



**Politecnico  
di Torino**

# Behavioural contagion in animal collectives.

Master's Degree in Physics of complex systems.

Supervisors:

**Luca Dall'Asta**

**Pawel Romanczuk**

Candidate:

**Bianca Pacini**

Academic year 2023/2024



# Contents

<b>1</b>	<b>Introduction:</b>	<b>4</b>
<b>2</b>	<b>Background:</b>	<b>7</b>
<b>3</b>	<b>Study system:</b>	<b>11</b>
<b>4</b>	<b>Network Model:</b>	<b>19</b>
4.1	Positioning algorithms: . . . . .	20
4.1.1	Nearest-Neighbors distance distribution (NND): . . . . .	22
4.2	Linking algorithms: . . . . .	23
4.2.1	Metric networks: . . . . .	24
4.2.2	Voronoi networks: . . . . .	25
4.2.3	Topological networks (or k-nearest neighbours): . . . . .	26
4.3	Weights: . . . . .	28
4.4	Network characterization: . . . . .	29
4.5	In degree: . . . . .	29
4.6	Clustering coefficient: . . . . .	30
4.7	Number of triangles: . . . . .	34
<b>5</b>	<b>Simulated Dynamics:</b>	<b>36</b>
5.1	Introduction: . . . . .	36
5.2	Dynamics description: . . . . .	39
5.2.1	Role of weights: . . . . .	43
5.3	Model Implementation: . . . . .	43
<b>6</b>	<b>Percolation transition: SIR model on spatial networks.</b>	<b>45</b>
6.1	Introduction: . . . . .	45
6.2	Metric Network: . . . . .	48
6.3	Voronoi Network: . . . . .	52
6.4	Knn Network: . . . . .	53
6.5	Conclusions: . . . . .	54
<b>7</b>	<b>Mean-Field approximation on SIS dynamics:</b>	<b>56</b>
7.1	Introduction to simplicial complexes: . . . . .	56
7.2	Mean field: . . . . .	59
7.3	Closures: . . . . .	60
7.4	Results: . . . . .	63
7.5	Conclusions: . . . . .	72

<b>8</b>	<b>Using empirically observed data from Mexico trip:</b>	<b>74</b>
8.1	Introduction: . . . . .	74
8.2	Network characterization: . . . . .	77
8.3	Dynamics of the SIR model: . . . . .	81
8.4	Mean-Field analysis: . . . . .	84
8.5	The hexnut experiment: . . . . .	85
	8.5.1 Scope: . . . . .	85
	8.5.2 Set-up: . . . . .	85
	8.5.3 Analysis: . . . . .	87
8.6	Conclusions: . . . . .	92
<b>9</b>	<b>Discussion:</b>	<b>93</b>
<b>10</b>	<b>Conclusions:</b>	<b>99</b>
<b>11</b>	<b>Appendices:</b>	<b>102</b>
	11.1 Square Lattice and Hexagonal Lattice: . . . . .	102
	11.2 Susceptibility measure for metric and KNN: . . . . .	102

# 1 Introduction:

The study of collective decision-making in large, distributed groups of animals is a truly fascinating problem. How often, when looking at the coordinate movements of birds [48], or the intricate trails of ants [3], do we find ourselves wondering: *how do they do it?* The answer to this question bridges the fields of physics and biology, offering insight from neuronal networks [5], genetic regulatory networks [6], all the way to crowds of people [29].

The common feature of all these systems is that they are, in fact, collective. What is most surprising is the emergence of complex phenomena, such as flocking and synchronization, which can be seen as forms of order arising from the interaction of individuals (agents), each of whom has access only to local information. However, what really characterizes these systems is not order, but response [10]. A defining property of living systems, distinguishing them from most inanimate matter, is their ability to adapt reactively. When this adaptation occurs at the group level, it often confers advantages to the participants, and can be seen as an evolutionary trait.

The precise mechanism enabling collectives to achieve such remarkable efficiency, however, remains largely unknown. In the early twentieth century, the British ornithologist Edmund Selous<sup>1</sup>, referring to flocks of birds, said: “*They must think collectively, all at the same time.. a flash out of so many brains*” [52]. Later, the attention had turned from telepathy to synchrony arising from the rapid transmission of local behavioural response to neighbours. This perspective has led to the view of such groups as intelligent entities, laying the foundations of the field now known as *swarm intelligence*.

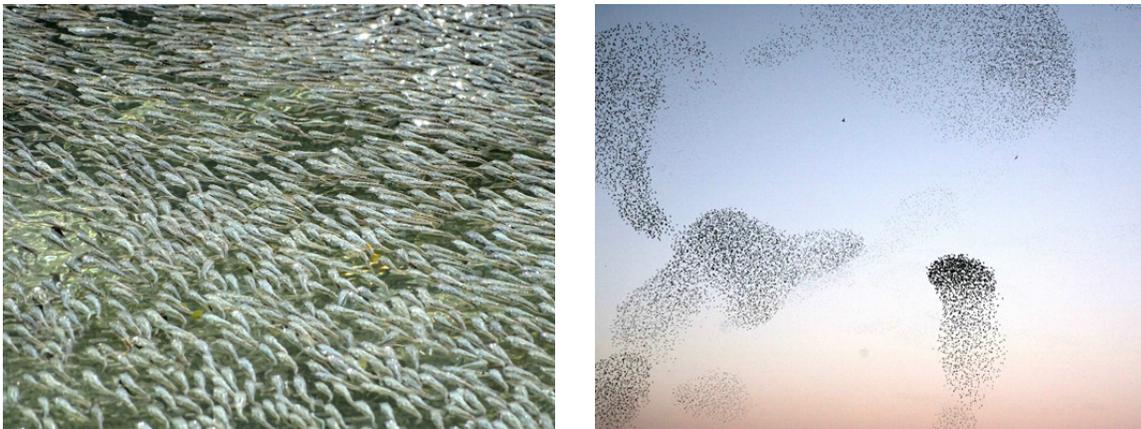


Figure 1: Shoal of sulphur mollies from Wikipedia and starling flocks from [48].

---

<sup>1</sup>Edmund Selous (14 August 1857 – 25 March 1934) was a British ornithologist and writer.

In this work, I focus on a curious predator-prey system involving large fish shoals of sulphur mollies (*Poecilia sulphuraria*), which constitute a perfect example of complex, adaptive system [46, 22, 95]. These shoals are unique in both their behaviour and the environment in which they live. Indeed, these mollies are one of only two species of fish known to inhabit the *Baños del Azufre* in Southern Mexico, a freshwater spring characterized by high temperatures and low oxygen levels, which requires them to swim near the surface for aquatic respiration. While positioned at the surface, they become vulnerable to fish-eating birds; in response, the shoals display repeated synchronized diving behaviour as a defence mechanism, generating ripples on the water surface, which lead to the formation of wave-like patterns, and have been shown to reduce the frequency of attacks [8].

My goal is to identify and analyse the core mechanisms underlying this biological system; a central question is “*How can animal collectives perceive, propagate, process and store information?*”.

To address this matter, and describe the specific biological system under consideration (section 3), I will introduce an SIR quadratic contagion model on a spatially embedded network (section 4, section 5). Here, the term “contagion” refers specifically to behavioural contagion, and a more complex mechanism is proposed to go beyond pairwise connections, considering group interactions too. This accounts for the importance of reinforcement-mechanism in social influence.

The results clearly show that social contagion and spatial organization are deeply interconnected. To explore this interplay, I will define various possible algorithms on which to build the networks, and point-out the differences. By systematically studying the properties and dynamics of this interacting network of fish, we aim to gain deeper insights into the principles that govern collective behaviour in social animals. This model could also provide a foundation for understanding similar processes in other animal groups, or even artificial agent-based systems, designed to mimic biological swarm behaviour.

Given that one of the peculiarities of the system is its ability to respond to danger, one might ask: “*How can fish discriminate between signal and noise?*”. To analyse this point, I will propose a definition for signal and one for noise, by focusing on the concept of spatially correlated activations. Based on this, I will introduce a measure of the system’s susceptibility, which quantifies its responsiveness by comparing perturbations induced by either true or false positives. Let us note here, that a key assumption in the study of collective biological systems is that they operate at (or near) the so-called *critical point*, and this is known in literature as the *criticality hypothesis*. The idea stems from the fact that critical systems exhibit distinct properties, such as maximal responsiveness to external stimuli and optimal propagation of information, making them well-suited in describing what we observe in animal collectives. In this context,

the biological system will be most accurately described by the model when the combination of parameters maximizes susceptibility.

Moreover, given the importance of spatial organization in these phenomena, it is crucial to understand *how the topology of the network affects the dynamics*. To this purpose, I will analyse the Mean-Field approximation (section 7). The results in the case of spatially correlated networks differ from those of the theoretical mean field case, and this confirms the crucial role of heterogeneities in shaping the final outcome of the contagion dynamics.

Finally, in a fieldwork trip to Mexico in May 2024, I obtained some empirical observations of the study system (section 8). The question rises immediately: “*How can we compare these results to real data?*”. In this direction, I will build new networks using the empirically observed positions of the fish. This will of course affect the topology and the dynamics in a non-trivial way. In particular, the quadratic contagion mechanism, which was suggested to be a possible way of discriminating between signal and noise, has no more (or little) influence. In the last section, I will also introduce the hexnut-experiment (section 8.5), whose goal is to test the response of the shoals by increasing an external stimulus and measuring the size of the waves. The results will be compared with those of the model, to estimate its ability to reproduce the behaviour of the system.

When examining such a complex study system, various levels of analysis can arise, depending on the initially posed questions. For this reason, the progression is not strictly linear, but rather follows an approach aimed at shedding light on specific points of interest, ranging from a more theoretical analysis, to a qualitative comparison with experimental data.

To conclude, these results would contribute shifting the fascinating “*collective mind*” metaphor to a more quantitative context.

## 2 Background:

The emergence of a global order is one distinctive trait of animals living in collectives: all individuals within the group synchronize to some extent their behavioural state. This phenomenon is visible in the motion of large groups of animals [7], as well as in crowds of people, like in the case of the famous Mexican wave, *La Ola*, which rose to fame during the 1986 World Cup in Mexico [29]. The presence of this order is easy to detect, but it may have radically different origins, and discovering the underlying mechanism is not straightforward [10].

In general, large-scale collective biological systems have been suggested to operate in a special parameter region where the system’s behaviour undergoes a qualitative change, or phase transition, at a so-called *critical point* [28, 10, 46] between two structurally and functionally different states. More convincing support for the “*criticality hypothesis*” can be obtained by identifying the mechanisms which enable biological systems to self-organize towards criticality. For example, in neuronal systems, synaptic plasticity has been shown to provide such a mechanism [31, 32]. In general, the concept of Self Organized Criticality (SOC), introduced by Per Bak, Chao Tang and Kurt Wiesenfeld (“BTW”) [50], is considered to be one of the mechanisms by which complexity arises in nature [94]. Such type of systems spontaneously organize into a critical state. Bak and colleagues introduced this concept when studying a cellular automaton model, known today as the sand-pile model. Here, we will refer to self-organization in a general sense for macroscopic (group-wide) collective behaviours, which lead to formation of spatio-temporal dynamics and patterns that cannot be attributed solely to the behaviour of isolated individuals or to environmental factors.

In the context of animal groups, the key concepts related to this hypothesis include the following [75]: *collective decision making*, *behavioural contagion*, *synchronization* and *flocking*. Let us explain these concepts one by one, and this will naturally lead us to the introduction of the main ideas of the thesis.

Starting from an initially undecided state, *collective decision making* is the commitment of a majority of a group to a single option [72]. Examples of this process include social insects choosing a new nest site [71], fish schools or baboon troops [73] deciding where to forage. Collective decision making can take place across a variety of spatial and temporal scales but the fundamental dynamics are often captured by simple, non spatial models. Most decision models consider binary decision tasks, which can be modelled by Ising-type models [74] whereas, in case of multi-choice decisions, the Potts model [76] is commonly used. These models typically assume some sort of quorum or threshold interactions, in which agents update their states based on the choices made by their neighbours; these situations are modelled with behavioural algorithms in discrete

time. It has been shown that speed and accuracy of collective decisions are strongly modulated by the distance from a critical point on networks [77].

With the term *behavioural contagion* we indicate the spread of behaviour or information through a group, resulting in behavioural cascades that can rapidly encompass the entire collective. This type of collective dynamics does not correspond to a spontaneous symmetry breaking, but rather to percolation, and is directly related to the non-equilibrium phase transition studied in epidemic spreading.

We have *synchronization* when individuals in a group engage in periodic behaviours [78]. A classic model used to describe this phenomenon is the Kuramoto model [79, 80] in which oscillators influence one another via pairwise coupling that depends on difference between phases.

One of the most easily observed phase transition is the emergence of orientational order due to spontaneous symmetry breaking, known as *flocking*. This transition separates a disordered state, where many individuals move in random directions with a vanishing centre of mass speed, from an ordered flocking state, in which a preferred direction of the group is noticeable, with a non vanishing average momentum of the entire system. A paradigmatic case is provided by European starlings, *Sturnus vulgaris*, which can be observed in many Italian cities where they establish their roosting sites. These birds are able to gather at dusk and swirl with extraordinary spatial coherence. They have been largely studied by Ballerini et al. [26].

The parameter space in which such collective biological systems operate is vast, both due to the complexity of single individuals, and to the large number of individuals that constitute a functional group. Then the relevant question that arises is whether certain combinations of parameters or specific regions within a system are especially suited to a biological function, enabling their collective behaviour to approach an optimal state. In the context of the criticality hypothesis, at such optimal points, statistical physics predicts maximal susceptibility, i.e. the sensitivity of the collective dynamics to any external stimuli. It is worth noticing that the applicability of the concepts of statistical physics is limited by two main aspects:

- Biological systems are out of equilibrium;
- The dimension of biological system is finite, typically of the order of  $10 - 10^3$  individuals, only in some rare cases can reach  $10^6$  individuals, such as large migration movements in animals like pelagic fish e.g. sardines [58] or desert locust [57].

These two observations, while correct, do not refute the significance of phase transitions and criticality when describing finite-sized biological systems. Various aspects of collective information processing, such as correlation lengths and susceptibility still become maximal at quasi-critical points in finite-sized systems [69, 67]. Moreover, even if over the past century, following the introduction of the Ising model, much of the research has focused on understanding systems at thermal equilibrium, the study of self-organization and pattern formation in non-equilibrium systems [30], relevant for biological systems, has gained increasing importance. In this context, criticality might be a plausible principle of distributed information processing in large animal collectives, offering a balance between robustness and adaptability [55]. Indeed, an intriguing property of living systems that distinguishes them from most (if not all) inanimate matter is their ability to react adaptively to changing environments. In general, this capacity relies on distributed processing of information at various levels, characterized by the collective dynamics of a large number of interacting components, or agents, that make up the complex biological system. These groups are able to perform efficient collective information processing, even though each unit or agent typically has sole access to local information. They can perform highly coordinated collective behaviours that confer benefits to the participating individuals by facilitating social information exchange and providing protection from predators [1, 4]. However, in multi-agent systems, group-level and individual-level evolutionary optima are often different, leading to so-called *social dilemmas*, emerging in a broad range of multi-agent evolutionary game theoretic problems. For this reason, understanding how consensus is reached and information is processed within a collective is fundamental to many aspects of social dynamics in animals and humans.

In this context, the social transmission of behavioural change is central to collective behaviour and collective response is the way a group as a whole reacts to its environment; it is crucial for a group (or for subsets of it) to respond coherently to perturbations. For many mobile groups, such as schooling fish and flocking birds, social contagion can be fast, resulting in dramatic waves of response [51]. Waving is indeed a vivid example of collective behaviour occurring in insects, birds, mammals and fish is the formation of global waves, which are direct consequences of the aforementioned processes, and has been typically interpreted as an anti-predator response [59]. These escape waves are an example of more general avalanche processes, which play an important role in the collective dynamics in many biological systems, including spike avalanches in neuronal networks [34] and disease transmission in human or animal populations [38]. In such processes a local change in the state of an individual unit (i.e. spiking of a neuron or the infection of an individual) can trigger the same change in its neighbours, thus spreading through the system like an avalanche [51, 1], with a rate of spreading that

can decrease (subcritical) or increase (supercritical) as the avalanche grows. Across many living systems, we have evidence that the degree of such behavioural spreading is regulated. In neural cultures, adding biochemical regulators that modify excitation and inhibition can force the system to supercritical and subcritical states [53], while macaques, key individuals have been shown to influence how conflict spreads through a colony [54].

One of the first experimental studies of cascading behavioural change was undertaken by Dimitrii Radakov [60]. Radakov hand-traced the paths of each fish, frame-by-frame, revealing that the speed of the “wave of agitation” could propagate much faster than the maximum swim speed of individuals. Using similar methodology, Treherne<sup>2</sup> and Foster [61] studied rapid waves of escape response in marine skaters, describing what they saw as “*the Trafalgar effect*” in reference to the speed of communication, via signalling flags, among ships in the British Navy’s fleet of the Admiral Nelson at the battle of Trafalgar in 1805. Signals observable at a distance allowed information to travel much faster than the ships could move themselves. In many situations, such as when a predator attacks a group [59], or when artificial stimuli are used, it is not possible to differentiate between the propagation of behaviour through social contagion and the propagation of behaviour resulting from direct response to the stimulus, or some combination of both. A fundamental aspect of the self-organization of collectives—from collective decision-making to consensus formation, including coordinated movements is that individuals are limited in terms of perception and cognition. Without direct access to the state of the whole group they must rely on local information. It has been shown [81] that in heterogeneous environments, strongly limited attention capability of individual agents results in higher accuracy with respect to large scale coordination; this is caused by a dynamical spatial “echo chamber”-like effect, where individual attention becomes saturated by social information and non-social cues are largely ignored.

In conclusion, the investigation of collective behaviour offers valuable insights into the mechanisms that drive self-organization, adaptability, and complex patterns across diverse biological systems.

---

<sup>2</sup>John Edwin Treherne (15 May 1929 – 23 September 1989) was an English entomologist who specialized in insect biochemistry and physiology and conducted extensive experimental studies.

### 3 Study system:

The defining property of living systems is their capacity for adaptive responses. Throughout their lifespan, nearly all organisms encounter environmental changes and these variations may occur at different temporal scales or arise unpredictably. For many animals, the initial response to altered conditions is often behavioural adaptation [13]. This behavioural flexibility is particularly vital in highly dynamic environments. These types of fluctuations have the potential to affect not only individuals and populations but also higher community levels through interactions between species, such as competition and predator-prey relationship [95].

In aquatic ecosystems oxygen is often a limiting factor [15].

As part of a research project at the Science of Intelligence Cluster, we<sup>3</sup> are studying the behaviour of naturally occurring schools of the freshwater fish *Poecilia sulphuraria* or Sulphur molly.

Their distinctive behaviour is closely linked to the unique characteristics of the environment in which they live. The specific conditions of their habitat, such as water temperature, and the presence of predators, play a crucial role in shaping their actions and responses. Over time, these environmental factors have influenced their behavioural adaptations, allowing them to thrive in their particular ecological niche. This connection between their behaviour and habitat underscores the importance of their surroundings in driving their evolution and survival strategies.

These fish are endemic to Mexico, specifically to the Baños del Azufre (Grijalva River basin) near Teapa, Tabasco (17°33 N, 93°00 W) 2. The Baños del Azufre are sulfidic springs fed by sulfidic groundwater aquifers with high concentrations of hydrogen-sulfide ( $H_2S$ ) generated from volcanic deposits and bacterial sulfate reduction (e.g., drainages Pichucalco, Tacotalpa and Puyacatengo) [95].

---

<sup>3</sup>When referring to *we* I mean Pawel Romanckuz, Yunus Sevinchan, Carla Vollmoeller, who have participated in the discussion of the model and its implementation, and David Bierbach, Jens Krause and Korbinian Pacher, biologists who helped me understand the biological system and set up the experiments in Mexico.

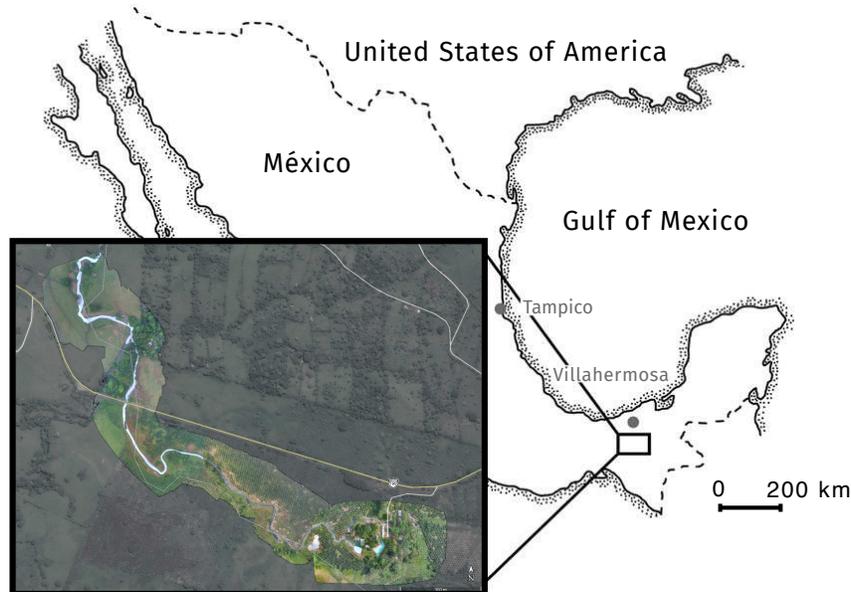


Figure 2: Map of Mexico to see the location of the study system from [95] and [96].

In the case of our study species, the sulphur mollies, sulfidic and non-sulfidic lineages genetically diverged between 15 and 30 ky ago [11]. While several morphological and physiological adaptations allow this species to persist in these conditions, for examples mouth and gill area enlargement for enhanced oxygen uptake, modified toxicity targets and detoxification pathways [112, 12], they are still dependent on aquatic surface respiration (ASR) for survival. A previous study estimated that sulphur mollies spend up to 84% of their time performing ASR [18], which was three times higher than a closely related species (*Poecilia mexicana*) from a nearby, slightly less sulfidic, spring system. The environment in which they live was characterized [19] (see also Fig.fig. 3) as a freshwater habitat with high temperatures ( $31.9 \pm 0.7$ ) and sulfide content ( $H_2S$ :  $190.4 \pm 119.7 \mu\text{mol/L}$ ), low oxygen ( $1.06 \pm 0.92 \text{ mg/L}$ ) and pH ( $6.9 \pm 0.1$ ), and high conductivity ( $2.7 \pm 0.2 \text{ mS/cm}$ ), which shows little variation across the year but some spatial variation due to differences in habitat structure and spring discharge.

Due to their reliance on ASR, the fish spend a majority of their time near the surface, often leading to dense aggregations that result in a quasi-2D system; however, they frequently dive under water to engage in benthic foraging as well as aggressive or reproductive activities (Fig.4). While at water surface, they are more visible and thus susceptible to predation by fish-eating birds; main predators are kingfishers, herons and egrets (see Fig.5). The fish have developed a defence mechanism: the shoals perform

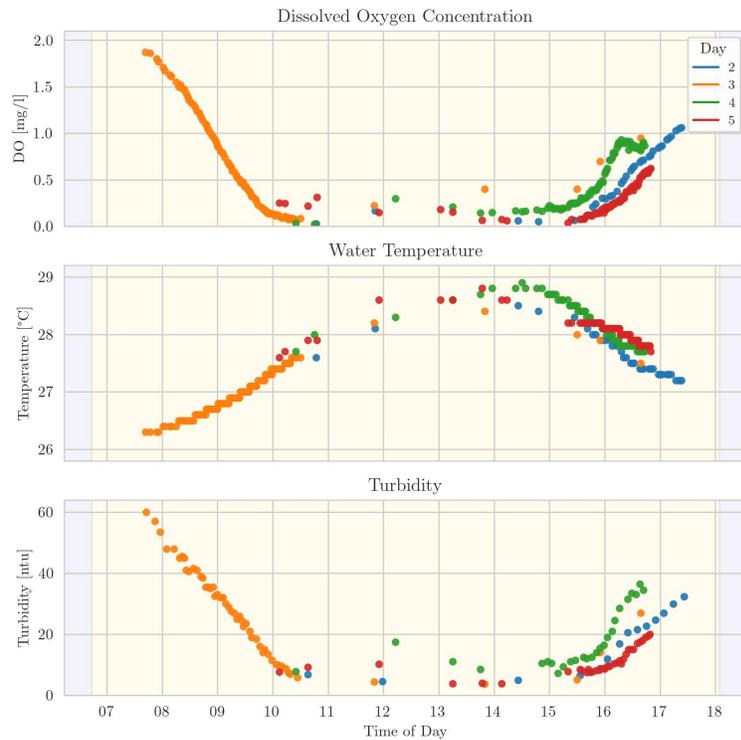


Figure 3: Oxygen concentration, Water temperature and turbidity over the daytime. Data from January 2023, plot made by Yunus Sevinchan and David Bierbach.

collective and synchronized diving as a response to bird attacks, causing ripples on the water surface and wave-like patterns (“surface waves”, see Videos)). The visibility of these waves can be attributed to a two-fold mechanism: individual splashes obviously play a role, but additionally, when the fish perceive risk, they often adopt a C-shape, revealing their light-coloured bellies.

This collective diving may last for an extended period of time and is presumed to reduce the frequency of attacks by predatory birds [8].

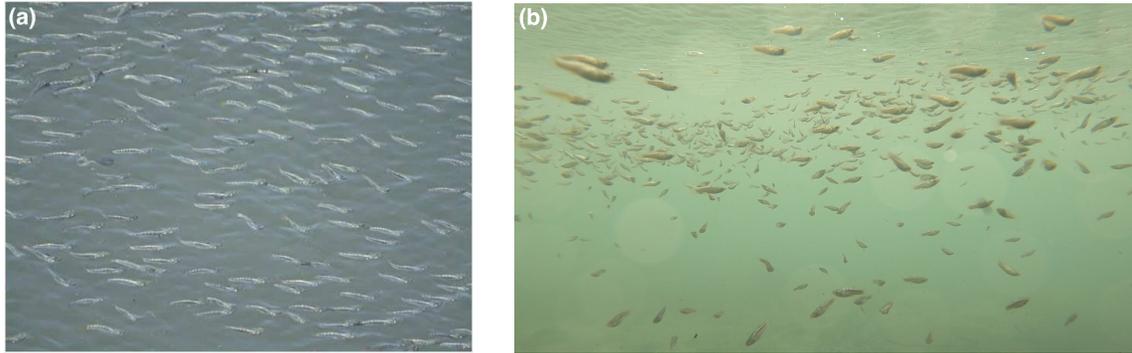


Figure 4: Images from the surface and underwater of the study system. These photos were taken and kindly made available from Juliane Lukas (a) and David Bierbach (b).

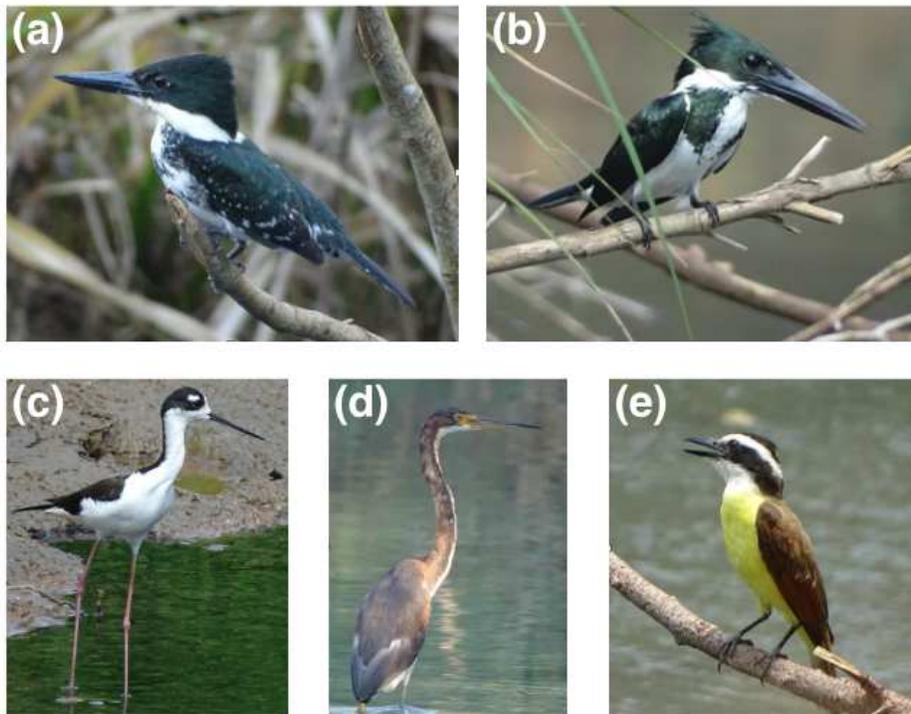


Figure 5: Some of the main predators. On the top left corner (a) the Green Kingfisher (*Chloroceryle americana*) and on the right (b) the Amazon Kingfisher (*Chloroceryle amazona*). On the bottom left (c) the Black-necked stilt (*Himantopus mexicanus*), in the centre (d) the Tricolored heron (*Egretta tricolor*) and on the right (e) the Great kiskadee (*Pitangus sulphuratus*). These photos were taken and kindly made available by Korbinian Pacher (a,b,c,e) and Jens Krause (d).



Figure 6: Image from Wikipedia page “Lateral line”. Oblique view of a goldfish (*Carassius auratus*), showing pored scales of the lateral line system.

Ultimately, these fish face two distinct pressures, which can be described as habitat constraints. On one hand, they must fulfil their need for oxygen, requiring them to engage in active respiration, often by surfacing or staying in oxygen-rich areas; this process demands a significant portion of their time and energy, as maintaining effective respiration is critical for their survival. On the other hand, they are under constant threat from predators, particularly birds, which forces them to remain vigilant and develop strategies to avoid attacks. Balancing these two pressures — ensuring sufficient oxygen intake, while minimizing their vulnerability to predation — presents a complex challenge that shapes their behaviour and overall survival strategies.

These fish have thus developed the remarkable ability of responding to two different types of stimuli. They can detect visual stimuli through their sense of sight, allowing them to process changes in their surroundings; additionally, they can perceive stimuli through the lateral line organ, a specialized sensory system that enables them to detect vibrations and movements in the water [68]. This dual sensory capability enables them to interact effectively with their environment. While the primary characteristics and limitations of their eyesight are relatively intuitive — being influenced by factors such as distance, occlusion, and light availability — the lateral line organ remains less well-known and understood. This organ plays a crucial, yet often overlooked, role in their sensory perception; to better appreciate its significance, let us introduce the structure and function of this fascinating sensory system, which complements their vision and enhances their ability to interact with their environment.

