Mean Field Behaviour of Collaborative Multi-Agent Foragers

Daniel Jarne Ornia*, *Student Member, IEEE,* Pedro J. Zufiria†, *Senior Member, IEEE,* Manuel Mazo, Jr.*, *Senior Member, IEEE*

***Abstract*—Collaborative multi-agent robotic systems where agents coordinate by modifying a shared environment often result in undesired dynamical couplings that complicate the analysis and experiments when solving a specific problem or task. Simultaneously, biologically-inspired robotics rely on simplifying agents and increasing their number to obtain more efficient solutions to such problems, drawing similarities with natural processes. In this work we focus on the problem of a biologically-inspired multi-agent system solving collaborative foraging. We show how mean field techniques can be used to re-formulate such a stochastic multi-agent problem into a deterministic autonomous system. This de-couples agent dynamics, enabling the computation of limit behaviours and the analysis of optimality guarantees. Furthermore, we analyse how having finite number of agents affects the performance when compared to the mean field limit and we discuss the implications of such limit approximations in this multi-agent system, which have impact on more general collaborative stochastic problems.**

***Index Terms*—Agent-Based Systems, Learning and Adaptive Systems, Swarms, Mean Field Models.**

I. Introduction

Smaller processors and faster communications are pushing towards larger multi-agent systems with simple agents for solving large complex problems in a decentralised fashion. Be it large groups of autonomous cars driving in an urban setting or groups of nano-agents used in biomedical applications, there is a drive to increase the amount of collaborative agents in such settings, for either necessity (as in the case of vehicle traffic) or efficiency (in problems where more agents translate in better solutions). In the past two decades there has been growing interest in biologically inspired methods for solving decentralised coordination problems for large groups of simple agents. Inspiration is often drawn from the behaviour of ants, birds, bees or fish, for example [1]–[4]. These biological systems seem to have developed inherent robustness towards problems such as individual agent errors, malfunction or communication disruptions. Furthermore, some of them have evolved to be extremely resource-efficient, that being in time, energy or information transmission.

We focus our attention in this paper on a particular subclass of bio-inspired multi-agent stochastic coordination problems: foraging. Foraging is the problem of locating an unknown target, e.g. a source of food, in an unknown environment, and exploiting the shortest path to such target from a given initial location, e.g. a nest, with the goal of depleting the food source as fast as possible. For an extensive set of stochastic multi-agent methods the foraging problem has served as both a benchmark but also a study subject on itself, given the combined nature of exploration plus optimization that the problem presents [5]. Naturally, many of the biological systems capable of solving foraging problems present some (degree of) de-centralised behaviour. Ants, for example, communicate with each-other only by depositing pheromones on the environment, and achieve global coordination by combining the individual contributions of all members of the swarm without the need of centralised instructions. The mechanism of communicating indirectly through enviromental marking is known as *stigmergy*.

Ants and bees make use of stigmergy methods to coordinate with other members of the swarm [6], [7] to solve specific tasks, for example foraging for food [8], [9]. This kind of cooperative behaviour has been modelled for social insects such as ants [10]–[12], but also bees [13], [14] in more general frameworks (see as well the work of Resnick [15] for an extensive analysis and application of such behaviours).

Ant-inspired heuristics have been widely used to solve foraging problems in a distributed fashion, sparking a whole branch of stochastic optimization algorithms: Ant Colony Optimization [16], [17]. Ant-inspired swarm coordination has also been applied to foraging problems in distributed robotic systems. Authors in [18] propose a stochastic ant-inspired approach to distribute swarm agents among different target regions. In [19] the authors present some early experiments on how robots can lay and follow pheromones to explore a space and collect targets, and [20]–[24] have presented similar robotic systems, either by using a *digital* pheromone field [22]–[24], using real chemicals [21], or fluorescent floors [20] (see also [25]–[27] among others). Despite the many models proposed in entomology, and the implementations on robotic systems, little is known about the convergence guarantees of

This work was partially supported by the ERC Starting Grant SENTIENT #755953, and the Spanish Ministry of Science and Innovation, grant PID2020-112502RB / AEI / 10.13039/501100011033.

*Delft Center for Systems and Control, Delft University of Technology, Delft, 2628 CD, The Netherlands. *d.jarneornia@tudelft.nl, m.mazo@tudelft.nl*

‡Departamento de Matemática Aplicada a las TIC, Information Processing and Telecommunications Center, ETSI Telecomunicación, Universidad Politécnica de Madrid, Avda. Complutense 30, 28040 Madrid, Spain. *pedro.zufiria@upm.es*

such systems. The main goal of this paper is to investigate the convergence properties of a simple version of a stigmergy-based solution to the foraging problem.

Drawing a parallelism between the pheromones of ant swarms and Q-values [28] one can formulate the dynamics of a stigmergy-based system as a problem that resembles traditional reinforcement learning (RL) approaches. Agents explore an environment with a set of actions to choose from, and reward (deposit pheromones) their current state (spatial location) depending on the set goal. In the works of Monekosso [29] a first approach was taken to mix traditional Q-learning and pheromone based interaction in foraging swarms. In [30] a variation of such utility function learning approach is presented, where the swarm uses two kinds of pheromones to distinguish between the food search and the nest search.

Nevertheless, complications arise when trying to apply $Q-$learning related strategies to study the convergence of utility-based foraging swarms. In these stigmergy-based foraging problems, solutions to the iterative utility values are coupled to the agent trajectories. This prevents us from using these sort of stochastic dynamic programming techniques to study the *trajectories* of the agents in such learning stigmergy-based swarms. To address this problem, one can rely on the work of [31], [32] about convergence of stochastic sequences of probability transition matrices.

Looking into stochastic systems of interacting agents, in [33] the authors propose decentralised stochastic controllers to allocate tasks in an interacting multi-robot system given a desired target distribution. To get rid of the stochasticity, one can look at the limiting behaviour of such systems when the number of interacting agents is taken to infinity, in what are known as *mean field models*. Mean field models have been extensively used in fluid mechanics and particle physics, and more recently in game theory and control [34], [35]. In recent years mean field formulations of large multi-agent systems or swarms have gained increased popularity [36]–[38] (see also an extensive survey in [39]) as these models abstract away the stochasticity in systems where the number of interacting agents becomes very large.

The main goals of this work are twofold:

1) First, to approximate a multi agent stochastic system for a foraging problem as a mean-field non-stochastic process. Additionally, to provide intuition on the role of the different parameters in a large multi-agent stigmergy swarm.
2) Second, to derive convergence guarantees that can be attributed to the mean field model of a stigmergy swarm, and provide insight on the resulting shape of the stationary solutions, both for the agent trajectories and the pheromone field.

II. Preliminaries

A. Notation

A set whose elements depend on a parameter is indicated as $\mathcal{S}(\cdot)$. Sequences are represented as $\{A(t)\} \equiv A_t \equiv \{A(0), A(1), ..., A(t)\}$, and the union of two different sequences is computed set-wise: $A_1 \cup A_2 := \{i : i \in A_1 \vee i \in A_2\}$. We consider only discrete time systems, i.e. $t \in \mathbb{N}_0$. Unless stated otherwise, upper case letters are used for matrices ($B \in \mathbb{R}^{n\times n}$) and (bold) lower case letters for (vectors) scalars ($\mathbf{b} \in \mathbb{R}^n$, $b \in \mathbb{R}$). We use superscripts to distinguish between related vectors, and subscripts to indicate entries in a vector. That is, $\mathbf{a}^1_k$ is the $k-$th entry of the vector $\mathbf{a}^1$. For two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$, we say $\mathbf{a} \geq (\leq) \mathbf{b}$ iff $\mathbf{a}_i \geq (\leq) \mathbf{b}_i \ \forall i$. We use $|\cdot|$ for the cardinality of a set, and $\|\cdot\|_k$ for the k-th norm of a vector or the $k-$th induced norm of a matrix. We define the set of all probability vectors of size n as $\mathbb{P}^n := \{\mathbf{v} \in [0,1]^n : \sum_{i=1}^n \mathbf{v}_i = 1\}$. Vectors $\mathbf{1}^n$, $\mathbf{0}^n$ are the one and zero vectors of size n, respectively. The function $\mathrm{sgn}(\cdot)$ is the sign operator, with $\mathrm{sgn}(\mathbf{0}) = \mathbf{0}$.

We say a function $f : \mathbb{R}_+ \to \mathbb{R}_+$ is in the class of functions $\mathcal{K}$ if f is continuous, monotonically increasing and $f(0) = 0$. We say a function $f : \mathbb{R}_+ \to \mathbb{R}_+$ is in class $\mathcal{K}_\infty$ if $f(\cdot) \in \mathcal{K}$ and $\lim_{a\to\infty} f(a) = \infty$.

A function assigning to each instant of time a value on each edge of a graph can be written as a matrix, and the subscript indicates both edges and entries in the image of the function. That is, let $|\mathcal{V}|$ be the number of vertices in a graph, and $f : \mathbb{N} \to \mathbb{R}^{|\mathcal{V}|\times|\mathcal{V}|}$. Then, $f_{ij}(k)$ is the i,j-th entry in the image $f(k)$.

When talking about stochastic processes, we use Ω as the set of outcomes in a probability space, $\mathcal{F}$ as the measurable algebra (set) of events, and P as a probability function $P : \mathcal{F} \to [0,1]$. We use $E[\cdot]$ and $\mathrm{Var}[\cdot]$ for the expected value and the variance of a random variable. We say a result holds *almost surely* (*a.s.*) when it holds with probability 1. When two (or more) random variables follow the same probability distribution and are independent from each other we use *independent and identically distributed* (*i.i.d.*).

B. Weighted Graphs

In this work we discretise geometrical (bi-dimensional) spaces using connected planar graphs.

Definition 1. *We define a vertex weighted graph with time varying weights $\mathcal{G} := (\mathcal{V}, \mathcal{E}, \mathbf{w}(t))$ as a tuple including a vertex set $\mathcal{V}$, edge set $\mathcal{E}$ and weights $\mathbf{w} : \mathbb{N}_0 \to \mathbb{R}^{|\mathcal{V}|}_{\geq 0}$, where each value $\mathbf{w}_i(t)$ is the weight assigned to vertex $i \in \mathcal{V}$ at time t. Furthermore, the graph is connected if for every pair $i \neq j \in \mathcal{V}$ there exists a set of edges $\{(iu_1), (u_1u_2), ..., (u_nj)\} \subseteq \mathcal{E}$ that connects i and j.*

We refer to an edge connecting i to j as $(ij) \equiv e \in \mathcal{E}$. Additionally, the graph is undirected if $(ij) \in \mathcal{E} \iff (ji) \in \mathcal{E}$.

The adjacency matrix $A \in \mathbb{R}^{|\mathcal{V}|\times|\mathcal{V}|}$ and (out)weight matrix $W : \mathbb{N} \to \mathbb{R}^{|\mathcal{V}|\times|\mathcal{V}|}$ are

$$A_{ij} \coloneqq \begin{cases} 1 & \forall (ij) \in \mathcal{E}, \\ 0 & \text{else.} \end{cases}, W_{ij}(t) \coloneqq \begin{cases} \mathbf{w}_j(t) & \forall (ij) \in \mathcal{E}, \\ 0 & \text{else.} \end{cases}$$

Note that the transition weight for transition $i \to j$ is always the out-going weight $\mathbf{w}_j(t)$. For simplicity in expressions, we use the functions $D : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ and $V : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ such that $D_{ii}(B) = \sum_j B_{ij}$, $D_{ij}(B) = 0 \; \forall i \neq j$, $V(B) = \operatorname{diag}(\max_k B_{ik})$. That is, $D(B)$ is a diagonal matrix of the row sums of B, and $V(B)$ is a diagonal matrix containing the maximum value of every row of B in the diagonal terms.

Definition 2. *[40] A path $p_{ij} = \mathcal{V}' \subseteq \mathcal{V}$ in $\mathcal{G}$ is any ordered subset of vertices satisfying*

$$\mathcal{V}' = \{i, k, l, ..., z, j\} : \; (ik), (kl), ..., (zj) \in \mathcal{E},$$

where no vertex appears twice. An i-cycle is then a path p_{ii} starting and ending in the same vertex $i \in \mathcal{V}$. We refer to $\{p_{ij}^k\}$ as the set of all paths connecting i, j.

We make use of the minimum distance between two vertices $\delta : \mathcal{V}^2 \to \mathbb{N}_0^+$, $\delta(i, j) := \min_k\{|p_{ij}^k|\}$, defined only for connected pairs, and the set of minimum length paths between two vertices,

$$\pi_{ij} := \{p_{ij}^* \in \{p_{ij}^k\} : |p_{ij}^*| = \delta(i, j)\}.$$

The diameter of the graph is $\delta^* := \max\{\delta(i, j)\} \quad \forall i, j \in \mathcal{V}$. At last, it is useful to define the set of vertices in all minimum length paths as $\cup p_{ij}^* := \{v \in p_{ij}^* : p_{ij}^* \in \pi_{ij}\}$.

C. Stochastic Processes and Limit Theorems

The following definitions and theorems related to stochastic processes and random variables are used throughout this work.

Definition 3 (Almost Sure Convergence [41])**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space equipped with a σ-algebra of measurable subsets of Ω, with $\omega \in \Omega$ being any outcome. We say a sequence of random variables $h_0, h_1, ..., h_t$ converges almost surely (a.s.) to a random variable h^* as $t \to \infty$ iff*

$$\Pr\left[\{\omega : h_t(\omega) \to h^*(\omega) \text{ as } t \to \infty\}\right] = 1.$$

Theorem 1 (Strong Law of Large Numbers [42])**.** *Let $(\Omega, \mathcal{F}, p)$ be a probability space equipped with a σ-algebra of measurable subsets of Ω. Let h_n be a sequence of n i.i.d. random variables defined over the probability space, with expectation $E[h_i]$. Let $s_n = h_1 + h_2 + ... + h_n$. Then,*

$$\lim_{n\to\infty} \frac{s_n}{n} = E[h_i] \quad \text{a.s.}$$

We present here a simplified version of the Perron-Frobenius theorem that we use through this work.

Theorem 2 (Perron-Frobenius Theorem [43])**.** *Let $P \in \mathbb{R}_{\geq 0}^{n\times n}$ be a non-negative column stochastic irreducible matrix. Then,*

- *$\lambda_1(P) = 1$, all other eigenvalues are smaller in norm.*
- *The eigenvector $Pv = v$ defines a dimension 1 subspace with some basis vector having strictly positive entries.*

At last, the following Theorem is a simplified version of results extracted from [31], [32].

