulación. Los robots a los que se les aplicaria esto son los JetBots ya que al contar con una tarjeta NVIDIA® Jetson Nano™, se le puede implementar un algoritmo de RL a cada robot y asi darles una tarea a realizar, y que estos puedan ir aprendiendo comportamientos individuales para lograr el objetivo de forma colectiva.

Por esto, los resultados y el enfoque de este trabajo no solo nos ayudan a entender un poco mejor los fenomenos naturales, también pueden ser llevados a aplicaciones de distintas áreas, pero debe adaptarse el modelo de RL y el MBA para cada caso en específico.

# 5. CONCLUSIONES Y TRABAJO A FUTURO

El uso de modelos basados en agentes utilizando aprendizaje por refuerzo resulta una forma adecuada de describir la dinámica colectiva que presentan algunos grupos sociales de animales. Esto debido a que toman en cuenta la interacción entre los componentes del sistema y el ambiente, lo cual le permite a los agentes adaptarse a variaciones del medio e ir modificando sus acciones, dando lugar a comportamientos emergentes que no pueden ser descritos a partir únicamente de conductas individuales.

Asimismo, el uso de aprendizaje por refuerzo permite hacer más entendible el modelo, es decir, permite entender qué es lo que están buscando lograr los agentes de manera individual (tener más vecinos) y permite determinar qué acciones son las que les permiten cumplir ese objetivo. Además, el modelo de aprendizaje por refuerzo no tiene reglas fijas, a diferencia del modelo de Boids, en el que a pesar de que el sistema tiene el mismo comportamiento colectivo, no tiene razón alguna para que los agentes sigan las reglas fijas del modelo.

La principal aportación de este trabajo fue lograr ir un paso más a profundidad en la forma de modelar a los enjambres, es decir el enfoque de modelado que se utilizó, ya que de esta forma los agentes tienen objetivos individuales en lugar de seguir reglas fijas, lo que le lleva a aprender comportamientos individuales, de los cuales emerge el comportamiento colectivo. Esto es importante ya que puede ser utilizado como un enfoque para entender los fenómenos que suceden en la naturaleza, y darles una mejor explicación. También resulta muy importante en áreas como la robótica, en la que un enjambre de robots puede desenvolverse en un ambiente dinámico con un mejor desempeño que con la implementación de reglas fijas. Al utilizar este enfoque de modelado, los agentes se pueden adaptar a cambios en el ambiente y lograr cumplir su objetivo.

Como ya se mencionó en la sección anterior, en futuros trabajos, puede explorarse el comportamiento del sistema bajo diferentes algoritmos de aprendizaje por refuerzo implementados en los agentes, diferentes métricas a optimizar y diferentes ambientes en los que se desenvuelven.  Por ejemplo, se podría agregar algún agente en el sistema

que actúe como depredador, y que los agentes busquen optimizar la probabilidad de sobrevivir o la energía que consumen al moverse en el ambiente. Sería interesante ver si un modelo con estas consideraciones también pudiera llegar a un comportamiento colectivo del sistema, ya que esto nos podría permitir entender mejor lo que sucede en los sistemas reales, y podría brindarnos una idea de porqué varios grupos de animales se comportan de la forma en la que lo hacen. Otro posible trabajo a futuro podría ser la implementación de un modelo similar a un conjunto de robots que tengan una tarea a resolver, y evaluar si las políticas que aprendan los llevan a interactuar colectivamente porque les resulte más beneficioso. En caso de que esto suceda, sería interesante determinar qué parámetros o configuraciones del ambiente ocasionan que estos sistemas presenten este tipo de comportamientos colectivos.

# 6. REFERENCIAS BIBLIOGRÁFICAS


Azuma, S. I., Yoshimura, R., & Sugie, T. (2013). Broadcast control of multi-agent systems. Automatica, 49(8), 2307-2316

Bak, P. (2013). How nature works: the science of self-organized criticality. Springer Science & Business Media.

Baker, G. L., & Gollub, J. P. (1996). Chaotic dynamics: an introduction. Cambridge university press.

Baldassarre, G., Nolfi, S., & Parisi, D. (2003). Evolving mobile robots able to display collective behaviors. Artificial life, 9(3), 255-267.