The mechanosensory lateral line (LLO) is a system of sensory organs found in all fish (see Fig.6) and some amphibia, which responds to hydrodynamic stimuli, such as water movement and vibration. The sensory ability is achieved via modified epithelia cells,

known as hair cells, which respond to displacement caused by motion, and transduce these signals into electrical impulses via excitatory synapses. This response is much faster than visually-induced reaction. The time-scales are estimated to be: for LLO (in the case of hydrodynamic sensory input) - 5-20 ms circa, for visual reaction - approximately 120 ms.

In a dense fluid environment, such as water, disturbances are created by anything that moves. Of course, lateral line organs can be very different; sensory structures and central processing have been shaped by evolution to exploit physical properties of water disturbances in order to extract biologically significant information. It is important to notice that LLO appears to be functionally important to all fishes, even those who rely most on sight, like billfish. We also note that other sensory systems are stimulated by mechanical disturbance: tactile receptors and external taste buds, for example. The range of cutaneous senses in fish also reinforces the notion that most behaviour is probably mediated via a synergy of diverse sensory information, rather than being mediated by a single discrete sensory modality such as the lateral line system. In both fish and amphibians, the lateral line input is connected to *Mauthner cells*, a pair of big and easily identifiable neurons that are responsible for a very fast escape reflex, known as *C-start response*, which mediates escape behaviour; this is clear evidence of the importance of the LLO to prey fish, though it must be said that vision is the dominant Mauthner cell input [35]. During active swimming the oscillatory movement of the tail produces a substantial turbulent wake, that can persist in the water for some time after the fish has passed; this vortex wake will produce a potent lateral line stimulus allowing fish that swim into the wake to detect the fish which has passed. In this view, water movements produced by swimming are also used as a form of communication, and the coordination of schooling fish is in part maintained by lateral line stimuli.

A detailed distinction between these two communication pathways lies beyond the scope of this thesis. For the purposes of this work, we will generally assume that they interact through a specific mechanism. However, when discussing the potential linking algorithms (section 4.2), we will explore which approaches are better suited to replicating visual interactions and which are more appropriate for modelling the functioning of the lateral line organ (LLO). This distinction will help identify algorithms that best capture the nuances of each sensory modality.

To conclude this section, it is important to highlight a significant study conducted on our same system. In a notable work by Gomez-Nava et al. [46], the authors examined the baseline behaviour of these fish shoals in the absence of bird attacks. This study is particularly relevant as it strongly supports the criticality hypothesis, which forms a cornerstone of my research. Given its importance and its role in motivating

this work, I will briefly outline the key findings of Gomez-Nava et al. to provide context for the following discussions.

By combining the experimental results with an agent-based model of the system, they were able to show that these fish shoals might operate at a critical point between a state of high individual diving activity and low overall diving activity. To test the criticality hypothesis they acquired videos over multiple days in absence of bird attacks, binarized the original videos into active pixels that represent the diving of fish and non active pixels corresponding to fish at surface, or underwater, and analysed the fraction of active pixels as a proxy for the number of fish diving at a given moment in time. They observed peaks of activity (spikes) corresponding to waves spreading through the system, separated by long periods of low activity corresponding to small-scale non-propagating surface activity. By defining the activity clusters as the number of active pixels connected in time and space corresponding to a single wave, they found that the empirical cluster size distribution is consistent with a power-law distribution with exponent  $\alpha = 2.3$ . However, as we know, the existence of a power-law distribution in empirical data is not sufficient to conclude that a system is at criticality. They also devised a cellular automata model for the spatio-temporal propagation of the surface-activity waves, in which each cell can be one of three states. In this way they simulated the emergent collective dynamics and identified parameters best fitting the observed macroscopic behaviour.

To identify the phase transition they:

- Computed the average correlation of fluctuations between neighbouring cells, which peaks at the phase transition;
- Computed the susceptibility of the surface-activity signal as a function of system size ( $L \in [25, 50, 100, 200, 400, 800]$ ): they found that the susceptibility increases as a function of system size in a way that is consistent with a power-law with exponent  $\approx 1.7$ ;
- Computed the susceptibility in the empirical videos using different window sizes, finding that it increases with said size. The exponents vary across measurement days — potentially due to environmental factors such as temperature, or lighting conditions — but they are consistent with the exponent in the model for at least three measurement days.

This means basically that susceptibility is scale-free, which is one of main indicator of criticality in biological systems [10].

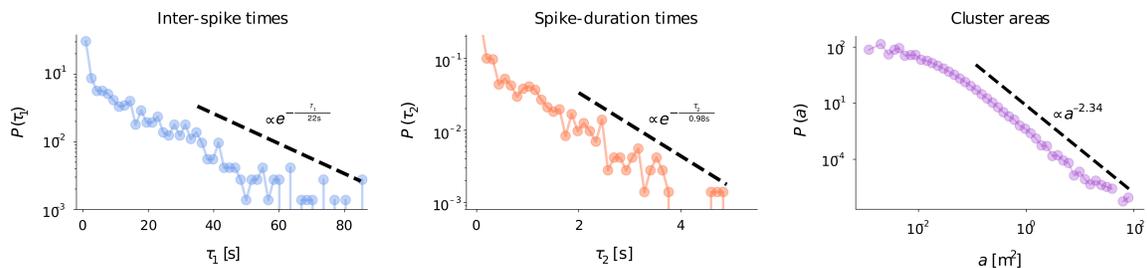


Figure 7: Results for inter-spike times, spike duration and cluster areas by Gomez et al. [46].

These findings provide strong evidence that the collective dynamics of the fish shoals are governed by critical phenomena, where the system is poised between high and low activity states.

Building on that research, we aim to better understand how these fish respond in different ecological contexts, maximizing their ability to discriminate among different environmental stimuli. We are combining experimental observations with computational approaches to understand how these fish tune the responsiveness of the school in a self-organised way, e.g. through individual-level changes in sensitivity to predator stimuli and social cues.

## 4 Network Model:

We will propose a contagion dynamics on a network model. It is necessary to introduce some of basic concepts of network theory and then explain what a contagion model is. Loosely speaking, a network is a collection of nodes connected in some way. Representing the system as a network means mapping elements of the system to nodes and relations among such elements, called links. A network is thus an interconnected structure where nodes and link have some meaning. The object obtained by stripping a network of the meaning of its elements is what in mathematics is called a graph, nodes are known as vertices and links as edges.

A graph  $G=(V,E)$  is a pair formed by a finite set  $V=\{1,\dots,N\}$  of vertices and a set  $E=\{e_1,\dots,e_m\}$  of pairs of vertices called edges. As a combinatorial object a graph is one of the possible ways in which  $N=|V|$  elements can be arranged into  $M = |E|$  subsets of 2 elements each. The graph is encoded in its adjacency matrix  $A$ , a  $N \times N$  matrix whose elements  $A_{ij}$  are such that:

$$A_{ij} = \begin{cases} 1 & \text{if } e = \{i,j\} \text{ exists} \\ 0 & \text{otherwise} \end{cases}$$

In general for a weighted graph instead of 1's or 0's there would be some positive or negative numbers. Moreover, there are also directed graphs where links have a direction.

A subset of vertices where each vertex is connected to any other vertex is called clique. If a clique is not included in any larger clique is called maximal. A n-cycle is a closed chain of n edges joining n vertices. A tree is a graph with no cycles. A graph is bipartite if the vertex set is partitioned into two subsets representing two types of vertices and so the edges only connect vertices of different type.

In this project, we define a network where nodes represent individual fish, and links represent interactions or communication pathways among them. The network is static since neither node positions nor links change. Indeed, there exists a time-scale separation where the spread of information is much faster than any change in the interaction network [47]. Moreover, we assume that individual fish in the shoal are interchangeable and the macroscopic structure of the fish shoal remains similar over time, despite fish changing their position within the shoal [115].

Under this hypothesis, fish are treated as agents in a spatially embedded network, allowing us to analyse how they interact within a shared environment. A key focus of our research is to examine the response of this network to specific stimuli, which, in this case, involves the activation of selected nodes within the network. By activating certain nodes, we simulate a stimulus that might represent a natural trigger, such as

the perception of a predator or an environmental change, and observe how this influence propagates throughout the network. This analysis is done by defining a contagion model, which describes the way a stimulus can spread within the network (section 5). To thoroughly understand this phenomenon, our study will involve:

- Characterizing the structural properties of the network: This includes analysing metrics such as degree distribution and clustering, which provide insights into the connectivity and resilience of the network;
- Exploring the dynamics of information flow: This analysis will help us identify key patterns in response dynamics that are crucial for coordinated group behaviours.

One of the central challenges is establishing a consistent and biologically plausible rule for constructing the network. Coordination among social animals, like fish, relies on efficient and rapid information transfer between individuals, which may depend on the structural characteristics of the network they form [89]. The topology of the network, specifically its connectivity patterns and levels of heterogeneity, plays a significant role in shaping how effectively information can spread. For example, a densely connected network might facilitate rapid information dissemination but could also lead to congestion or excessive sensitivity to local disturbances. Furthermore, the structure and heterogeneity of the network have profound implications for the dynamics of contagious processes [114, 17], such as the spread of information, behaviours, or alerts within the group. Accurately predicting the scale and speed of contagion, as well as identifying nodes that are particularly influential in these processes, is critical for understanding the group’s coordinated behaviour.

By systematically studying the properties and dynamics of this fish interaction network, we aim to gain deeper insights into the principles that govern collective behaviour in social animals. This model could also provide a foundation for understanding similar processes in other animal groups or even artificial agent-based systems designed to mimic biological swarm behaviour.

## 4.1 Positioning algorithms:

We can define different algorithms to generate the positions of the network. Prior to having the empirical positions of the fish, we estimated them via a *Poisson Disk Sampling* method [21]. This method is an iterative sampling strategy. Starting from a seed sample,  $n$ -candidates are sampled in the hypersphere surrounding the seed. Candidates below a certain radius or outside the domain are rejected. New samples are added in a pool of sample seeds. The process stops when the pool is empty or when the number of required samples is reached. In the end, this algorithm allows us to obtain a set

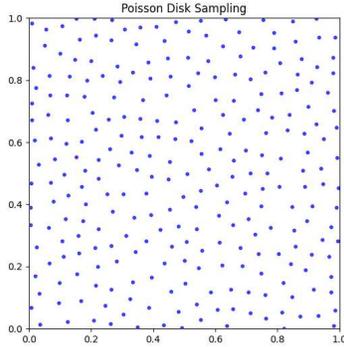


Figure 8: Poisson disk sampling to place 100 nodes in a square.

of uniformly distant nodes with a fixed density, which is spatially homogeneous but non-regular (see Fig.8). Positional noise can also be added, but we opted not to in order to maintain the physical constraint that distance between fish must be greater than zero, given the very size of the fish. Additionally, the introduction of noise makes the network highly heterogeneous. While this heterogeneity is beneficial for creating a more realistic representation of the network, it can lead to undesirable outcomes when analysing borderline cases. For example, in networks with very few links, certain specific configurations may arise where the contagion cannot effectively propagate. This choice allows us to concentrate on the role of linking.

We set the density to be consistent with the real system; typical densities are around 1000- 2000 individuals per square meter.

In the last part of the thesis (section 8) we will introduce the empirically measured positions of the fish, which will be denoted as *fish positions*. We will see just how deeply modified the outcome of the dynamics results by simply using these new positions.

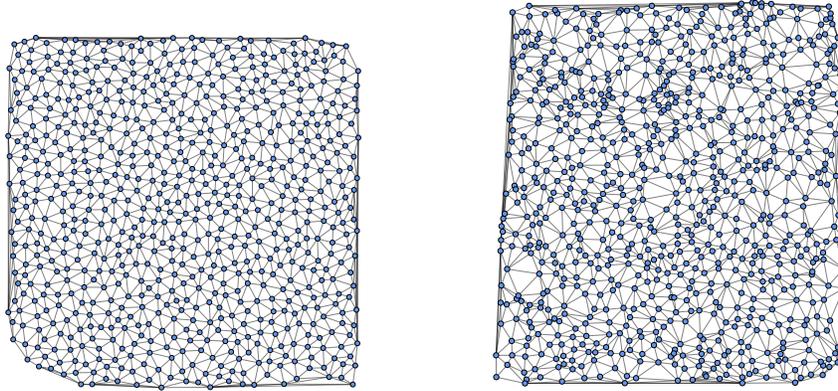


Figure 9: Two examples of networks: in the first, the positions are estimated via Poisson sampling, in the second, empirically observed positions from annotated images are used.

#### 4.1.1 Nearest-Neighbors distance distribution (NND):

A clear characterization of the structure group is given by the spatial distribution of nearest-neighbours. As we know, the Poisson sampling method allows us to maintain a lower limit for the Nearest Neighbors Distance (NND).

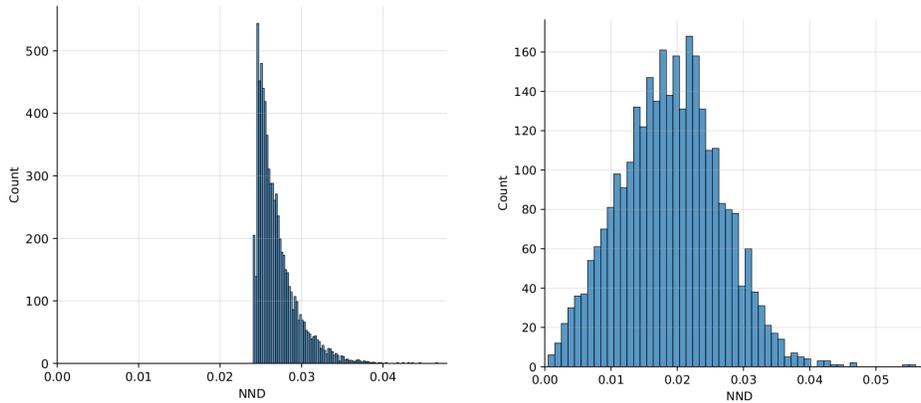


Figure 10: Nearest Neighbour distribution in absence of noise (left) and with positional noise (right).

In the first plot we see a cut-off. Adding positional noise, the cut-off disappears and the nearest-neighbour distance can be 0.

## 4.2 Linking algorithms:

A major challenge in the study of collective animal behaviour is that the pathways of communication are not directly observable. It has long been recognized in the study of isolated organisms that mapping the physical and functional connectivity of neural networks is essential for developing a quantitative and predictive science of how individual behaviour is generated [14]. By contrast, in the study of mobile animal groups, the analogous issue of determining the structure of the sensory networks by which interactions, and the resulting group behaviour, are mediated remains to be explored. Typically, spatially embedded interaction networks between biological agents are modelled either via metric network models, where the probability of a link (or its strength) depends only on the inter-individual distance, or by topological models where a focal agent is connected to a set of spatial neighbours based on their distance rank in comparison to all others, as e.g., in the  $k$ -nearest neighbour network model, but where the actual link probability (or strength) does not depend on the absolute physical distance. In the past, most agent-based models assumed metric interaction networks, but after evidence for topological interaction in starling flocks has been presented by Ballerini et al. [48], corresponding topological interaction networks have received increased attention in the context of collective animal behaviour. In the study of the starling flocks, they discovered that each bird interacts on average with a fixed number of neighbours equal to either six or seven individuals. Their explanation for this phenomenon is that a topological interaction is indispensable for maintaining the flock's cohesion against the large density changes caused by external perturbations, typically predation. Finally, topological interaction networks with a limited number of nearest neighbours have been recently discussed in the context of cognitive constraints regarding the number of neighbours (or objects) a focal individual can pay attention to [81].

However, the discussion of these two idealized models largely ignores the constraints set by different sensory mechanisms underlying social interactions. In the system under analysis, the two most relevant mechanisms are: visual and LLO communications (see section 3). This second mechanism is not affected by constraints, as is vision (e.g. occlusion at high densities), but mainly depends on sensory limits and properties of the medium. The LLO in fish is indeed involved in detecting mechanical vibrations and water movements, which can indirectly facilitate acoustic-like interactions. While it is not a traditional “acoustic” system like the ears of terrestrial animals, it shares similarities with acoustic detection because it is sensitive to pressure waves and vibrations in the surrounding water.

We will briefly illustrate the three different types of network we are going to analyse

from a structural and dynamical point of view.

#### 4.2.1 Metric networks:

Metric interaction networks are the most common in the study of animal collectives and in models of collective movement. This is due to the typical assumption that collective behaviour emerges from simple rules of interaction among individuals [48]. The main theoretical assumptions are:

- attraction among individuals;
- short range repulsion;
- alignment of the velocities.

These are reasonable, but generic, and there are as many different models as there are different ways of implementing these assumptions. To grant cohesion, models make the sound assumption that individuals align and attract to each other, and that such interaction decays with increasing distance between them. These models adopt metric distance, since animals have developed many ways of estimating distance, including stereo-vision, retinal image size, and optic flow [49]. Thus, a metric interaction seems natural. However, it is important to underline that little is known about the mechanism and accuracy with which fish perceive distance [4]. Li *et al.* [2], for example, suggested that fish use parallax while moving through water. In the case of fish, metric interaction may provide a simple model for social interactions via LLO, which is reasonably dependent on the sole measure of distance, given the medium.

In these models, we could define a strength of interaction as a decreasing function of the Euclidean distance  $d_{ij}$  between two individuals  $i$  and  $j$ . Here we will just set a threshold distance: two nodes are connected only if their Euclidean distance is smaller than the metric threshold  $\theta_{metric}$ . The adjacency matrix is then defined as:

$$A = \begin{cases} 1 & \text{if } d_{ij} \leq \theta_{metric} \\ 0 & \text{otherwise} \end{cases}$$

This algorithm strongly depends on the density and on the radius ( $d_{ij}^4$ ) we assume.

---

<sup>4</sup>This radius, even if not explicitly specified, will be given in metres.

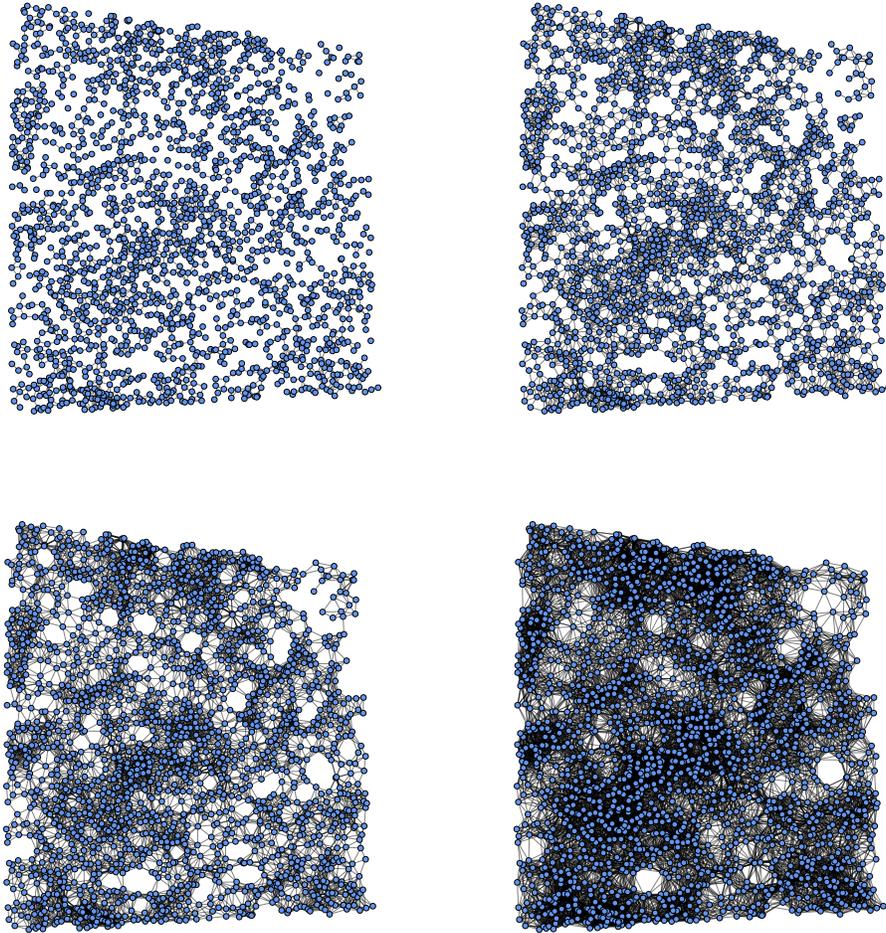


Figure 11: Metric network for radius 0.03, 0.04, 0.05, 0.07 m on a network built with the empirically observed positions (section 8).

All these measures are referred in terms of metres.

#### 4.2.2 Voronoi networks:

This algorithm is based on *Voronoi tassellation*<sup>5</sup>, which for a given set of node positions, divides the space into cells, each containing one node. A cell around a node is made up

---

<sup>5</sup>Invented by the russian mathematician Georgij Feodos'evič Voronoj (Žuravka, 1868 – Žuravka, 1908), should be more correctly written *Voronoj* tassellation.

of all points in space which, using Euclidean distance and a metric, are closer to this node than to any other.

The first shell of Voronoi neighbours of a node is given by all nodes for which the containing cells share a common border. The second and higher order shells are then defined similarly by having a common cell border with the previous shell. A Voronoi network is constructed by connecting each node to its (usually first shell of) Voronoi neighbours. As a result, network neighbours are more evenly distributed around the focal individual than in the KNN topological network, described in the next section. By construction, Voronoi networks always connect all individuals within a group. As observed by Strandburg et al. [90], Voronoi models may better approximate visual ones, as their transitivity (which is the extent to which individuals who share a neighbour are neighbours themselves, and reflects the redundancy of information among neighbours) is similar.

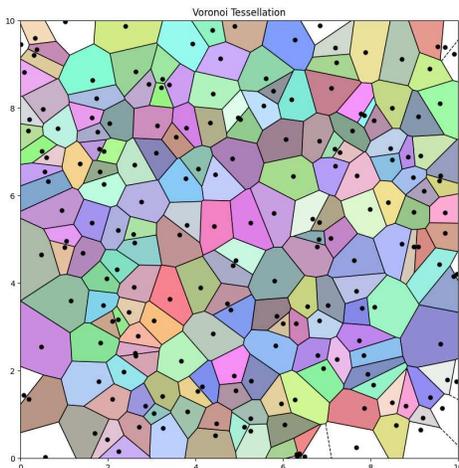


Figure 12: Example of Voronoi tassellation with randomly located points.

### 4.2.3 Topological networks (or k-nearest neighbours):

Motivated by empirical research and considerations of cognitive constraints that limit the number of objects or individuals an animal can keep track of [81], these network prescribe a fixed number of links ( $k$ ) to each node. For comparison, we now recall that human observers are unable to accurately report a number of objects shown in a flash whenever this number exceeds 4, a well studied effect, known as numerosity in the human psychophysical literature [24, 23]. For example, the percentage of incorrect counting when eight objects are briefly presented to a human observer is known to exceed 50%. Based on these observations, Krause *et al.* [4] showed through an ex-

periment on Three-spined sticklebacks (*Gasterosteus aculeatus*) frightened via a light stimulus, that they cannot make any accurate size distinction between shoals with  $n > 3$ . Of course, it is important to keep in mind that we are working on a totally different system, and then these results can not be directly used in our model. Finally, systems based on metric distance may be unable to reproduce the density changes typical of animal aggregation [48]. In this case, defining as  $k$  the number of nearest neighbours to consider, the adjacency matrix gets rewritten:

$$A = \begin{cases} 1, & \text{if } kNN \\ 0, & \text{otherwise} \end{cases}$$

We will analyse how the variation of  $k$  affects the dynamic.

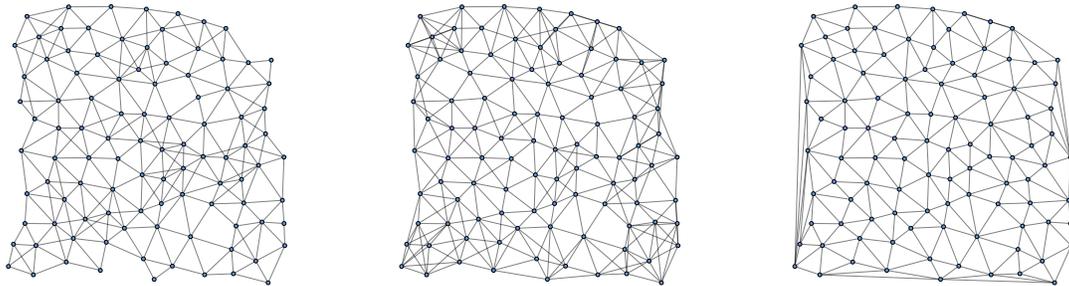


Figure 13: Generated networks with same positions (100 nodes, with Poisson sampling positioning ) in the case of Metric, KNN and Voronoi networks.

In conclusion, let us show an image from Strandburg et al. [90], where we clearly see the difference between the above illustrated networks and the visual network. In this work they showed explicitly how the Voronoi network was a good estimation of the visual one.

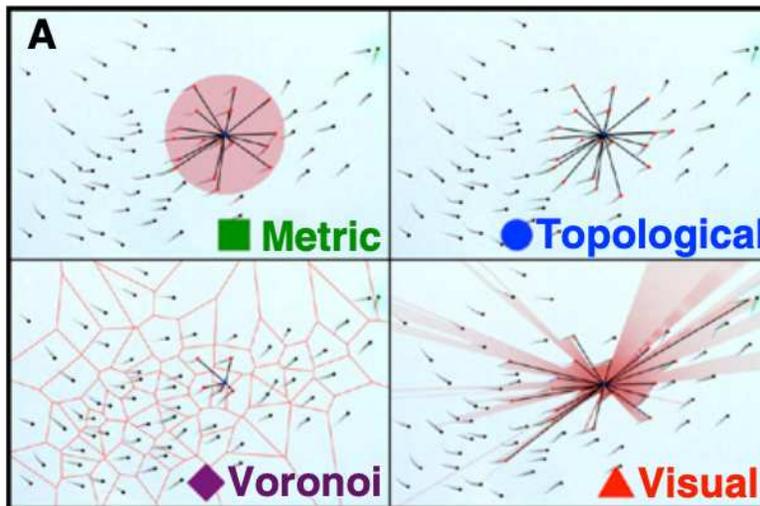


Figure 14: Comparison between networks from Strandburg et al. [90].

### 4.3 Weights:

In the following discussion we will set the weights to 1, assuming an undirected network. This modelling choice was made to keep the model as simple as possible, with a minimal number of variables and parameters.

However, there are alternative descriptions of the study system that include the effects of weights. For example, Rosenthal et al. [82] computed a planar representation of each fish's visual field using ray casting to approximate the pathways of light onto the retina, based on automated estimation of the body posture and eye position of each individual. This representation reveals the underlying visual information available to each fish. Then, determining the initiator and first responder of a startle, they investigate the social contagion, analysing how individuals translate sensory information to motor response (evasion) and revealing the social cues that inform individual decision making in this behavioural context.

They finally formulated a probability of response  $P(i|j)$  in fish  $i$  given that fish  $j$  is startled:

$$P(i|j) = (1 + \exp(-\beta_1 - \beta_2 LMD - \beta_3 AR))^1$$

where  $\beta_1, \beta_2, \beta_3$  are the model coefficients to fit the data (0.302, -1.421, -0.126), LMD is the log of the metric distance between fish  $i$  and  $j$ , AR is the ranked angular area of fish  $j$  subtended on the eye of fish  $i$ . This mapping between sensory input and behavioural

response, constructed based only on the behaviour of first responders, allows to build a hypothetical network, predictive of how behaviour will spread through the group. Here, the weight of an edge between individuals  $i$  and  $j$ ,  $w_{ij} = P(i|j)$ , is the probability of a behavioural response by individual  $i$  if individual  $j$  exhibits behavioural change.

#### 4.4 Network characterization:

The emergent collective behaviour of animal groups, or more generally multi-agent systems, is decisively shaped by the underlying networks of social interactions. A variety of measures have been introduced to characterize the structural properties of networks, often with the intention to predict the average outcome of a certain dynamical process on the network [102]. Within this thesis several well-established measures are used to characterize the different types of networks and understand the behaviour of different dynamical processes: the average in-degree, the clustering coefficient and the number of triangles. These quantities, are relevant in order to describe the contagion processes on a network [89]. It has been shown [62] that infectious disease dynamics can be profoundly influenced by two key network properties: the number of contacts per individual (degree distribution) and the transitivity, or clustering, of contact. Moreover, although it is widely accepted that high connectivity among individuals facilitates group consensus, and that being in a group provides benefits to individuals through social information about environment provided by other group members, it has been shown [81] that large scale coordination can be maximized by strongly limiting the cognitive capacity of individuals, where due to self-organized dynamics the collective self-isolates from disrupting information. In conclusion, within this thesis, it is possible to show that the outcome of the contagion process is ultimately determined not by the actual links, but mainly by these general measures, and in particular the distributions of degrees and triangles.

#### 4.5 In degree:

The in-degree of a node is defined as the number of its incoming links. Since we consider undirected networks, it coincides with the degree. Given the adjacency matrix, the degree of node  $i$  is defined as:

$$k_i = \sum_{j=1}^N A_{ji}$$

Counting the fraction of nodes with degree  $k$  in the graph gives the degree distribution. This distribution is typically used as a measure of heterogeneity, as it provides insights into how connections are distributed across the network.

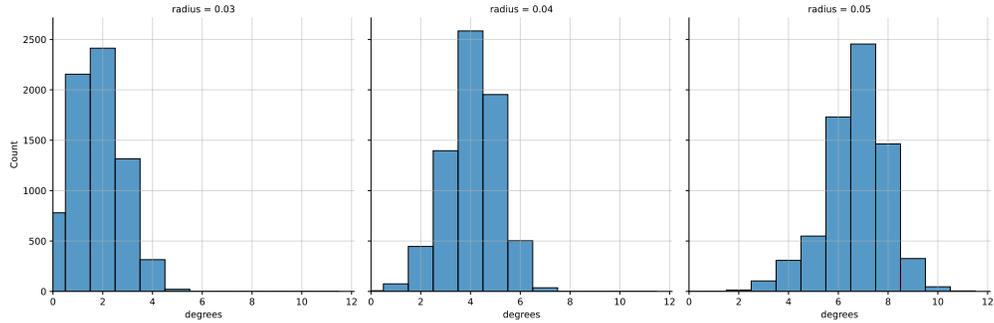


Figure 15: Histogram of degrees on a metric network for different radii (0.03, 0.04, 0.05).

The peak obviously shifts by changing the radius of influence. Here we see that it goes from  $\sim 3$  to  $\sim 7$ .