Theorem 3 (Swarm Distribution Convergence [31], [32])**.** *Let $P(t)$ be a column stochastic time dependent probability transition matrix with at least one odd length cycle at $t = 0$, that follows stochastic dynamics. Let for some scalar $\varepsilon > 0$ and $\forall\, t \geq 0$, $P_{ij}(0) > 0 \Rightarrow P_{ij}(t) \geq \varepsilon$. Then, the limit product*

$$\lim_{t\to\infty} \prod_{t_k=0}^{t} P(t_k) = \boldsymbol{\zeta}\mathbf{1}^T \quad \text{a.s.},$$

where $\boldsymbol{\zeta}$ is a probability vector, and does so exponentially fast with a rate no slower than $\alpha = (1 - \varepsilon^{1+2\delta^*})^{\frac{1}{1+2\delta^*}}$.

III. Foraging Swarm Model

We state in this section the statement of a foraging problem over a graph, and present the dynamics of a proposed finite multi-agent system trying to solve the foraging problem.

A. Foraging Problem

Consider a swarm of n agents moving over an undirected weighted graph $\mathcal{G}$ trying to solve a *foraging problem*: the graph has a source vertex $\mathcal{S} \in \mathcal{V}$ where the agents are initialised, and a target vertex $\mathcal{T} \in \mathcal{V}$ they are supposed to find, converging to trajectories following the shortest path between $\mathcal{T}$ and $\mathcal{S}$. *Foraging* concerns, in general, both finding the shortest path between two points and depleting a food source as fast as possible. Given the discretised form of the problem, we consider in this work the *foraging problem* to be solved if agents reach a state of steadily following the shortest path between $\mathcal{S}$ and $\mathcal{T}$, back and forth, since, for real agents that move at constant speed and are able to carry a limited amount of food per trip, this would be the desired scenario for maximal depletion of the food source.

The swarm does not have accurate individual position information (GPS-like data). They can only receive measurements of a weight field from the vertices immediately next to them. Additionally, assume the agents are not able to communicate with any other member of the swarm. The agents are only able to send information to the vertex they are located at, and to receive information only from the neighbouring vertices.

B. Agent Dynamics

We are interested in solving the foraging problem using only indirect communication through the environment (the graph). It is convenient now to introduce the assumptions that are used throughout this work.

Assumption 1. *Any undirected graph $\mathcal{G}$ is strongly connected and has at least one odd length cycle.*

Assumption 2. *We assume there is only one $\mathcal{S} \in \mathcal{V}$ and $\mathcal{T} \in \mathcal{V}$, and the distance between them is larger than one.*

Remark 1. *Since we use graphs to discretise physical space, we can consider graphs to be triangular grids, and Assumption 1 is always satisfied.*

Now, let $\mathcal{G}$ be a vertex weighted undirected graph as in Definition 1. Let $A \in \{0,1\}^{|\mathcal{V}|\times|\mathcal{V}|}$ be its adjacency matrix. We define $\mathcal{A} := \{1, 2, ..., n\}$ as a set of agents walking from vertex to vertex. The position of agent a at time t is $(t) = v,\ v \in \mathcal{V}$, and we group them as $x(t) := \{\mathbf{x}_a(t) : a \in \mathcal{A}\}$. We define the vector of proportion of agents $\hat{\mathbf{q}}(t,n) : \mathbb{N}_0 \times \mathbb{N} \to \mathbb{P}^{|\mathcal{V}|}$ such that $\hat{\mathbf{q}}_i(t,n) = \frac{1}{n}|\{a \in \mathcal{A} : \mathbf{x}_a(t) = i\}|\ \forall i \in \mathcal{V}$.

The position of the agents evolves depending on some probability transition matrix $P(\cdot)$. That is, for $i, j \in \mathcal{V}$,

$$\Pr\{\mathbf{x}_a(t+1) = j \,|\, \mathbf{x}_a(t) = i\} = P_{ji}(\cdot),\ \mathbf{x}_a(0) = \mathcal{S},\ \forall\, a \in A. \tag{1}$$

In stigmergy algorithms, the transition probabilities are usually defined as the normalised weights around a vertex. In our case, drawing inspiration from experimental examples in literature, we model the probabilities of transitioning between vertices with an $\varepsilon-$greedy approach. Define the gradient matrix:

Definition 4. *Let $m_i := |\operatorname{argmax}_k W_{ik}(t)|$. The gradient matrix $P^{\nabla}(\mathbf{w}(t))$ is a stochastic matrix such that*

$$P^{\nabla}_{ji}(\mathbf{w}(t)) = \begin{cases} \frac{1}{m_i} & \text{if } W_{ij} = \max_k\{W_{ik}(t)\}, \\ 0 & \text{else.} \end{cases} \tag{2}$$

Definition 5. *Let $\mathbf{w}(t)$ be the corresponding time dependent weight matrix of a connected graph $\mathcal{G}$. Let A be the adjacency matrix of the graph. For a minimum probability $\varepsilon > 0$ we define the $\varepsilon-$greedy matrix as*

$$P^{\mathcal{G}}(t,\varepsilon) := \varepsilon(D(A)^{-1}A)^T + (1-\varepsilon)P^{\nabla}(\mathbf{w}(t)).$$

Then, the foraging agent dynamics are as follows.

Definition 6. *The distribution of agents $\hat{\mathbf{y}}(t) \in \mathbb{P}^{|\mathcal{V}|}$ is the probability of a given agent a being on vertex $i \in \mathcal{V}$ at time t. This probability evolves as a random walk,*

$$\hat{\mathbf{y}}_i(t+1) = \Pr\{\mathbf{x}_a(t+1) = i\} = \left(P^{\mathcal{G}}(t,\varepsilon)\hat{\mathbf{y}}(t)\right)_i \quad \forall a \in \mathcal{A}, \tag{3}$$

with $\hat{\mathbf{y}}_{\mathcal{S}}(0) = 1\ \forall a$.

The probability matrix $P^{\mathcal{G}}$ needs to be column stochastic to satisfy (3). Therefore, when we consider transitions $i \to j$, the corresponding probability is $P^{\mathcal{G}}_{ji}(t,\varepsilon)$ to avoid using the transposed matrix. From (3) we define the indicator vectors

$$\boldsymbol{\zeta}^a_i(t) = \begin{cases} 1 & \text{if } \mathbf{x}_a(t) = i, \\ 0 & \text{else,} \end{cases} \quad \hat{\mathbf{q}}(t,n) = \frac{1}{n}\sum_{a=1}^{n} \boldsymbol{\zeta}^a(t). \tag{4}$$

Since $P^{\mathcal{G}}(t,\varepsilon)$ is the same for all agents, all $\boldsymbol{\zeta}^a(t)$ share the same probability distribution for all $t \geq 0$ if $\boldsymbol{\zeta}^a(0) = \hat{\mathbf{y}}(0)\ \forall a$. Then, $\hat{\mathbf{q}}(t,n)$ is a sum of identically distributed random variables with probability distribution $\hat{\mathbf{y}}(t)$.

Remark 2. *Equation* (3) *can be read as "The probability of having an agent in some vertex i at time $t+1$ is equal to the probability of being in a neighbourhood of i at time t times the probability of moving to i". However, this raises some complications. In our case $P^{\mathcal{G}}(t,\varepsilon)$ is a stochastic sequence with respect to t and $P^{\mathcal{G}}(t,\varepsilon) = f(\mathcal{Q}_t(n))$. Therefore, the transition probabilities depend on the entire event history. A way of dealing with this challenge is proposed in Section IV.*

C. Weight Dynamics

The agents also modify the weights in the graph, similar to ants laying pheromones on the ground. Let $R(\cdot)$ be the amount of weight added to each vertex (to be properly defined below), such that $R_i(\cdot)$ is the weight added per agent to vertex i at time t. Then, the weights in $\mathcal{G}$ evolve as

$$\mathbf{w}_i(t+1) = (1-\rho)\mathbf{w}_i(t) + \rho\hat{\mathbf{q}}_i(t,n)R_i(\cdot),$$

where $\rho \in (0,1)$ is a chosen discount factor. The weights are initialised such that $\mathbf{w}(0) = \mathbf{1}\mathbf{w}_0$ with $\mathbf{w}_0 \geq 0$.

Remark 3. *Keeping in mind that these systems are defined over a continuous space in reality, and to avoid over-accumulation of communication (or marking) events in one single vertex, it is useful to consider a saturated form of reinforcement, where we write* (6) *as*

$$\mathbf{w}_i(t+1) = (1-\rho)\mathbf{w}_i(t) + \rho R_i(\cdot)\operatorname{sgn}\left(\hat{\mathbf{q}}_i(t,n)\right).$$

Effectively, this saturates the agent vector such that at every vertex there can be only one "reinforcement" event at a given time. From a real implementation point of view, this is logical since the reinforcement needs to be processed as some form of aggregated signal by an interacting environment or infrastructure, and otherwise such environment would need to process arbitrarily large amount of signals in finite time. Additionally, unbounded accumulation of weights may be undesirable. From this point on, we will retain this formulation.

In order for the swarm to solve the foraging problem, we draw similarities with reinforcement learning approaches to design our reward function R. Let $r \in \mathbb{R}_{\geq 0}$ be some positive constant, and the vector $\boldsymbol{\gamma} \in \mathbb{R}^{|\mathcal{V}|}_{\geq 0}$ take values

$$\boldsymbol{\gamma}_v(r) = \begin{cases} r & \text{if } v = \mathcal{T}, \mathcal{S} \\ 0 & \text{else.} \end{cases}$$

Then let $\lambda \in (0,1)$, and let $\Gamma(r) := \operatorname{diag}(\boldsymbol{\gamma}(r))$. Then, we can write the reward function in diagonal matrix form as

$$R(t,r,\lambda) := \left(I + \Gamma(r) + \lambda V\left(\mathbf{w}(t)\right)\right), \tag{5}$$

and the weight dynamics are simply

$$\mathbf{w}(t+1) = (1-\rho)\mathbf{w}(t) + \rho R(t,r,\lambda)\operatorname{sgn}\left(\hat{\mathbf{q}}(t,n)\right). \tag{6}$$

The intuition about this is as follows. The reward diagonal matrix has three explicit terms in each component. First, a

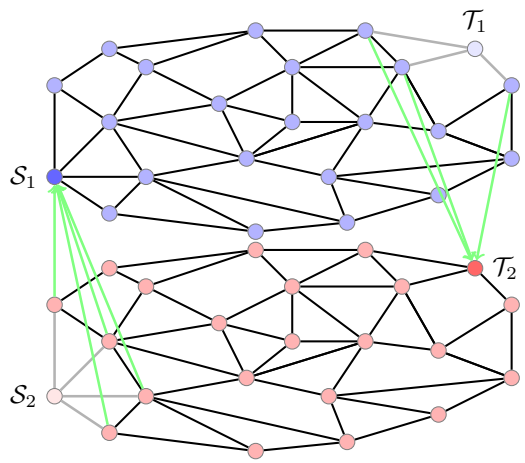

Fig. 1. Doubled interconnected Graph resulting from constructing $P(t,\varepsilon)$.

constant reward 1 to all vertices to replicate the behaviour of ants: ants add pheromones to every position they are located at, with at least a minimum amount (1 in our case), reflecting that vertex has been visited before. Second, the term $\Gamma(r)$ where the agents reward with an additional amount r the specific goals of our problem: finding $\mathcal{T}$ and returning to $\mathcal{S}$. This is also inspired in entomology; ants may mark the ground with different intensities if they have found food [44], [45]. At last, the third term is a diffusivity term (pheromones diffuse through the air to their immediate surroundings), and this term makes ants reinforce more or less based on neighbouring weights. Additionally, diffusivity is a commonly used strategy in value function learning problems. When using Q-values, diffusivity represents the maximum utility to be obtained at the next (or previous) step.

With (3), (4) and (6) the stochastic dynamics of the agents and weights are fully defined. We can now present how to use such a model to obtain a foraging swarm.

D. Foraging Swarm

For the swarm to produce emerging behaviour solving the foraging problem, some additional conditions must be added into the agent behaviour and weight update rules. As proposed in several examples in the literature [12], [25], [30], one way to achieve this is to make use of two different pheromones (or weights $\mathbf{w}^1(t)$, $\mathbf{w}^2(t)$). In such situation, agents looking for $\mathcal{T}$ follow $\mathbf{w}^1(t)$ (move according to $P^{\mathcal{G}^1}(t,\varepsilon) \equiv P^1(t,\varepsilon)$) and modify $\mathbf{w}^2(t)$, while agents looking for $\mathcal{S}$ follow $\mathbf{w}^2(t)$ (move according to $P^{\mathcal{G}^2}(t,\varepsilon) \equiv P^2(t,\varepsilon)$) and modify $\mathbf{w}^1(t)$. This implies a certain "memory" condition in the agent behaviour that may look non-Markovian: the agents follow a set of pheromones and reward another depending on their trajectory. But in fact, we can modify the system by duplicating the graph size so it remains "memoryless". The effects of this can be seen in Figure 1. In blue we have the weights $\mathbf{w}^1(t)$ and in red the weights $\mathbf{w}^2(t)$. The green edges represent the new directed edges added to the graph as a result of the interconnection of the two *original* sub-graphs. The intuition behind this "duplication" of the graph is to translate into the size of the state-space the fact that there are 2 simultaneous goals in the foraging problem (finding $\mathcal{S}$ and finding $\mathcal{T}$). We can retain the memoryless condition of the swarm by duplicating the size of the state-space and interconnecting sub-graphs. For details on how to construct this interconnected graph, see Appendix A.