Ball, P. (1999). The Self-Made Tapestry: Pattern Formation in Nature. Oxford University Press.

Bedau, M. A. (2003). Artificial life: organization, adaptation and complexity from the bottom up. Trends in cognitive sciences, 7(11), 505-512.

Beni, G. (2004). From swarm intelligence to swarm robotics. In International Workshop on Swarm Robotics (pp. 1-9). Springer, Berlin, Heidelberg.

Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. Proceedings of the national academy of sciences, 99, 7280-7287.

Bonabeau, E., Dorigo, M., & Theraulaz, G. (2000). Inspiration for optimization from social insect behaviour. Nature, 406(6791), 39-42.

Bonabeau, E., Theraulaz, G., Deneubourg, J. L., Aron, S., & Camazine, S. (1997). Self-organization in social insects. Trends in ecology & evolution, 12(5), 188-193.

Bonabeau, E., Theraulaz, G., Dorigo, M., Theraulaz, G., & Marco, D. D. R. D. F. (1999). Swarm intelligence: from natural to artificial systems (No. 1). Oxford university press.

Broer, H. W., & Takens, F. (2011). Dynamical systems and chaos (Vol. 172, pp. 133-133). New York: Springer.

Camazine, S., Deneubourg, J. L., Franks, N. R., Sneyd, J., Theraula, G., & Bonabeau, E. (2020). Self-organization in biological systems. In Self-Organization in Biological Systems. Princeton university press.

Cavagna, A., Giardina, I., Grigera, T. S., Jelic, A., Levine, D., Ramaswamy, S., & Viale, M. (2015). Silent flocks: constraints on signal propagation across biological groups. Physical review letters, 114(21), 218101.

Cohen, I. R., & Harel, D. (2007). Explaining a complex living system: dynamics, multi-scaling and emergence. Journal of the royal society interface, 4(13), 175-182.

Couzin, I. D. (2009). Collective cognition in animal groups. Trends in cognitive sciences, 13(1), 36-43.

Couzin, I. D., & Krause, J. (2003). Self-organization and collective behavior in vertebrates. Advances in the Study of Behavior, 32(1), 10-1016.

Das, S., Abraham, A., & Konar, A. (2008). Swarm intelligence algorithms in bioinformatics. In Computational Intelligence in Bioinformatics (pp. 113-147). Springer, Berlin, Heidelberg.

Del Valle, Y., Venayagamoorthy, G. K., Mohagheghi, S., Hernandez, J. C., & Harley, R. G. (2008). Particle swarm optimization: basic concepts, variants and applications in power systems. IEEE Transactions on evolutionary computation, 12(2), 171-195.

Devaney, R. L. (2018). A first course in chaotic dynamical systems: theory and experiment. CRC Press.

Dobson, S., Hutchison, D., Mauthe, A., Schaeffer-Filho, A., Smith, P., & Sterbenz, J. P. (2019). Self-organization and resilience for networked systems: Design principles and open research issues. Proceedings of the IEEE, 107(4), 819-834.

Domenico, M., Brockmann, D., Camargo, C., Gershenson, C., Goldsmith, D., Jeschonnek, S., Kay, L., Nichele, S., Nicolás, J.R., Schmickl, T., Stella, M., Brandoff, J., Martínez, A.J., Sayama, H. (2019). Complexity Explained.

Duan, H. & Yu, X. (2007). Progresses and challenges of ant colony optimization-based evolvable hardware. En: 2007 IEEE Workshop on Evolvable and Adaptive Hardware (WEAH2007) (pp. 67-71). IEEE.

Ecoffet, P., André, J. B., & Bredeche, N. (2020, July). Learning to Cooperate in a Socially Optimal Way in Swarm Robotics. En: ALIFE 2020: The 2020 Conference on Artificial Life (pp. 251-259). MIT Press.

Faassen, E. J., Veraart, A. J., Van Nes, E. H., Dakos, V., Lürling, M., & Scheffer, M. (2015). Hysteresis in an experimental phytoplankton population. Oikos, 124(12), 1617-1623.

Fleischer, M. (2005). Foundations of swarm intelligence: From principles to practice. arXiv preprint nlin/0502003.

Giardina, I. (2008). Collective behavior in animal groups: theoretical models and empirical studies. HFSP journal, 2(4), 205-219.