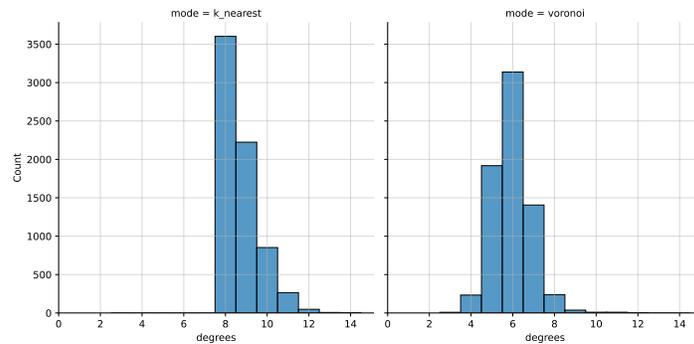


Figure 16: Histogram of degrees on topological ( $k=8$ ) and Voronoi networks.

KNN and Voronoi linking algorithms are independent on the density of the nodes. Let us note that in the topological network, even if we pose  $k=8$ , we will have a non negligible fraction of nodes with higher degree, because we built it in a such a way that all nodes have *at least*  $k$  neighbours. In the Voronoi network, the average degree is around 6.

## 4.6 Clustering coefficient:

Clustering, also known as transitivity, is a typical property of acquaintance networks, where two individuals with a common friend are likely to know each other. The clustering coefficient is a measure of how frequently any two neighbours of a node are themselves neighbours. In terms of a generic graph, transitivity means the presence of

high number of triangles. For unweighted graphs, the clustering of a node is the fraction of possible triangles through that node ( $T(u)$  is the number of triangles through that node). The clustering coefficient per node is defined as:

$$c_u = \frac{2T(u)}{k(u)(k(u) - 1)}$$

It is assigned to 0 if the degree is lower than 2.

We can analyse the distribution of clustering coefficients among the different networks.

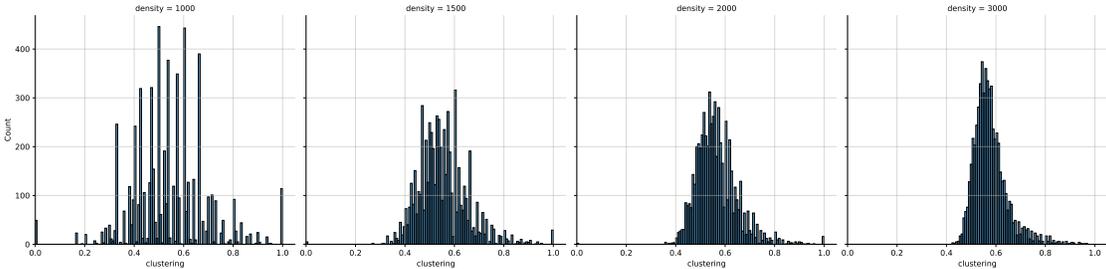


Figure 17: Distribution of clustering coefficients for a metric network with increasing density (from 1000 individuals per square metre to 3000).

As we can see, by increasing the density, the distribution tends to be more peaked. For low densities it is wider and takes on some discrete values, while for bigger densities it is a continuous distribution with a well defined peak.

If instead we analyse the variation of the radius, the distribution looks like:

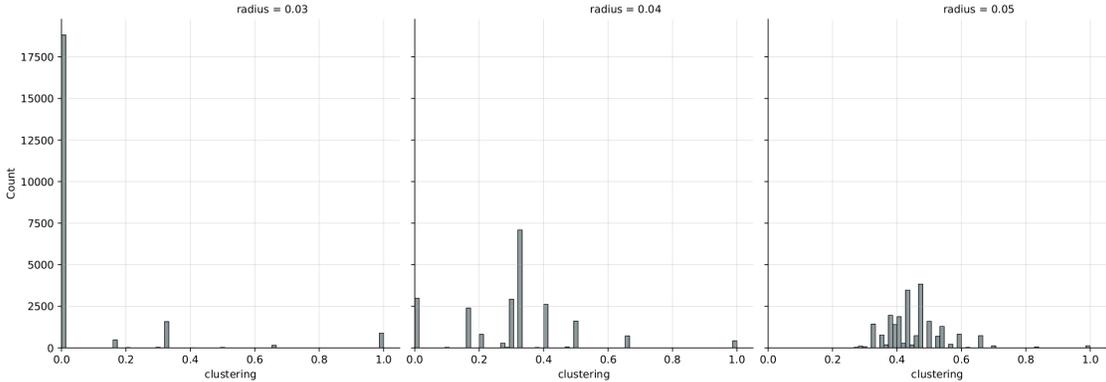


Figure 18: Distribution of clustering coefficients for a metric network with increasing radii (0.03, 0.04, 0.05).

In case of smaller radius, being the network almost disconnected, the distribution is highly peaked in a clustering coefficient equal to 0. Finally, the KNN and the Voronoi networks:

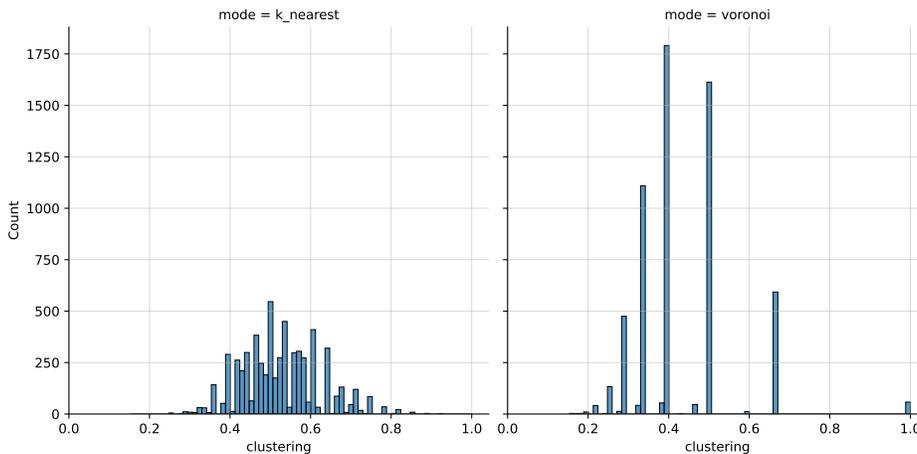


Figure 19: Distribution of clustering coefficients for KNN (left) and Voronoi networks.

The Voronoi network is characterized by some discrete values associated with a very high number of counts, while the KNN has a more Gaussian-shaped distribution.

This coefficient can have a significant impact on the epidemic size or endemic state. For example, it is widely accepted that the value of the transmission rate needed to generate an epidemic is larger for networks which are clustered when compared to an equivalent network with the same degree distribution but no clustering [102].

We could define a generalized definition accounting for weights. There are several ways to define clustering [33], one example is to use the geometric average of the sub-graph edge weights:

$$c_u = \frac{1}{deg(u)(deg(u) - 1)} \sum_{vw} (w_{uv}w_{uw}w_{vw})^{1/3}$$

where  $w_{uv}$  are the edge weights, normalized by the maximum weight in the network.

Representing the clustering coefficient with different colours from white to dark green, we can visualize how this coefficient changes by varying the radius in the construction of the metric networks.

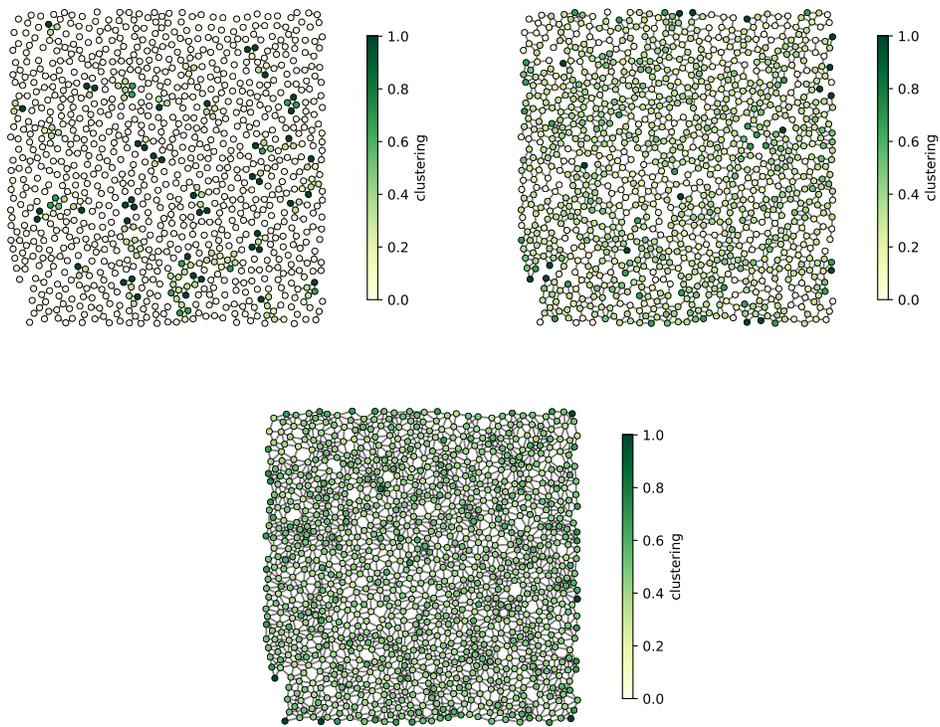


Figure 20: Metric networks for radius 0.03, 0.04 and 0.05 fixing density at 1000 individuals/square meters. The green colours represent the clustering coefficients.

If in the quasi-disconnected case (radius = 0.03) all nodes have clustering equal to zero apart for some isolated nodes with clustering 1, by increasing the radius, the clustering seems to be more uniformly distributed among the network.

Finally, let us see the same representation for the KNN and Voronoi networks:

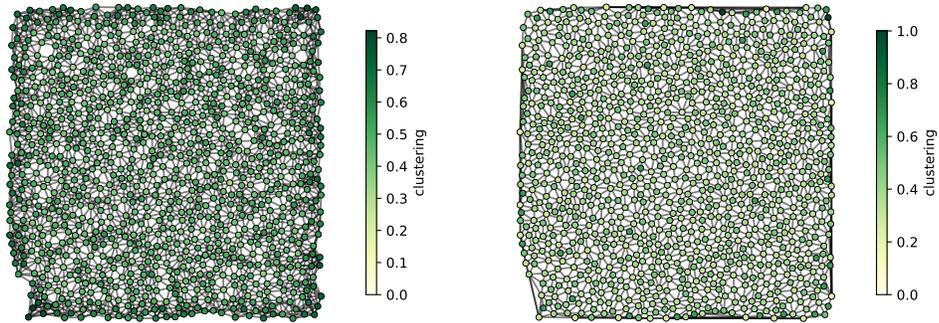


Figure 21: KNN and Voronoi networks for fixed density 1000. The green colours represent the clustering coefficients.

The Voronoi network tends to be more homogeneous in terms of spatial distribution of the clustering coefficient per node with some few nodes with clustering 1, the KNN never reaches clustering equal to 1.

#### 4.7 Number of triangles:

Another useful measure, which will turn out to be relevant in the description of complex contagion, is the number of triangles, the configurations of 3 strongly connected nodes. In the case of metric networks, this measure depends on the chosen density.

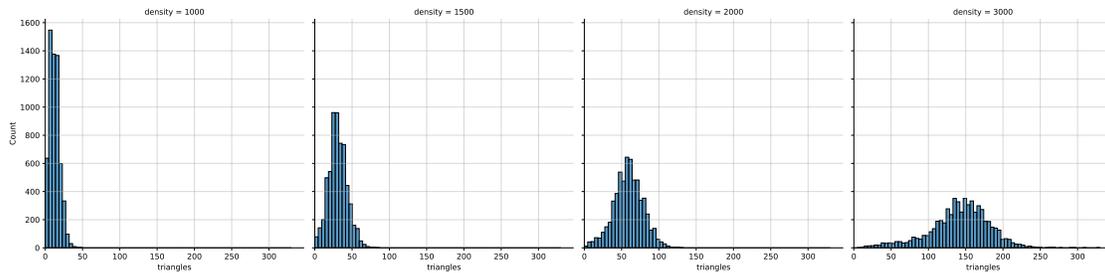


Figure 22: Distribution of number of triangles for metric networks of increasing density.

The distribution from being very peaked relaxes with increasing density. It also depends strongly on the radius. Let us see how it varies by changing it from 0.03 to 0.05 m.

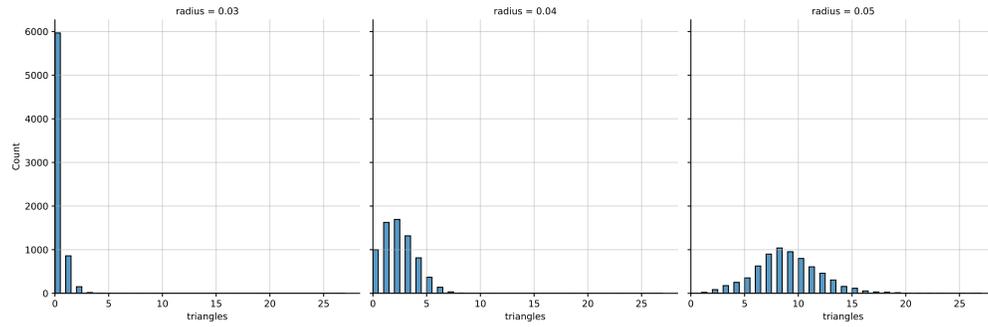


Figure 23: Distribution of triangles for metric with radius 0.03, 0.04, 0.05 for fixed density 1000.

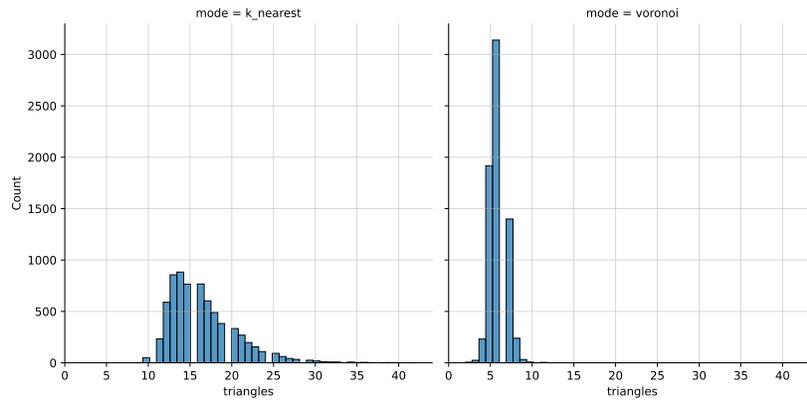


Figure 24: Distribution of triangles for the KNN and the Voronoi networks.

As compared to metric and KNN networks, Voronoi tends to have, on average, a smaller number of triangles.

## 5 Simulated Dynamics:

### 5.1 Introduction:

To model our physical system, we will consider a contagion dynamic over a network model. Let us now better characterize the contagion models.

The term 'contagion' itself has roots in the Latin verb 'contingere', formed by the combination of *com* ("with, together") and *tangere* ("to touch"), thus conveying the idea of spreading through contact. The earliest recorded usage of this word dates back to ancient Rome. For instance, it appears in Lucretius'<sup>6</sup> epic poem *De Rerum Natura* (1st century BCE circa), where he describes the transmission of disease from one person to another through direct contact. This disease-related meaning of "contagion" persisted for centuries and was largely confined to describing the spread of pathogens. It was not until the late 19th century that the concept began to extend beyond the domain of disease. Sociologists Gabriel Tarde<sup>7</sup> and Gustave Le Bon<sup>8</sup> were among the first to examine contagion in a social context, with Le Bon introducing the term behavioural contagion to describe how behaviours spread within groups [39]. Later, in 1939, Herbert Blumer<sup>9</sup>— a foundational figure in the study of collective behaviour—coined the term social contagion [42]. By the mid-20th century, contagion had become a widely used concept in sociology, capturing the spread of behaviours, opinions, rumours, and norms across societies [43].

In the late 20th and early 21st centuries, the application of contagion theory expanded further, encompassing financial and economic systems. Economists and financial analysts began using the term to describe the transmission of financial crises or shocks, which could propagate between interdependent markets and economies through mechanisms such as investor behaviour and information cascades [44]. Nowadays, this term also describes the rapid spread of information, trends, and visual content across online platforms [41].

What brings together all the described phenomena is the transmission of some quantity-

---

<sup>6</sup>Titus Lucretius Carus (c.99 – c.55 BC) was a Roman poet and philosopher. His only known work is the philosophical poem *De Rerum Natura*, a didactic work about the tenets and philosophy of Epicureanism.

<sup>7</sup>Gabriel Tarde (12 March 1843 – 13 May 1904) was a French sociologist, criminologist and social psychologist who conceived sociology as based on small psychological interactions among individuals (much as if it were chemistry), the fundamental forces being imitation and innovation.

<sup>8</sup>Charles-Marie Gustave Le Bon (7 May 1841 – 13 December 1931) was a leading French polymath whose areas of interest included anthropology, psychology, sociology, medicine, invention, and physics. He is best known for his 1895 work *The Crowd: A Study of the Popular Mind*, which is considered one of the seminal works of crowd psychology.

<sup>9</sup>Herbert George Blumer (March 7, 1900 – April 13, 1987) was an American sociologist whose main scholarly interests were symbolic interactionism and methods of social research.

information-entity throughout a population, facilitated by a network of interactions-communication. In this way, contagion dynamics in a network model capture how certain entities—whether biological, social, or informational—propagate through these interactions, regardless of whether the “contact” is physical or simply a network-mediated influence.

Classical contagion models, like SIR, SIS, and SIRS, describe these phenomena by attributing to all agents a state. In these models S means susceptible, that is to say that an individual can be infected, I means infected, R recovered, died, or no-more prone to infection. The SIR model assumes permanent immunity after recovery, useful for diseases like measles [63]; the SIS model has individuals return to a susceptible state, describing well the common cold for example [64]; and the SIRS model includes temporary immunity, conveys effectively infections like certain influenza strains [65]. Extensions, like SEIR and SEIRS, add an “Exposed” stage for incubation periods and are used for diseases with delayed infectiousness, such as COVID-19 [66].

Contagion processes can be classified as simple or complex. There are different non-equivalent criteria to define the two. We can distinguish between them by assuming that a contagion is simple if each exposure acts independently from any other exposure, while complex contagions are trivially those in which multiple exposures do not act independently, but their effect is instead interdependent.

In particular, when dealing with social contagion phenomena, such as adoption of norms, simple epidemic-like contagions do not provide a satisfactory description, especially if reinforcement mechanisms are at work [92]. Complex contagion mechanisms have been proposed to account for these effects. As defined by Centola and Macy [93], “*a contagion is complex if its transmission requires an individual to have contact with two or more sources of activation*”, i.e. if a “*contact with a single active neighbour is not enough to trigger adoption*”. Here, we propose to go further and take into account that contagion can occur in different ways, either through pairwise interactions (the links of a network) or through group interactions, i.e., through higher-order structures. Consequently we can define a minimum “critical mass” of adopters, often required for many social entities to spread widely. Finally, we observe that higher order interactions enable more nodes to exchange information simultaneously, thus allowing more efficient communication and leading to enhanced synchronization [83].

Since our model will be of the SIR kind with a quadratic contagion mechanism, let us explain briefly how to approach the standard SIR model. This model, formulated by

Kermack<sup>10</sup> and McKendrick<sup>11</sup> [27] in 1927 to describe the irreversible propagation of an infectious disease, is the most popular and used contagion model. Given the rate of infection  $\beta$ ,  $\beta\Delta t$  is the probability that infection occurs during the small time interval  $\Delta t$ . Clearly this probability should be proportional to the number of contacts  $k$  an individual had during the interval. Moreover, infected individuals recover and obtain permanent immunity with a constant rate  $\mu$ . Let  $S(t)$ ,  $I(t)$ ,  $R(t)$  be the fraction of susceptible, infected and recovered at time  $t$ , the dynamics is described by a set of ordinary differential equations (ODEs):

$$\begin{cases} \dot{S}(t) = -\beta k S(t) I(t) \\ \dot{I}(t) = \beta k S(t) I(t) - \mu I(t) \\ \dot{R}(t) = \mu I(t) \end{cases}$$

with the constraint  $S(t)+I(t)+R(t)=1$ , only two of these equations are independent. Rewriting the above equations, we get the following:

$$\frac{dS(t)}{S(t)} = -(\beta k/\mu)dR(t)$$

integrating, we obtain a solution for the evolution of susceptible nodes:

$$S(t) = S(0)\exp\left(\frac{-\beta k}{\mu}(R(t) - R(0))\right)$$

then  $I(t)$  is just a function of  $R(t)$  with parameters  $S(0)$  and  $R(0)$ .

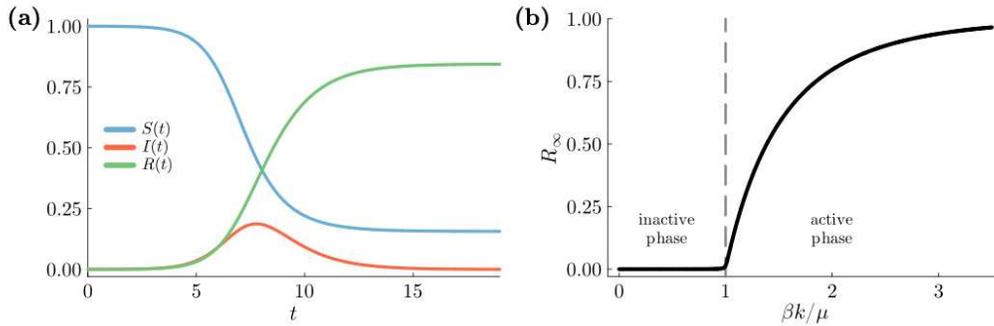


Figure 25: Figure from the Ph.D. thesis of Burgio G. [45]. Phenomenology of the SIR model. (a) Time evolution of the fraction of susceptible, infected and recovered individuals in the population. (b) Final attack rate  $R_\infty$  as function of  $\beta k/\mu$ .

<sup>10</sup>William Ogilvy Kermack (26 April 1898 – 20 July 1970) was a Scottish biochemist.

<sup>11</sup>Lt Col Anderson Gray McKendrick (8 September 1876 – 30 May 1943) was a Scottish military physician and epidemiologist.

In this analysis the quantity  $\beta k/\mu$  plays a central role. Here,  $1/\mu$  represents the average infection period, which is the average time interval between an individual becoming infected and recovering. During this period, an infected individual transmits the infection to each of its susceptible contacts with a probability of  $\beta/\mu$ . In a population where all other individuals are susceptible, an infected individual will, on average, infect  $\beta k/\mu$  others. This value is known as the basic reproduction number, commonly denoted as  $R_0$ . It represents the number of secondary cases generated by a single primary case. If  $R_0 > 1$ , the first case will cause more than one secondary case, on average, before recovering, leading to exponential growth in the number of infections.

It is important to note that the deterministic dynamics describing the evolution of the system's average state are valid only in the infinite-size limit. For any finite population size, statistical fluctuations around the average become significant, and there is a non-zero probability that the number of infected individuals deviates from the expected infinite-size behaviour, potentially fluctuating down to zero. Once the system reaches this state with no infected individuals, it cannot return to an active state. This state is referred to as an absorbing state.

In the above description there is no contact network in the model; contacts occur uniformly at random between any two individuals in the population. In fact, we only specified the number or rate of contacts  $k$ , the degree in the network, which is the same for each node, thus the network is  $k$ -regular.

## 5.2 Dynamics description:

Contagion models have broad applications, ranging from epidemiology, where they help model the spread of infectious diseases [85], to social contagion [86] and the diffusion of rumours [87]. Traditionally, these models operate under the assumption that contagion spreads via pairwise interactions, often represented through network graphs where nodes signify individuals and edges represent interactions between them. In these conventional models, the spread is limited to interactions between two individuals at a time. However, this framework has recently been expanded to capture more complex dynamics, as researchers have introduced models that incorporate higher-order interactions, which are essential in scenarios involving group-based interactions or collective behaviour [92, 45].

To model these higher-order effects in behavioural contagion networks, we introduce a complex contagion model that accounts for interactions within groups of three (triangles) and includes a fractional contagion perspective. This fractional model emphasizes the relative influence of contagious neighbours, based on their proportion, rather than

their absolute number. Thus, in this formulation, the contagion mechanism depends on both the presence of immediate contacts and their clustering in small groups. Ultimately, this model adapts the Susceptible-Infected-Recovered (SIR) framework to incorporate both link-based and group-based contagion. Here Susceptible corresponds to the fish at the surface of the water, Infected means in the act of diving and moving in this movement the water and Recovered corresponds to a fish underwater which cannot infect any more the others. In this sense, the contagion corresponds to the adoption of the diving-behaviour. In the complex SIR model, the **infection rate** for an individual  $i$  is defined as follows:

$$\alpha_i(S \rightarrow I) = \alpha_1[k_i^1]^{-\gamma} \sum_j A_{ij}x_j + \alpha_2[k_i^2]^{-\gamma} \sum_{jk} A_{ij}A_{ik}A_{jk}x_jx_k \quad (1)$$

where:

- $\alpha_1$  and  $\alpha_2$  are respective infection rates for pairwise and triangle-based interactions;
- $k_i^1$  is the degree of individual  $i$  in terms of pairwise (single-link) connections, while  $k_i^2$  represents the count of triangles involving  $i$ ;
- $A_{ij}$  is the adjacency matrix element indicating the presence (1), or absence (0), of a link between nodes  $i$  and  $j$ ;
- $x_i$  is a binary indicator of the infectious state of neighbour  $i$ ;
- the exponent  $\gamma$  modulates the influence of the connection strength based on degree. Specifically  $\gamma = 0$  for numeric contagion =1 for fractional.

In the following we will assume fractional contagion and we have two different mechanisms for the recovery:

- constant time recovery: Individuals recover after a fixed number of steps after  $t$  steps;
- constant rate recovery: Each individual has a fixed probability  $\mu$  of recovering at each time step.

Seeing as our goal is to model a biological system, it would be reasonable to assume a deterministic recovery (constant time) mechanism, reflecting the observation that fish typically need a consistent amount of time to dive underwater. If not specified explicitly, it will be a deterministic model with  $t=10$ . Changing this time simply shifts the transition, but does not deeply affect the dynamics (see section 6.1). Nonetheless, we

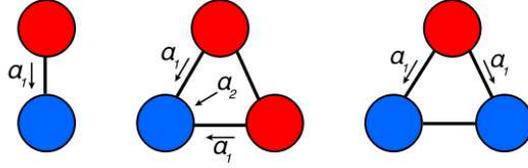


Figure 26: Explanation of the various channels of infection, with the corresponding rates. If we only have one link between an infected (red) node and a susceptible (blue) node, the infection rate is just  $\alpha_1$ . In the case of triangles, either 2 nodes are infected, in which case we can have infection via both the links and the triangle itself with rate  $\alpha_2$ , or just one node is infected and the contagion travels along the links only.

will also explore the probabilistic recovery case to consider possible variations and to analytically analyse the model.

Let us observe that moreover, if we wanted to analyse the repeated wave behaviour, we would also need to consider the transition from R to S again. We know from the data that the period they spend underwater is around 2 seconds.

As seen in Formula 1, we are working within a fractional contagion framework. Here, the activation of each node is influenced by the activity level of its neighbouring nodes. However, unlike independent link channels, activation depends on the fraction of active neighbours relative to the maximum possible activity level. This fractional approach is particularly suited for modelling fish school behaviour, as studies have shown that behavioural transmission in these groups aligns closely with fractional contagion dynamics [82]. In general, for all types of networks and initial conditions with increasing  $\alpha_1$  (simple contagion), we observe a clear transition from a small fraction ( $n_{rec}$ ) of recovered to  $n_{rec} \sim 1$ , and this can be defined as single avalanche spreading through the network. However the corresponding transition for increasing  $\alpha_2$  (complex contagion) depends strongly on network properties and initial conditions. In the next section, section 6.1, and in the final section, section 8, we will try to understand the role that network properties play in this transition.

Another relevant aspect to be analysed is the ability of filtering out noise from signal in collective networked systems. Since birds must enter the water to catch fish, a combination of visual and mechano-acoustic cues (multimodal) characterize an immediate attack, while single cues (unimodal) may represent less dangerous disturbances.

In a field survey, Lukas et al. [95] confirmed that the sulphur mollies frequently have to distinguish between these two stimuli. Sensitivity was high regardless of stimulus type and number, but fish dove deeper, faster, and for longer when both stimuli were available simultaneously. Based on the system's high rates of bird activity, we argue that the fish adopt a non-selective dive initiation strategy, quickly starting their descent without detailed evaluation, and then fine-tune their diving parameters based on received cues. This behaviour allows them to conserve energy, which is essential for responding to future predatory attacks. The model does not reproduce this mechanism, but what is relevant is that it retains the ability to discriminate.

We have focused on two biologically relevant and arbitrary definitions of signal and noise:

- **Coincidence detection:** We know from fish schools that individuals can occasionally startle spontaneously, even without an actual predation threat, but it is unlikely for two individuals to startle simultaneously by chance. Therefore we focus on the case where the signal is defined as two nodes activated simultaneously, and noise as only one activated node. This is a reasonable approximation in the case that the rate of spontaneous startles is low, such that it becomes vanishingly unlikely for two spontaneous startles to occur in the same cascade event.
- **Spatially correlated activations:** In some cases a true positive, such as a predator approaching from a given spatial position, is likely to induce activations, which are correlated in space. Of course, the opposite could also be true, that simultaneous activations of distant individuals are less likely to be correlated and are more likely to give uncorrelated information about the environment. To model this, we compare simultaneous initial activations of individuals close and far apart, in network terms.

The question is now whether complex contagion could be a robust mechanism for filtering out false positive signals (spontaneous startles). Additionally, both signal and noise may be introduced at certain places within the group's structure, and we have separated this into boundary and bulk conditions (i.e. introduced only at the periphery or in the core of the group). This is important to consider because biological systems have boundary conditions that, on the scale of the systems we are interested in, will be relevant to the dynamics of the process. We consider this factor because it is relevant for real animal groups and would generate empirically testable predictions.

An important observable to be analysed is the fraction of recovered individuals, which basically defines the dimension of the infection in the network.

### 5.2.1 Role of weights:

Another important factor to take into account when modelling the contagion process is the role of weighted interactions, which can capture the intensity or strength of connections between nodes. To represent this, we introduced two distinct infection rates for the two different contagion mechanism. Specifically, if we wanted to account for the presence of weights ( $w_i$ ):

$$\alpha'_1 = \alpha_1 \cdot w_1$$

$$\alpha'_2 = \alpha_2 \cdot w_2$$

in which the weights are combinations of weights associated to each link. Specifically, we distinguish between absolute and fractional modes:

- simple-weight ( $w_1$ ) can be:
  - absolute: simply the weight of the associated link;
  - fractional: weight divided by the total weight for links including the node;
- quadratic-weight ( $w_2$ ) can be defined in different ways:
  - mean: the average of the weights involved in the triangle;
  - mean fractional: average divided by the total weight for triangles including the node;
  - max: max between weights of links in the triangle;
  - max fractional: max divided by the total weight for triangles including the node;
  - min: min between weights;
  - min fractional: min divided by the total weight for triangles including the node.