System 1. *Given two (original) undirected graphs $\mathcal{G}^1 = \mathcal{G}^2 = \mathcal{G}$, we define a foraging swarm as the tuple $\phi := (\mathcal{G}, \mathcal{S}, \mathcal{T}, \hat{\mathbf{q}}(t,n), n, \varepsilon, \lambda)$ with $\mathcal{S} \in \mathcal{V}, \mathcal{T} \in \mathcal{V}, \hat{\mathbf{q}} : \mathbb{N}_0 \times \mathbb{N}_0 \to \mathbb{P}^{|\mathcal{V}|}, n \in \mathbb{N}_0, \varepsilon \in \mathbb{R}_{\geq 0}, \lambda \in (0,1)$ such that*

$$\hat{\mathbf{q}}(t,n) := \begin{pmatrix} \hat{\mathbf{q}}^1(t,n) \\ \hat{\mathbf{q}}^2(t,n) \end{pmatrix}, \hat{\mathbf{y}}(t) := \begin{pmatrix} \hat{\mathbf{y}}^1(t) \\ \hat{\mathbf{y}}^2(t) \end{pmatrix},$$
$$\mathbf{w}(t) := \begin{pmatrix} \mathbf{w}^1(t) \\ \mathbf{w}^2(t) \end{pmatrix}, W(t) := \begin{pmatrix} W^1(t) & 0) \\ 0 & W^2(t) \end{pmatrix},$$
$$P(t,\varepsilon) := \begin{pmatrix} (I-T)P^2(t,\varepsilon) & SP^1(t,\varepsilon) \\ TP^2(t,\varepsilon) & (I-S)P^1(t,\varepsilon) \end{pmatrix},$$

initialised as $\mathbf{w}(0) = \mathbf{0}^{2\times|\mathcal{V}|}$, $\hat{\mathbf{y}}_{\mathcal{S}}(0) = 1$, $\hat{\mathbf{q}}(n,0) = \mathbf{y}(0)$ which follows the dynamics

$$\begin{aligned} \hat{\mathbf{y}}(t+1) &= P(t,\varepsilon)\hat{\mathbf{y}}(t), \\ w(t+1) &= (1-\rho)\mathbf{w}(t) + \rho R(t,r,\lambda)\,\mathrm{sgn}\left(\hat{\mathbf{q}}(t,n)\right), \\ R(t,r,\lambda) &:= (I + \Gamma(r) + \lambda V(W(t))) \end{aligned} \quad (7)$$

Remark 4. *The resulting connected graph in System 1 has some edges removed with respect to the original graph $\mathcal{G}$. It effectively disconnects $\mathcal{T}_1$ and $\mathcal{S}_2$ from the rest of the graph. Nevertheless, the density of agents initialised in these vertices is 0, and since this is a virtual duplication of the graph, we can simply consider $\mathcal{G} \in \phi$ to have edges $\mathcal{E} = \{(ij) \in \mathcal{E}_1 \cup \mathcal{E}_2 : i,j \neq \mathcal{T}_1 \cup \mathcal{S}_2\}$ and vertices $\mathcal{V} = \{i \in \mathcal{V}_1 \cup \mathcal{V}_2 : i \neq \mathcal{T}_1 \cup \mathcal{S}_2\}$. This does not affect the dynamics, and results again in a strongly connected graph. We will refer to $\mathcal{T}_2 \equiv \mathcal{T}$ and $\mathcal{S}_1 \equiv \mathcal{S}$ as the resulting target and starting vertices in ϕ.*

Observe the weight dynamics in (7) present coupled terms between the weights and the agents position, which is a random variable. For this reason, it becomes extremely challenging to analyse the solutions to which the system converges in such finite agent form. One way to solve this is to study what happens when we consider very large number of agents.

With the presented framework of stochastic foraging swarm, we can specify the first problem to solve in further sections.

Problem 1. *Let ϕ be a foraging swarm communicating based on a double pheromone stigmergy method. Construct a non-stochastic mean field model of the system as $n \to \infty$.*

IV. Mean Field Swarm

In mean field models for Swarm Robotics, the number of agents is assumed to be large enough ($n \to \infty$) so that random variables can effectively be replaced by a mean valued deterministic variable. We show here how to do this in the foraging swarm presented in System 1. Recall that the state of our system is fully defined by the $\sigma-$algebra generated by the proportion of agents in each vertex,

$$\mathcal{F}_t = \sigma(\hat{\mathbf{q}}(0,n), \dots \hat{\mathbf{q}}(t,n)).$$

Let us define the sequence $\mathcal{Q}_t(n) := \{\hat{\mathbf{q}}(0,n), ..., \hat{\mathbf{q}}(t,n)\}$. In this case, an event $\mathcal{Q}_t(n) \in \mathcal{F}_t$ is a sequence of agent proportion vectors until time t resulting in the generator sequence of random variables $\hat{\mathbf{q}}(0,n), \dots \hat{\mathbf{q}}(t,n)$. Now, observe that the conditional expected value of $\boldsymbol{\zeta}^a(t)$ is

$$E[\boldsymbol{\zeta}^a(t+1) = 1 | \mathcal{F}_t] = P(t,\varepsilon)\boldsymbol{\zeta}^a(t). \tag{8}$$

Recall that $\boldsymbol{\zeta}^a(0) = \hat{\mathbf{y}}(0)\ \forall a$, and note that while all $\boldsymbol{\zeta}^a(t)$ follow the same probability distribution for all $t \geq 0$ they are not independent from each other. The evolution of the probability distribution of every $\boldsymbol{\zeta}^a$ follows a product of probability matrices that resembles the dynamics of a Markov process. From (3),

$$\Pr[\{\boldsymbol{\zeta}_i^a(t) = 1\}] = \hat{\mathbf{y}}_i(t) = \left(\prod_{t_k=0}^{t} P(t_k,\varepsilon)\hat{\mathbf{y}}(0) \right)_i.$$

But in this case, $P(t_k,\varepsilon) = f(\mathcal{Q}_k)$. That is, the sequence of probability transition matrices is a function of the agent positions for all previous times. This means that, in general, for two different events $\mathcal{Q}_t^m(n), \mathcal{Q}_t^l(n) \in \mathcal{F}_t$,

$$\Pr[\{\boldsymbol{\zeta}_i^a(t+1) = 1 | \mathcal{Q}_t^m(n)\}] \neq \Pr[\{\boldsymbol{\zeta}_i^a(t+1) = 1 | \mathcal{Q}_t^l(n)\}].$$

Furthermore, observe that the dependence is on the entire sequence until time t. Therefore, in general, the probability of finding agents in each vertex will depend as well on the position of other agents (making their indicator random vectors dependent). Despite this complexity, we can show convergence of the agent proportion vector to its distribution when $n \to \infty$.

Theorem 4. *Let ϕ be a foraging swarm. Let $\mathbf{y}(t) := \lim_{n\to\infty} \hat{\mathbf{y}}(t)$, and $\mathcal{Y}_t := \{\mathbf{y}(0), \mathbf{y}(1), ..., \mathbf{y}(t)\}$. Then,*

$$\mathcal{Q}_t^\infty := \lim_{n\to\infty} \mathcal{Q}_t(n) = \mathcal{Y}_t \quad \textit{a.s.} \quad \forall t \geq 0.$$

Proof (Theorem 4). We show this by induction. Let us look first at $t = 0$. For a fixed set of initial conditions $\hat{\mathbf{q}}(0,n)$, $\mathbf{w}(0)$, observe that $\hat{\mathbf{q}}(0,n) = \hat{\mathbf{y}}(0)$, and we have $\forall a \in \mathcal{A}$

$$\Pr[\{\boldsymbol{\zeta}_i^a(1) = 1 \,|\, \hat{\mathbf{q}}(0,n), \mathbf{w}(0)\}] = \hat{\mathbf{y}}_i(1) = (P(0,\varepsilon)\hat{\mathbf{q}}(0,n))_i.$$

Observe that in this case, the initial conditions are fixed, therefore we can consider the total probability

$$\Pr[\{\boldsymbol{\zeta}_i^a(1) = 1\}] = (P(0,\varepsilon)\hat{\mathbf{q}}(0,n))_i = (P(0,\varepsilon)\hat{\mathbf{y}}(0))_i. \tag{9}$$

Observe that (9) does not depend on a. Therefore, all agents have the same marginal probability distribution for $t = 1$. Additionally, the transition probabilities at $t = 0$ have not been affected by agent trajectories, therefore for the first time step $\boldsymbol{\zeta}_i^a(1)$ are *i.i.d.* $\forall a$. We can then specify the joint distribution of having k agents in vertex i at time $t = 1$: this is the joint probability of events resulting in k agents moving to i, and $n - k$ agents moving elsewhere. Recall $\hat{\mathbf{q}}(1,n) = \frac{1}{n}\sum_{a=1}^{n} \boldsymbol{\zeta}^a(1)$. Since $\boldsymbol{\zeta}_i^a$ are indicator variables,

$$E[\boldsymbol{\zeta}^a(1)] = P(0,\varepsilon)\hat{\mathbf{y}}(0) = \hat{\mathbf{y}}(1).$$

Let us now consider the case where $n \to \infty$, and define $\mathbf{y}(1) := \lim_{n\to\infty} \hat{\mathbf{y}}(1)$. Since at $t = 0$ the initial conditions are fixed and all agents are initialised in the same vertex, it also holds that $\hat{\mathbf{y}}(0) = \mathbf{y}(0)$. Additionally, $P(0,\varepsilon)$ is not affected by the limit $n \to \infty$, and $\hat{\mathbf{y}}(1) = P(0,\varepsilon)\hat{\mathbf{y}}(0) = P(0,\varepsilon)\mathbf{y}(0) = \mathbf{y}(1)$. Therefore, by Theorem 1 we have

$$\lim_{n\to\infty} \hat{\mathbf{q}}(1,n) = E[\boldsymbol{\zeta}^a(1)] = \mathbf{y}(1) \quad \text{a.s.} \tag{10}$$

That is, with probability 1, the agent proportion converges to the marginal probability distribution as $n \to \infty$ for $t = 1$. From (10) it holds that any event $\mathcal{Q}_1 \in \mathcal{F}_1$ (*i.e.* any possible combination of agent positions until time $t = 1$) satisfies $\mathcal{Q}_1 \in \mathcal{F}_1 \Rightarrow \mathcal{Q}_1 = \{\mathbf{y}(0), \mathbf{y}(1)\}$ *a.s.* That is, $\Pr\{\mathcal{Q}_1 \in \mathcal{F}_1 : q(1) = \mathbf{y}(1)\} = 1$ (the union of events has measure 1). Then, the update of $P(1,\varepsilon)$ depends on $\mathbf{w}(1)$, and in the limit $\lim_{n\to\infty} \mathbf{w}(1) = f(\lim_{n\to\infty} \hat{\mathbf{q}}(1,n), \mathbf{w}(0)) = f(\mathbf{y}(0), \mathbf{w}(0))$. Now for $t = 2$,

$$\begin{aligned} E[\boldsymbol{\zeta}^a(2)] =& E[E[\boldsymbol{\zeta}^a(2) | \mathcal{F}_1]] = E[P(1,\varepsilon)E[\boldsymbol{\zeta}^a(1) | \mathcal{F}_1]] = \\ =& P(1,\varepsilon)E[\boldsymbol{\zeta}^a(1)] = P(1,\varepsilon)\mathbf{y}(1) = \mathbf{y}(2). \end{aligned} \tag{11}$$

Therefore, with probability 1, the marginal probability distributions $\boldsymbol{\zeta}^a(2)$ are determined by $\mathbf{y}(1)$ (since they depend on $\mathcal{Q}_1$, and this occurs *a.s.*). Therefore, the variables are *i.i.d.* in the limit $n \to \infty$, and by the law of large numbers,

$$\lim_{n\to\infty} \hat{\mathbf{q}}(2,n) = E[\boldsymbol{\zeta}^a(2)] = \mathbf{y}(2) \quad \text{a.s.}$$

By induction, it holds that there is only one possible sequence of outcomes $\mathcal{Q}_t = \{\mathbf{y}(0), \mathbf{y}(1), ..., \mathbf{y}(t)\}$ where $\Pr\{\lim_{n\to\infty} \mathcal{Q}_t(n) = \mathcal{Q}_t\} = 1\ \forall t \geq 0$. Therefore $E[\boldsymbol{\zeta}^a(t+1)] = P(t,\varepsilon)\mathbf{y}(t) = \mathbf{y}(t+1)$, thus

$$\lim_{n\to\infty} \hat{\mathbf{q}}(t,n) = \mathbf{y}(t) \quad \text{a.s.} \quad \forall t \geq 0.$$

□

By making use of Theorem 4, one can take the mean field limit and approximate the behaviour of $\hat{\mathbf{q}}(t,n)$ with $\mathbf{y}(t)$ as $n \to \infty$. Additionally, the indicator variables $\boldsymbol{\zeta}^a(t)$ become *i.i.d.* When $n \to \infty$, there is only one possible sequence $\mathcal{Q}_t$ occurring with probability one. In other words, $\Pr[\{\mathcal{Q}_t \in \mathcal{F}_t : \mathcal{Q}_t = \mathcal{Y}_t\}] = 1$, so $\mathcal{Q}_t = \mathcal{Y}_t$ happens for a set of outcomes of measure 1, and the evolution of the agent density becomes deterministic. This means that the sequence of matrices $P(t,\varepsilon)$ is also deterministic, and independent

of every single $\boldsymbol{\zeta}^a(t)$. Therefore, the probability distribution $\mathbf{y}(t)$ of all $\boldsymbol{\zeta}^a(t)$ becomes independent from individual agent trajectories. This translates into the indicator vectors being *i.i.d.* with respect to each other.

Remark 5. *Observe the difference between $\hat{\mathbf{y}}(t)$ (probability distribution of agent positions for a finite number n) and $\mathbf{y}(t)$ (probability distribution of agent positions when $n \to \infty$). In all cases, $\hat{\mathbf{y}}(0) = \mathbf{y}(0)$, but they can be different from each other for $t > 0$ since they evolve according to $P(t, \varepsilon)$, which implicitly depends on n.*

We can now define our *mean field swarm system*.

System 2. *Let $\mathcal{G}^1 = \mathcal{G}^2 = \mathcal{G}$ be two identical connected weighted graphs. A mean field foraging swarm system is defined as the tuple $\Phi := (\mathcal{G}, \mathcal{S}, \mathcal{T}, \mathbf{y}(t), \varepsilon, \lambda)$ with $\mathcal{S} \in \mathcal{V}, \mathcal{T} \in \mathcal{V}, \mathbf{y} : \mathbb{N}_0 \to \mathbb{P}^{|\mathcal{V}|}, \varepsilon \in \mathbb{R}_{\geq 0}, \lambda \in (0, 1)$. The state variables are initialised as $\mathbf{w}(0) = \mathbf{0}^{|\mathcal{V}|}$, $\mathbf{y}_{\mathcal{S}}(0) = 1$, and:*

$$\begin{aligned} \mathbf{y}(t+1) &= P(t, \varepsilon)\mathbf{y}(t), \\ \mathbf{w}(t+1) &= (1-\rho)\mathbf{w}(t) + \rho R(t, r, \lambda) \operatorname{sgn}(\mathbf{y}(t)), \end{aligned} \tag{12}$$

These concepts lead us to the second goal of this work.

Problem 2. *Let Φ be a mean field foraging swarm. Do the mean field dynamics converge to a (sub-optimal) fixed point? Additionally, what can we say (experimentally) about the deviation from the mean field case when choosing a finite number n?*

V. Convergence Guarantees

We study next the convergence properties of the mean field foraging swarm Φ of System 2.