Goldstein, J. (2011). Emergence in complex systems. The sage handbook of complexity and management, 65-78.

Green, D. G., Sadedin, S., & Leishman, T. G. (2008). Self-Organization. Encyclopedia of Ecology, 3195–3203.

Helbing, D. (2012). Agent-based modeling. In Social self-organization (pp. 25-70). Springer, Berlin, Heidelberg.

Holland, J.H. (1992). Adaptation in Natural and Artificial Systems. MIT press.

Hüttenrauch, M., Adrian, S., & Neumann, G. (2019). Deep reinforcement learning for swarm systems. Journal of Machine Learning Research, 20(54), 1-31.

Janssen, A. B., Teurlincx, S., & Mooij, W. M. (2015). Research summary: Alternative stable states in large shallow lakes. Lake Scientist.

Johnson, C., Venayagamoorthy, G. K. & Palangpour, P. (2008). Hardware implementations of Swarming Intelligence—a survey. En: 2008 IEEE Swarm Intelligence Symposium (pp. 1-9). IEEE.

Kabiraj, L. I. P. I. K. A. (2012). Intermittency and route to chaos in thermoacoustic oscillations (Doctoral dissertation, Ph. D. thesis (Indian Institute of Technology, Madras, 2012)).

Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement learning: A survey. Journal of artificial intelligence research, 4, 237-285.

Krasnosel'skii, M. A., & Pokrovskii, A. V. (2012). Systems with hysteresis. Springer Science & Business Media.

Kuremoto, T., Obayashi, M., Kobayashi, K., Adachi, H., & Yoneda, K. (2008). A reinforcement learning system for swarm behaviors. En: 2008 IEEE International

Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence) (pp. 3711-3716). IEEE.

Lan, X., Liu, Y., & Zhao, Z. (2020). Cooperative control for swarming systems based on reinforcement learning in an unknown dynamic environment. Neurocomputing, 410, 410-418.

Lichtenstein, B. B., & Plowman, D. A. (2009). The leadership of emergence: A complex systems leadership theory of emergence at successive organizational levels.

Lim, C. P., & Jain, L. C. (2009). Advances in swarm intelligence. In innovations in swarm intelligence (pp. 1-7). Springer, Berlin, Heidelberg.

Litimi, H., BenSaida, A., Belkacem, L., & Abdallah, O. (2018). Chaotic behavior in financial market volatility. Journal of Risk, Forthcoming.

Liu, Y., & Passino, K. M. (2000). Swarm intelligence: Literature overview. Department of electrical engineering, the Ohio State University.

Manson, S. M., Sun, S., & Bonsal, D. (2012). Agent-based modeling and complexity. En: Agent-based models of geographical systems (pp. 125-139). Springer, Dordrecht.

Martinoli, A. (1999). Swarm intelligence in autonomous Collective robotics: from tools to the analysis and synthesis of distributed control strategies. PhD Thesis. Lausanne: Computer Science Department, Ecole Polytechnique Fédérale de Lausanne.

Mataric, M. J. (1992). Designing emergent behaviors: From local interactions to collective intelligence. En: J. A. Meyer, H. L. Roitblat, and S. W. Wilson (Eds.), From animals to animats 2: Proceedings of the Second International Conference on Simulation of Adaptive Behavior (pp. 432–441). Cambridge, MA: MIT Press.

May, R. M. (1977). Thresholds and breakpoints in ecosystems with a multiplicity of stable states. Nature, 269(5628), 471-477.

Milinski, M., & Heller, R. (1978). Influence of a predator on the optimal foraging behaviour of sticklebacks (Gasterosteus aculeatus L.). Nature, 275(5681), 642-644.

Mitchell, M. (2009). Complexity: A guided tour. Oxford University Press.

Olfati-Saber, R. (2006). Flocking for multi-agent dynamic systems: Algorithms and theory. IEEE Transactions on automatic control, 51(3), 401-420.

Parrish, J. K., & Edelstein-Keshet, L. (1999). Complexity, pattern, and evolutionary trade-offs in animal aggregation. Science, 284(5411), 99-101.

Phelan, S. E. (2001). What is complexity science, really?. Emergence, A Journal of Complexity Issues in Organizations and Management, 3(1), 120-136.