We will concentrate on the case in which weights are constant, and the chosen modes for computing probabilities are fractional and mean fractional.

## 5.3 Model Implementation:

To run our simulations we used Utopia, a comprehensive modelling framework for complex and adaptive systems [99, 100, 101]. Utopia handles simulation configuration, parallelized parameter sweeps, as well as efficient reading, writing, and evaluating of high-dimensional simulation output. It is designed to facilitate collaborative research

and flexible model development while maintaining high individual freedom in implementation and analysis. Utopia includes a C++ library for model implementations and data writing, and a Python interface for simulation control, data analysis, and plotting. The model itself is implemented in Python, using NetworkX for representing spatially embedded networks.

## 6 Percolation transition: SIR model on spatial networks.

### 6.1 Introduction:

By defining the model as a three-state SIR system (section 5.2), we aim to describe the diving behaviour of fish while excluding the possibility of repeated waves. Instead, the focus is solely on the so-called escape wave. Observations of the biological system (see empirical data in section 8) reveal that when externally stimulated by a sufficiently strong input, an initial wave is generated. The size of this wave depends on the stimulus intensity. Subsequently, additional waves may occur, either in the same or different locations. In this context, the stimuli correspond to bird attacks.

The final state of an SIR model is characterized by a fraction of susceptible nodes, while the remaining nodes are classified as recovered. Recovered nodes (denoted as R) are all those that were infected by the contagion. Thus, when comparing with the biological system, it is reasonable to associate these recovered nodes with the fish that participated in the escape wave (see Fig.27).

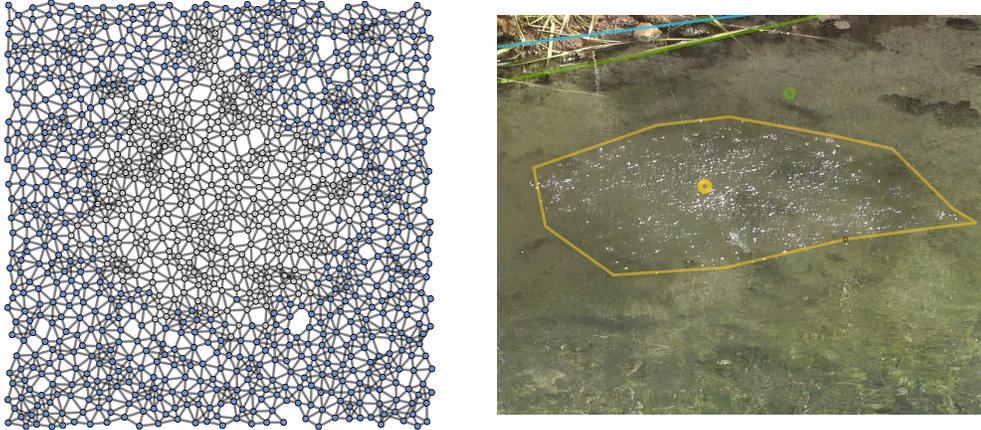


Figure 27: In the Figure above, we see on the left an example of the final state of the network, where recovered nodes are in grey and susceptible nodes are in blue. On the right, a screenshot of the annotation tool we used to analyse empirical data (section 8). The small yellow circle indicates the impact point of the external artificial stimulus, the outer yellow line outlines the clearly visible escape wave.

The fraction of R relative to the total number of individuals serves as the order

parameter, a key measure that allows us to distinguish between two distinct phases: one in which the contagion spreads across the entire network, and another where only a subset of the fish dive. This perspective naturally aligns with a percolation model, where the critical observable is the size of the giant cluster—the single, large connected component of nodes that dive in response to the stimulus.

What we will do, in this first part, is study the behaviour of the system by varying the strength of the contagion process, or rate of infection, not the initial input. This can be seen as a phase transition between a system that does not have the capability to spread the diving behaviour, and a fully connected system in which all fish dive. This transition would strongly depend on the topology of the network.

To better understand what we will be studying, let us briefly introduce the main concepts of percolation theory [117, 102].

Percolation theory is one of the simplest statistical-mechanics models showing non-trivial phase transition. The purely geometrical nature of these transitions makes them fundamental ingredients of more complex collective phenomena observed in many physical systems. We can distinguish between bond percolation and site percolation. In general each bond (or link) of the model is assigned to a binary variable  $x_i$ , with probability  $p$  (typically a constant unique value) of being equal to 1, and with probability  $(1-p)$  of being 0. The probability  $p$  is called *occupation probability*, and  $x_i = 1$  means that the site (the link) is either occupied, present, or active, depending on the context. The percolation problem involves analysing the conditions on the occupation probability required for the emergence of connected groups of sites (clusters), whose size is comparable to that of the entire system, corresponding to infinite clusters in the thermodynamic limit. The existence of a macroscopic aggregate of connected sites would ensure the existence of a path between boundaries. We will thus observe a singularity at a certain  $p_c$ , called *critical probability*, minimum value over which percolating clusters exist. Generally bond and site percolation are different, i.e. the value of  $p_c$  is not the same for a given topology, but there are some trivial cases in which they coincide, such as the one-dimensional case. Moreover, this model is strongly affected by the dimensionality of the embedding space, which is however, not a variable in this case.

Given the presence of a phase transition, many properties will exhibit power-law behaviour with respect to the distance from the critical point,  $(p - p_c)$ . The behaviour of the system close to criticality is expected to be characterized by universal properties. In this context, when exact results are not accessible, approximation methods can be employed; Real Space Renormalization Group [40] is particularly useful because it allows to extract qualitatively correct information on critical behaviour and determine the critical exponents.

In the following section, we will give a general analysis of the model in terms of the ability of the networks to distinguish between noise and signal. In the context of epidemic modelling on networks, measuring the susceptibility of a population to infection is essential for understanding the dynamics of disease spread and the potential for outbreaks. A useful way of quantifying this susceptibility is by assessing how sensitive the final epidemic outcome is to the initial number of infected individuals. For this purpose, we introduce a susceptibility measure based on the difference in the fraction of individuals in the recovered state under two initial conditions in the SIR model; namely the difference between a minimal infection -one infected individual only- versus a slightly larger initial infection -two infected individuals. This assumption is just an arbitrary choice. Even though we assume that the peaks of this measure correspond to the so-called critical region of parameters, this could be profoundly different for a different definition of susceptibility. In networked populations, however, this measure can reveal aspects of the structure that influence epidemic spread, such as clustering, connectivity, or the presence of highly connected nodes (hubs).

Of course it is useful to analyse other cases, like the difference between infections in the periphery and in the centre, or in the case of more initially infected agents.

In the final part of this thesis (section 8), we will see why this analysis is relevant for our specific study system by introducing an experiment we conducted in the field (section 8.5).

The different network topologies will be discussed separately.

## 6.2 Metric Network:

In the beginning, we focus on a metric network to observe how varying connectivity, through changes in the interaction radius, affects the system's susceptibility to disease spread. By tuning the interaction radius, we control the average in-degree of nodes, thereby moving the network between states of near-isolation (with a low radius and few connections), where widespread disease transmission improbable and independent on the quadratic contagion, and highly connected configurations (with a higher radius), in which the system becomes sufficiently connected, allowing an infection to potentially activate the entire network and here we've a strong dependence on  $\alpha_2$ . The transition point while varying the radius is of particular interest, as it marks the critical connectivity where disease spread dynamics change significantly. As explained, a key aspect of our analysis is the system's response to different initial conditions, specifically comparing the outcomes when the infection begins with only one infected node versus when it begins with a pair of connected infected nodes. This comparison is essential for assessing the system's sensitivity to initial infection conditions and understanding the degree to which network structure amplifies or dampens the spread from a minimal initial infection.

We will always, if not explicitly said, apply the stimulus in the centre to a number  $N$  of individuals. This analysis will be focused on the heatmaps describing, by varying  $\alpha_1$  and  $\alpha_2$ , the fraction of recovered nodes and the susceptibility measure.

Let us comment the results. As observed in Fig.28, the dynamics' sensitivity to variations in  $\alpha_2$  appears to increase with the radius. This behaviour is intuitive, as for smaller radii (e.g., 0.03), there is practically no dependence on quadratic contagion effects. In such cases, the number of triangles in the network is essentially zero, which limits the influence of higher-order interactions.

In the more connected cases, the susceptibility peaks at high values of  $\alpha_2$  (see Fig. 29), highlighting how important the quadratic contagion mechanism is for these values.

This result could appear straightforward. The assumption of having two infected connected nodes and  $\alpha_2 \sim 1$  implies that these two nodes will probably infect other nodes via both channels of contagion. On the contrary, for small values of  $\alpha_2$ , only the links spread the epidemic, thus the two analysed initial conditions are not so different, since being linked to another infected node does not favour the contagion via triangles.

However, as will we see in the analysis of the dynamics on the networks built using the empirically observed positions (section 8), this is not always the case. The heterogeneities change non-trivially the results.

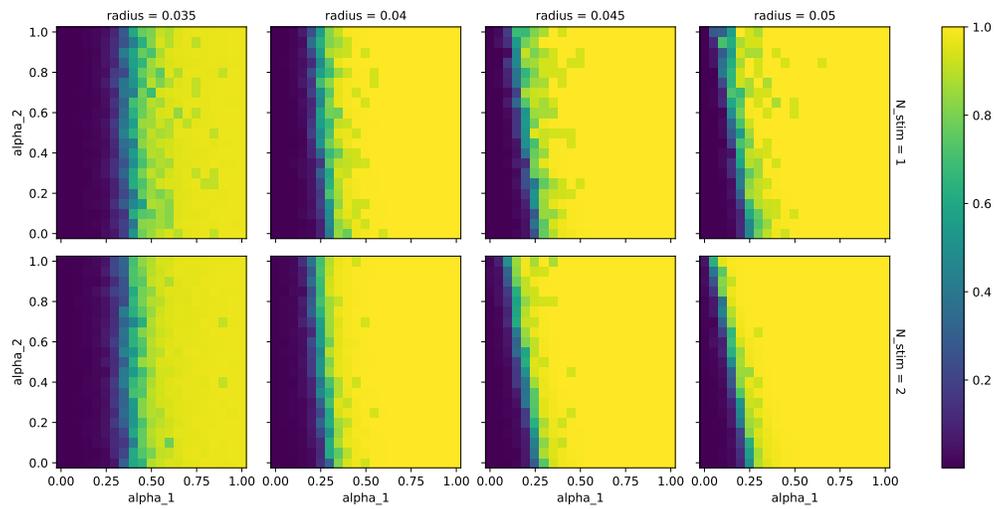


Figure 28: Fraction of recovered for  $N=1$  (above) and  $N=2$  (below) in the case of metric network by varying the radius. Yellow means all nodes recovered, purple none of them (just few).

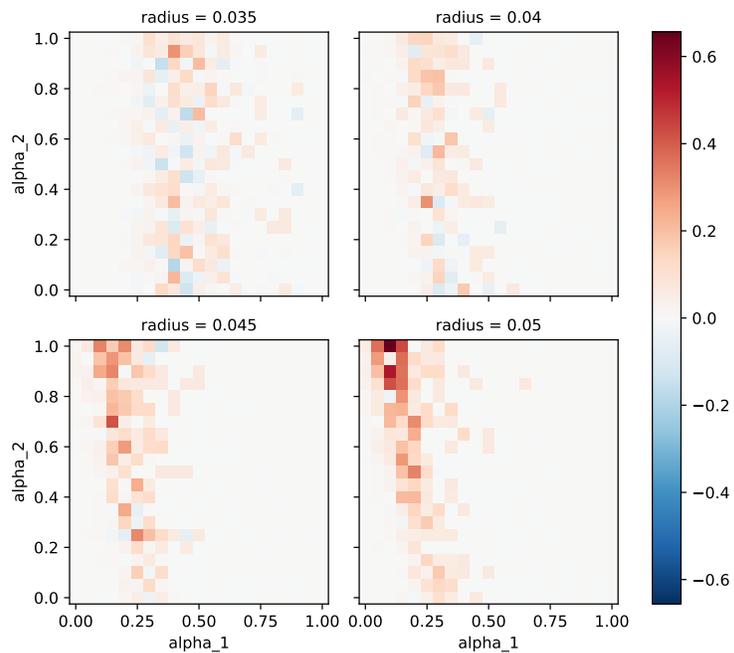


Figure 29: Susceptibility heatmap. It is obtained by subtracting, for each radius, the  $N=1$  and  $N=2$  cases.

Let us focus for a moment on the last case, with radius fixed to 0.05 and we analyse the susceptibility. In the context of the criticality hypothesis (section 2), identifying the parameters that maximize the susceptibility measure allows us to conclude that the system, under that specific combination of parameters, is at the critical state. Of course, this conclusion must be justified within the framework of the chosen model and, specifically, in terms of the defined susceptibility. We are not concluding that the biological system is critical. Instead, we state that if the system is well described by the model under that specific combination of parameters, and if it can be demonstrated through experimental evidence that the system is indeed in a critical state (which is something partly done in [46]), then the defined susceptibility could be considered a good description of the sensitivity of the system.

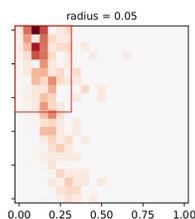


Figure 30: This plot represents the susceptibility heatmap for metric network with radius 0.05. The red square represents the region of parameters where susceptibility assumes its highest values.

It is worth noticing that:

- This is just one possible measure of susceptibility, we could have defined a different susceptibility. Here, for the purpose of studying the discrimination ability of the shoals, we decided to keep just this and use it to describe the system phase;
- The susceptibility will strongly depend on the network size;
- The susceptibility strongly depends on the network topology.

Another aspect to analyse is that the infection dynamics are notably influenced by the location where the stimulus is applied. Specifically, when the stimulus is initiated at the centre, the infection spreads more effectively. This is likely due to the higher connectivity of central nodes, which facilitates the propagation of the signal. In contrast, nodes located near the boundary are less connected, and as a result, the wave signal tends to dissipate before reaching the edges. This highlights the critical role of network structure and geometry in shaping the spread of contagion processes.

Indeed, if we assume that the initially infected nodes are chosen in the left-down corner

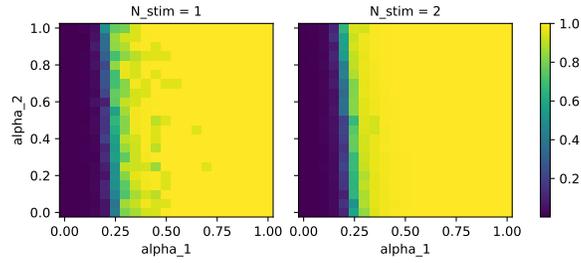


Figure 31: Fraction of recovered for  $N=1$  and  $N=2$ . Here we have a metric network with radius 0.05 and the initially infected nodes are in the left-down corner.

instead of the centre, we will obtain Fig.31. In this plot is evident that the dependence on the quadratic contagion is much less important in the transition. This can be justified by the fact that in the border it is less probable to be involved in triangles.

Finally, we observe that the duration of the Infected to Recovered ( $I \rightarrow R$ ) transition affects the system’s behaviour by shifting the overall transition line but does not fundamentally change the dynamics of the spread. Here, the “time” refers to the fixed duration that an individual remains in the Infected (I) state before transitioning to the Recovered (R) state. This duration directly influences the rate at which infected individuals recover, thus shaping the timing of the epidemic’s progression through the network. To simplify our analysis, we fixed this Infected to Recovered transition time to  $t=10$ . By holding this transition time constant, we can focus more effectively on the effects of two key parameters, which control other aspects of the infection dynamics.

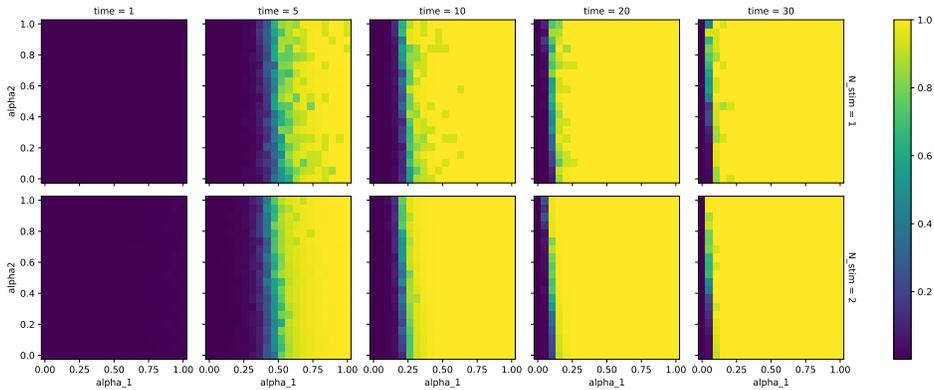


Figure 32: Fraction of recovered for the metric network (radius 0.05) as function of the parameters for  $N=1,2$  in the metric network by changing the time of recovery.

### 6.3 Voronoi Network:

Assuming now the Voronoi linking algorithm, we want to analyse the same quantities as in the metric case, specifically the fraction of recovered and the susceptibility measure.

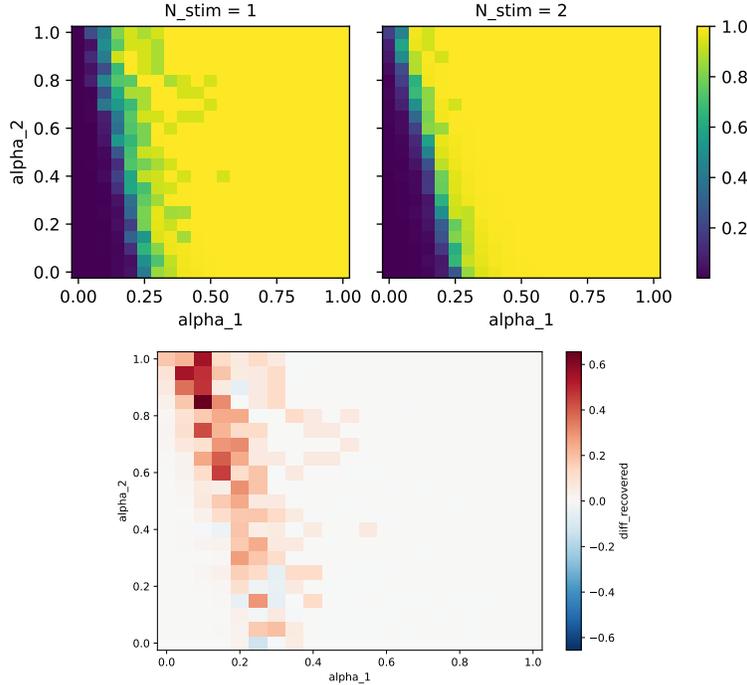


Figure 33: Fraction of recovered for the Voronoi network for  $N=1$  and  $N=2$ .

The region of parameters where the susceptibility maximizes is again concentrated in the upper-left portion of the heatmap, emphasizing the critical role of the quadratic contagion mechanism (see Fig.33). These results are comparable to those observed in the metric network with a radius of 0.05, as both exhibit similar degree distributions.

We now turn finally our attention to the probabilistic case of the model, which introduces randomness in the transition dynamics. Analysing this probabilistic approach offers valuable insights, as it allows us to draw more analytically tractable models. In the probabilistic model, the likelihood of an infected individual passing the infection or transitioning to recovery at any given time step is based on predefined probabilities rather than fixed time intervals. This approach aligns well with real-world scenarios, where the exact duration of infection and recovery times vary among individuals. As expected, compared to the deterministic case, there is significantly more noise (especially for  $N=1$ ) due to the additional stochasticity introduced (see Fig.34). Nevertheless,

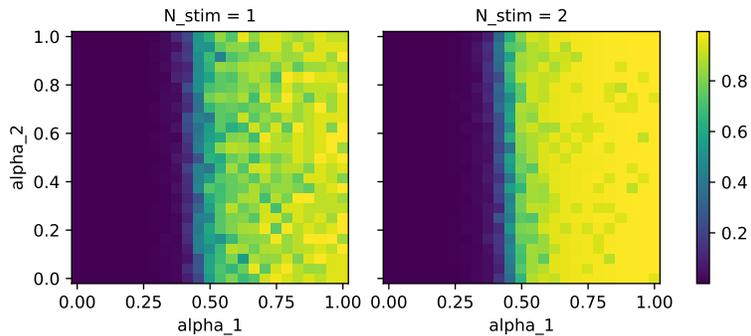


Figure 34: Fraction of recovered for the Voronoi network for  $N=1$  and  $N=2$  in the probabilistic case assuming  $p=0.1$

it is still possible to observe a dependency on  $\alpha_2$ .

We are able to show that a key distinguishing factor between these two cases lies in the evolution of the number of “active” triangles over time, which means triangles with two infected nodes. In the probabilistic case, this count is generally lower (for example, comparing  $t=10$  with  $p=0.1$ ) but also active for more time.

This suggests that a form of “synchronized” behaviour is essential to maintain the structure of active triangles. Consequently, the probabilistic case is more sensitive to chance deactivations of nodes, as a single early deactivation can disrupt triangle connections, while a chance extension of activity in one node has a less pronounced effect. Let us show explicitly the number of active triangles over time:

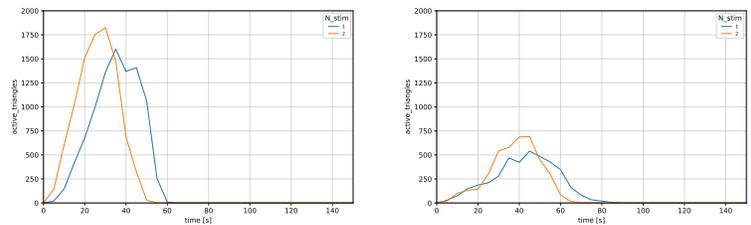


Figure 35: Comparison of the number of active triangles over time between the deterministic case with time of recovery  $t=10$  and the probabilistic case with  $p=0.1$  for  $N=1,2$ . In both plots  $\alpha_1 = 1$  and  $\alpha_2 = 1$ .

## 6.4 Knn Network:

Considering a topological linking algorithm we obtain the knn network. Let us analyse again the fraction of recovered nodes and the susceptibility. It is interesting to show how

these two observables are dependent on the chosen  $k$  (to build the network); similarly to the metric case, increasing the connectivity we observe an increase in dependency with respect to  $\alpha_2$ . Highly connected graphs imply greater dependency on the quadratic contagion mechanism.

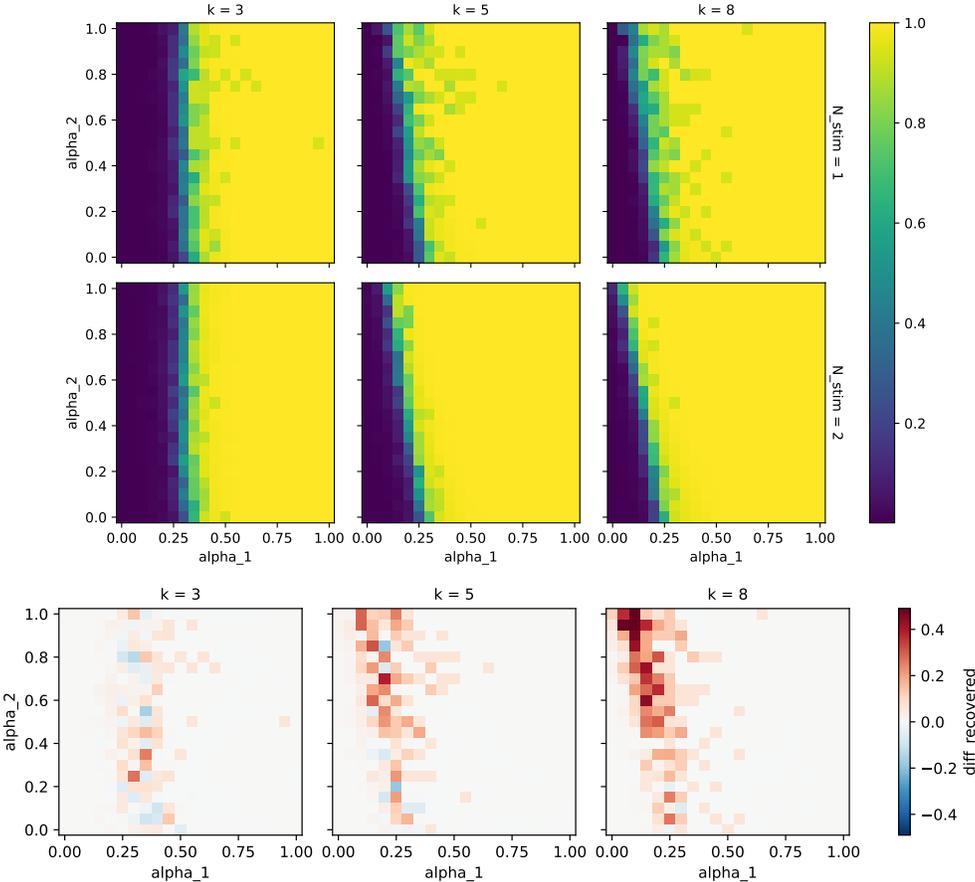


Figure 36: The first image shows the fraction of recovered for the Knn network for  $N=1$  and  $N=2$  by varying the value of  $k$ . The second represents the susceptibility measure.

## 6.5 Conclusions:

To summarize, the analysis of the percolation transition allows us to conclude that the collective response ultimately depends heavily on the interplay between the dynamics and the structure of the underlying network. We were in fact able to show how the choice of the linking algorithm is determinant in the outcome of the contagion and in the identification of the values of the parameters that maximized the susceptibility; in

the framework of the criticality hypothesis these ultimately correspond to the critical values.

In the deterministic case, where the infection time is fixed, the entire network exhibits a sharp transition characterized by a well-defined critical line. While this transition is less pronounced in the probabilistic case, it remains observable. This distinction can, in part, be explained by examining the evolution of the number of active triangles over time. Additionally, the degree distribution plays a pivotal role: only in sufficiently connected networks does the quadratic contagion mechanism exhibit a clear and explicit influence on the system's behaviour.

## 7 Mean-Field approximation on SIS dynamics:

### 7.1 Introduction to simplicial complexes:

Complex networks have proven highly effective in describing disease spread within populations of interacting individuals [56].

One recent advancement in this field is the formalization of interaction patterns using hypergraphs [92, 114, 109], which generalize traditional graphs by allowing edges, or hyperedges, to connect *any* number of vertices. In this work, we will introduce the main concepts of hypergraph theory and demonstrate how they can be applied to analyse the mean-field approximation of the SIS model. However, to facilitate a comparison with our model, we will ultimately return to the framework of traditional graph theory.

A hypergraph is encoded either in a series of adjacency tensors each one accounting for the connectivity at the level of m-edges or through an incidence matrix  $H$ , defined as follows:

$$H(v, e) = \begin{cases} 1 & \text{if } v \in e \\ 0 & \end{cases}$$

where  $H$  indicates the presence or absence of vertices in each hyper-edge, allowing for a detailed and flexible representation of the relationships within complex social systems.

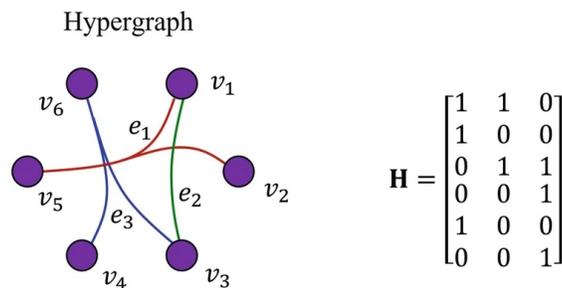


Figure 37: Image from *Mathematical Foundations of Hypergraph* [110].

This flexibility makes hypergraphs particularly suited to capturing correlations in complex, multi-way interactions beyond simple pairwise connections.

To better understand these structures, we can define an equivalence between hypergraphs and bipartite graphs mapping the two subset of vertices of the bipartite graph to the vertex set and the edge set of the hypergraph.

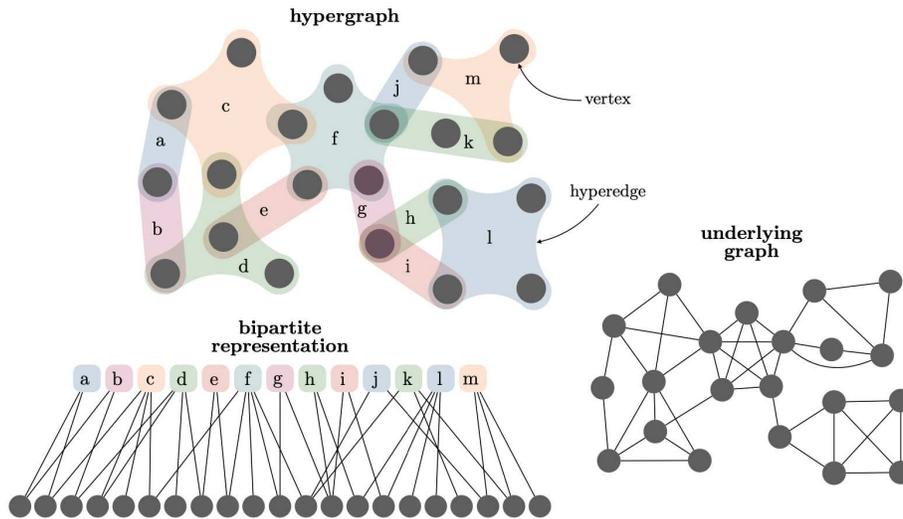


Figure 38: Image from [45]. A hypergraph, its equivalent representation as a bipartite graph, and its underlying graph. Note that the hypergraph in this example is linear.

A hypergraph is called simple if no hyperedges is subset of any other. If  $H$  is simple and two hyperedges share no more than one vertex, then  $H$  is said to be linear.

This framework allows us to introduce the concept of *simplicial complexes*.

We know that a simplex is simply a generalization of the concept of a triangle (in 2 dimensions) or a tetrahedron (in 3 dimensions) to any number of dimensions. For simplicity, we call nodes (or vertices) the 0-simplices and links (edges) the 1-simplices of a simplicial complex  $K$ , 2-simplices correspond to the full triangles, 3- simplices to the tetrahedra of  $K$  and so on (see Fig.39).

A simplicial complex  $K$  on the vertex set  $V$  is a hypergraph endowed with the additional hereditary property: given  $k \in K$  and  $k' \subseteq k$ , then  $k' \in K$ .