Proposition 1. *Let Φ be a mean field foraging swarm system. Let Assumption 1 hold. Let $1 \geq \varepsilon > 0$. Then, the agent density $\mathbf{y}(t)$ converges exponentially to a stationary density $\mathbf{y}(\infty) \in \mathbb{P}^{|\mathcal{V}|}$, unique for given initial conditions $\mathbf{y}(0)$ and $\mathbf{w}(0)$, that satisfies $\mathbf{y}_i(\infty) > 0 \ \forall i \in \mathcal{V}$.*

Proof. See Appendix B. □

Additionally, we can show the following result. Recall δ^* is the diameter of the underlying graph.

Lemma 1. *Let Φ be a mean field foraging swarm system. With $t_\delta = 2\delta^*$, it holds that $\operatorname{sgn}(\mathbf{y}(t)) = \mathbf{1} \quad \forall t > t_\delta$.*

Proof. See Appendix B. □

Since there is a minimum probability of accessing any vertex in the graph (and the graph has odd cycles), eventually there is a non-zero amount of agents in every vertex, regardless of the foraging dynamics. Going back again to the relation finite-infinite agents, this is equivalent to saying that agents have a non-zero probability of accessing every vertex of the graph for all times greater than t_δ.

Remark 6. *In fact, $2\delta^*$ is an upper bound for the required time t_δ. It represents the case where $\mathcal{G}$ is one edge away from being bipartite, and to reach some even vertex in odd time it takes δ^* time steps to reach the (only) odd length cycle plus δ^* time steps to reach the target vertex. In practice, $t_\delta \in [\delta^*, 2\delta^*]$.*

With these preliminary results, we can present the main contribution of this section.

Proposition 2. *There is a unique weight vector $\mathbf{w}(\infty)$ and corresponding matrix $W(\infty)$ satisfying $\mathbf{w}(\infty) := (I + \Gamma(r) + \lambda V(W(\infty)))\mathbf{1}$ for a fixed reward matrix $\Gamma(r)$ and $\lambda \in [0, 1)$.*

Proof. Let $B \in \{0, 1\}^{|\mathcal{V}| \times |\mathcal{V}|}$ be the selector matrix satisfying $B\mathbf{w}(\infty) = V(W(\infty))\mathbf{1}$. Since B is a row stochastic matrix by Theorem 2 it has all its eigenvalues in the unit disc, and $(I - \lambda B)$ has all its eigenvalues in a disc of radius λ centred at 1. Therefore, its inverse is properly defined and $\mathbf{w}(\infty) = (I - \lambda B)^{-1}(I + \Gamma(r))$ has a unique solution if $\lambda \in [0, 1)$. □

Theorem 5. *The weight dynamics in Φ have a fixed point $\mathbf{w}(\infty)$, and*

$$\lim_{t \to \infty} \mathbf{w}(t) = \mathbf{w}(\infty) := (I + \Gamma(r) + \lambda V(W(\infty)))\mathbf{1}.$$

Proof (Theorem 5). Recall the weight dynamics:

$$\begin{aligned} \mathbf{w}(t+1) =& (1-\rho)\mathbf{w}(t) + \\ &+ \rho \left(I + \Gamma(r) + \lambda V(W(t))\right) \operatorname{sgn}(\mathbf{y}(t)). \end{aligned} \tag{13}$$

Let $\mathbf{w}(\infty) = (I + \Gamma(r) + \lambda V(W(\infty)))\mathbf{1}$, $\mathbf{z}(t) := \mathbf{w}(t) - \mathbf{w}(\infty)$. Subtract $\mathbf{w}(\infty)$ at each side of (13):

$$\mathbf{z}(t+1) = (1-\rho)\mathbf{z}(t) + \rho \left(R(t, r, \lambda) \operatorname{sgn}(\mathbf{y}(t)) - \mathbf{w}(\infty)\right). \tag{14}$$

Define $\mathbf{e}_{\mathbf{y}}(t) := \operatorname{sgn}(\mathbf{y}(t)) - \mathbf{1}$ to obtain:

$$\begin{aligned} &\left(I + \Gamma(r) + \lambda V(W(t))\right) \operatorname{sgn}(\mathbf{y}(t)) - \mathbf{w}(\infty) = \\ =&\left(I + \Gamma(r) + \lambda V(W(t))\right)\mathbf{e}_{\mathbf{y}}(t) + \\ &+ \lambda \left(V(W(t)) - V(W(\infty))\right)\mathbf{1}. \end{aligned}$$

Taking the ∞-norm at each side of (14):

$$\begin{aligned} \|\mathbf{z}(t+1)\|_\infty =& \|(1-\rho)\mathbf{z}(t) + \rho(R(t, r, \lambda)\mathbf{e}_{\mathbf{y}}(t) + \\ &+ \lambda \left(V(W(t)) - V(W(\infty))\right)\mathbf{1})\|_\infty \leq \\ \leq& (1-\rho)\|\mathbf{z}(t)\|_\infty + \rho \|R(t, r, \lambda)\|_\infty \|\mathbf{e}_{\mathbf{y}}(t)\|_\infty + \\ &+ \rho\lambda \|V(W(t)) - V(W(\infty))\|_\infty \|\mathbf{1}\|_\infty. \end{aligned} \tag{15}$$

Recall the induced ∞-norm of a matrix is its maximum absolute row sum. Then,

$$\begin{aligned} &\|V(W(t)) - V(W(\infty))\|_\infty \|\mathbf{1}\|_\infty = \\ =& \max_i |\max_j \mathbf{w}_{ij}(t) - \max_j \mathbf{w}_{ij}(\infty)| \leq \\ \leq& \max_i |\max_j |\mathbf{w}_{ij}(t) - \mathbf{w}_{ij}(\infty)|| = \max_i |\mathbf{z}_i(t)| = \|\mathbf{z}(t)\|_\infty. \end{aligned} \tag{16}$$

Now from Lemma 1, $\|\mathbf{e}_{\mathbf{y}}(t)\|_\infty = 0 \ \forall t > 2\delta^*$, therefore substituting (16) in (15):

$$\|\mathbf{z}_i(t+1)\|_\infty \leq (1-\rho)\|\mathbf{z}(t)\|_\infty + \rho\lambda\|\mathbf{z}(t)\|_\infty = \\ = (1-\rho(1-\lambda))\|\mathbf{z}(t)\|_\infty \leq (1-\rho(1-\lambda))^2\|\mathbf{z}(t-1)\|_\infty \leq \\ \leq (1-\rho(1-\lambda))^{t-2\delta^*}\|z(2\delta^*)\|_\infty \Rightarrow \lim_{t\to\infty}\|\mathbf{z}_i(t)\|_\infty = 0. \tag{17}$$

Finally, $\lim_{t\to\infty}\|\mathbf{z}(t)\|_\infty = 0 \Rightarrow \lim_{t\to\infty}\mathbf{w}(t) = \mathbf{w}(\infty)$, and the proof is complete. □

Corollary 1. *The probability transition matrix converges to a unique matrix* $\lim_{t\to\infty}P(t,\varepsilon) = P(\infty,\varepsilon)$*, and the stationary distribution of agents* $\lim_{t\to\infty}\mathbf{y}(t) = \mathbf{y}(\infty)$ *is the eigenvector corresponding to the eigenvalue 1. That is,*

$$P(\infty,\varepsilon) := \lim_{t\to\infty}P(t,\varepsilon), \quad \mathbf{y}(\infty) = P(\infty,\varepsilon)\mathbf{y}(\infty). \tag{18}$$

Proof. See Appendix B. □

A. On the optimality of solutions

Let us examine now what do the agent distributions look like in a mean field swarm system Φ. To this end, we define first a few useful concepts to characterize the state variables.

Definition 7. *Let* $\mathbf{w}$ *be the weight vector in a system* Φ*. We define a maximum (weight) gradient set of paths* $\pi_{ij}^{\nabla}(\mathbf{w})$ *as the set of all unique paths between vertices* i*,* j *satisfying*

$$p_{ij}^{\nabla} \in \pi_{ij}^{\nabla}(\mathbf{w}) \iff p_{ij}^{\nabla} := \{i, i_2, i_3, ..., i_k, j\}, \\ i_2 = \text{argmax}_v(W_{iv}(t)),\ i_3 = \text{argmax}_v(W_{i_2v}(t)), ..., \\ j = \text{argmax}_v(W_{i_kv}(t)).$$

In other words, let the weight vector be $\mathbf{w}(t)$. Then, $\pi_{ij}^{\nabla}(\mathbf{w}(t))$ is the set of all paths obtained from following the maximum neighbouring weights at each step when going from i to j. Note that, for any two $i, j \in \mathcal{V}$, it can be that $\pi_{ij}^{\nabla}(\mathbf{w}(t)) = \emptyset$ if picking the maximum weight neighbour at every step does never connect i with j.

Definition 8. *We define the set of optimal weight vectors* $\mathcal{W}^* \subset \mathbb{R}_{\geq 0}^{|\mathcal{V}|}$ *for a mean field foraging system* Φ *as*

$$\mathcal{W}^* := \{\mathbf{w}^* : \pi_{ij}^{\nabla}(\mathbf{w}^*) \equiv \pi_{ST} \wedge \pi_{ij}^{\nabla}(\mathbf{w}^*) \equiv \pi_{TS}\}.$$

That is, for every weight vector $\mathbf{w}^* \in \mathcal{W}^*$*, the set of paths resulting from starting at* $\mathcal{S}$ $(\mathcal{T})$ *and following the maximum gradient vertices lead to* $\mathcal{T}$ $(\mathcal{S})$*, and is equal to* π_{ST} (π_{TS})*.*

This interpretation of an optimal set of weights is entirely pragmatic. We call a weight distribution optimal if, when starting at $\mathcal{S}$ and following the maximum weight vertex at every step, we end up at $\mathcal{T}$ and we obtain a minimum length path between the two (and vice-versa from $\mathcal{T}$ to $\mathcal{S}$). Additionally, observe that the optimal weight set $\mathcal{W}^*$ is defined for the *doubled* graph in Figure 1. Nevertheless, given the symmetry of the graph (the sub-graphs satisfy $\mathcal{G}^1 = \mathcal{G}^2$), any optimal weight vector $\mathbf{w}^* \in \mathcal{W}^*$ generates optimal paths on the original (unweighed) graph too, but it does so separately for paths $\mathcal{S} \to \mathcal{T}$ and for paths $\mathcal{T} \to \mathcal{S}$. Intuitively, constructing a weight vector $\mathbf{w}^*$ means the swarm has solved the foraging problem by building a weight function whose gradient always leads towards an optimal path.

Proposition 3. *Let* Φ *be a mean field stigmergy swarm. Then,* $\lim_{t\to\infty}\mathbf{w}(t) = \mathbf{w}(\infty) \in \mathcal{W}^*$*.*

Proof. See Appendix B. □

From Definition 5, abusing the notation for the variable ε we can decompose $P(\infty,\varepsilon)$ in two matrices such that

$$P(\infty,\varepsilon) = (1-\varepsilon)P(\infty,0) + \varepsilon P(\infty,1), \tag{19}$$

where $P(\infty,0)$ is the transition matrix corresponding to moving according to the gradient of the weights $\mathbf{w}(\infty)$. Observe as well that $P(\infty,1)$ depends only on the adjacency matrix A, which guarantees the decomposition (19) to be unique. We define then the following sets.

Definition 9. *We define* $N_v^{out} = \{j \in \mathcal{V} : P_{jv}(\infty,0) > 0\}$*, and* $N_v^{in} = \{j \in \mathcal{V} : P_{vj}(\infty,0) > 0\}$ *as the out and in neighbour vertices connected to* v *by following* $P(\infty,0)$*.*

Observe that in Definition 4 we use m to count the number of (out) neighbours that have maximum weight around a vertex, and therefore $m_v \equiv |N_v^{out}|$. Now let $k = \delta(\mathcal{S},\mathcal{T})$, and recall $p_{ST}^* \in \pi_{ST}$ is any path in the set of optimal paths. Let $\overline{\mathbf{y}} \in \mathbb{P}^{|\mathcal{V}|}$ be a probability vector taking values

$$\overline{\mathbf{y}}_i := \begin{cases} \frac{1}{2k} & \text{if } i = \mathcal{S}, \mathcal{T}, \\ \frac{1}{2k}\sum_{p\in\pi_{Si}}\prod_{u\in p\setminus i}\frac{1}{|N_u^{out}|} & \text{if } i \in \cup p_{ST}^* \setminus \mathcal{S}, \mathcal{T}, \\ \frac{1}{2k}\sum_{p\in\pi_{Ti}}\prod_{u\in p\setminus i}\frac{1}{|N_u^{out}|} & \text{if } i \in \cup p_{TS}^* \setminus \mathcal{S}, \mathcal{T}, \\ 0 & \text{else.} \end{cases}$$

The term $\frac{1}{|N_u^{out}|}$ can be interpreted as the probability of moving out of u towards a specific neighbour, therefore the product $\Pr\{p\} := \prod_{u\in p\setminus i}\frac{1}{|N_u^{out}|}$ can be interpreted as the probability of following a path p until vertex i, starting at $\mathcal{S}, \mathcal{T}$. Then, we obtain the following result.

Proposition 4. *Let* Φ *be a mean field stigmergy swarm. Let the system converge as* $t \to \infty$ *for a fixed* $1 > \varepsilon > 0$ *and let* $P(\infty,0)$ *defined in* (19)*. Then,* $P(\infty,0)\overline{\mathbf{y}} = \overline{\mathbf{y}}$*.*

Proof. See Appendix B. □

Proposition 4 indicates that the vector $\overline{\mathbf{y}}$ is the first eigenvector of the "gradient" matrix $P(\infty,0)$. That is, as the weights converge, when the agents move by selecting the maximum weight vertex around them, the only stationary distribution is the one that spreads all agents equally across the optimal paths between $\mathcal{S}$ and $\mathcal{T}$.