Pitcher, T. J., Magurran, A. E., & Winfield, I. J. (1982). Fish in larger shoals find food faster. Behavioral Ecology and Sociobiology, 10(2), 149-151.

Ranganathan, A., & Kira, Z. (2003). Self-organization in artificial intelligence and the brain. College of Computing, Georgia Institute of Technology.

Rasband, S. N. (2015). Chaotic dynamics of nonlinear systems. Courier Dover Publications.

Ren, W., & Cao, Y. (2011). Distributed coordination of multi-agent networks: emergent problems, models, and issues (Vol. 1). London: Springer.

Reynolds, C. W. (1987, August). Flocks, herds and schools: A distributed behavioral model. In Proceedings of the 14th annual conference on Computer graphics and interactive techniques (pp. 25-34).

Rivkin, J. W., & Siggelkow, N. (2007). Patterned interactions in complex systems: Implications for exploration. Management science, 53(7), 1068-1085.

Sayama, H. (2015). Introduction to the Modeling and Analysis of Complex Systems. Open SUNY Textbooks.

Scheffer, M. (1989). Alternative stable states in eutrophic, shallow freshwater systems: a minimal model. Hydrobiological Bulletin, 23(1), 73-83.

Scheffer, M., & Carpenter, S. R. (2003). Catastrophic regime shifts in ecosystems: linking theory to observation. Trends in ecology & evolution, 18(12), 648-656.

Serra, R., & Zanarini, G. (2013). Complex systems and cognitive processes. Springer Science & Business Media.

Siegenfeld, A. F., & Bar-Yam, Y. (2020). An introduction to complex systems science and its applications. Complexity, 2020.

Stonier, R. J., & Yu, X. H. (Eds.). (1994). Complex systems: mechanism of adaptation. IOS Press.

Strogatz, S. (1994). Nonlinear Dynamics and Chaos. CRC Press.

Sumpter, D. J. (2010). Collective animal behavior. In Collective Animal Behavior. Princeton University Press.

Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction. MIT press.

Theraulaz, G., & Bonabeau, E. (1995). Coordination in Distributed Building. Science, 269, 686-688.

Touma, J. R., Shreim, A., & Klushin, L. I. (2010). Self-organization in two-dimensional swarms. Physical Review E, 81(6), 066106.

Vicsek, T., Czirók, A., Ben-Jacob, E., Cohen, I., & Shochet, O. (1995). Novel type of phase transition in a system of self-driven particles. Physical review letters, 75(6), 1226.

Waldrop, M. M. (1993). Complexity: The emerging science at the edge of order and chaos. Simon and Schuster.

Ward, C. R., Gobet, F., & Kendall, G. (2001). Evolving collective behavior in an artificial ecology. Artificial life, 7(2), 191-209.

Watkins, C. (1989). Learning from delayed rewards. PhD Thesis, University of Cambridge, England.

Watkins, C. J., & Dayan, P. (1992). Q-learning. Machine learning, 8(3), 279-292.

Werner, G. M., & Dyer, M. G. (1992). Evolution of herding behavior in artificial animals. En: J. A. Meyer, H. L. Roitblat, and S. W. Wilson (Eds.), From animals to animats 2: Proceedings of the Second International Conference on Simulation of Adaptive Behavior (pp. 393 399). Cambridge, MA: MIT Press.

Wilensky, U. (1999). NetLogo. http://ccl.northwestern.edu/netlogo/. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL.

Wimsatt, W. C. (1972, January). Complexity and organization. In PSA: Proceedings of the biennial meeting of the Philosophy of Science Association (Vol. 1972, pp. 67-86). D. Reidel Publishing.

Wolfram, S. (2002). A new kind of science (Vol. 5, p. 130). Champaign: Wolfram media.

Yackinous, W. S. (2015). Understanding complex ecosystem dynamics: A systems and engineering perspective. Academic Press.

Young, I. M., & Crawford, J. W. (2004). Interactions and self-organization in the soil-microbe complex. Science, 304(5677), 1634-1637.

Zhang, W. (2015). Selforganizology: The Science of Self-Organization. World Scientific.

Zhu, B., Xie, L. H., & Han, D. (2017) A survey on recent progress in control of swarm systems. Sci China Inf, 60(070201), 1–24.