The elements of  $K$  are called faces and a  $n$ -dimensional face ( $n$ -face) is a subset of  $V$  made of  $n+1$  vertices. The hereditary property says that given a face or hyperedge which is included in the complex, then all its sub-faces or sub-hyperedges are included too. For instance:  $k = \{i, j, k\}$  generates the simplex  $s = \{\{i, j, k\}, \{i, j\}, \{i, k\}, \{j, k\}, \{i\}, \{j\}, \{k\}\}$ .

In the context of networks and complex systems, a simplex is often used to represent a basic unit of interaction or relationship between entities in a higher-dimensional space, where interactions involve more than just pairs of individuals. Introduced by Iacopini et al. [92], the simplicial contagion model recognizes group, many-body interactions as an alternative mechanism responsible for complex contagions.

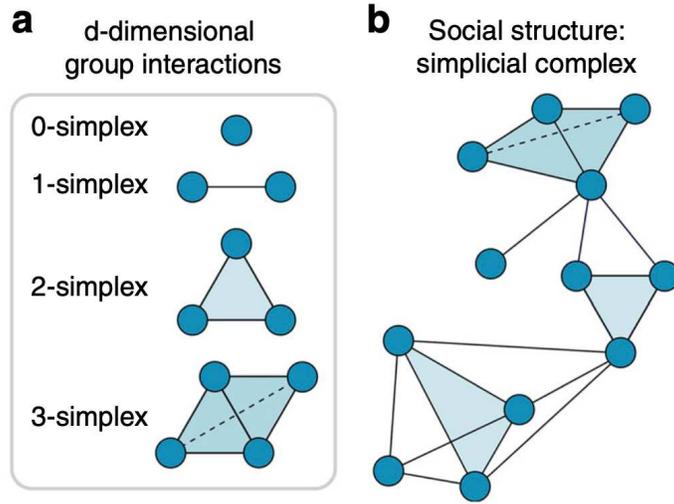


Figure 39: Image from Iacopini et al. ([92])

Unlike generic hypergraphs, simplicial complexes received substantial attention from a broader community. In particular, they recently proved to be useful in describing the architecture of complex networks: functional [103] and structural [104] brain networks, protein interactions [105], semantic networks [106], and co-authorship networks in science [107]. Building on these principles, we model a social group as a simplex and employ simplicial complexes to represent the underlying structure of the social system under study.

In a recent paper, Iacopini et al. [92] showed that in the case of exponential random simplicial complex and comparing to some social interaction based networks, higher order interactions lead to the emergence of new phenomena, changing the nature of the transition at the epidemic threshold from continuous to discontinuous. Furthermore, Burgio et al. [108], demonstrated that the outcome of the contagion process is closely tied to how interactions of different orders are structured within the system. These works offer a more quantitative analysis that reinforces the previous qualitative findings [36, 37] and provide a qualitative understanding through mean-field approximation.

To conclude, these findings underscore the crucial role of higher-order interactions in shaping the dynamics of complex systems.

## 7.2 Mean field:

In this section, to gain additional insights into the interaction network we have constructed and to better characterize our model, we will introduce the widely used Mean-Field approximation (MF). Contagion processes are known to fundamentally depend on the structural properties of the interaction networks through which they propagate. Many real-world networked systems exhibit clustered substructures, comprising either collections of all-to-all pairwise interactions (cliques) or group interactions involving multiple members simultaneously. Therefore, it is crucial to examine how the underlying structure and its heterogeneities influence the dynamics of these processes.

In particular, differently from the other sections of this thesis, we will map now the social contagion to a SIS process. To keep this difference evident we will call now  $\beta_1$  and  $\beta_2$  the infection rates, instead of  $\alpha_1$  and  $\alpha_2$ .

Systems with many (sometimes infinite) degrees of freedom are generally hard to solve exactly or compute in closed, analytic form, except for some simple cases. In this way, MF is an approximation method that often makes the original problem to be solvable and open to calculation, and in some cases it may give very accurate approximations. In general, the MF theory assumes that to some extent degree of freedom interacts equally with the others, as if it was interacting just with some average of the others and neglecting correlations between the agents.

Here, the approximation consists on considering each agent being affected in the same way by the presence of the neighbourhood. This removes local variability and we keep only some features of the original network, specifically the average degree and the average number of triangles.

We will be able show how the outcome of the contagion process is fundamentally linked to how the interactions of different order are arranged in the system and then how this approximation is weak in our case.

Generically, we will define as  $\beta^{(n)}$  the rate of which a node involved in a  $(n+1)$ -body interaction with  $n$  simultaneously infected individuals gets infected.

Suppose  $m$  is the size of the largest groups, the evolution of the fraction of infected individuals in the population under the assumption of homogeneous mixing hypothesis is given by:

$$\dot{I}(t) = -\mu I(t) + (1 - I(t)) \sum_{n=1}^{m-1} \beta^{(n)} k^{(n)} I(t)^n$$

where  $k^{(n)}$  is the average generalized simplicial degree.

That corresponds essentially to the rate of change of the average density of active

sites. Under this homogeneous mixing hypothesis, all sites play an identical role in the dynamics of the well-mixed system, therefore the dynamics is completely specified by the number of active sites.

We will develop a slightly more accurate approximation. In any case, if we want now to define the first-order moments and derive the evolution of it, we see that it depends on moments of higher orders. Moment equations does not close, so we need to introduce moment closures.

### 7.3 Closures:

Let us start from the closure approximation which we apply to the exact microscopic equations on hypergraphs. We know that there are several choices of closures. Closing, for example, the equations at the level of nodes requires the assumption that the states of the neighbouring nodes are statistically independent. In our description we will introduce instead the triadic approximation. This choice implies a more qualitatively correct model, capable to predict new qualitative effects. The need of introducing such approximation is to account for dynamical correlations between the individuals.

To better understand this point let us introduce an illustrative example from [102] of a SIR model on a simple line network with three nodes. Let us denote as:  $\langle S_i \rangle(t)$  the probability that node  $i$  is susceptible at time  $t$ ,  $\langle I_i \rangle(t)$  the probability that  $i$  is infected at time  $t$ ,  $\langle R_i \rangle(t)$  recovered,  $(\langle I_i \rangle + \langle R_i \rangle)(t)$  the probability that  $i$  is either infected or recovered,  $\langle I_i \rangle \langle R_j \rangle$  the probability that  $i$  is susceptible and  $j$  infected at time  $t$ . Each node can be infected if it is susceptible from a neighbour with rate  $\tau$ , while the recovery rate is called  $\gamma$ .

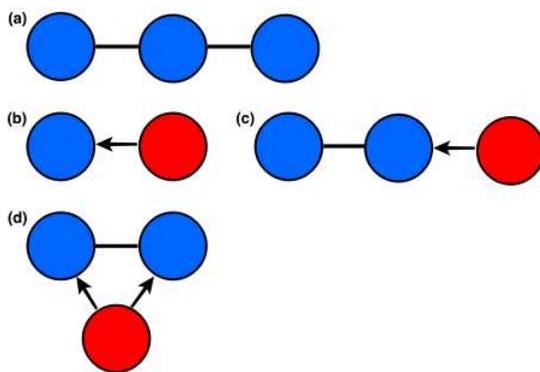


Figure 40: (a) Simple line network with three nodes. (b) Creation of new infected node. (c,d) SI links are created when SS pairs are invaded by infected nodes, which requires us to develop equations for triplets such as SSI.

If one would be to write down evolution equations at level of nodes, pairs, triplets, etc, the procedure would need to continue until full network size is reached.

$$\left\{ \begin{array}{l} \langle \dot{S}_1 \rangle = -\tau \langle S_1 I_2 \rangle \\ \langle \dot{I}_1 \rangle = \tau \langle S_1 I_2 \rangle - \gamma \langle I_1 \rangle \\ \langle \dot{R}_1 \rangle = \gamma \langle I_1 \rangle \\ \langle \dot{S}_2 \rangle = -\tau \langle I_1 S_2 \rangle - \tau \langle S_2 I_3 \rangle \\ \langle \dot{I}_2 \rangle = \tau \langle I_1 S_2 \rangle + \tau \langle S_2 I_3 \rangle - \gamma \langle I_2 \rangle \\ \langle \dot{R}_2 \rangle = \gamma \langle I_2 \rangle \\ \langle \dot{S}_3 \rangle = -\tau \langle I_2 S_3 \rangle \\ \langle \dot{I}_3 \rangle = \tau \langle I_2 S_3 \rangle - \gamma \langle I_3 \rangle \\ \langle \dot{R}_3 \rangle = \gamma \langle I_3 \rangle \end{array} \right.$$

The equations for R can be eliminated because  $\langle R_i \rangle = 1 - \langle S_i \rangle - \langle I_i \rangle$ . This reduces the number of differential equations. However, we have to study the evolution of the new terms such as  $\langle S_1 I_2 \rangle$  and this requires additional equations. Using similar arguments, we find:

$$\left\{ \begin{array}{l} \langle \dot{S}_1 I_2 \rangle = \tau \langle S_1 S_2 I_3 \rangle - (\tau + \gamma) \langle S_1 I_2 \rangle \\ \langle \dot{I}_1 S_2 \rangle = -\tau \langle I_1 S_2 I_3 \rangle - (\tau + \gamma) \langle I_1 S_2 \rangle \\ \langle \dot{S}_2 I_3 \rangle = -\tau \langle I_1 S_2 I_3 \rangle - (\tau + \gamma) \langle S_2 I_3 \rangle \\ \langle \dot{I}_2 S_3 \rangle = \tau \langle I_1 S_2 S_3 \rangle - (\tau + \gamma) \langle I_2 S_3 \rangle \end{array} \right.$$

These equations requires further information about specific triples:

$$\left\{ \begin{array}{l} \langle S_1 \dot{S}_2 I_3 \rangle = -(\tau + \gamma) \langle S_1 S_2 I_3 \rangle \\ \langle S_1 \dot{S}_2 I_3 \rangle = -(2\tau + 2\gamma) \langle I_1 S_2 I_3 \rangle \\ \langle I_1 \dot{S}_2 S_3 \rangle = -(\tau + \gamma) \langle I_1 S_2 S_3 \rangle \end{array} \right.$$

Ultimately, by neglecting the equations for R, we reduce the system to just 13 differential equations. However, this approach quickly becomes impractical as the network size increases, due to the rapid growth in the number of required equations. For a generic network, the number of terms needed would equal the total number of nodes, making the model cumbersome to handle.

To simplify the model, we aim to express pairs in terms of singles, triples in terms of pairs and singles, and so on. These simplifications are typically performed to reduce higher-order structures into lower-order ones. For instance, if we can represent all triples in terms of pairs and singles, such as:

$$\langle S_i S_j I_k \rangle = \langle S_i S_j \rangle \langle S_j I_k \rangle / \langle S_j \rangle$$

then the total number of equations is  $O(N)$  provided that the number of edges scales as such. This yields systems simple enough to be studied numerically and sometimes even analytically.

In practice, the common thread across all these approaches is the goal of expressing higher-order structures in terms of lower-order ones, either exactly or approximately. This strategy relies on the assumption that, at some scale, certain variables can be treated as independent. For instance, consider two neighbouring sites with either identical or different statuses. The correlation between these two types can be expressed as:

$$C_{A_i, B_j} = \frac{P(A_i B_j)}{P(A_i)P(B_j)} = \frac{\langle A_i B_j \rangle}{\langle A_i \rangle \langle B_j \rangle}$$

where  $C=1$  is equivalent with assuming independence.

For both SIS and SIR models, this assumption does not hold, as infected nodes transmit to their susceptible neighbours, leading to an increased probability that infected nodes will be connected to other infected nodes. In contrast, susceptibles and infected nodes are negatively correlated. In this context, knowing that node  $i$  is infected provides information about the status of node  $j$ . This correlation can be expressed as:

$$\langle A_i B_j \rangle = \langle A_i \rangle \langle B_j \rangle C_{A_i, B_j}$$

Assuming independence at the pair level means:

$$\langle A_i B_j \rangle \approx \langle A_i \rangle \langle B_j \rangle$$

Thus simplification is the assumption of *weak correlation* which is often reasonable in large networks where nodes interact only locally and their states are not strongly dependent on each other. However, as the epidemic progresses, correlations tend to increase, making this assumption less accurate. This can lead to an underestimation of the epidemic's spread.

For this reason, this approach is generalized to higher order structures. For example, following van Baalen ([9]), the probability of an open triple  $(i, j, k)$  where  $i$  and  $k$  not connected can be written as:

$$\langle A_i B_j C_k \rangle = \langle A_i \rangle \langle B_j \rangle \langle C_k \rangle C_{A_i, B_j} C_{B_j, C_k} T_{A_i, B_j, C_k}$$

where  $T_{A_i, B_j, C_k}$  is the triple-level correlation. Note that if  $\langle B_j \rangle = 0$  then this is zero. If  $\langle B_j \rangle > 0$  then plugging in the corresponding expressions for the pair-level correlations and neglecting triple-level correlations ( $T_{A_i, B_j, C_k}$ ) then:

$$\langle A_i B_j C_k \rangle \simeq \frac{\langle A_i B_j \rangle \langle B_j C_k \rangle}{\langle B_j \rangle}$$

Equivalently, if i,j,k form a triangle (closed triple), yields:

$$\langle A_i B_j C_k \rangle = \langle A_i \rangle \langle B_j \rangle \langle C_k \rangle C_{A_i B_j} C_{B_j C_k} C_{A_i C_k} T_{A_i B_j C_k}$$

which leads:

$$\langle A_i B_j C_k \rangle \simeq \frac{\langle A_i B_j \rangle \langle B_j C_k \rangle \langle A_i C_k \rangle}{\langle A_i \rangle \langle B_j \rangle \langle C_k \rangle}$$

This approach can be extended to higher-order interactions by considering more complex structures, such as quadruples, and beyond. In these cases, the same principle of expressing higher-order dependencies in terms of lower-order correlations is applied, but now the relationships between multiple nodes must be carefully considered. This extension allows for a more detailed approximation of the system's behaviour, accounting for increasingly intricate interactions and incorporating the information about network topology, that cannot be captured by pairwise correlations alone. As such, the method enables a more accurate description of the dynamics in systems where higher-order connections play a significant role.

This leads to some questions:

- At which level (pairs, triplets) should the closure be applied?
- How well will the solutions of the closed system compare to results based on the original exact system?
- How do all the above depend on the structure/topology of the network and properties of the dynamics?

Let us see a possible approach.

## 7.4 Results:

Going back to our model, we need to analyse the SIS dynamics on the networks built. Considering up to three-body interactions, we account for the evolution of:

- the state probability  $P_i^{\sigma_i}$  for node i to be in state  $\sigma_i$ ;
- the maximal link probability  $P_{ij}^{\sigma_i \sigma_j}$  for the link to be in state  $\sigma_i \sigma_j$ ;
- the maximal (i.e. cliques not subsets of larger ones) 3-clique probability  $P_{ijl}^{\sigma_i \sigma_j \sigma_l}$ .

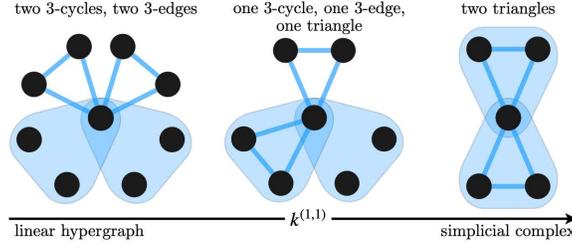


Figure 41: Image from Burgio et al. [108] showing the difference between the 3 cycles, 3-edge, triangle.

We observe that a 3-clique, when projected back onto the hypergraph, can manifest in one of three forms (see Fig:41): a length-3 cycle, representing three 2-body interactions; a 3-edge, representing a 3-body interaction; or a 2-simplex (triangle), which encapsulates all of these interactions. Although we will initially present the general analysis considering all three types of 3-cliques, we will ultimately simplify our model by assuming that only the first form—essentially a triangle in a traditional graph structure—is present.

The state probability of the other local structures is approximated in terms of the maximal cliques composing it. We thus need a closure only for the following local structures: two maximally connected links, a maximal link connected to a 3clique, 2 connected 3cliques.

The closure approximation we apply to the exact microscopic equations on hypergraphs is the **triadic approximation**:

$$\begin{cases} P_{ijl}^{\sigma_i \sigma_j \sigma_l} \approx P_{ij}^{\sigma_i \sigma_j} P_{jl}^{\sigma_j \sigma_l} / P_j^{\sigma_j} \\ P_{ijlh}^{\sigma_i \sigma_j \sigma_l \sigma_h} \approx P_{ij}^{\sigma_i \sigma_j} P_{jlh}^{\sigma_j \sigma_l \sigma_h} / P_j^{\sigma_j} \\ P_{ijlhk}^{\sigma_i \sigma_j \sigma_l \sigma_h \sigma_k} \approx P_{ijl}^{\sigma_i \sigma_j \sigma_l} P_{lhk}^{\sigma_l \sigma_h \sigma_k} / P_l^{\sigma_l} \end{cases}$$

The triadic approximation helps to improve the accuracy of epidemic predictions and is particularly important in networks with many triangles. In lattice networks without triangles, such as the square or hexagonal lattice, the triadic approximation often provides less additional information because triads are absent, meaning that the pairwise approximation alone can often suffice.

The higher-order interaction structure is encoded in the following binary tensors:  $A^{(1)}$  s.t.  $A_{ij}^{(1)} = 1$  if link  $ij$  exists;  $A^{(1,0)}$  and  $A^{(0,1)}$  s.t.  $A_{ijl}^{(1,0)} = 1$  ( $A_{ijl}^{(1,0)} = 0$  and  $A_{ijl}^{(0,1)} = 0$  ( $A_{ijl}^{(0,1)} = 1$ ) if  $ijl$  is a 3-cycle (3-edge), and such that  $A_{ijl}^{(1,0)} A_{ijl}^{(0,1)} = 1$  if  $ijl$  is a triangle. Rescaling by  $\mu$ , the general process is described by the following system of microscopic

equations:

$$\dot{P}_i^I = -P_i^I + \beta^{(1)} \sum_j A_{ij}^{(1)} P_{ij}^{SI} + \frac{1}{2} \sum_{jl} [A_{ijl}^{(1,0)} \beta^{(1)} (P_{ijl}^{SSI} + P_{ijl}^{SIS} + 2P_{ijl}^{SII}) + A_{ijl}^{(0,1)} \beta^{(2)} P_{ijl}^{SII}]$$

$$\begin{aligned} \dot{P}_{ij}^{SI} &= P_{ij}^{II} - (1 + \beta^{(1)}) P_{ij}^{SI} - \beta^{(1)} \sum_{l \neq j} A_{il}^{(1)} P_{jil}^{ISI} + \beta^{(1)} \sum_{l \neq i} A_{jl}^{(1)} P_{ijl}^{SSI} + \\ &\quad - \frac{1}{2} \sum_{l,h} [A_{ilh}^{(1,0)} \beta^{(1)} (P_{jih}^{ISIS} + P_{jih}^{ISSI} + 2P_{ilh}^{ISII}) + A_{ilh}^{(0,1)} \beta^{(2)} P_{jih}^{ISII}] + [i - j] \end{aligned}$$

$$\begin{aligned} \dot{P}_{ijl}^{SSI} &= -(1 + 2A_{ijl}^{(1,0)} \beta^{(1)}) P_{ijl}^{SSI} + P_{ijl}^{ISI} + P_{ijl}^{SII} - \beta^{(1)} \sum_{h \neq j,l} A_{ih}^{(1)} P_{jih}^{SISI} + \\ &\quad - \beta^{(1)} \sum_{h \neq i,l} A_{jh}^{(1)} P_{iljh}^{SISI} + \beta^{(1)} \sum_{h \neq i,j} A_{th}^{(1)} P_{ijlh}^{SSSI} - \frac{1}{2} \sum_{h,k \neq j,l} [A_{ihk}^{(1,0)} \beta^{(1)} (P_{jihk}^{SISIS} + \\ &\quad + P_{jihk}^{SISSI} + 2P_{jihk}^{SISII}) + A_{ihk}^{(0,1)} \beta^{(2)} P_{jihk}^{SISII}] - \sum_{h,k \neq i,l} [i - j] + \sum_{h,k \neq i,j} [i - l] \end{aligned}$$

$$\begin{aligned} \dot{P}_{ijl}^{SII} &= -(2 + 2A_{ijl}^{(1,0)} \beta^{(1)} + A_{ijl}^{(0,1)} \beta^{(2)}) P_{ijl}^{SII} + A_{ijl}^{(1,0)} \beta^{(1)} (P_{ijl}^{SSI} + P_{ijl}^{SIS}) + P_{ijl}^{III} \\ &\quad - \beta^{(1)} \sum_{h \neq j,l} A_{ih}^{(1)} P_{jih}^{IISI} + \beta^{(1)} \sum_{h \neq i,l} A_{jh}^{(1)} P_{iljh}^{SISI} + \beta^{(1)} \sum_{h \neq i,j} P_{ijlh}^{SISI} - \frac{1}{2} \sum_{h,k \neq j,l} [A_{ihk}^{(1,0)} \beta^{(1)} (P_{jihk}^{IISIS} \\ &\quad + P_{jihk}^{IISSI} + 2P_{jihk}^{IISII}) + A_{ihk}^{(0,1)} \beta^{(2)} P_{jihk}^{IISII}] + [i - j] \end{aligned}$$

where  $[i - j]$  denoted that obtained by swapping  $i$  and  $j$  in the explicit term.

The other state probabilities are found as:

$$\begin{cases} P_i^S = 1 - P_i^I \\ P_{ij}^{SS} = 1 - P_i^I - P_{ij}^{SI} \\ P_{ij}^{II} = P_i^I - P_{ij}^{IS} \\ P_{ijl}^{SSS} = 1 - P_i^I - P_{ijl}^{SII} - P_{ijl}^{SSI} - P_{ijl}^{SIS} \\ P_{ijl}^{III} = P_i^I - P_{ijl}^{IIS} - P_{ijl}^{ISS} \end{cases}$$

To make the model more analytically tractable, we perform a Mean Field approximation by regarding all the nodes and cliques as equivalent to their average counterparts, equivalently by considering the network to be homogeneous. Accordingly, every node

is assumed to be part of the same number of maximal links  $k^{(1)}$ , 3-cliques  $k^{(1,0)}$ , 3-edges  $k^{(0,1)}$  and triangles  $k^{(1,1)}$ . The states probabilities  $P_i^\sigma$ ,  $P_{ij}^{\sigma\sigma'}$ ,  $P_{ijl}^{\sigma\sigma'\sigma''}$  with  $\sigma, \sigma', \sigma'' \in \{S, I\}$  are taken equal to their respective averages. Considering  $\sigma = S, I, R$ , we can write:

$$P^\sigma = \sum_i \frac{P_i^\sigma}{N}$$

Similarly for the edges:

$$P^{\sigma\sigma'} = \sum_{ij} A_{ij}^{(1)} \frac{P_{ij}^{\sigma\sigma'}}{Nk^{(1)}}$$

And for triangles:

$$P_x^{\sigma\sigma'\sigma''} = \sum_{i,j,l} A_{ijl}^x \frac{P_{ijl}^{\sigma\sigma'\sigma''}}{2Nk^x}$$

where  $x \in \{(1,0), (0,1), (1,1)\}$  indicating the type of the considered 3-clique. Then the system of equations is:

$$\dot{P}^I = -P^I + \beta_1 k^{(1)} P^{SI} + 2\beta^{(1)} [k^{(1,0)} (P_{(1,0)}^{SSI} + P_{(1,0)}^{SII}) + k^{(1,1)} (P_{(1,1)}^{SSI} + P_{(1,1)}^{SII})] + \beta_2 [k^{(0,1)} P_{(0,1)}^{SII} + k^{(1,1)} P_{(1,1)}^{SII}]$$

$$\begin{aligned} \dot{P}^{SI} = & -(1 + \beta_1) P^{SI} + P^{II} - \beta_1 (k^{(1)} - 1) P^{SI} \frac{P^{SI} - P^{SS}}{PS} + \\ & - 2\beta_1 [k^{(1,0)} (P_{(1,0)}^{SSI} + P_{(1,0)}^{SII}) + k^{(1,1)} (P_{(1,1)}^{SSI} + P_{(1,1)}^{SII})] \frac{P^{SI} - P^{SS}}{PS} + \\ & + \beta_2 [k^{(0,1)} P_{(0,1)}^{SII} + k^{(1,1)} P_{(1,1)}^{SII}] \frac{P^{SI} - P^{SS}}{PS} \end{aligned}$$

$$\begin{aligned} \dot{P}_x^{SSI} = & -2(1 + \beta_1 1_{x \neq (0,1)}) P_x^{SSI} + 2P_x^{SII} - \beta_1 k^{(1)} P^{SI} \frac{2P_x^{SSI} - P_x^{SSS}}{PS} - \\ & + 2\beta_1 [(k^{(1,0)} - 1_{x=(1,0)}) (P_{(1,0)}^{SSI} + P_{(1,0)}^{SII}) + (k^{(1,1)} - 1_{x=(1,1)}) (P_{(1,1)}^{SSI} + \\ & + P_{(1,1)}^{SII})] \frac{2P_x^{SSI} - P_x^{SSS}}{PS} - \beta_2 [(k^{(0,1)} - 1_{x=(0,1)}) P_{(0,1)}^{SII} + (k^{(1,1)} + \\ & - 1_{x=(1,1)}) P_{(1,1)}^{SII}] \frac{2P_x^{SSI} - P_x^{SSS}}{PS} \end{aligned}$$

$$\begin{aligned}
\dot{P}_x^{SII} = & -(12 + 2\beta_1 1_{x \neq (0,1)} + \beta_2 1_{x \neq (1,0)}) P_x^{SII} + 2\beta_1 1_{(x \neq (0,1))} P_x^{SSSI} + P_x^{III} + \\
& - \beta_1 k^{(1)} P^{SI} \frac{P_x^{SII} - 2P_x^{SSSI}}{PS} - 2\beta_1 [(k^{(1,0)} - 1_{x=(1,0)}) (P_{(1,0)}^{SSSI} + P_{(1,0)}^{SII}) + (k^{(1,1)} + \\
& - 1_{x=(1,1)}) (P_{(1,1)}^{SSSI} + P_{(1,1)}^{SII})] \frac{P_x^{SII} - 2P_x^{SSSI}}{PS} - \beta_2 [(k^{(0,1)} - 1_{x=(0,1)}) P_{(0,1)}^{SII} + (k^{(1,1)} + \\
& - 1_{x=(1,1)}) P_{(1,1)}^{SII}] \frac{P_x^{SII} - 2P_x^{SSSI}}{PS}
\end{aligned}$$

where:

$$\begin{cases}
P^S = 1 - P^I \\
P^{SS} = 1 - P^I - P^{SI} \\
P^{II} = P^I - P^{SI} \\
P^{SSS} = 1 - P^I - P^{SII} - 2P^{SSSI} \\
P^{III} = P^I - P^{SSSI} - 2P^{SII}
\end{cases}$$

Now, since we have a network model without hyper-edges we can now assume that  $k_{11}$  and  $k_{10}$  are both equal to zero and use as  $k_1$  and  $k_{01}$  the average degree and quadratic degree, respectively. Once the equations are defined, it is possible to find a numerical solution. To this end, I implemented a Python code to simulate the dynamics of the system of differential equations, specifically focusing on the stationary state of  $P_I$ , associated to the infected individuals which represents the proportion of infected individuals. This stationary state depends on the parameters  $\beta_1$  and  $\beta_2$  for the contagion dynamics which govern the contagion dynamics, as well as the initial condition, defined by the initial probability of being infected,  $P_I(\epsilon)$ . We considered two values for  $\epsilon$ , specifically:  $[0.0000001, 0.7]$ , to explore how the initial condition influences the final outcome.

The results from this analysis are reported in Fig.42.

Let us comment these results.

Firstly, we were able to reproduce the effect for which increasing  $\beta_2$  (then the non-linearity of the transmission process) the transition becomes discontinuous, which is well known in literature [108, 114, 92].

On the opposite, if we consider now the particular case of simple contagion, or  $\beta_2 = 0$ , we have only the continuous transition as expected for a SIS model (see blue curves in Fig.42). To interpret this situation we have to reflect about the difference between a discontinuous and continuous transition. Here, we are looking at a phase transition between the inactive, when the infection dies, and the active phase, where you have a constant number of infected individuals over time. In this context, the order parameter, is just the fraction of infected individuals. We know that, whenever you have

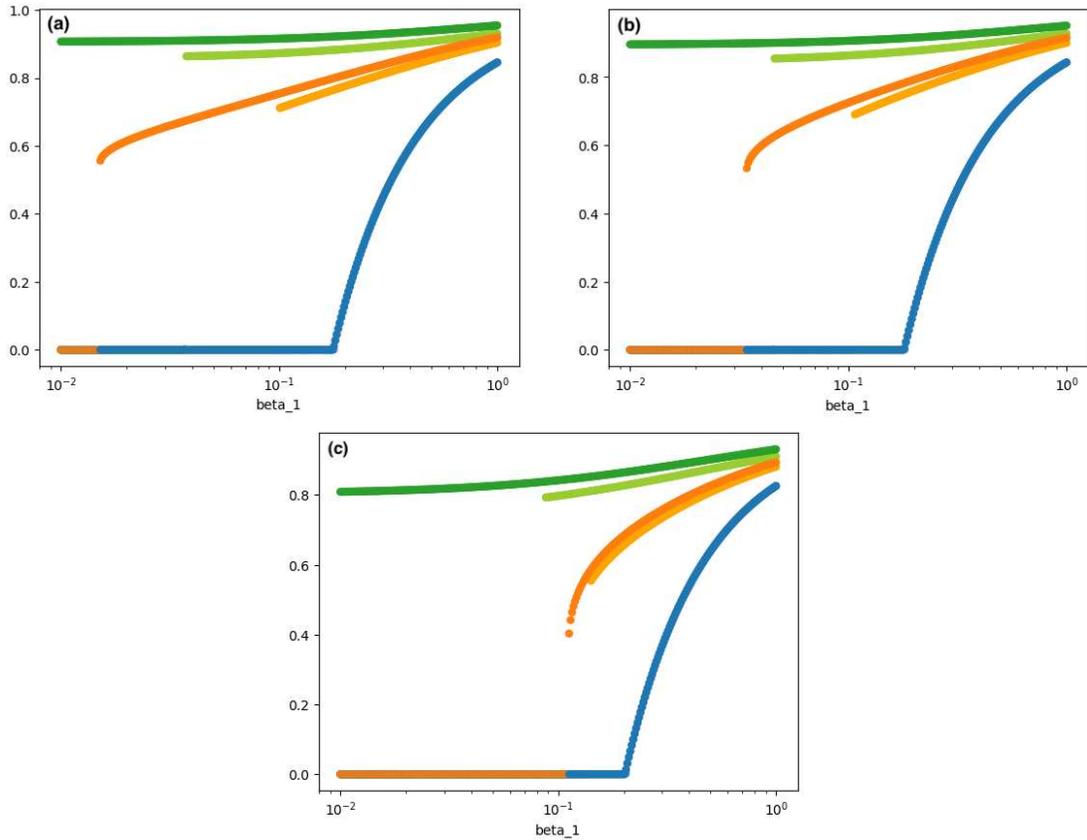


Figure 42: Numerical results for the metric (a), knn (b) and Voronoi (c) networks. These plots represent the stable state of infected individuals in a mean-field approximation for different initial conditions and varying the values of  $\beta_1$  and  $\beta_2$ . The three networks are characterized by different average degree and quadratic degree (number of triangles). Specifically: metric degree: 6.66 and metric quadratic degree: 8.55, Voronoi degree=5.94 and Voronoi quadratic degree=5.92, knn degree= 6.55 and knn quadratic degree 8.04. The dark-orange and dark-green lines correspond to  $\epsilon=0.7$ . The values of  $\beta_2$  are 0 (blue), 0.5 (orange) and 1 (green).

an abrupt change in the order parameter by varying the control parameter, now the infection strength (or better  $\beta_1$  for each selected  $\beta_2$ ), then we call it first order phase transition or discontinuous phase transition. Whenever we have such transition we have a sort of threshold mechanism, to filter out false positive which implies the existence of a critical mass value needed for the activation of the transition. On the opposite, for slowly varying order parameter we have a second order or continuous phase transition. Here, even if we have a non-heterogeneous network (thanks to the mean-field hypothesis), we have a clustered network. Indeed, we see that the average number of triangles per node is quite high (comparable to the number of links). Thanks to this high clustering whenever a node becomes infected, its neighbours are likely connected to each other via triangles. This means that an infected node can simultaneously reinforce the infection in multiple neighbours through their shared connections, amplifying the spread locally. At low infection probability this reinforcement mechanism is weak, and the infection struggles to sustain itself. However, as the infection probability crosses a critical threshold, the clustering effect “activates” and triggers a rapid cascade of infections. This feedback mechanism creates a situation where system abruptly jumps from a low-infection state to a high-infection state, skipping intermediate levels. This is analogous to a bistable system in physics or engineering, where the system has two stable states (low and high infection), and small changes in input lead to dramatic shifts between states.