Theorem 6. *Let* Φ *be a mean field stigmergy swarm. Let* $\beta : [0,\infty) \to [0,\infty)$ *be* $\beta \in \mathcal{K}_\infty$*. Then, it holds that*

$$\|\mathbf{y}(\infty) - \overline{\mathbf{y}}\|_1 \leq \beta(\varepsilon).$$

That is, the stationary agent distribution of Φ *gets arbitrarily close to the optimal distribution as* $\varepsilon \to 0$*.*

Proof (Theorem 6). Recall $P(\infty,\varepsilon) = (1-\varepsilon)P(\infty,0) + \varepsilon P(\infty,1)$. Additionally, from Corollary 1 and Proposition 4,

$$P(\infty,\varepsilon)\mathbf{y}(\infty) = \mathbf{y}(\infty), \quad P(\infty,0)\overline{\mathbf{y}} = \overline{\mathbf{y}}.$$

Now let $L := (I - P(\infty,0))$, $\Delta P := P(\infty,1) - P(\infty,0)$. Then, we can expand

$$\begin{aligned}
&\mathbf{y}(\infty) - \overline{\mathbf{y}} = P(\infty,\varepsilon)\mathbf{y}(\infty) - P(\infty,0)\overline{\mathbf{y}} = \\
&= (1-\varepsilon)P(\infty,0)\mathbf{y}(\infty) + \varepsilon P(\infty,1)\mathbf{y}(\infty) - P(\infty,0)\overline{\mathbf{y}} = \\
&= P(\infty,0)(\mathbf{y}(\infty) - \overline{\mathbf{y}}) + \varepsilon\Delta P\mathbf{y}(\infty) \Rightarrow \\
&\Rightarrow L(\mathbf{y}(\infty) - \overline{\mathbf{y}}) = \varepsilon\Delta P\mathbf{y}(\infty).
\end{aligned} \tag{20}$$

The null space of L is given by $Lv = 0 \iff P(\infty,0)v = v$, and by Theorem 2 we know v is unique, therefore $\operatorname{rank}(L) = |\mathcal{V}| - 1$. But to solve the system of equations $L(\mathbf{y}(\infty) - \overline{\mathbf{y}}) = \varepsilon\Delta P\mathbf{y}(\infty)$, L needs to be invertible. For this we can add the following additional equation: We know it must hold that $\mathbf{1}^T(\mathbf{y}(\infty) - \overline{\mathbf{y}}) = 0$, and this equation is linearly independent from all rows in L if and only if $\nexists \mu \in \mathbb{R}^{|\mathcal{V}|}$ that satisfies $L\mu = \mathbf{1}$. Let us show that there does not exist such a μ by contradiction. Assume $\exists \mu : L\mu = \mathbf{1}$. Recall π_{ST}, π_{TS} are the sets of optimal paths between $\mathcal{S}, \mathcal{T}$ and $\mathcal{T}, \mathcal{S}$, with $p^*_{ST} \in \pi_{ST}$. Then, $\forall i_1 \in N^{out}_{\mathcal{S}}$,$L_{\mathcal{S}\mathcal{S}} = L_{i_1 i_1} = 1, \quad L_{i_1\mathcal{S}} = -\frac{1}{|N^{out}_{\mathcal{S}}|}$. Adding the rows of L $\forall i_1$:

$$\left(\sum_{i_1 \in N^{out}_{\mathcal{S}}} L_{i_1}\right)_j = \begin{cases} 1 & \text{if } j \in N^{out}_{\mathcal{S}}, \\ -1 & \text{if } j = \mathcal{S}, \\ 0 & \text{else.} \end{cases} \tag{21}$$

Now let $\cup_{i_1} N^{out}_{i_1} := \{k : k \in N^{out}_{i_1} \forall\, i_1 \in N^{out}_{\mathcal{S}}\}$ be the set of all vertices at distance 2 from $\mathcal{S}$ when following optimal paths, and $i_2 \in \cup_{i_1} N^{out}_{i_1}$. Adding the rows of L $\forall i_2$:

$$\left(\sum_{i_2 \in \cup_{i_1} N^{out}_{i_1}} L_{i_2}\right)_j = \begin{cases} 1 & \text{if } j \in \cup_{i_1} N^{out}_{i_1}, \\ -1 & \text{if } j \in N^{out}_{\mathcal{S}}, \\ 0 & \text{else.} \end{cases} \tag{22}$$

Now it is clear that adding (21) and (22):

$$\left(\sum_{i \in N^{out}_{\mathcal{S}}} L_{i_1} + \sum_{i_2 \in \cup_{i_1} N^{out}_{i_1}} L_{i_2}\right)_j = \begin{cases} 1 & \text{if } j \in \cup_{i_1} N^{out}_{i_1}, \\ -1 & \text{if } j = \mathcal{S}, \\ 0 & \text{else.} \end{cases} \tag{23}$$

Extending the sum until vertex $\mathcal{T}$, we add rows $\forall i \in \cup p^*_{ST}$:

$$\left(\sum_{i \in \cup p^*_{ST} \setminus \mathcal{S}} L_i\right)_j = \begin{cases} 1 & \text{if } j = \mathcal{T}, \\ -1 & \text{if } k = \mathcal{S}, \\ 0 & \text{else.} \end{cases} \tag{24}$$

Analogously, considering the reverse paths π_{TS} one obtains

$$\left(\sum_{i \in \cup p^*_{TS} \setminus \mathcal{T}} L_i\right)_j = \begin{cases} 1 & \text{if } j = \mathcal{S}, \\ -1 & \text{if } j = \mathcal{T}, \\ 0 & \text{else.} \end{cases} \tag{25}$$

Define $\theta := |\cup p^*_{ST}| = |\cup p^*_{TS}|$ as the number of vertices in all optimal paths, and from (24) and (25) one obtains

$$\sum_{i \in \cup p^*_{ST} \setminus \mathcal{S}} L_i\mu = \theta - 1, \sum_{j \in \cup p^*_{TS} \setminus \mathcal{T}} L_j\mu = \theta - 1 \Rightarrow -\theta = \theta,$$

which is a contradiction. Then, $\nexists \mu : L\mu = \mathbf{1}$, and there is a row in L, ΔP such that replacing it (assuming it is the last row, without loss of generality) we obtain

$$\tilde{L} := \begin{pmatrix} L_1 \\ \dots \\ \mathbf{1}^T \end{pmatrix}, \tilde{\Delta P}\mathbf{y}(\infty) := \begin{pmatrix} \Delta P_1\mathbf{y}(\infty) \\ \dots \\ \mathbf{0}^T \end{pmatrix},$$

where $\operatorname{rank}(\tilde{L}) = |\mathcal{V}|$. Now, observe

$$\tilde{L}(\mathbf{y}(\infty) - \overline{\mathbf{y}}) = \varepsilon\tilde{\Delta P}\mathbf{y}(\infty) \Rightarrow \mathbf{y}(\infty) - \overline{\mathbf{y}} = \varepsilon\tilde{L}^{-1}\tilde{\Delta P}\mathbf{y}(\infty).$$

At last, since $\|\tilde{L}^{-1}\|_1$ is bounded and does not depend on ε and $\|\tilde{\Delta P}\mathbf{y}(\infty)\|_1 \leq 2$, $\exists c \in \mathbb{R}_{\geq 0}$ and $\beta(\varepsilon) \in \mathcal{K}_\infty$ such that

$$\|\mathbf{y}(\infty) - \overline{\mathbf{y}}\|_1 \leq \varepsilon\|\tilde{L}^{-1}\tilde{\Delta P}\mathbf{y}(\infty)\|_1 \leq \varepsilon c =: \beta(\varepsilon). \tag{26}$$

□

We can now reflect on the similarities between a Q-Learning (or other value function iteration) strategy for finding optimal policies on a MDP and our weight-based foraging problem, as discussed in the introduction. There is a parallelism between the Q values associated to state-action pairs and the weight values (pheromones) associated to vertices on our graph: In both cases they represent the "utility" of a state. However, in our case, by taking the mean field limit we can study the limit distribution of agents interacting with this utility field, as well as the utility values themselves. Additionally, in the mean field limit we can derive deterministic guarantees about both the *distribution* of agents around the graph (i.e. the distance $\|\mathbf{y}(\infty) - \overline{\mathbf{y}}\|_1$) and the *trajectories* of the agents given by the matrix $P(\infty,\varepsilon)$.

VI. Experiments

Some experiments are here presented to verify the results presented in Section V. All experiments were performed over a 20×20 triangular lattice graph, which has $\min_i g^{out}_i = 2$, $\max_i g^{out}_i = 6$ and $\delta^* = 31$. The parameters used are presented in Table I. It is worth mentioning that the amount of "parameter tuning" applied is minimal. The guarantees from Section V ensure the mean field process converges to (a neighbourhood of) a set of vertices along the shortest path, and we choose parameters to obtain representative results when having a finite amount of agents in the graph. We picked $\lambda \approx 1$ to have high diffusion, $r = 5$ to be significantly higher than the unitary reinforcement in $R(t,r,\lambda)$, $\varepsilon = 0.5$ to have an average exploration rate and $\rho = 0.005$ since this yields an evaporation of $(1-\rho)^{4\delta^*} \approx 1/2$.

A. Mean Field Process

In Figures 2 and 3 we present the results for two different scenarios of simulations for a mean field system Φ:

Parameters			
ρ	λ	r	ε
0.005	0.9	5	0.5

TABLE I
SIMULATION PARAMETERS

1) One without obstacles where the shortest path is a perfect line between nest (red triangle) and food source (upside red triangle).
2) One with a sample (non-convex) obstacle where the shortest path (or collection of paths) has to go around it on the right side.

At every vertex we plot the value $\mathbf{w}_i^1(t) - \mathbf{w}_i^2(t)$, with the color bar representing the values of the last plot. In this way we can see the vertices that have a higher overall weight corresponding to each goal. The number of agents is then proportional to the size of the red markers on the vertices. The behaviour of the system is specially interesting in the case of the obstacles in Figure 3. In the first few time-steps there is a random exploration taking place, and very early (around $t = 40$) the agent density starts accumulating in the diagonal line outwards from the nest, indicating that the shortest path is starting to be exploited. Soon after (around $t = 120$) the shortest connecting path can already be observed, but the agent distribution presents oscillations. After enough time-steps the oscillations dampen (since the graph has enough odd length cycles) and the distribution converges to $\mathbf{y}(\infty)$. It is important to remark the convergence speed of the mean field dynamics; $\mathcal{G}$ has 452 vertices, and the agent density has converged to the shortest path in about 300 time-steps.

In Figure 4 we present the temporal trajectories for the mean field system compared to the optimal vector $\overline{\mathbf{y}}$, for different values of ε. Plots for different ρ and λ values are not included since these parameters did not have an impact on the mean field results. To better isolate the effect of every parameter in the system, the influence of ε is only studied on the mean field system, and later a fixed value of ε is applied to the rest of the experiments.

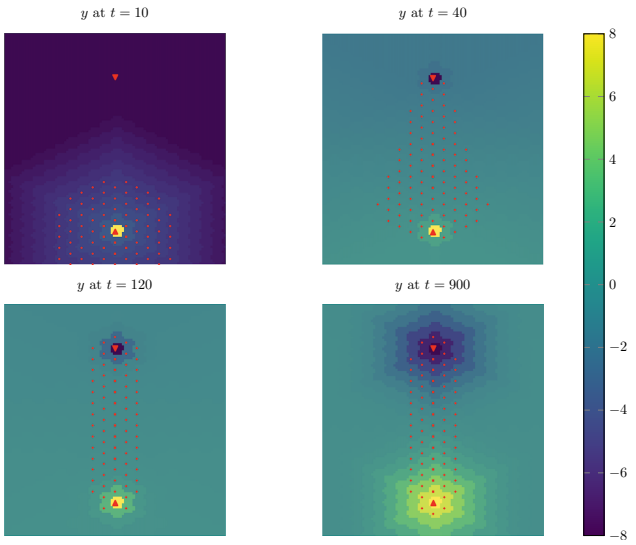

Fig. 2. Mean Field results without obstables

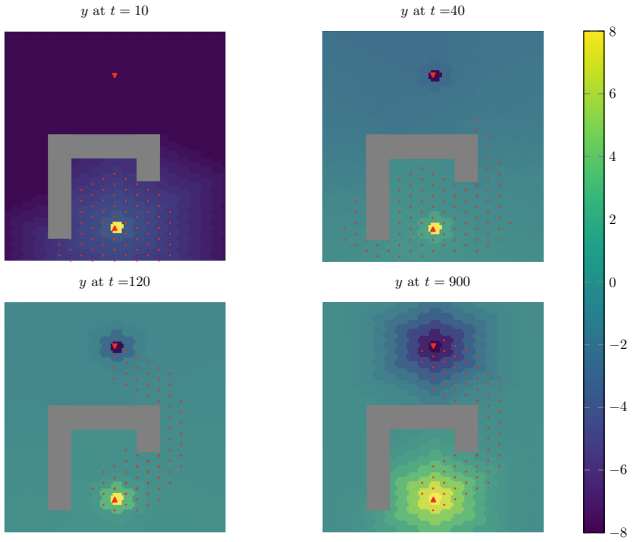

Fig. 3. Mean Field results with obstacles

B. Finite Agents vs. Mean Field

We compare now the results obtained from a mean field approximation system to the ones obtained when using a finite number of agents. Figures 5 and 6 show a similar scenario from the mean field case, but in this case for a finite number of agents. As it can be seen, both cases take longer to achieve convergence to the shortest path. Additionally the agents concentrate over wider regions, and some are "trapped" in irrelevant parts of the graph. Other examples in literature [30] solve this by re-setting the agent position if they have not found the goal vertices over a too long period of time.

We now study the impact of having a reduced number of agents compared to the optimal solutions obtained in the mean field case. Let us for this define an error random variable:

$$\nu(t, n) := \hat{\mathbf{q}}(t, n) - \mathbf{y}(\infty),$$

and, finite sample expectation and variance as

$$\hat{E}[\nu(t,n)] := \frac{1}{K}\sum_{k=1}^{K}\nu(t,n),$$

$$\hat{\mathrm{Var}}[\nu(t,n)] := \frac{1}{K}\sum_{k=1}^{K}(\nu(t,n) - \hat{E}[\nu(t,n)])^2.$$

In Figures 7 and 8 show the results over $K = 5000$ runs. As expected from Theorem 4, both the mean and the variance approach zero for large times as n increases. Interestingly, they both exhibit a peak value after a few time-steps into the

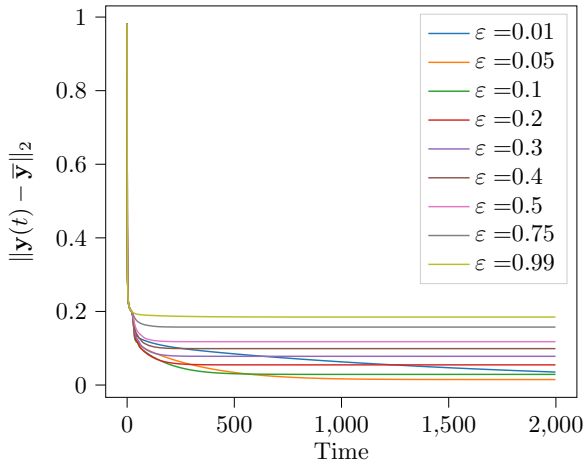

Fig. 4. Mean Field trajectories compared to $\overline{\mathbf{y}}$

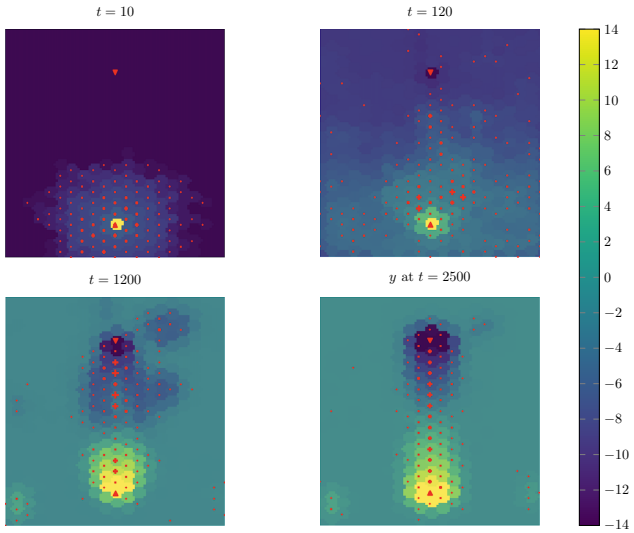

Fig. 5. Discrete agent swarm with $n = 600$

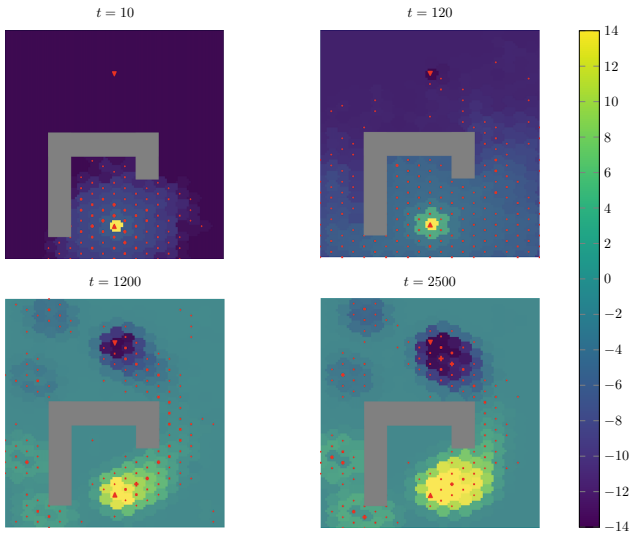

Fig. 6. Discrete agent swarm with $n = 600$ and obstacles

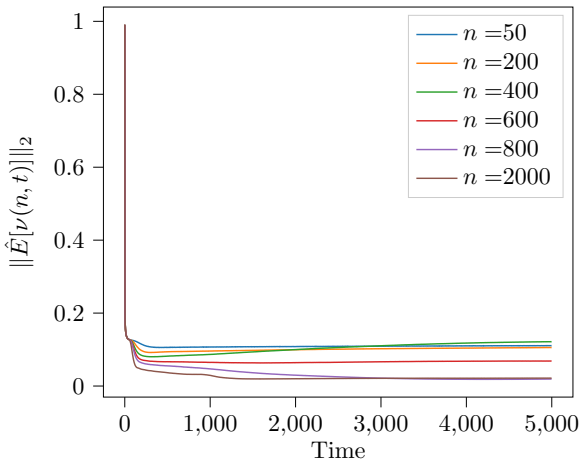

Fig. 7. Sample expectation of $\nu(t, n)$ for different n

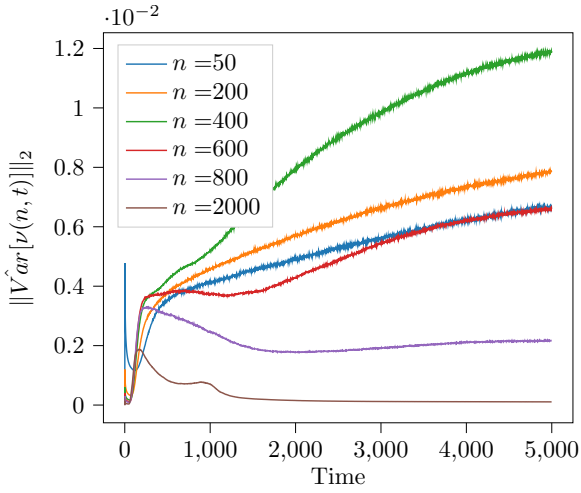

Fig. 8. Sample variance of $\nu(t, n)$ for different n

runs. This is likely due to the fact that when agents find $\mathcal{T}$ the weights change quite fast since reward is added to $\mathcal{G}^2$ suddenly, and the stochastic system runs may be prone to differ more from each other.

C. Interpretation of Variance results

Figures 7 and 8 show the norms of the finite sample (error) expectation and the finite sample variance. As expected, for large numbers of agents the plots go to zero relatively quickly. However, it is interesting to note that the variance and expectation of error increase with n until $n \approx 600$. A possible justification for this is that there is a threshold under which more agents cause more disorder, but not necessarily better solutions. Looking at the variance values at $t = 5000s$, the first curve for $n = 50$ settles around $0.6 \cdot 10^{-2}$, and the following curves for $n = 200$, $n = 400$ go up until around 10^{-2}. This indicates that the variance increases for a range of n values, until a certain threshold where it decreases until approaching 0 for $n > 1000$.

r	ρ	n	$\|\hat{E}[\nu(n,\overline{t})]\|_2$	$\|\hat{\mathrm{Var}}[\nu(n,\overline{t})]\|_2$
20	0.005	200	0.112	0.0097
5	0.05	200	0.121	0.0140
5	0.005	200	0.105	0.0078
20	0.005	800	0.028	0.0006
5	0.05	800	0.022	0.0003
5	0.005	800	0.018	0.0021

TABLE II
SIMULATION RESULTS

Table II displays the end results of different combination of parameters over 5000 runs, for $\overline{t} = 5000$ and fixed $\varepsilon = 0.5$, $\lambda = 0.99$. In general, lower ρ values and larger agent numbers seem to cause smaller variances and smaller $\nu(t,n)$ values. However, for large swarms ($n = 800$) decreasing the evaporation results in an increase in variance. This effect seems to be caused by the fact that for large enough swarms, higher evaporation actually pulls agents towards the optimal solutions faster, therefore decreasing the variance (or diversity) in trajectories. Interestingly, the impact of r in $\nu(t,n)$ seems to be small for the tested cases. Further study of this issue is left for future work, since it may have implications on other multi-agent stochastic systems where stochastic processes exhibit couplings that vanish for large number of agents.

VII. DISCUSSION

We have shown throughout this work how a multi-agent collaborative system solving a foraging problem can be approximated by a mean field formulation of the problem when $n \to \infty$. In section V we developed formal results on the convergence and optimality of such mean field foraging system. We are able now to draw a set of conclusions from these results, combined with the experiments in section VI.

First, the mean field foraging system converges to a unique stationary solution, and does so exponentially fast, under the proposed conditions. In fact, the distance between the mean field agent distribution $\mathbf{y}(\infty)$ and the optimal distribution $\overline{\mathbf{y}}$ seems to only depend on the exploration rate ε (see Figure 4, Theorem 6). That is, the evaporation (or learning) rate ρ and the discount factor λ do not have an effect in the stationary solutions, nor in the convergence speed of the mean field system. This can be explained by the fact that ρ and λ act as scaling parameters that do not change the shape of the weight gradients, thus not having an impact on the matrices $P(t,\varepsilon)$. Note as well from the results in Figure 4 that the distance between $\mathbf{y}(t)$ and $\overline{\mathbf{y}}$ shows some linearity with ε as obtained in the bounds of Theorem 6. This indicates that, for collaborative multi-agent systems in stochastic settings, learning rates and discount factors cease to have an impact when considering large numbers of agents. Therefore, the study of mean field limits on such a multi-agent system allows us to de-couple the influence of some parameters, that may only come into play when considering small agent numbers.

Second, lower exploration rates seem to cause a much slower spread of agents along the optimal path, resulting in a slowly damped "wave-like" behaviour, as it can also be seen in the simulation examples in the supplementary multimedia file[1]. These waves are caused by the initial conditions of agents, since all agents start at $\mathcal{S}$ on the first sub-graph, but they are more quickly damped (agents spreading out faster) for higher values of ε. This may have an impact when considering finite agent numbers; if we observe fast oscillations for a set of parameters as $n \to \infty$, there may be reasons to believe that these can result in non-convergent behaviour for finite agents.

We should also remark the interpretation of the mean field limit. By considering $\mathbf{y}(\infty) := \lim_{t\to\infty}(\lim_{n\to\infty}\hat{\mathbf{q}}(t,n))$ we are computing the (limit) behaviour in time of an infinitely large system of agents. Our results do not guarantee, however, that the alternate limit $\lim_{n\to\infty}(\lim_{t\to\infty}\hat{\mathbf{q}}(t,n))$ exists as well. The problem of studying this second limit corresponds to the limitations of a mean field approximation, and the study of stochastic trajectories of the finite agent system. Such study would shine some more light on how agents affect the limit distributions in these systems. It is worth mentioning that the impact of mean field solutions on discrete time MDPs is in itself a whole subject of study (see [46]–[48]), and the interest on such mean field solutions applied to reinforcement learning problems seems to be growing fast in the last years. Knowing more about the relation between the distributions of finite agent systems and their mean field limits will give us tools to design multi-agent systems with guarantees concerning the number of agents needed to solve a specific problem.

ACKNOWLEDGEMENTS

Authors would like to thank G. Delimpaltadakis, G. Gleizer and C. Verdier for the useful discussions. This work was supported by the ERC Starting Grant SENTIENT 755953. Pedro J Zufiria wants to acknowledge financial support from the Spanish Ministry of Science and Innovation, grant PID2020-112502RB / AEI / 10.13039/501100011033.

APPENDIX

A. Graph Doubling Procedure

Consider two identical graphs $\mathcal{G}^1$, $\mathcal{G}^2$ at $t = 0$. Let $\hat{\mathbf{q}}_i^1(t)$ be the agent proportion in vertex $i \in \mathcal{G}^1$ with probability distribution $\hat{\mathbf{y}}^1(t)$, and $\hat{\mathbf{q}}_i^2(t)$ be the agent proportion in $\mathcal{G}^2$ such that $\hat{\mathbf{y}}^2(0) = \mathbf{0}$. Since the agents follow opposite weight fields (agents in graph 1 follow weights of graph 2, and

[1]This work has an attached multimedia file, available at http://ieeexplore.ieee.org, provided by the authors, showing the entire runs of the mean field system for different ε parameters.

viceversa), let us write the following matrices by considering the union of both systems, $\mathcal{G}^1 \cup \mathcal{G}^2$:

$$\mathbf{w}(t) := \begin{pmatrix} \mathbf{w}^1(t) \\ \mathbf{w}^2(t) \end{pmatrix}, P^{\cup}(t,\varepsilon) := \begin{pmatrix} P^2(t,\varepsilon) & 0 \\ 0 & P^1(t,\varepsilon) \end{pmatrix},$$

and $\hat{\mathbf{y}}(t) := \begin{pmatrix} \hat{\mathbf{y}}^1(t) \\ \hat{\mathbf{y}}^2(t) \end{pmatrix}$. Note the reordering of the blocks in $P^{\cup}(t,\varepsilon)$, reflecting the fact that the agents follow opposite weights. The weight dynamics as written in (6) are then

$$\begin{aligned} \mathbf{w}_i^1(t+1) &= (1-\rho)\mathbf{w}_i^1(t) + \rho\hat{\mathbf{q}}_i^1(t,n)R_i(t,r,\lambda), \\ \mathbf{w}_i^2(t+1) &= (1-\rho)\mathbf{w}_i^2(t) + \rho\hat{\mathbf{q}}_i^2(t,n)R_i(t,r,\lambda), \end{aligned} \tag{27}$$

with $\mathbf{w}(0) = \mathbf{w}_0\mathbf{1}$. Agents enter graph $\mathcal{G}^2$ when they find the vertex $\mathcal{T}^1$, and go back to graph $\mathcal{G}^1$ when they find vertex $\mathcal{S}^2$, resulting in two interconnected systems having each an inflow $(\mathbf{u}^1(t), \mathbf{u}^2(t) \in \mathbb{P}^{|\mathcal{V}|})$ and outflow $(\mathbf{v}^1(t), \mathbf{v}^2(t) \in \mathbb{P}^{|\mathcal{V}|})$ of agents exiting and entering the graphs. The dynamics for the agent distribution can be written as

$$\begin{aligned} \hat{\mathbf{y}}^1(t+1) &= P^2(t,\varepsilon)\hat{\mathbf{y}}^1(t) + \mathbf{u}^1(t) + \mathbf{v}^1(t) \\ \hat{\mathbf{y}}^2(t+1) &= P^1(t,\varepsilon)\hat{\mathbf{y}}^2(t) + \mathbf{u}^2(t) + \mathbf{v}^2(t). \end{aligned} \tag{28}$$

Define now the selector matrices $S \in \mathbb{R}^{|V|\times|V|}$ and $T \in \mathbb{R}^{|V|\times|V|}$ as diagonal matrices with $S_{ii}, T_{jj} = 1$ for $i = \mathcal{S}, j = \mathcal{T}$, zero otherwise. If $\mathbf{u}^1(t)$ is the distribution of agents entering graph $\mathcal{G}^1$ from graph $\mathcal{G}^2$ at time t and $\mathbf{v}^1(t)$ is the density of agents leaving $\mathcal{G}^1$, both graphs are interconnected and closed to external inputs, and then:

$$\begin{aligned} \mathbf{u}^1(t) \equiv -\mathbf{v}^2(t) &= SP^1(t,\varepsilon)\hat{\mathbf{y}}^2(t) \\ \mathbf{v}^1(t) \equiv -\mathbf{u}^2(t) &= TP^2(t,\varepsilon)\hat{\mathbf{y}}^1(t). \end{aligned} \tag{29}$$

Therefore, substituting (29) in (28), the agent probability distribution dynamics are given by

$$\begin{aligned} \hat{\mathbf{y}}(t+1) = &\begin{pmatrix} (I-T)P^2(t,\varepsilon) & SP^1(t,\varepsilon) \\ TP^2(t,\varepsilon) & (I-S)P^1(t,\varepsilon) \end{pmatrix}\hat{\mathbf{y}}(t) = \\ &=: P(t,\varepsilon)\hat{\mathbf{y}}(t). \end{aligned} \tag{30}$$

Furthermore, observe that the matrix $P(t,\varepsilon)$ in (30) is also column stochastic. Effectively, we have interconnected the two graphs by the vertices $\mathcal{S}$ and $\mathcal{T}$, and made the agents move according to the opposite pheromones.