Now, in order to compare the mean field approximation to our model, we can plot the fraction of infected nodes in the stationary state regime, obtained waiting a sufficiently large number of steps (5000 steps). Let us show from the time series why it is possible to conclude that the stationarity is effectively reached.

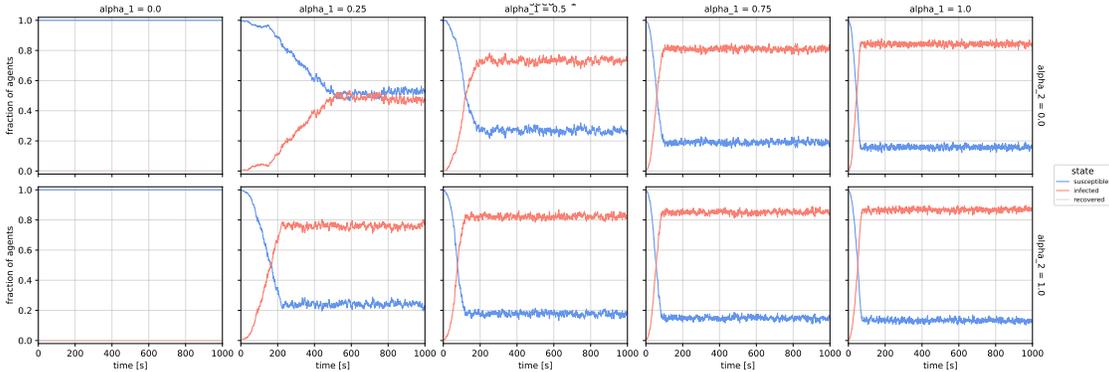


Figure 43: Example of a time series for the fraction of infected nodes in Voronoi network, only one node is initially infected, deterministic recovery time.

The results of the model simulation are illustrated in Fig.44.

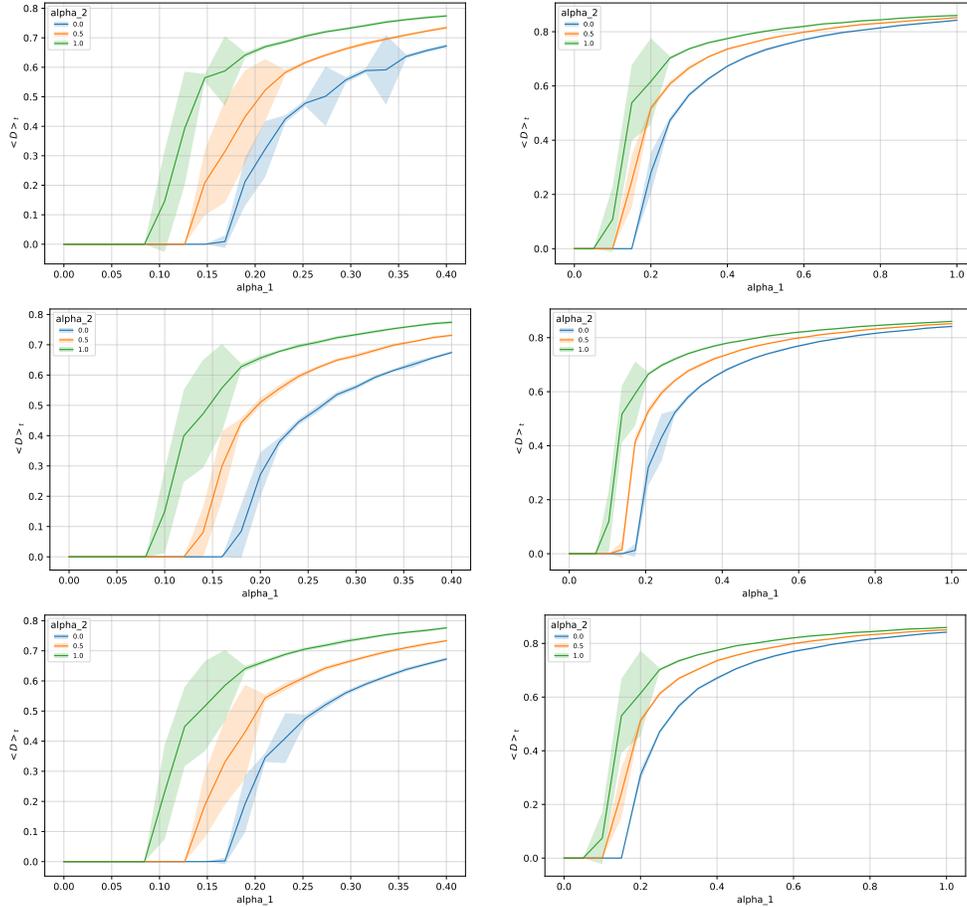


Figure 44: Stable state of the number of infected in the simulation of the model. We reproduced the results for all three networks as a comparison to the above mean-field results. Specifically, the first line represents the Metric network (radius 0.05), the second the topological network ( $k=6$ ) and the third is the Voronoi network. Same number of random seeds (26),  $N=10$  infected initially,  $p=0.1$  for transition I to S, averaging over the last 100 time-steps (over 5000 steps). In the left to the right we just show a zoom-in of the right.

We also include one example of the above plots in the log-x scale to compare with Fig.42.

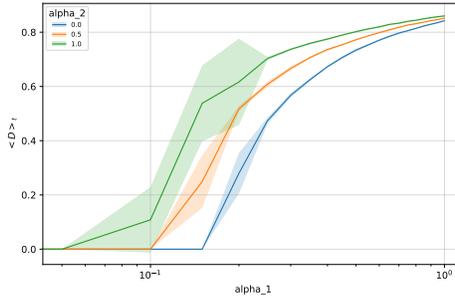


Figure 45: This is Fig.44 in log-x scale.

Let us comment these results.

First of all we observe that of course, we get a great variance only in correspondence of the sharp transition.

Moreover, we see that the transition is only continuous now. It is known that typically the heterogeneities can increase the values of  $\beta_2$  where we have the destruction of the discontinuous transition increases until when we just have continuous transition [114]. Let us observe that real networks, like the ones we are analysing, are heterogeneous, with some nodes (hubs) having many connections and others having very few, especially close to the border. This variability dilutes the effect of clustering because hubs can act as bridges between otherwise unconnected clusters, destroying in some way the clustering effect above. In such networks, even when triangle-based contagion is strong, the influence of hubs ensures a smoother progression of infection, preventing abrupt changes. Here hubs act as “super-spreaders” ensuring that the infection spreads more uniformly throughout the network, regardless of local clustering effects. This uniform spread smooths out the transition, as the infection can propagate effectively even at lower probabilities, bypassing the need for a critical threshold to activate clustering-based feedback. The presence of hubs diminishes the abrupt “jump” caused by triangles, as infection pathways via hubs dominate over localized clustering effects. Even if the network built with the Poisson disk sampling and the linking algorithms defined are not very heterogeneous, this effect is still present.

In conclusion, in heterogeneous graphs, the infection does not rely solely on triangle-based pathways. Instead, it can propagate through multiple mechanisms: direct infection via long-range connections, indirect infection via weakly clustered or loosely connected parts. This diversity of pathways ensures that the infection spreads incrementally, with a smooth increase in prevalence as infection probability rises. The feedback loop necessary for a discontinuous transition becomes diluted by these alternative routes.

Another factor to consider is that the mean-field approximation is quite crude, making it reasonable for the results to show discrepancies.

It is worth also noticing that the above results do not tell us whether there are some cases in which we actually have bi-stability. To see if this was the case the plotted the density of the stable state in a heatmap.

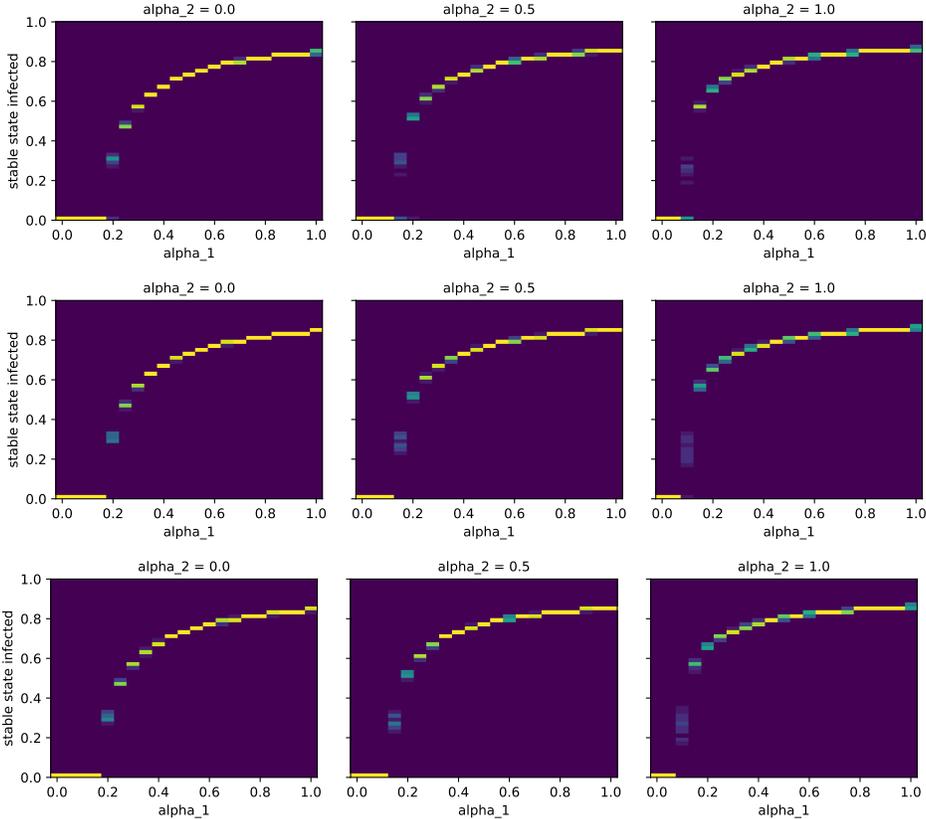


Figure 46: Density of stable state for the Metric network with  $N=10, 200, 800$  initial infected.

The results are consistent also for the other two networks. We can then conclude that we do not have bi-stability.

### 7.5 Conclusions:

We conclude this section by reiterating that these results clearly demonstrate the deep interconnection between social contagion behaviour and spatial organization within animal collectives. The scope of conducting this Mean-Field analysis was to understand

just how important the actual topology of the network was in determining the dynamics. It turns out to be fundamental. The outcome of the contagion process is deeply modified by assuming that all nodes are, in terms of degree and number of triangles, equivalent. The heterogeneities destroy the discontinuous transition we observe through the analysis of the mean field equations; the infection spreads incrementally, propagating through multiple mechanisms (direct infection via long-range connections, indirect infection via weakly clustered or loosely connected parts, for example). This naturally leads us to the introduction of the next section (section 8), where we will analyse even more heterogeneous cases; we will build more realistic networks using the actual positions of the fish instead of the approximate we used so far.

## 8 Using empirically observed data from Mexico trip:

### 8.1 Introduction:

In this last section we want to try to link the theoretical model with the real system. This can be done using the empirically observed data. In the second half of May 2024, during the local dry season, we went to Mexico to study the system in its natural environment. First of all we wanted to understand how much our estimation of the positions of the fish, through the Poisson disk sampling (we will denote these networks as *Poisson networks*), was correct. Then we took photos of the shoals in different positions and at different times and analysed the underlying network. In this way the density fluctuations are more reproduced. It is important to note that, as we could understand from the introduction of the study system above, there are several aspects that could affect the dimension and the topology of the shoals. Primary, the shoals are in different positions in the river. Then, and this is really relevant, the presence of predators of external stimuli in general (as our presence in the field) strongly affect the behaviour. Sometimes it was evident that they basically escaped from us. Moreover, there are several other factors as the time of the day and the temperature. For these reasons it is really important that we develop now, using our field data, a systematic analysis of the real network in different external conditions (which are not accounted in the model). From now on we will refer to the empirically observed positions of the fish as the *fish positions*. Of course, we can not have the real pathway on interaction since we do not develop a specific analysis about the mechanism of communication. However, it is reasonable to assume a Voronoi network as it approximate the visual network [90]. This network is also quite convenient since it does not depend explicitly on some parameters, as the metric and the topological networks which depend on the radius the first and on the number  $k$  the latter.

An example of an annotated image is shown in Fig.47, where the heads and tails of the fish are highlighted with red and green dots, respectively. For more details regarding these annotations see section 8.5.3.

After annotating the positions from frames of the fish in the field we built new network using the actual positions. Let us note that since our model do not account for the dimension of the nodes, to build the network we consider only the position of the head and forget about the actual size of the fish and about the orientation.

As an example, in Fig.48, two networks obtained from different locations are shown, displaying significantly different densities. Apart from the links, which are now set via metric rule, what is evident is that the node-positions are really different with respect to

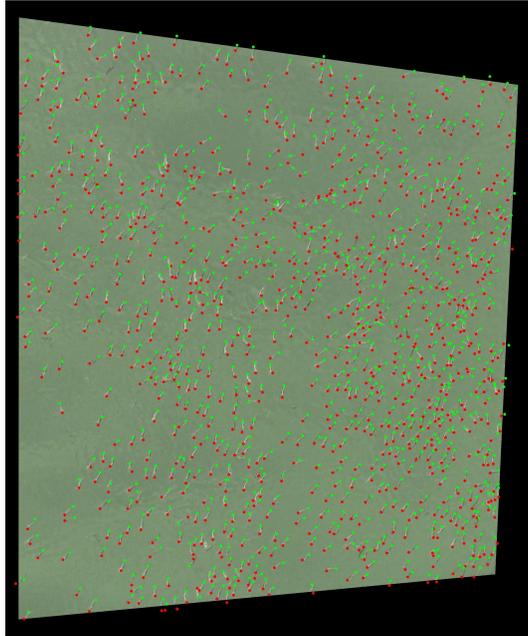


Figure 47: Example of annotated image. The red dot are the heads of the fish, the green dot the tails.

the network we were using until now. We indeed observe graphs that are far less homogeneous, featuring highly clustered regions (hubs) adjacent to sparsely connected areas, with even isolated, disconnected nodes. This structural heterogeneity will undoubtedly have a significant impact on the dynamics.

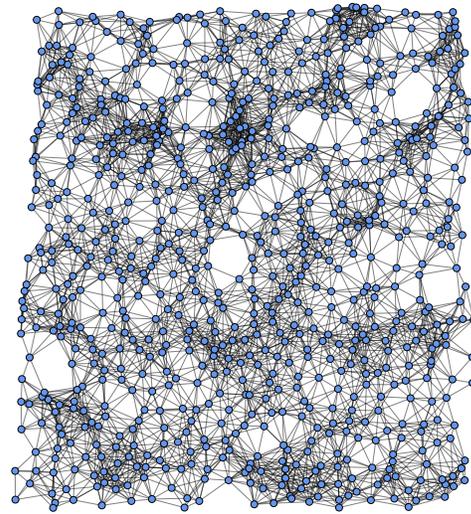
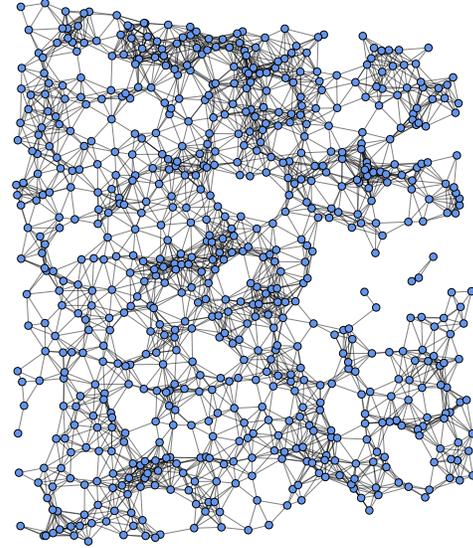
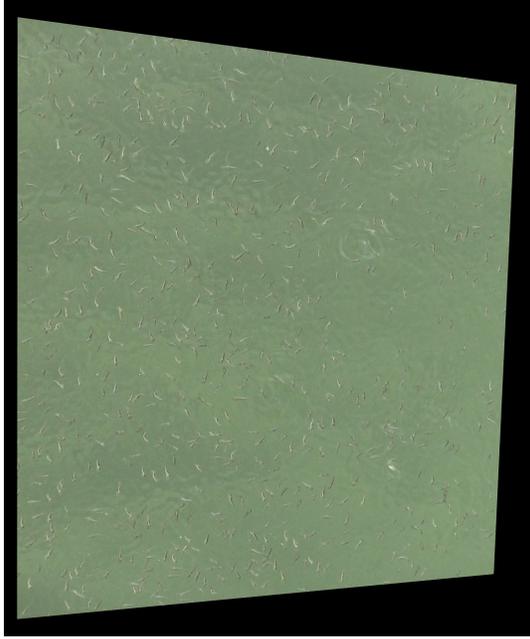


Figure 48: Example of images from the field (left) and the corresponding networks constructed via metric rule (right). There are 704 and 791 fish, respectively.

## 8.2 Network characterization:

To characterize the network built with the fish positions, we use the same measures introduced in first part of the thesis for the Poisson networks. Let us first visualize how different the network structures here is by changing the linking algorithm.

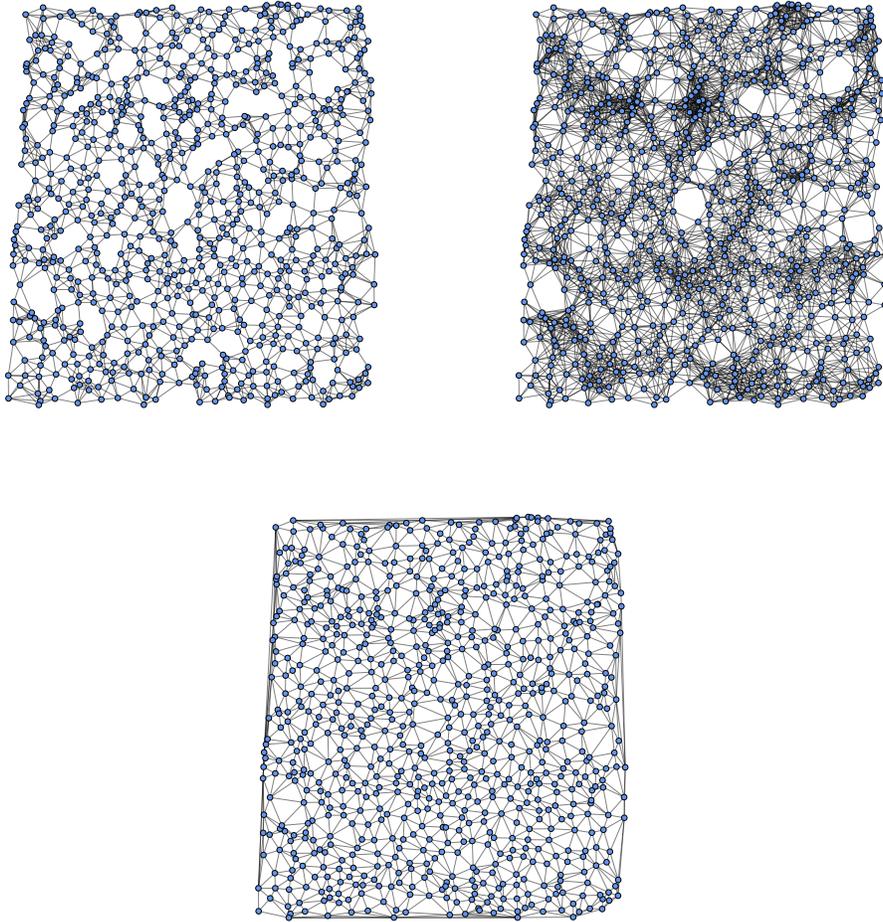


Figure 49: Example of KNN, metric and Voronoi networks built on the same network structure from empirical data.

Here, this difference becomes more pronounced compared to homogeneous networks, particularly in the case of the metric network.

The main measures of interest, given the definition of the dynamics, are [62]:

- the degree distribution;
- the clustering coefficient;
- the distribution of the number of triangles.

Let us show what these distributions look like, then we will analyse a bit deeper the comparison with the Poisson network.

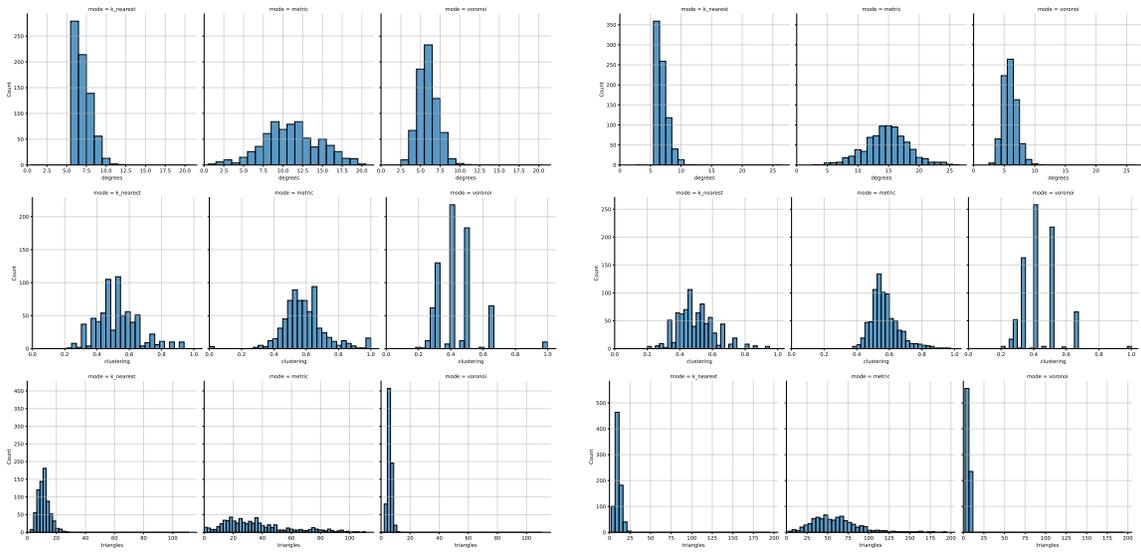


Figure 50: Degree distribution (first line), clustering coefficient (second line), triangles (third line), for two empirically built positions (on the left and on the right). For both we show the results for the three linking algorithms.

Although the two networks discussed above are quite distinct, their distributions exhibit a similar overall shape. Notably, the metric network differs significantly from both the Voronoi and knn networks, particularly in the distribution of the number of triangles.

To deepen this analysis, we directly compare the number of triangles in these networks with the distributions observed in the approximated Poisson networks examined earlier in this thesis. This comparison proves highly valuable for understanding the dynamics, especially in evaluating how the measure of susceptibility differs when applied to fish position networks.

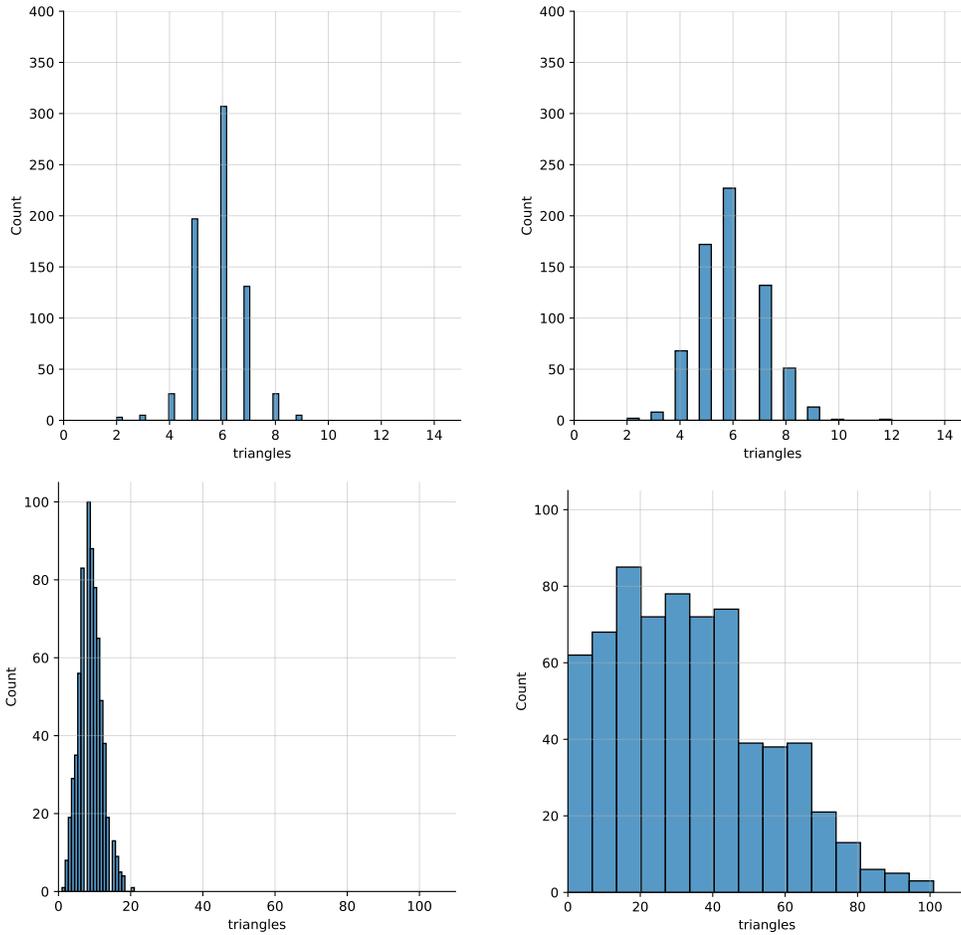


Figure 51: These plots represent the number of triangles for the network with the estimated positions with Poisson sampling and for the empirically observed network with same number of nodes. The above plots correspond to the Voronoi network while below they correspond to the metric network. Note the different scales for Voronoi compared to metric. Note that the width of the lines does not have a meaning.

It is evident that in both cases, and this is drastic in the case of metric network, the distributions related to the fish positions are less peaked. In the metric network we see that there exist nodes involved in even 100 triangles, compared to the Poisson network distribution that goes to zero after 20 triangles.

Homogeneous structures, being more regular, tend to have the same number of triangles (then very peaked distributions). On the opposite, in empirical networks we have both highly connected nodes (hubs) and peripheral or sparsely connected nodes that are

less likely to participate in triangles, leading to a wider distribution. This feature is expected to have significant implications for the study of the dynamics. Specifically, we will find that the final outcome is largely independent of the quadratic contagion mechanism. This can be partially attributed to the heterogeneous nature of the graph. Such heterogeneity creates structural challenges that hinder the spread of the epidemic. To characterize the degree of heterogeneity we computed the variance of the degree distribution, denoted as  $Var(k)$ .

$$Var(k) = \langle k^2 \rangle - \langle k \rangle^2$$

We report some numerical values for graphs with same number of nodes built on the estimated network, denoted as Poisson network, and with empirical positions, called the empirical network.

Linking algorithm	Poisson network:	Empirical network:
Voronoi:	0.91	1.49
KNN:	0.5	0.98
metric:	1.7	13.75

Table 1: Table of the values of variance for the three linking algorithms for both the Poisson network and the empirical network.

Comparing this results with an homogeneous network, as the hexagonal lattice (0.06), it is quite evident how these two networks are heterogeneous. Values bigger than 1 means that there are a few highly connected nodes (hubs) and many nodes with lower degrees. This is common in real-world networks like social networks, the internet, or biological networks, where a small number of nodes (hubs) dominate the connections, and most nodes have a much lower degree.

Finally it is interesting to observe the change in shape of the nearest neighbour distribution (Fig.52). This indicates that the assumption of removing positional noise in the positioning algorithm, though reasonable, results in a significantly altered distribution. However, when accounting for the presence of noise, the distribution becomes notably more symmetrical, aligning better with expected patterns.