B. Proofs

Proof (Proposition 1). Given the bounded probability matrix $P(t,\varepsilon)$, for any edge $(ij) \in \mathcal{E}$, we have $P_{ji}(t,\varepsilon) \geq \varepsilon$. Furthermore, since by Assumption 1 there is at least one odd length cycle, the graph is aperiodic and we can directly invoke results from [31] on convergence of stigmergy swarm probability distributions. In particular, from Theorem 3

$$\exists \mathbf{y}(\infty) : \lim_{t\to\infty} \left(\prod_{t_k=0}^{t} P(t_k,\varepsilon) \right) \mathbf{y}(0) = \mathbf{y}(\infty).$$

Since all positive terms in matrices $P(t,\varepsilon)$ are lower bounded, the product matrix $P^{\infty}(\varepsilon) := \lim_{t\to\infty} \prod_{t_k=0}^{t} P(t_k,\varepsilon)$ is irreducible, and from Theorem 2 the eigenvector $\mathbf{y}(\infty)$ is unique and has strictly positive entries. Additionally, from Theorem 3 we know that the convergence is exponential, with a rate bounded by $\alpha = (1 - \frac{\varepsilon}{1+(g^*-1)\varepsilon}^{1+2\delta^*})^{\frac{1}{1+2\delta^*}}$. □

Proof (Lemma 1). First, since all $P(t,\varepsilon)$ have the positive entries lower bounded by ε and the graph is connected, they are all irreducible and we can infer

$$\begin{aligned} &\left(\prod_{t_k=t_0}^{t_0+2\delta^*} P(t_k,\varepsilon) \right)_{ji} = \\ &= (P(t_0+2\delta^*,\varepsilon)...P(t_0,\varepsilon))_{ji} \geq \varepsilon^{2\delta^*} \ \forall i,j \in \mathcal{V}. \end{aligned} \tag{31}$$

In other words, any vertex is reachable from any other vertex for times larger than $2\delta^*$. Now, making use of (31), and $l_1^T, l_2^T, ..., l_{|\mathcal{V}|}^T$ being the rows:

$$\begin{aligned} &\prod_{t_k=t_0}^{t_0+2\delta^*} P(t_k,\varepsilon) = \begin{pmatrix} l_1^T \\ ... \\ l_{|\mathcal{V}|}^T \end{pmatrix} \Rightarrow \\ &\Rightarrow \prod_{t_k=t_0}^{t_0+2\delta^*} P(t_k,\varepsilon)\mathbf{y}(t_0) = \begin{pmatrix} l_1^T\mathbf{y}(t_0) \\ ... \\ l_{|\mathcal{V}|}^T\mathbf{y}(t_0) \end{pmatrix} \geq \varepsilon^{2\delta^*}\mathbf{1}. \end{aligned}$$

Therefore, for $t_0 = 0$ and $t > 2\delta^*$ we have $\mathbf{y}(t) = \prod_{t_k=0}^{t} P(t_k,\varepsilon)\mathbf{y}(0) \geq \varepsilon^t\mathbf{1}$. Last, from Proposition 1, $\lim_{t\to\infty}\mathbf{y}(t) = \mathbf{y}(\infty) > \mathbf{0}$, therefore

$$t > 2\delta^* \Rightarrow \mathbf{y}(t) > \mathbf{0} \iff \operatorname{sgn}(\mathbf{y}(t)) = \mathbf{1}.$$

□

Proof (Corollary 1). From Theorem 3 we know that the limit $\lim_{t\to\infty}\mathbf{y}(t+1) = \lim_{t\to\infty}\mathbf{y}(t) = \mathbf{y}(\infty)$ exists. Additionally, from Theorem 5 we know that the limit $\lim_{t\to\infty} P(\varepsilon,t) = P(\varepsilon,\infty)$ also exists. Therefore, using the limit product rule:

$$\lim_{t\to\infty}\mathbf{y}(t+1) = \lim_{t\to\infty} P(t,\varepsilon)\mathbf{y}(t) = P(\infty,\varepsilon)\mathbf{y}(\infty) = \mathbf{y}(\infty).$$

□

Proof (Proposition 3). From Theorem 5, the fixed point is $\mathbf{w}(\infty) = (I + \Gamma(r) + \lambda V(W(\infty)))\mathbf{1}$, and recall from Proposition 2 that it is unique. Additionally, $\boldsymbol{\gamma}_{\mathcal{S},\mathcal{T}}(r) = r$ and is 0 for all other vertices, and it can be shown by contradiction (not added here for brevity) that $\operatorname{argmax}_i(\mathbf{w}_i(\infty)) = \mathcal{S}, \mathcal{T}$. Now, to prove the proposition we assume the following structure for $\mathbf{w}(\infty)$, and later show it is indeed a solution (and therefore the only one, since it is unique). Let us assume for $\mathbf{w}(\infty)$:

$$v,u \in \mathcal{V}^1 : \delta(\mathcal{S},v) > \delta(\mathcal{S},u) \Rightarrow \mathbf{w}_v(\infty) < \mathbf{w}_u(\infty), \tag{32}$$

and the same holds for the converse $v,u \in \mathcal{V}^2$ with the distance to $\mathcal{T}$. That is, if v is one step further away from $\mathcal{S}$ than u, then it has a smaller weight value. Now recall

$$\mathbf{w}_i(\infty) = (1 + \boldsymbol{\gamma}_i(r) + \lambda \max_{j\in\mathcal{V}} \mathbf{w}_{ij}(\infty)), \tag{33}$$

and $\boldsymbol{\gamma}_i(r) = 0 \; \forall \; i \neq \mathcal{S}, \mathcal{T}$. Then, $\forall j \in \mathcal{V} : \delta(\mathcal{S}, j) = 1$:

$$\begin{aligned} \mathbf{w}_j(\infty) =& (1 + \lambda \max_{k \in \mathcal{V}} \mathbf{w}_{jk}(\infty)) = (1 + \lambda \mathbf{w}_{\mathcal{S}}(\infty)), \\ \mathbf{w}_{\mathcal{S}}(\infty) =& (1 + r + \lambda \max_{k \in \mathcal{V}} \mathbf{w}_{ik}(\infty)) = (1 + r + \lambda \mathbf{w}_j(\infty)). \end{aligned} \tag{34}$$

Solving (34) for both weights we obtain

$$\mathbf{w}_{\mathcal{S}}(\infty) = \frac{1 + r + \lambda}{1 - \lambda^2}, \; \mathbf{w}_j(\infty) = \frac{1 + \lambda(1 + r)}{1 - \lambda^2}. \tag{35}$$

Therefore, $r > 0 \Rightarrow \mathbf{w}_{\mathcal{S}}(\infty) > \mathbf{w}_j(\infty) \forall j : \delta(i, j) = 1$. Then, for any $k \in \mathcal{V}^1$, $k \neq \mathcal{S}$,

$$\begin{aligned} &\mathbf{w}_k(\infty) = 1 + \lambda \max_{l \in \mathcal{V}} \mathbf{w}_{kl}(\infty) = 1 + \lambda + \lambda^2 \max_{m \in \mathcal{V}} \mathbf{w}_{lm}(\infty) \\ =& ... = \sum_{a=1}^{\delta(\mathcal{S},k)} \lambda^{a-1} + \lambda^{\delta(\mathcal{S},k)} \mathbf{w}_i(\infty) = \\ =& \sum_{a=1}^{\delta(\mathcal{S},k)} \lambda^{a-1} + \frac{\lambda^{\delta(\mathcal{S},k)} (1 + r + \lambda)}{1 - \lambda^2} = \frac{1 + \lambda + \lambda^{\delta(\mathcal{S},k)} r}{1 - \lambda^2}, \end{aligned} \tag{36}$$

and the same holds for any $k \in \mathcal{V}^2$ with the distance $\delta(\mathcal{T}, k)$. Observe (36) yields an explicit solution to the fixed point $\mathbf{w}(\infty)$ that satisfies the assumption in (32). From Proposition 2, this is the only solution, thus (32) indeed holds for the fixed point and graphs considered. Finally, by construction (36) guarantees that picking the neighbouring maximum weight $\mathbf{w}(\infty)$ from any $v \in \mathcal{V}$ leads to $\mathcal{S}$ (or $\mathcal{T}$) through the minimum distance path, i.e. $\mathbf{w}(\infty) \in \mathcal{W}^*$. □

Proof (Proposition 4). From Definition 5, if $\overline{\mathbf{y}}$ is the eigenvector of $P(\infty, 0)$ corresponding to the eigenvalue 1,

$$\begin{aligned} &P(\infty, 0)\overline{\mathbf{y}} = \overline{\mathbf{y}} \Leftrightarrow \\ \Leftrightarrow &\begin{cases} (I - T)P^{\nabla}(\mathbf{w}^2(\infty))\overline{\mathbf{y}}^1 + SP^{\nabla}(\mathbf{w}^1(\infty))\overline{\mathbf{y}}^2 = \overline{\mathbf{y}}^1 \\ TP^{\nabla}(\mathbf{w}^2(\infty)))\overline{\mathbf{y}}^1 + (I - S)P^{\nabla}(\mathbf{w}^1(\infty))\overline{\mathbf{y}}^2 = \overline{\mathbf{y}}^2. \end{cases} \end{aligned} \tag{37}$$

Recall Remark 4. Since we are considering the full doubled graph with $|\mathcal{V}| = 2|\mathcal{V}_1| = 2|\mathcal{V}_2|$ (that is, with all $\mathcal{T}_1, \mathcal{S}_1, \mathcal{T}_2, \mathcal{S}_2 \in \mathcal{V}$), there are two vertices in the graph that are effectively disconnected from the rest, namely $\mathcal{T}_1$ and $\mathcal{S}_2$. Therefore, $y_{\mathcal{T}_1}(t) = y_{\mathcal{S}_2}(t) = 0 \quad \forall t$. Similarly,

$$\mathcal{T}_1, \mathcal{S}_2 \notin (\cup p^*_{ST}) \cup (\cup p^*_{TS}) \Rightarrow \overline{\mathbf{y}}_{\mathcal{T}_1} = \overline{\mathbf{y}}_{\mathcal{S}_2} = 0. \tag{38}$$

Let us focus on the first equality in (37). Recall $\overline{\mathbf{y}}^1_{\mathcal{S}} = \overline{\mathbf{y}}^2_{\mathcal{T}} = \frac{1}{2k}$ and $\overline{\mathbf{y}}^1_{\mathcal{T}} = \overline{\mathbf{y}}^2_{\mathcal{S}} = 0$. Let us now verify that $P^{\nabla}(\mathbf{w}^1(\infty))\overline{\mathbf{y}}^2 = \overline{\mathbf{y}}^2$ for all vertices $v \neq \mathcal{S}, \mathcal{T}$. Recall from Definition 4 that $P^{\nabla}_{ji}(\mathbf{w}^1(\infty)) = \frac{1}{|N_i^{out}|} \; \forall j \in N_i^{out}$. Then, for any $v \neq \mathcal{S}, \mathcal{T}$,

$$\left(P^{\nabla}(\mathbf{w}^1(\infty))\overline{\mathbf{y}}^2\right)_v = \sum_{j \in N_v^{in}} \frac{\overline{\mathbf{y}}^2_j}{|N_j^{out}|}. \tag{39}$$

Substituting now $\overline{\mathbf{y}}^2_j = \frac{1}{2k} \sum_{p \in \pi_{Tj}} \prod_{u \in p \setminus j} \frac{1}{|N_u^{out}|}$ in (39):

$$\begin{aligned} \sum_{j \in N_v^{in}} \frac{\overline{\mathbf{y}}^2_j}{|N_j^{out}|} =& \sum_{j \in N_v^{in}} \frac{1}{2k} \left(\sum_{p \in \pi_{Tj}} \prod_{u \in p \setminus j} \frac{1}{|N_u^{out}|} \right) \frac{1}{|N_j^{out}|} = \\ =& \frac{1}{2k} \sum_{j \in N_v^{in}} \sum_{p \in \pi_{Tj}} \prod_{u \in p} \frac{1}{|N_u^{out}|}. \end{aligned} \tag{40}$$

Since all $j \in N_v^{in}$ lead to v, (40) is simply

$$\frac{1}{2k} \sum_{j \in N_v^{in}} \sum_{p \in \pi_{Tj}} \prod_{u \in p} \frac{1}{|N_u^{out}|} = \frac{1}{2k} \sum_{p \in \pi_{Tv}} \prod_{u \in p \setminus v} \frac{1}{|N_u^{out}|} = \overline{\mathbf{y}}^2_v, \tag{41}$$

and $\left(P^{\nabla}(\mathbf{w}^1(\infty))\overline{\mathbf{y}}^2\right)_v = \overline{\mathbf{y}}^2_v$. Similarly,

$$\left(P^{\nabla}(\mathbf{w}^2(\infty))\overline{\mathbf{y}}^1\right)_v = \overline{\mathbf{y}}^1_v \quad \forall v \neq \mathcal{S}, \mathcal{T}, \tag{42}$$

and $\overline{\mathbf{y}}^1_{\mathcal{T}} = 0$. Now observe

$$\left(SP^{\nabla}(\mathbf{w}^1(\infty))\overline{\mathbf{y}}^2\right)_i = \begin{cases} \left(P^{\nabla}(\mathbf{w}^1(\infty))\overline{\mathbf{y}}^2\right)_i & \text{if } i = \mathcal{S}, \\ 0 \quad \text{else.} \end{cases} \tag{43}$$

From Proposition 3, we know that $\mathbf{w}^1_{\mathcal{S}}(\infty) = \max_j \mathbf{w}_j(\infty) \Rightarrow P^{\nabla}_{\mathcal{S}i}(\mathbf{w}^1(\infty)) = 1 \; \forall \; (i\mathcal{S}) \in \mathcal{E}$. Since all paths $p \in \pi_{TS}$ start and end at the same vertices and have the same length, recall $\frac{1}{|N_u^{out}|}$ can be interpreted as the probability of moving out of u, therefore the product $\Pr\{p\} := \prod_{u \in p \setminus \mathcal{S}} \frac{1}{|N_u^{out}|}$ is the probability of following the entire path p, and it holds that

$$\sum_{p \in \pi_{TS}} \prod_{u \in p \setminus \mathcal{S}} \frac{1}{|N_u^{out}|} = \sum_{p \in \pi_{TS}} \Pr\{p\} = 1, \tag{44}$$

Therefore, by making use of (44), we can compute (43):

$$\begin{aligned} &\left(SP^{\nabla}(\mathbf{w}^1(\infty))\overline{\mathbf{y}}^2\right)_{\mathcal{S}} = \sum_{j \in N_{\mathcal{S}}^{in}} \frac{\overline{\mathbf{y}}^2_j}{|N_j^{out}|} = \\ =& \sum_{j \in N_{\mathcal{S}}^{in}} \frac{1}{2k} \left(\sum_{p \in \pi_{Tj}} \prod_{u \in p \setminus j} \frac{1}{|N_u^{out}|} \right) \frac{1}{|N_j^{out}|} = \\ =& \frac{1}{2k} \sum_{j \in N_{\mathcal{S}}^{in}} \sum_{p \in \pi_{Tj}} \prod_{u \in p} \frac{1}{|N_u^{out}|} = \\ =& \frac{1}{2k} \sum_{p \in \pi_{TS}} \prod_{u \in p \setminus \mathcal{S}} \frac{1}{|N_u^{out}|} = \frac{1}{2k} = \overline{\mathbf{y}}^1_{\mathcal{S}}. \end{aligned} \tag{45}$$

At last, combining (42) and (45) we have

$$(I - T)P^{\nabla}(\mathbf{w}^2(\infty))\overline{\mathbf{y}}^1 + SP^{\nabla}(\mathbf{w}^1(\infty))\overline{\mathbf{y}}^2 = \overline{\mathbf{y}}^1, \tag{46}$$

and analogously one can show that the same holds for the second equation in (37). Therefore, $P(\infty, 0)\overline{\mathbf{y}} = \overline{\mathbf{y}}$. □

REFERENCES

[1] G. Beni and J. Wang, "Swarm intelligence in cellular robotic systems," in *Robots and biological systems: towards a new bionics?* Springer, 1993, pp. 703–712.

[2] M. Dorigo, M. Birattari *et al.*, "Swarm intelligence." *Scholarpedia*, vol. 2, no. 9, p. 1462, 2007.

[3] C. Blum and D. Merkle, "Swarm intelligence," *Swarm Intelligence in Optimization; Blum, C., Merkle, D., Eds*, pp. 43–85, 2008.

[4] J. Kennedy, "Swarm intelligence," in *Handbook of nature-inspired and innovative computing*. Springer, 2006, pp. 187–219.

[5] O. Zedadra, N. Jouandeau, H. Seridi, and G. Fortino, "Multi-agent foraging: state-of-the-art and research challenges," *Complex Adaptive Systems Modeling*, vol. 5, no. 1, pp. 1–24, 2017.

[6] P.-P. Grassé, "La reconstruction du nid et les coordinations interindividuelles chezbellicositermes natalensis etcubitermes sp. la théorie de la stigmergie: Essai d'interprétation du comportement des termites constructeurs," *Insectes sociaux*, vol. 6, no. 1, pp. 41–80, 1959.

[7] G. Bernasconi and J. E. Strassmann, "Cooperation among unrelated individuals: the ant foundress case," *Trends in Ecology & Evolution*, vol. 14, no. 12, pp. 477–482, 1999.

[8] C. R. Carroll and D. H. Janzen, "Ecology of foraging by ants," *Annual Review of Ecology and systematics*, vol. 4, no. 1, pp. 231–257, 1973.

[9] J. F. Traniello, "Foraging strategies of ants," *Annual review of entomology*, vol. 34, no. 1, pp. 191–210, 1989.

[10] J. Watmough and L. Edelstein-Keshet, "Modelling the formation of trail networks by foraging ants," *Journal of Theoretical Biology*, vol. 176, no. 3, pp. 357–371, 1995.

[11] F. Schweitzer, K. Lao, and F. Family, "Active random walkers simulate trunk trail formation by ants," *BioSystems*, vol. 41, no. 3, pp. 153–166, 1997.

[12] B. Meyer, "A tale of two wells: noise-induced adaptiveness in self-organized systems," in *2008 Second IEEE International Conference on Self-Adaptive and Self-Organizing Systems*. IEEE, 2008, pp. 435–444.

[13] H. de Vries and J. C. Biesmeijer, "Modelling collective foraging by means of individual behaviour rules in honey-bees," *Behavioral Ecology and Sociobiology*, vol. 44, no. 2, pp. 109–124, 1998.

[14] D. Sumpter and S. Pratt, "A modelling framework for understanding social insect foraging," *Behavioral Ecology and Sociobiology*, vol. 53, no. 3, pp. 131–144, 2003.

[15] M. Resnick, *Turtles, termites, and traffic jams: Explorations in massively parallel microworlds*. Mit Press, 1997.

[16] M. Dorigo, V. Maniezzo, and A. Colorni, "Ant system: optimization by a colony of cooperating agents," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 26, no. 1, pp. 29–41, 1996.

[17] M. Dorigo and C. Blum, "Ant colony optimization theory: A survey," *Theoretical Computer Science*, vol. 344, no. 2, pp. 243 – 278, 2005. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0304397505003798

[18] M. A. Hsieh, Á. Halász, S. Berman, and V. Kumar, "Biologically inspired redistribution of a swarm of robots among multiple sites," *Swarm Intelligence*, vol. 2, no. 2, pp. 121–141, 2008.

[19] A. Drogoul and J. Ferber, "Some experiments with foraging robots," in *From Animals to Animats 2: Proceedings of the Second International Conference on Simulation of Adaptive Behavior*, vol. 2. MIT Press, 1993, p. 451.

[20] K. Sugawara, T. Kazama, and T. Watanabe, "Foraging behavior of interacting robots with virtual pheromone," in *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)(IEEE Cat. No. 04CH37566)*, vol. 3. IEEE, 2004, pp. 3074–3079.

[21] R. Fujisawa, S. Dobata, K. Sugawara, and F. Matsuno, "Designing pheromone communication in swarm robotics: Group foraging behavior mediated by chemical substance," *Swarm Intelligence*, vol. 8, no. 3, pp. 227–246, 2014. [Online]. Available: https://doi.org/10.1007/s11721-014-0097-z

[22] A. Campo, Á. Gutiérrez, S. Nouyan, C. Pinciroli, V. Longchamp, S. Garnier, and M. Dorigo, "Artificial pheromone for path selection by a foraging swarm of robots," *Biological cybernetics*, vol. 103, no. 5, pp. 339–352, 2010.

[23] S. Alers, K. Tuyls, B. Ranjbar-Sahraei, D. Claes, and G. Weiss, "Insect-inspired robot coordination: foraging and coverage," *Artificial life*, vol. 14, pp. 761–768, 2014.

[24] A. Font Llenas, M. S. Talamali, X. Xu, J. A. R. Marshall, and A. Reina, "Quality-sensitive foraging by a robot swarm through virtual pheromone trails," in *Swarm Intelligence*, M. Dorigo, M. Birattari, C. Blum, A. L. Christensen, A. Reina, and V. Trianni, Eds. Cham: Springer International Publishing, 2018, pp. 135–149.

[25] D. Payton, M. Daily, R. Estowski, M. Howard, and C. Lee, "Pheromone robotics," *Autonomous Robots*, vol. 11, no. 3, pp. 319–324, 2001.

[26] S. Garnier, F. Tache, M. Combe, A. Grimal, and G. Theraulaz, "Alice in pheromone land: An experimental setup for the study of ant-like robots," in *2007 IEEE swarm intelligence symposium*. IEEE, 2007, pp. 37–44.

[27] B. Hrolenok, S. Luke, K. Sullivan, and C. Vo, "Collaborative foraging using beacons." in *AAMAS*, vol. 10, 2010, pp. 1197–1204.

[28] C. J. Watkins and P. Dayan, "Q-learning," *Machine learning*, vol. 8, no. 3-4, pp. 279–292, 1992.

[29] N. Monekosso and P. Remagnino, "Phe-q: a pheromone based q-learning," in *Australian Joint Conference on Artificial Intelligence*. Springer, 2001, pp. 345–355.

[30] L. Panait and S. Luke, "A pheromone-based utility model for collaborative foraging," in *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems, 2004. AAMAS 2004.* IEEE, 2004, pp. 36–43.

[31] D. J. Ornia and M. Mazo, "Convergence of ant colony multi-agent swarms," in *Proceedings of the 23rd International Conference on Hybrid Systems: Computation and Control*, ser. HSCC '20. New York, NY, USA: Association for Computing Machinery, 2020. [Online]. Available: https://doi.org/10.1145/3365365.3382199

[32] Y. Qin, M. Cao, and B. D. O. Anderson, "Lyapunov criterion for stochastic systems and its applications in distributed computation," *IEEE Transactions on Automatic Control*, pp. 1–1, 2019.

[33] S. Berman, A. Halász, M. A. Hsieh, and V. Kumar, "Optimized stochastic policies for task allocation in swarms of robots," *IEEE transactions on robotics*, vol. 25, no. 4, pp. 927–937, 2009.

[34] D. A. Gomes *et al.*, "Mean field games models—a brief survey," *Dynamic Games and Applications*, vol. 4, no. 2, pp. 110–154, 2014.

[35] J.-M. Lasry and P.-L. Lions, "Mean field games," *Japanese journal of mathematics*, vol. 2, no. 1, pp. 229–260, 2007.

[36] K. Lerman, A. Martinoli, and A. Galstyan, "A review of probabilistic macroscopic models for swarm robotic systems," in *International workshop on swarm robotics*. Springer, 2004, pp. 143–152.

[37] K. Elamvazhuthi, S. Biswal, and S. Berman, "Mean-field stabilization of robotic swarms to probability distributions with disconnected supports," in *2018 Annual American Control Conference (ACC)*. IEEE, 2018, pp. 885–892.

[38] L. Stella and D. Bauso, "Mean-field games for bio-inspired collective decision-making in dynamical networks," *arXiv preprint arXiv:1802.03435*, 2018.

[39] K. Elamvazhuthi and S. Berman, "Mean-field models in swarm robotics: A survey," *Bioinspiration & Biomimetics*, vol. 15, no. 1, p. 015001, 2019.

[40] R. Diestel, "Graph theory, volume 173 of," *Graduate texts in mathematics*, p. 7, 2012.

[41] A. Gut, *Probability: a graduate course*. Springer Science & Business Media, 2013, vol. 75.

[42] T. C. Gard, *Introduction to Stochastic Differential Equations. Monographs and Text-books in pure and applied mathematics*. Dekker, Inc, 1988.

[43] P. Brémaud, *Markov chains: Gibbs fields, Monte Carlo simulation, and queues*. Springer Science & Business Media, 2013, vol. 31.

[44] A. Dussutour, S. C. Nicolis, G. Shephard, M. Beekman, and D. J. Sumpter, "The role of multiple pheromones in food recruitment by ants," *Journal of Experimental Biology*, vol. 212, no. 15, pp. 2337–2348, 2009.

[45] T. J. Czaczkes, C. Grüter, L. Ellis, E. Wood, and F. L. Ratnieks, "Ant foraging on complex trails: route learning and the role of trail pheromones in lasius niger," *Journal of Experimental Biology*, vol. 216, no. 2, pp. 188–197, 2013.

[46] R. Carmona, M. Laurière, and Z. Tan, "Model-free mean-field reinforcement learning: mean-field mdp and mean-field q-learning," *arXiv preprint arXiv:1910.12802*, 2019.

[47] J. A. Carrillo, Y.-P. Choi, and M. Hauray, "The derivation of swarming models: mean-field limit and wasserstein distances," in *Collective dynamics from bacteria to crowds*. Springer, 2014, pp. 1–46.

[48] N. Gast, B. Gaujal, and J.-Y. Le Boudec, "Mean field for markov decision processes: from discrete to continuous optimization," *IEEE Transactions on Automatic Control*, vol. 57, no. 9, pp. 2266–2280, 2012.

Daniel Jarne Ornia is a PhD Student at the Delft Center for Systems and Control, Delft University of Technology (The Netherlands). He received a B.Sc. degree in Aerospace Engineering from the Polytechnical University of Catalonia (Barcelona, Spain) in 2015 and a M.Sc. in the same field from KTH Royal Institute of Technology (Stockholm, Sweden) in 2017. He started his PhD in 2018, and his research interests include communication and formal methods in multi-agent learning, cooperative robotic systems and mean field theory for stochastic systems.

Pedro J Zufiria was born in Donostia-San Sebastián (Spain) in 1962. He received the Ingeniero de Telecomunicación degree by the Universidad Politécnica de Madrid (UPM) in 1986, the M.Sc. in ME, M.Sc. in EE, and Ph.D. degrees from the University of Southern California in 1989. He also received the Doctor Ingeniero de Telecomunicación degree by the MEC in 1991 and the Licenciado en Ciencias Matemáticas degree by the Universidad Complutense de Madrid in 1997. Since 1987 until 1990 he was Teaching and Research Assistant in USC, collaborating in different projects of the National Science Foundation and TRW/NASA. Since 1990, he is in the Escuela Técnica Superior de Ingenieros de Telecomunicación (ETSIT) of the UPM, where he was Chairman of the Matemática Aplicada a las Tecnologías de la Información Department from 1999 to 2004. In the ETSIT-UPM, in 1999 he was Vice-Dean for Research and Graduate Studies, from 2009 to 2018 co-director of the Orange Chair, and since 2019 is co-director of the Cabify Chair. His research interests focus on the analysis, control and fault diagnosis of dynamical systems, the theory of complex networks, and the study of machine learning paradigms for applications in data processing and sustainable mobility, having authored over 100 international publications in these fields.

Manuel Mazo Jr is an associate professor at the Delft Center for Systems and Control, Delft University of Technology (The Netherlands). He received the Ph.D. and M.Sc. degrees in Electrical Engineering from the University of California, Los Angeles, in 2010 and 2007 respectively. He also holds a Telecommunications Engineering ”Ingeniero” degree from the Polytechnic University of Madrid (Spain), and a ”Civilingenjör” degree in Electrical Engineering from the Royal Institute of Technology (Sweden), both awarded in 2003. Between 2010 and 2012 he held a joint post-doctoral position at the University of Groningen and the innovation centre INCAS3 (The Netherlands). His main research interest is the formal study of problems emerging in modern control system implementations, and in particular the study of networked control systems and the application of formal verification and synthesis techniques to control.