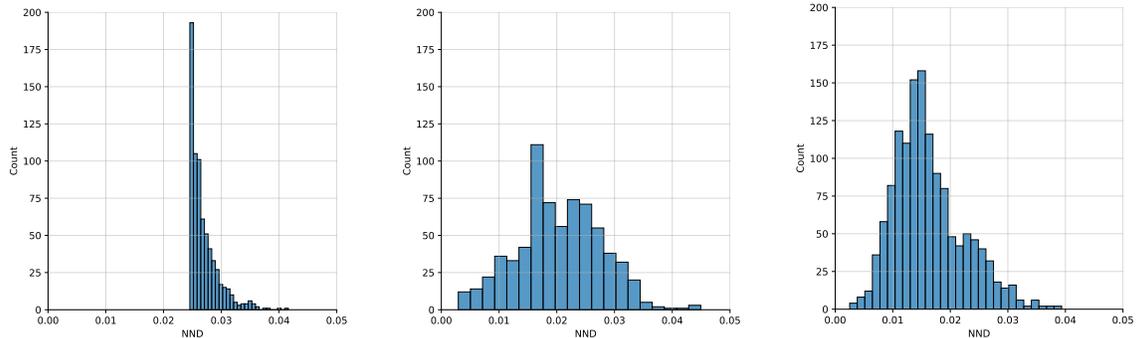


Figure 52: Nearest-neighbour distribution for the Poisson network without positional noise, with the normally distributed noise to the individual coordinates (standard deviation: 0.01) and the network using the fish positions.

### 8.3 Dynamics of the SIR model:

As outlined in the Percolation section (section 6.1), we define an SIR dynamics incorporating a quadratic contagion mechanism, which accounts for higher-order interactions. The objective is to estimate the escape wave size, corresponding to the extent of the contagion phenomenon or the number of final recovered nodes. Analysing this system by varying model parameters reveals behaviour akin to a percolation transition, distinguishing between parameter regimes where the contagion remains localized and those where it propagates across the entire network.

Of particular interest is the system’s ability to discriminate between noise and signal, where noise is characterized as a single random fish diving alone, and signal as two neighbouring fish diving together. This distinction enables the definition of a susceptibility measure, calculated as the difference in response between these two scenarios. The “critical” parameter values can then be identified as those that maximize susceptibility, highlighting the system’s optimal sensitivity to distinguishing noise from signal. Instead of repeating the whole analysis as in 6.1 let us analyse specifically only the susceptibility measure. In Fig.53 we compare directly the result we got with the Poisson network and the new susceptibility we obtain with the fish positions.

The region of “critical” parameters undergoes a noticeable shift. While in the first plot we observe a strong dependence on  $\alpha_2$ , the quadratic contagion parameter, this influence vanishes in the case of the empirical network. In this latter scenario, the critical region is determined almost exclusively by the contagion via links ( $\alpha_1$ ), with the maximum values of susceptibility occurring even at small values of  $\alpha_2$ . This highlights how the structural characteristics of the empirical network significantly alter the system’s dynamics. The presence of such a degree of heterogeneity greatly influences

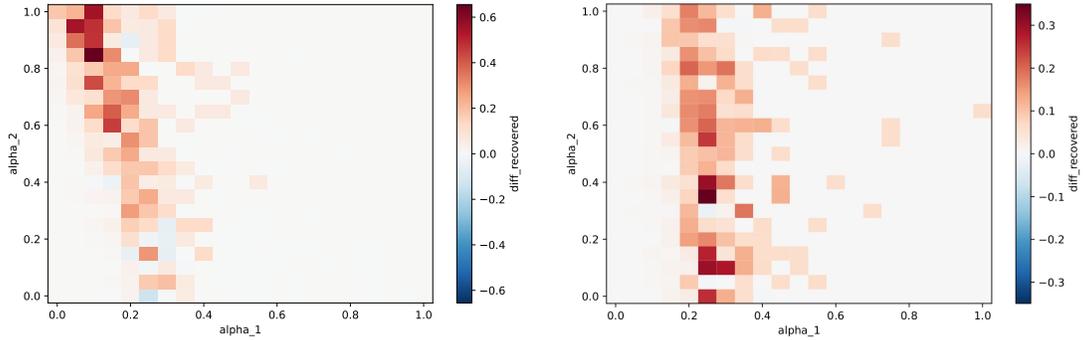


Figure 53: Susceptibility heatmap for a Voronoi network built with the estimated positions with the Poisson sampling and with the fish positions.

the dynamics, implying the existence of various pathways for the infection to spread: direct infection via long-range connections and indirect infection via weakly clustered or loosely connected regions. In this context, the model of quadratic contagion loses its relevance and only the rate of contagion via pairwise interactions affects the susceptibility.

From Fig.53 we can extract the values of the alpha's that maximizes the susceptibility and analyse the sub-critical, critical and super-critical cases. We are then mainly interested in the region of parameters indicated by the red square in the following image:

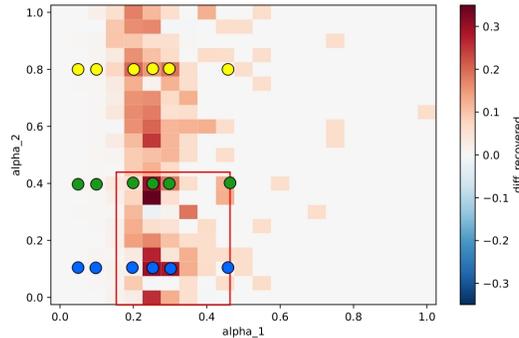


Figure 54: Susceptibility measure for network with empirically observed positions. The red square evidence the region where this measure maximizes. The points correspond to the values we will analyse. In particular we will consider 3 possible values of  $\alpha_2$  (0.1, 0.4, 0.8) and for each of them 6 possible values of  $\alpha_1$  (0.05, 0.1, 0.2, 0.25, 0.3, 0.45).

The blue, yellow and green dots correspond to the combinations of parameters we will consider. In particular, we will classify as sub-critical the first two values of  $\alpha_1$

for all three values of  $\alpha_2$ , critical the three central values and finally as super-critical the last point. Even if this last point does not seem *really* supercritical, it is evident from the plot of the fraction of recovered nodes that it already represents the upper bound of the dynamics. For this reason, we will not analyse any larger values. We want to study the response of the system by varying the external stimulus, which is simply represented by the number of initially infected nodes. After the activation of them, indeed, the contagion can diffuse in the network and the way it does it depends on the infection rate, of course, and on the network topology.

Let us show the fraction of final recovered nodes, which can be denoted as  $\langle R \rangle$ .

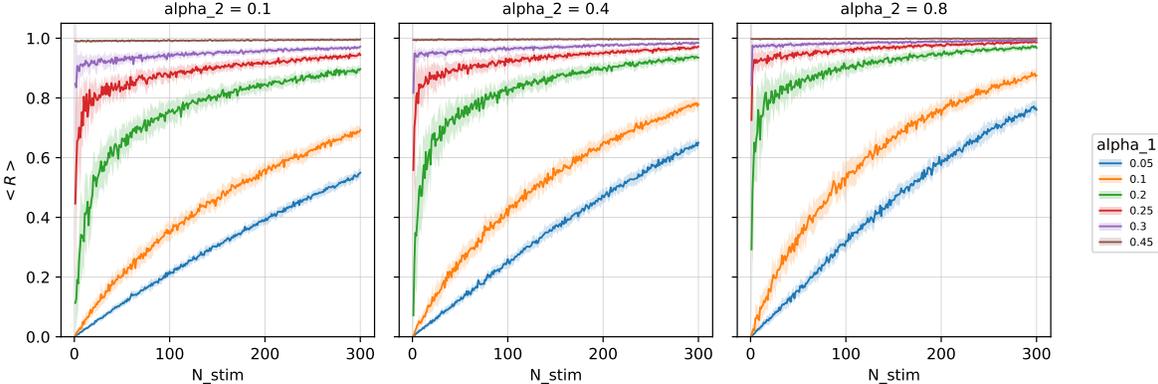


Figure 55: Voronoi network built with fish positions. Estimation of final number of recovered for some values of  $\alpha_1$  and  $\alpha_2$  as explained above and varying the number of initially infected nodes from 0 to 300.

We note a clear change of behaviour. The sub-critical lines are almost linear, the critical lines have a pronounced increase followed by a plateau and the super-critical line is almost flat. However, the dependence with respect to  $\alpha_2$  is almost absent. We see that all the curves are just a little bit steeper.

These results will be compared to the empirical results (section 8.5).

## 8.4 Mean-Field analysis:

Similarly with what we did for the networks with the estimated positions (section 7), we can perform the mean-field approximation of the SIS model under the triadic closure approximation. This analysis was motivated by the willing of acquiring more information about the network and specifically the role of the heterogeneities for the dynamics. Indeed, this approximation consists in assuming each agent of the network as equivalent in terms of degree and number of triangles. We find here comparable results as before, specifically:

1. The appearance of the discontinuous transition for increasing  $\beta_2$ , then the non linearity, in the solution of the mean field equations;
2. The presence of only continuous transitions from the simulation of the model.

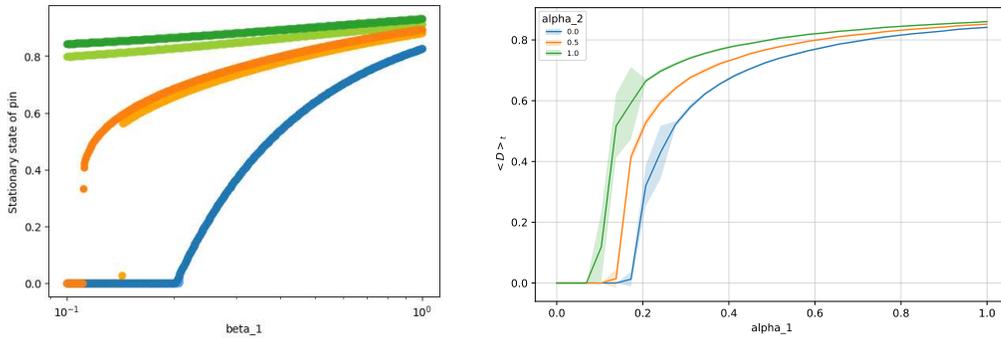


Figure 56: Voronoi network built with real positions. comparison of the estimated transition via mean field analysis and the results of the model.

The first point is justified by the fact that the high average number of triangles and then an high clustering induce a clustering effect for which the infection experiences a sharp increase in correspondence of a threshold value for  $\beta_2$  and then system abruptly change. For this reason we observe the appearance of a discontinuous transition (for more detailed explanation see section 7).

The second point, instead, is once a way, due to the heterogeneities of the network. This variability implies the presence of multiple channels of infection. We could have for example some long range connections or some indirected infection via weakly clustered parts connected to highly clustered hubs.

The topology plays a crucial role in determining the outcome of the contagion process, and as such, the mean-field (MF) approximation proves to be too simplistic to capture its full complexity.

## 8.5 The hexnut experiment:

We present here one of the experiments we conducted in Mexico in May 2024. We defined a easily reproducible experiment which main goal was the analysis of the escape wave. It is not easy in the field to account for all possible side-effects of an experiment in the same way as it is impossible in a model to account for all possible variables that play a role. In order to have more statistics we analysed different shoals in different parts of the river in different moments of the days with two filming sessions per day for 2 weeks.

### 8.5.1 Scope:

The scope of this experiment was to analyse the response of the fish to external artificial stimuli. Of course, it is now important to distinguish the 2 different mechanisms that could lead a fish to dive. The first one is the direct stimulation from the external, as the attack of a predator or some external stimuli. The second one is the behavioural contagion, the one modelled with the complex contagion mechanism. These two are difficult to disentangle, we can not really know if a fish “felt” the dangerous or if it is just copying the behaviour of its neighbours. However, as we will see, we are able in this specific setting to have an estimation of the fraction of individuals that immediately react after an external perturbation and of the fraction of fish involved in the whole escape wave. In the prospective of analysing the escape wave we filmed the reaction of the shoals when stimulated with an increasing the size of the stimulus, which now corresponds with objects of increasing weight.

### 8.5.2 Set-up:

We set a stick on the riverside from which a string was tied up (see Fig.57). The string terminated in the water, where it was anchored to the river floor. In this way we had a constant angle between the string and the water. This set-up allowed us to drop into water hexnuts of increasing weight which correspond to stimuli of increasing strength. Then with time intervals dependent on the weight of the stimulus (to ensure the system to come back to the “normal state”, then when repeated wave behaviour was finished) we stimulated the shoals and film the response. The stimuli were typically randomized and are all of different sizes and we can assume that each of them does not imply a visual stimulation of the shoal, compared to the typical size of the bird predators. Probably the external stimulus mainly affect the lateral line organ (LLO) for the first responding fish. We analysed three different locations along the river, which are denoted as: Small Bridge, BigPoolOutlet, Banana Bay. The river was filmed by two cameras (Sony FDR-AX53 and Canon XF-200 ) one with a close-up on where the

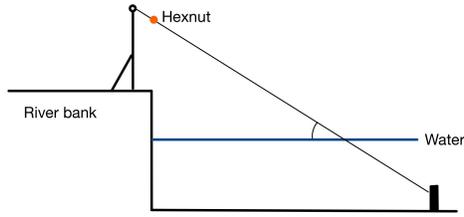


Figure 57: Scheme of the experiment. From the river bank we set the stick from which a string was kept in tension into the water with a fixed angle. From the opposite side of the bank there were two cameras used to record their reaction in response to the external stimulation.

string entered the water, to have a recording of the local density and to estimate the kinetic energy of the hexnuts when approaching the water, and the other for the whole shoal, to analyse the waves. A list of the different stimuli and the time interval to wait after each of them:

stimulus type	weight:	time interval:
1	0.3 g	2 min
1+1	0.6 g	2 min
2	0.8 g	2 min
3	1.1 g	2 min
4	1.2 g	2 min
4D	1.3 g	2 min
2+2	1.6 g	2 min
3+3	2.2 g	2 min
5	2.9 g	3 min
5+2	3.7 g	4 min
6	4.2 g	3 min
4x4D	5.2 g	3 min
7	6.5 g	4 min
6+6	8.4 g	4 min
7+5	9.4 g	4 min
8	12.8 g	4 min
9	31.3 g	4 min
10	49.0 g	4 min

Table 2: Table of the different stimuli we used in the experiment and the corresponding waiting times.

### 8.5.3 Analysis:

Thanks to an annotation tool implemented by Yunus Sevinchan, member of the research group, it is possible to analyse in a systematic way the response. *FishWaveAnalysis* (see [Link to the Zenodo page](#) ) is a Python-based toolkit for the quantitative analysis of so-called “fish waves”; to annotate and process video recordings made in the field and subsequently extract quantitative observables of the waving behaviour.

We manually marked the coordinates of fish positions in the shoal and then the escape wave shapes. Using a calibration object of known size and a perspective transformation we were able to obtain a 2D top-down view of the shoal, where a quantitative geometric analysis becomes possible. To see how it works, see Fig.58.

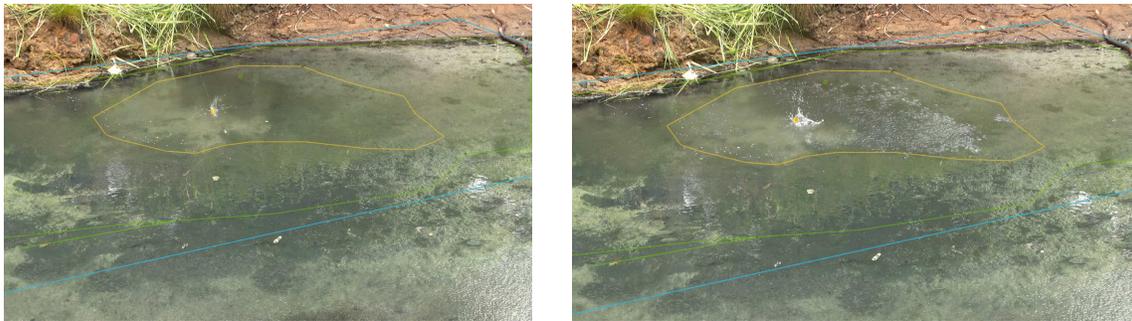


Figure 58: Screenshots from the annotation tool. In all images the yellow curve represents the final size of the escape wave. The small yellow circle is the impact point. In the first image the hexnut is approaching the water, showing no waves, the second is taken after the impact, showing the wave expanding.

Let us observe that the escape wave area is a sort of integration over time, in the sense that we do not have the whole area active simultaneously but rather an initial small region around the hexnut which propagates outwards.

What we can immediately visualize from the above experiment is the area of the waves, depending on the stimulus strength which is quantified by the kinetic energy of the hexnut approaching the water (Fig.59).

Of course we could repeat this plot by associating the different colours to different days, moments of the day for example, or find some other criteria. However, this would not be useful. There are so many observables we could account for and it is really impossible to disentangle all possible effects, like the relation between the time of the day and the location, which could be more/less exposed to the sun. This goes far beyond our analysis of the system.

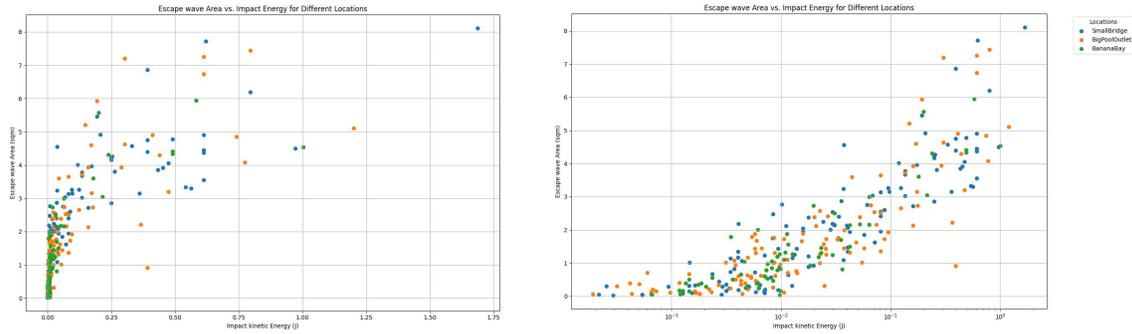


Figure 59: The plot above represent the empirical data of the hexnut experiment; they represent the area of diving fish with respect to impact kinetic energy. The second plot has a log-x scale. The different colours represent different locations along the river, namely: SmallBridge, BigPoolOutlet, BananaBay.

What we observe from these plots is that the escape wave area increases steeply with the external stimulus for low energies (below 0.1J) and then appears to plateau or exhibit a slower rate of increase. However, it is not entirely clear how this impact energy translates into a stimulus perceived by the animals. The pressure wave becomes complicated when looked at it in detail but as a first approximation, a behaviour of the pressure scaling with  $1/r^2$  is something reasonable to assume.

Even fish that do not react directly, would feel some shock wave, which may prime them. This type of stimulus is not encoded in the actual model so to gain deeper insights, we need quantities that are more directly comparable with the underlying model.

To this purpose, we analysed the *first response area*, which is defined as the area directly reacting to the external stimulus. We arbitrarily (heuristically) chose to analyse the size of the wave at 3 frames after impact ( $120 \pm 20$  ms). To understand this choice let us consider some frames (Fig.60).

In frame 2, 3 and 4, some very clear, small, individual splashes. It is even possible to see the light-coloured belly of the fish. In the later frames, ripples overlap, forming some kind of waving front. After four frames the social response sets in. As observed in section 3, the timescale of reaction is typically 5-20 ms for LLO, 120 ms circa for visual. It is therefore reasonable to assume that, 120 ms after the hexnut approaches the water, the only fish diving are those directly frightened by it.

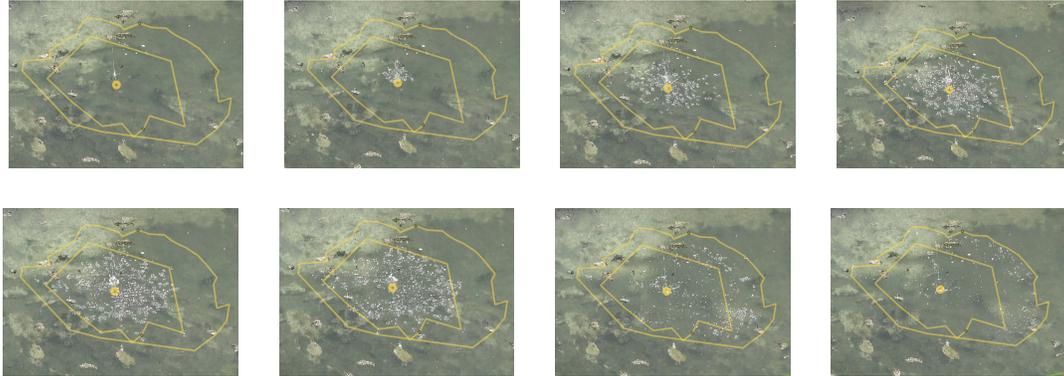


Figure 60: Frames from annotation tool.

Coming back to the analysis, let us show the plot of the escape wave from the empirical data and the related plot from the model analysis.

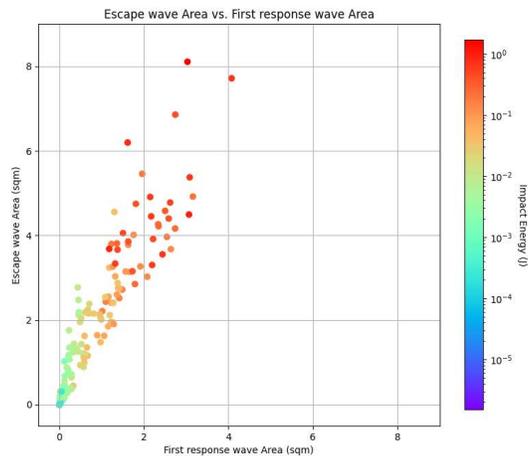


Figure 61: This plot represents the Escape wave area (y) with respect to the first response area (x). The different colours represent the impact energy of the hexnut.

This type of analysis is more related to Fig.55 where we analysed the fraction of recovered nodes by increasing the number of initially infected nodes. Specifically to have an even more clear correspondence between these two levels of analysis we could consider from the model point of view the fraction of recovered nodes by varying the fraction of initially infected nodes, from the empirical analysis the normalized area of the escape wave (with respect to the maximum area, which we can assume to correspond to the size of the shoal), as function of the normalized first response area (same normalization as before). Let us show these plots and then discuss the comparison (Fig.62, Fig.63).

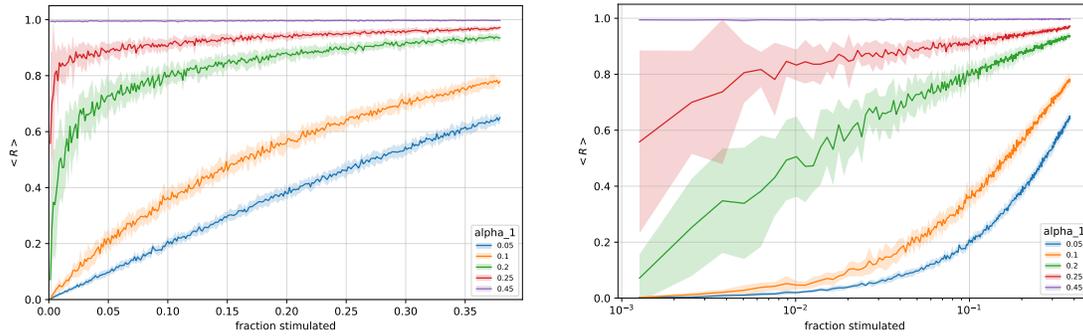


Figure 62: These plots represent (in lin-x scale and log-x scale) the fraction of recovered nodes w.r.t. the fraction of initially infected nodes. The value of  $\alpha_2$  is fixed equal to 0.4, while the different colours represent the values of  $\alpha_1$ .

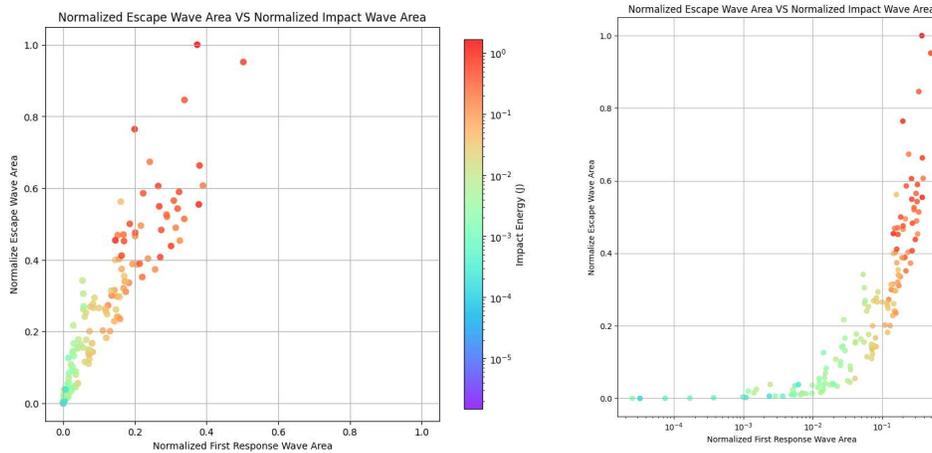


Figure 63: These plots represent (in lin-x scale and log-x scale) the escape wave area w.r.t. the fast response area. The different colours represent the impact energy of the hexnut.

By analysing the non-logarithmic plot in Fig.63, it becomes clear that the empirical data exhibits an increase that lacks the characteristic steepness observed in Fig.62 for high values of  $\alpha_1$ , particularly in the critical and super-critical cases. Furthermore, the data does not show the subsequent plateau, characteristic of green and red curves in Fig.62. Moreover, the log plot shows a qualitative correspondence only with the sub-critical case, where both plots feature an initial plateau, which can be interpreted as an activation threshold. Then, the biological system seems to be described by the model in the sub-critical regime.

As previously mentioned in section 6.1, this analysis alone does not allow us to draw strong conclusions about the applicability of the criticality hypothesis to this system. However, given arbitrary definition of the susceptibility we have given, we can only conclude that, within the specific parameter region and for that particular definition of susceptibility, the model can be defined critical. If we are able to show that the model is well-suited in describing the empirical observations and that the biological system shows some evidences of being at criticality (like the fact that the main properties are described by power-laws, as observed by Gomez et al., [46]), then this analysis would be meaningful.

With these results, however, we do not find it to be critical but rather sub-critical.

In conclusion, another interesting observable is the shape of the waves (Fig.64). As we will discuss in (section 9), this feature is not completely represented in the model. In particular, if the escape wave has typically an approximately-circular shape, similar to what we obtain in the model, the repeated waves will have much different forms. The shape of the wave depends on the shape of shoal, which is clearly modelled by the riverbank, presence of obstacles, the current and other external factors. Here we show all the escape waves for different filming sessions; the colours are related to the kinetic impact energy, so to the stimulus strength.

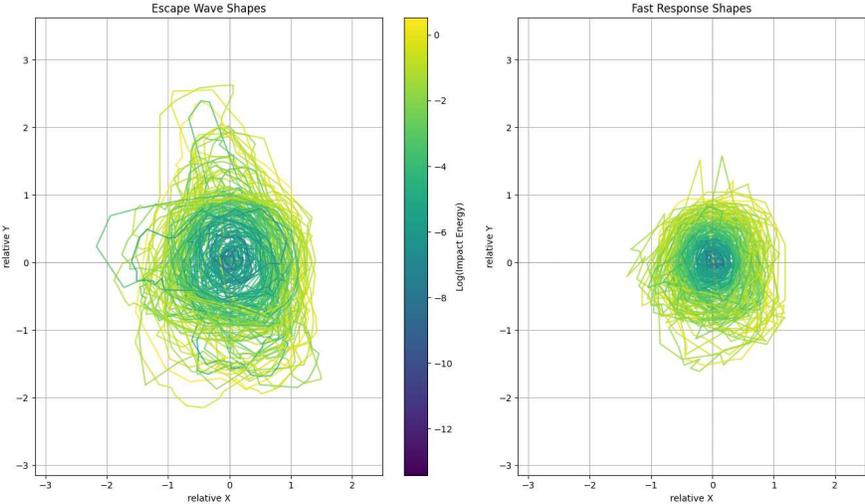


Figure 64: Shapes of the waves. The first plot represent all annotated waves shifting the location where the hexnut approaches the water to the origin for all. The second plot contains all the first-response waves. The colour-bar indicates the corresponding energies.

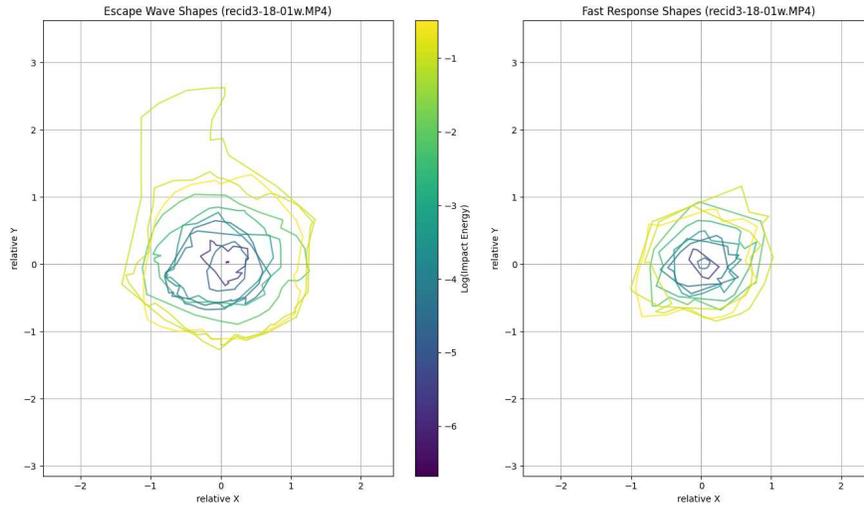


Figure 65: Same as above but for just one recording (recid3-18-01w.MP4). Note that the colour range is different.

## 8.6 Conclusions:

This section of thesis allows to understand how complex the analysis of this biological system really is. The introduction of the positions of the fish, which of course affects the topology but it is only one of the competing factors in shaping the dynamics, turned out to have a crucial role. In particular, the quadratic contagion mechanism, which was suggested to be a possible way to discriminate between signal and noise, has no more (or little) influence. The role of the linking is crucial. Moreover, with the hexnut experiment section 8.5, we aimed tested the response of the shoals. By varying the external stimulus, we estimated fraction of them reacting via the analysis of videos. This is then qualitatively compared to the model, which allows to predict exactly the same result. It seems that only in the "sub-critical" regime, the model is able to reproduce this behaviour.

## 9 Discussion:

We now want to revisit some of the key concepts explored throughout this thesis, and reflect on the underlying motivation that has driven our research. By summarizing the central themes and objectives, we can better appreciate how the various components of the study connect and contribute to a deeper understanding of the study system.

We started with the broad observation that social contagion behaviour within animal groups is closely linked to their spatial organization. While interactions within these collectives occur primarily on a local scale, the group as a whole exhibits emergent patterns of self-organization [82].

With this framework in mind, we focused on a predator-prey system, involving large fish shoals of sulphur mollies. These shoals inhabit a freshwater environment characterized by high temperatures and low oxygen levels [22]. These conditions force them to stay near the surface for aquatic respiration [70], making them more visible and thus vulnerable to fish-eating birds. As a defence mechanism, the shoals exhibit synchronized and repeated dives [8], generating waves on the water surface. Notably, this diving behaviour appears to act as a form of information transmission within the group. The spatio-temporal characteristics of these waves can be mapped onto a network contagion model, offering a valuable framework for analysing the dynamics of such collective behaviours.

To describe this system, we defined a spatially embedded network model with three potential linking algorithms (section 4.2): the metric, the Voronoi, and the topological network. Each of these algorithms successfully captures certain key properties of collective biological systems like the one we are analysing, but it remains unclear which gives the most accurate representation. It is possible, for instance, that the more suited network is a hybrid of these models. Moreover, particularly when the empirical positions of the fish are used, the metric linking can overestimate connections in certain areas due to the significant spatial variability in density.

Another possible description could be given by considering only the visual interaction, which can be studied systematically from the experiments [90]. This would give a more accurate linking, but as we know, sight is not the only mechanism of interaction.

The positions are initially estimated via a Poisson disk sampling which was designed to obtain uniformly spaced nodes, with the density value obtained empirically from the biological system. This approach generates a spatially homogeneous - which does not correctly reproduce density fluctuations - but irregular graph.

In the latter part of the thesis, where we compared the model to the empirical data (section 8), we focused on the Voronoi network, as it has been shown to closely resemble

the visual network of the fish [90]. Let us note that, while positions are central to shaping the dynamics, what truly makes a difference is the linking - a factor incorporated in the model and that we will never be able to observe directly.

Additionally, we assumed the interaction network to be static: neither the positions of the nodes, nor the links between them, change over time. This assumption is based on the idea that individual fish within the shoal are interchangeable, and that the macroscopic structure of the shoal remains relatively constant, even though individual fish may change their positions within the group. We implicitly assumed that the time scale at which the network changes is much lower than that at which contagion occurs. One natural evolution of this approach would be to track the positions over time, creating a dynamic network. This is extremely challenging, if not impossible, with empirical data, but can be done with simulations. Alternatively, another way to improve the model would be to incorporate an adaptive network, as is done, for example, in modelling neuronal processing networks [116].

Finally, a potential extension of the model involves implementing a weighted network, as discussed in section 4.3. For instance, Rosenthal et al. [82] computed the planar representation of each fish's visual field, uncovering the visual information accessible to each individual. By identifying the initiator and first responder of a startle event, they examined social contagion dynamics, analysing how sensory input translates into motor responses (evasion) and identifying the social cues that drive decision-making in this behavioural context. This approach enabled them to calculate the probability of a behavioural response by individual  $i$  when individual  $j$  exhibited behavioural change. These probabilities could be introduced as weights in the network, adding an empirical layer to the model that could significantly influence the dynamics, particularly in real-world networks where sparsely connected zones might result in longer links.

The dynamics we defined are based on a quadratic SIR contagion model applied to our spatially embedded, static network (section 5). Two analyses were developed: the study of the percolation transition (section 6.1) and the mean-field analysis (section 7). Regarding the latter, it is clear that the approximation of treating all individuals as identical agents is quite rough. This approach could be improved: by introducing an individual-based mean-field framework [16], we would achieve a better characterization of the network's final state. With this approach, the local topology of the network is taken into account.

The choice of SIR dynamics was primarily driven by our goal of studying the escape wave, which is the initial wave observed following the external stimulation of the system. Moreover, to capture a more realistic social contagion dynamic that includes reinforcement mechanisms, we incorporated a quadratic contagion mechanism in which contagion spreads both via links and through triangles, each with distinct rates.

That said, other modelling alternatives are also possible and may be better suited for describing certain aspects of the system. One such alternative is a drift-diffusion model [115]. In this approach, the system is defined by a three-state dynamics on a spatially embedded network, similar to our model, but with different state transition rules. Specifically, each node is associated with an excitatory variable that increases as a result of social interactions, generating a drift term. Once the excitation exceeds a certain threshold, the node changes state. This model is likely to align more closely with the biological system under study, as it offers a more individual-based perspective on the dynamics; however, it introduces a greater number of variables, which could complicate the analysis and make it more challenging to explore all possible combinations of these variables.

One important aspect to consider when justifying the choice of the model is that, except for some general information - such as typical densities and a few pre-annotated networks - the empirical data for this system only became available towards the end of this thesis. Initially, it was not anticipated that such data would be accessible, and thus the model was not designed to be directly data-driven. Instead, it was primarily an abstract, theoretical framework, which was only later compared to the available empirical data. Thus, the initial focus of this project was the analysis and exploration of the model itself, with a relatively distant connection to the specific biological system under investigation. This allowed us to work in a more controlled and simplified manner, without being constrained by the complexities of empirically measured data.

The comparison between the model and the biological system revealed that some aspects of real-world dynamics cannot be fully captured by the chosen model mechanisms. One notable discrepancy is the geometry of the waves. In the model, the escape waves are essentially circular in shape. This is due to the assumptions of an undirected interaction network and point-like agents, which do not account for agent orientation or possible asymmetric interactions, and also due to the fact that we typically assume the stimulus to be in the centre. In contrast, in the natural environment, waves often exhibit a variety of shapes, depending on the specific configuration of the shoal. However, it is worth noting that when focusing specifically on the escape wave, like in the case of our study, its shape approximates a circular pattern, as can be observed by inspecting the x-y shapes in Fig.64. This could suggest that, in shaping the escape wave, the polarization plays a marginal role. As a consequence, seeing as the visual interaction strongly depends on the orientation of the individuals, it can not be the only mechanism to consider when building the networks. It is worth noticing however, that using the empirical positions sometimes causes the waves to have highly irregular and non-circular shapes, even when the stimulus is applied at the centre (see Fig.66).

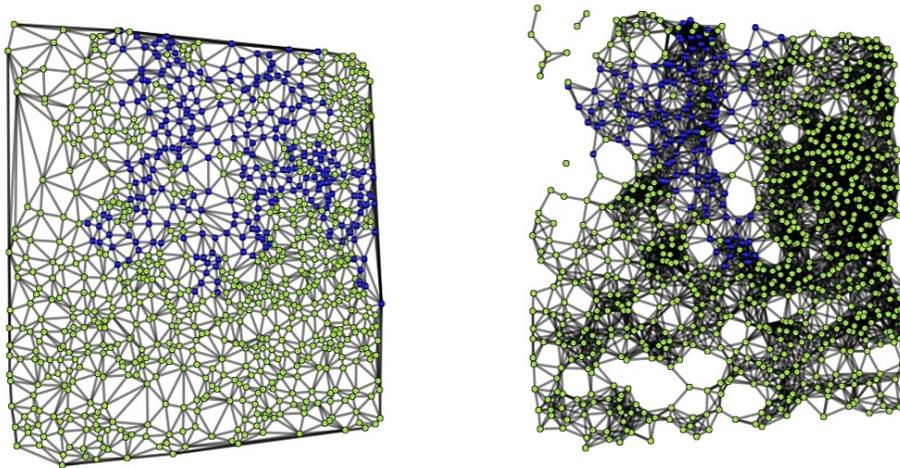


Figure 66: Final state of the network in case of Voronoi and metric network built on the same fish positions; blue indicates recovered nodes, while the green indicates susceptible nodes. In both cases, we started with only 4 individuals infected at the centre.

Another important limitation of the model concerns the size of the shoals. Our analysis was conducted on a small subset of the shoal, which may not fully represent the characteristics of a larger, more complex group. Shoals in nature are typically much larger, and it is often difficult to precisely define the boundaries of a single shoal, as more of them can exist in close proximity and even overlap. The physical boundary for the shoals is the river itself, and thus their shape is usually elongated along the riverbank. This elongated configuration is not accounted for in our model, which assumes a more abstract, simplified representation of the shoal. This limitation should be taken into account when interpreting the model's results in relation to the biological system. Despite all these observations and the simple assumptions we made in the model, we showed that a spatial interaction model can qualitatively reproduce some features of this complex biological system, and we demonstrated that the network topology plays a critical role in shaping its dynamics.

Certainly, there are several possible refinements we could introduce to the model. For example, one important direction for future work would be to extend the model to capture the dynamics of repeated waves. While the escape wave serves as an important starting point for understanding the collective behaviour of the fish shoals, focusing

solely on this initial wave represents a limitation of the current study. The repeated wave behaviour, which is crucial for the survival of these shoals, must be incorporated to provide a more comprehensive view of the system's dynamics. These waves are considered in other works on the same biological system [115, 46]. By considering these successive waves, we would be able to better model the ongoing, adaptive nature of the shoals' responses to external threats, which is essential for understanding the long-term survival strategies of the fish.

Another important aspect is the depth of the diving: thanks to some underwater filming, we noticed this could be a key parameter in allowing the fish to distinguish between signal and noise. In a field survey, Lukas et al. [95] showed that fish dive deeper, faster, and for longer periods of time when both visual and mechano-acoustic cues stimuli are available simultaneously. To incorporate this additional feature we could, for example, distinguish between two levels of infection, if we wanted to keep our model as it is.

This point is also relevant to another factor: the waves are spatially limited. Each time the shoal is stimulated, a reaction of varying intensity is observed. However, only for very strong stimuli—such as predator attacks—do all the fish perform a diving response. What determines the criterion at which the fish stop diving, causing the wave to halt? A decisive factor could lie in the dynamics of the diving behaviour itself, such as its speed and depth. It is worth noting that, despite not explicitly incorporating this type of characterization, the model successfully reproduces the finite nature of the contagion waves.

Finally, another significant aspect that could be integrated into the model is the consideration of fish orientation and the polarization of the shoals. As mentioned in the introduction, one of the key features of collective animal behaviour, particularly in the study of biological collectives as critical systems, is flocking—where individuals within a group align their movements and form orientational order. This emergent order plays a critical role in the dynamics of the collective, as it allows for coordinated responses and efficient movement. In the case of sulphur mollies, the orientation of individual fish within the shoal is likely influenced by external factors, such as the flow of water in their environment. Given that these fish live in a river, where there is a constant flow in one direction, it is plausible that they tend to align themselves with the current. This alignment could significantly affect the way the shoal reacts to stimuli and performs collective actions, such as diving. Investigating how this alignment and polarization evolve before and after a diving event would offer valuable insights into the dynamics of the shoal. It would also allow for a better understanding of the potential link between the spatial configuration of the fish and the overall collective response to external threats. By incorporating these elements into the model, we would be able to explore

how orientational order influences the system's behaviour and how it changes over time in response to environmental conditions or behavioural cues from other members of the shoal. This would help bridge the gap between the simplified model used in the current study and a more realistic representation of the biological system.

## 10 Conclusions:

Let us now summarize the key steps of our work. The first step in our analysis, after having understood the relevant aspects of the biological system, was to define the model. We implemented SIR contagion dynamics on a spatially embedded network. After defining this general framework, we investigated the percolation transition (section 6.1) - essentially, the conditions (in terms of dynamics parameters and network) under which contagion can spread throughout the entire network. This allowed us to identify a measure of susceptibility by examining the difference between noise detection (just one fish diving) and coincidence detection (two connected fish diving). Even if the spontaneous startling is probable, it is infrequent. Then it is reasonable to assume that two spontaneous startles occur in case of true danger.

By comparing these two cases, we were able to gain deeper insight into the contagion mechanisms at play. Specifically, this analysis helped us highlight the significance of the quadratic contagion mechanism, which operates via both links and triangles within the network. We showed that this quadratic contagion mechanism plays a crucial role in determining the final outcome of the network's response and could be a key mechanism in the discrimination ability of the agents between actual signal and noise.

Moreover, by systematically varying the parameters, we were able to pinpoint the conditions that maximize the susceptibility measure, shedding light on the factors that influence the spread of contagion within the system. These findings provided valuable insights into the dynamics of collective behaviour, especially in the context of how network structure and contagion mechanisms interact to shape the overall dynamics. Ultimately, our work demonstrates the importance of these interactions in the study of collective systems and lays the groundwork for further investigation into the factors that drive contagion processes in spatially embedded networks.

Linking this analysis to the biological system, the goal was to study the escape wave, the first wave after any attack, which can be described in terms of "diving" area, and may be compared to the size of the entire shoal to estimate the fraction of diving individuals. The analysis of the size of the epidemic as a percolation process allows us to obtain a directly comparable measure from the model.

By delving deeper into the role of the spatial network, we employed a mean-field analysis (section 7) by assuming that each agent is included in the same number of links and triangles, which were fixed to be equal to the corresponding averages of the network. We also introduced the triadic approximation as a closure for the equations. This analysis allowed us to uncover the critical role that local heterogeneities in interactions play in the system's overall responsiveness to external stimuli. The results showed that the mean-field is not able to describe the system under study. Indeed, while in the

mean-field description we observe the appearance of a discontinuous transition for large values of the quadratic contagion (thus the non-linearity), this is not present in the analysis of the model, which shows only continuous transition. This emphasizes once more the importance of accounting for spatial structure when studying the dynamics of collective systems.

Finally we introduced some empirical data (section 8). Firstly, the detected positions of the fish, which allowed us to build new networks and repeat the same analysis (percolation and mean-field) on these new structures. The comparison between the Poisson-sampled network (which represents a more uniform structure) and the real empirical network (which reflects the observed spatial heterogeneities) highlighted substantial differences in the way contagion spreads through the system. In particular, the empirical network, with its inherent local clustering and spatially structured interactions, exhibited a more nuanced response to external stimuli, compared to the more homogeneous network generated via Poisson sampling. This revealed that the spatial distribution of links—such as the clustering of interactions or the presence of isolated regions—can profoundly alter the dynamics of contagion processes, making the study of systems with explicit spatial embedding crucial for accurately capturing the behaviour of said processes.

Furthermore, our findings suggest that these local heterogeneities—whether in the form of tightly-knit clusters or sparsely connected regions—play a key role in determining the effectiveness of contagion transmission and the overall resilience of the network to disruptions. This underscores the importance of not only considering the global network structure, but also paying close attention to the localized interactions within the system. In conclusion, our analysis demonstrates that explicit spatial modelling is essential for a more realistic and detailed understanding of the dynamics which govern complex, spatially embedded systems, such as those observed in collective animal behaviour.

During field studies on the sulphur molly system, carried out during the local dry season in May 2024, we also set up an experiment whose goal was to test the response of the system under external stimulations of increasing strength. We called it the hexnut experiment (section 8.5) since the stimuli of increasing strength correspond to hexnuts of increasing weight. Defining an easily reproducible set-up and filming the response of the shoals at different location and time of day, we were able to estimate the escape wave area as a function of the strength of the stimulus, which was measured as the final kinetic energy of the hexnuts. Moreover, it was possible to estimate the first response area and gain a clear distinction between the fraction of fish reacting to the stimulus, and the fraction of fish diving due to behavioural contagion.

To connect these results to the model, we have estimated the final fraction of recovered nodes as a function of the number of initially infected nodes, which should be compared to an increasing external stimulus. What we found is that only by considering values of the infection rates in the so-called sub-critical region the curves are comparable; they present a linear increase. This observation raises the possibility that our measure of susceptibility may not be fully accurate or, more fundamentally, that the criticality hypothesis might not be applicable to this system. Naturally, our model is relatively simple, and it would be unrealistic to expect it to fully capture the complexity of the biological system or yield definitive conclusions. Nevertheless, with specific parameter combinations, the model successfully replicated the qualitative trends observed in the system. This outcome highlights the potential for further exploration in this direction, with the aim of refining the model and possibly achieving more quantitatively robust results.

These findings contribute to advancing our understanding of large animal collectives, showcasing the power of physics in predicting and explaining the behaviour of biological systems.

## 11 Appendices:

### 11.1 Square Lattice and Hexagonal Lattice:

We repeat the analysis we did in section 6.1 in two very known and regular structures: the square lattice and the hexagonal lattice.

A square lattice in 2D consists of nodes arranged in a regular grid pattern, where each node is connected to its four immediate neighbours and this structure lacks triangular connections. The absence of triangles simplifies the contagion dynamics, as interactions are primarily between directly adjacent nodes, reducing the likelihood of clustering effects that might accelerate contagion. On the other hand, a hexagonal lattice in 2D is formed by arranging nodes so each has six neighbours, creating a honeycomb pattern. Like the square lattice, the hexagonal lattice is triangle-free.

Both the square and hexagonal lattices provide valuable insight into how network structure affects the spread of infection. Due to their lack of triangles, these networks exhibit no quadratic contagion effects. In such networks, it is also common to observe some isolated nodes remaining susceptible even after the majority of nodes are infected, highlighting their unique contagion-resistance properties. Here we see that these models are not able to distinguish between  $N=1$  and  $N=2$  and we can not identify a region of parameters where the susceptibility peaks. I report only the results for the regular lattice, since the same results are obtained for the hexagonal lattice.

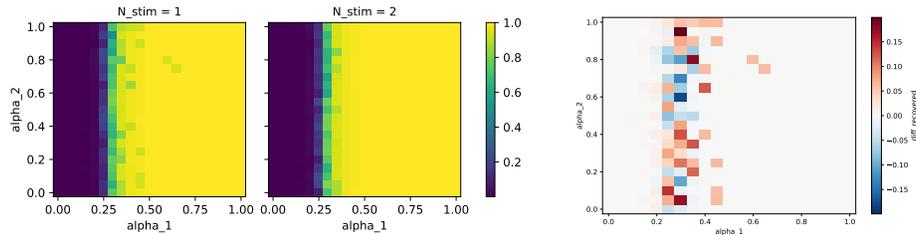


Figure 67: Results in lattice network with 2500 individuals.

When comparing cases, in the susceptibility measure, we observe that the network with  $N=1$  often experiences higher infection rates than the one with  $N=2$ .

### 11.2 Susceptibility measure for metric and KNN:

In section 8, I focused exclusively on the Voronoi linking algorithm to avoid making things too complex. However, it is equally interesting to analyse the other two algorithms, particularly the metric linking, which exhibits markedly different characteristics

in terms of degree and triangle distributions. In both cases, the results remain consistent with the Voronoi network, showing little to no dependence on  $\alpha_2$ . However, while in the Voronoi case susceptibility is maximized at low values of  $\alpha_2$ , the parameter region for maximum susceptibility becomes broader for both the metric and alternative algorithms, leading to an even weaker dependence on the quadratic contagion.

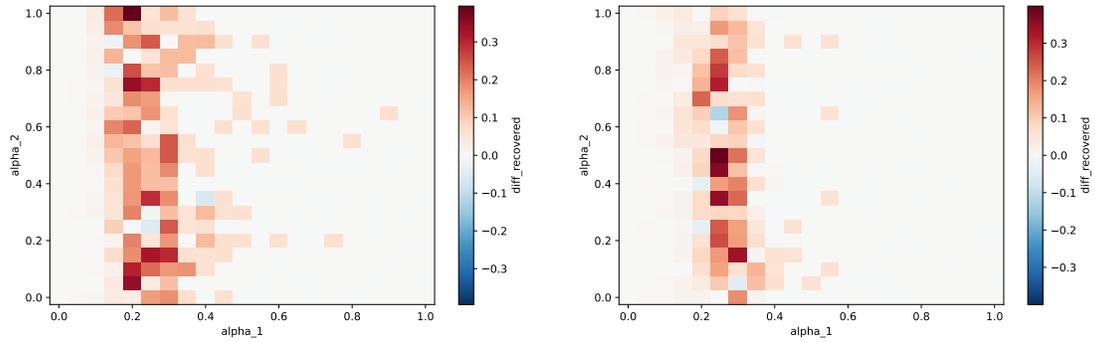


Figure 68: Suscpetibility measure for metric (radius 0.05) and knn (k=6) networks built with fish positions.



## References

- [1] Krause J. and Ruxton G. D. *Living in Groups*, Oxford Univ. Press, 2002.
- [2] Li, K. T., et al, *Fish size, visual resolution and prey selectivity*, Ecology 66, 1729-1735 (1985).
- [3] Cristín, J., et al, *Spatiotemporal organization of ant foraging from a complex systems perspective*, Sci Rep 14, 12801 (2024).
- [4] Krause J. and Tegeder R. W. *Density dependence and numerosity in fright stimulated aggregation behaviour of shoaling fish*, The royal society publishing, 1995.
- [5] Chialvo, D.R. (2010) *Emergent complex neural dynamics*, Nat. Phys. 6, 744-750.
- [6] Balleza, E. et al (2008) *Critical dynamics in genetic regulatory networks: Examples from four kingdoms*, PLoS ONE 3, e2456.
- [7] Vicsek, T. and Zafeiris, A. *Collective motion*, Phys. Rep. 517, 71-140, 2012.
- [8] Doran et al *Fish waves as emergent collective antipredator behavior*, Current Biology, 2022.
- [9] Van Baalen, M. *Pair approximations for different spatial geometries*, The Geometry of Ecological Interactions: Simplifying spatial Complexity, pp. 359-387. Cambridge University Press, Cambridge (2000).
- [10] Cavagna, A. et al *Scale-free correlations in starling flocks*, PNAS, 2009.
- [11] Greenway, R. et al *Correlated divergence of female and male genitalia in replicated lineages with ongoing ecological speciation*, Evolution, 2019.
- [12] Greenway, R. et al *Convergent evolution of conserved mitochondrial pathways underlies repeated adaptation to extreme environments*, PNAS, 2020.
- [13] Tuomainen, U., et al, *Behavioral responses to human-induced environmental change*, Biol. Rev. (2011), 86, pp. 640–657.
- [14] Grace W. Lindsay, *Grounding neuroscience in behavioral changes using artificial neural networks*, Current Opinion in Neurobiology Volume 84, February 2024, 102816.

- [15] Pollock, M.S., et al, *Survival in an extreme habitat: the roles of behaviour and energy limitation*. *Naturwissenschaften* 94, 991–996. doi: 10.1007/s00114-007-0279-2, 2007.
- [16] Qu, B. *The Accuracy of Mean-Field Approximation for Susceptible-Infected-Susceptible Epidemic Spreading with Heterogeneous Infection Rates*, *Complex networks and their application*, 2016.
- [17] Zixiang, Y., et al, *Investigation on the influence of heterogeneous synergy in contagion processes on complex networks*, *Focus Issue on Disruption of Networks and System Dynamics*, *Chaos* 33, 073147 (2023).
- [18] Tobler, M. et al *Compensatory behaviour in response to sulphide-induced hypoxia affects time budgets, feeding efficiency, and predation risk*, *Evol. Res.*, 11, 935-948.
- [19] Tobler, M. et al *Evolution in extreme environments: replicated phenotypic differentiation in livebearing fish inhabiting sulfidic springs*, *Evolution* 65, 2213-2228.
- [20] Tobler, M. et al *Extreme environments and the origins of biodiversity: adaptation and speciation in sulphide spring fishes*, *Mol. Ecol.* 27(4), 843-859 (2018).
- [21] Bridson, R., *Fast poisson disk sampling in arbitrary dimension*, ACM SIG- GRAPH 2007 sketches. SIGGRAPH07. ACM (Aug 2007).
- [22] Pacher, K et al, *Thermal tolerance in an extremophile fish from mexico is not affected by environmental hypoxia*, *Biol. Open* 13(2) (2024).
- [23] Miller, G.A, *The magical number seven, plus or minus two: Some limits on our capacity for processing information*, *Psychological Review*. 63 (2): 81–97.
- [24] Jevons, S., *The power of numerical discrimination*, *Nature*, pp 281-283 (1871)
- [25] Ashcroft, N., et al, *N.D. Solid state physics*, Holt-Saunders (1976).
- [26] Ballerini M. et al, *Empirical investigation of starling flocks: a benchmark study in collective animal behaviour.*, ScienceDirect, 2008.
- [27] Kermack, W.O., McKendrick, A.G. *Contributions to the mathematical theory of epidemics. II.- The problem of endemicity*, *Proceedings of the Royal Society of London, Series A, Containing Papers of a Mathematical and Physical Character* 115, 700-721 (1927).
- [28] Munoz Miguel A. *Colloquium: Criticality and dynamical scaling in living systems*, *REVIEWS OF MODERN PHYSICS*, 2013.

- [29] Farkas I. et al. *Mexican waves in an excitable medium* , Brief communications, Nature, 2002.
- [30] Cross, M., *Pattern Formation and Dynamics in Nonequilibrium Systems.* , Cambridge University Press (2012).
- [31] Bornholdt S. and Rohlf T. *Topological evolution of Dynamical Networks: Global Criticality from Local Dynamics* , Physical review letters, 2000.
- [32] Bornholdt S. and Rohlf T. *Self-organised critical neural networks* , physical review, 2003.
- [33] Saramäki J. et al. *Generalizations of the clustering coefficient to weighted complex networks* , physical review, 75 027105,2007.
- [34] Beggs J. M. and Timme N. *Being critical of criticality in the brain* , Frontiers of physiology 3 (2012), p. 163.
- [35] Steven J. Zottoli and Paul D. Danielson *Lateral Line Afferent and Efferent Systems of the Goldfish with Special Reference to the Mauthner Cell* , SpringerNature, 1989.
- [36] Bodò, A. et al, *Sis epidemic propagation on hypergraphs* , Bull. Math. Biol. 78, 713-735 (2016).
- [37] Jhul B., et al, *Simplicial sis model in scale-free uniform hypergraphs* , J, Stat. Mech. Theory Exp. 2019, 123207 (2019).
- [38] Davidson J.D. et al, *Collective detection based on visual information in animal groups* , Journal of the Royal Society Interface 18 (2021), p. 20210142.
- [39] Tarde, G. *The laws of imitation* , Henry Holt and Company, New York, 1903.
- [40] Robert, M., et al, *Real space renormalization group theory of the percolation model* , Volume 76, pages 477–495, (1994).
- [41] Romero, D.M., et al. *Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter* , Proceeding of the 20th international conference on Worls wide web (2011), 695-704.
- [42] Blumer, H. *Principles of Sociology* , New York, Barnes and Noble, 1939.
- [43] Granovetter, M.S. *The strength of weak ties* , American Journal of Sociology 78, 1360-1380 (1973).

- [44] Dornbusch, R., et al, *S. Contagion: Understanding how it spreads.* , The world Bank Research Observer 15, 177-197 (2000).
- [45] Burgio, G. *Contagion processes on higher-order networks.* , Universitat Rivira i Virgili, Doctoral Thesis, 2024.
- [46] Gomez-Nava L. et al. *Fish shoals resemble a stochastic excitable system driven by environmental perturbations*, nature physics, 2023.
- [47] Poel et al. *Spatial Structure and Information Transfer in Visual Networks*, frontiers in physics, 2021.
- [48] M. Ballerini, N. Cabibbo, R. Candelier, A. Cavagna, E. Cisbani, I. Giardina, V. Lecomte, A. Orlandi, G. Parisi, A. Procaccini, et al. *Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study*, Proceedings of the national academy of sciences 105.4 (2008), pp. 1232–1237.
- [49] Esch H. E. and Burns J. E. *Distance estimation by foraging honeybees*, J. Exp. Biol. 199, 155-162, 1996.
- [50] P. Bak, C. Tang, K. Wiesenfeld *Self-organized criticality*, Physical review, 1988.
- [51] Iain D. Couzin and Jens Krause *Self-Organization and Collective Behavior in Vertebrates*, Advances in the study of behaviour, 2003.
- [52] Selous E. *Thought-Transference (or What?) in Birds*, Constable and Company, London, 1993.
- [53] Shew, W.L., et al, *Information Capacity and Trasmision Are Maximized in Balanced Cortical Networks with Neuronal Avalanches*, Journal of Neuroscience 31 (1 2011), pp. 55-63.
- [54] Daniels, B.C., et al, *Control of finite critical behaviour i a small-scale social system*, Nature Communications 8 (20217), o. 14301.
- [55] Hidalgo, J. et al, *Cooperation, competition and the emergence of criticality in communities of adaptive systems*, J. Stat. Mech. Theory Exp. 2016.
- [56] Pastor-Satorras, R., Castellano, C., Van Mieghem, P. Vespignani, A. *Epidemic processes in complex networks*, Rev. Mod. Phys. 87, 925 (2015).
- [57] Buhl J., et al, *From disorder to order in marching locusts*, Science, 312

- [58] Misund O.A., *Schooling behaviour of sardine sardinops sagax in false bay, south africa*, African Journal of Marine Science. 25, 185-193 (2003).
- [59] Procaccini A, et al *Propagating waves in starling, Sturnus vulgaris, flocks under predation.*, Animal Behaviour 82(4):759–765, 2011.
- [60] Radakov DV *Schooling in the Ecology of Fish*, 1973.
- [61] Treherne JE, Foster WA *Group transmission of predator avoidance behaviour in a marine insect: The Trafalgar effect.*, Animal Behaviour 29(3):911–917, 1981.
- [62] Volz EM, Miller JC, Galvani A, Ancel Meyers L *Effects of heterogeneous and clustered contact patterns on infectious disease dynamics*, PLOS Comput Biol 7(6): e1002042, 2011.
- [63] Allen, L.J.S, et al, *A discrete-time model with vaccination for a measles epidemic*, Mathematical Biosciences Volume 105, Pages 111-131 (1991).
- [64] Nur Farhana Hazwani Abdul Shamad, et al., *Deterministic and Stochastic SIS Model of Common Cold in Universiti Malaysia Perlis*, Applied Mathematics and Computational Intelligence Volume 6, 2017 [29-40].
- [65] Bucyibaruta, G., et al., *A discrete-time susceptible-infectious-recovered-susceptible model for the analysis of influenza data.*, Infectious Disease Modelling Volume 8, Issue 2, June 2023, Pages 471-483.
- [66] Zifeng Yang, et al., *Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions.*, Journal of thoracic disease, Vol 12, No 3 (March 23, 2020).
- [67] Kinouchi O. and Copelli M. *Optimal dynamical range of excitable networks at criticality*, Nature Physics. 2(5), 348-351 (2006).
- [68] D. S. Calovi et al, *Collective response to perturbations in a data-driven fish school model*, Reviews in Fish Biology and Fisheries, 5, 399-416 (1995).
- [69] D. S. Calovi et al *Collective response to perturbations in a data-driven fish school model*, Journal of the Royal Society Interface. 12(104), 20141362 (2015).
- [70] Lukas, J. et al, *Diurnal changes in hypoxia shape predator-prey interaction in a bird-fish system*, Front. Ecol. Evolu. 9 (2021).
- [71] P. K. Visscher, *Group decision making in nest-site selection among social insects*, Annual Review of Entomology. 52, 255-275 (2007).

- [72] I. D. Couzin et al, *Uninformed individuals promote democratic consensus in animal groups*, Science. 334(6062), 1578-1580 (2011).
- [73] A. Strandburg-Peshkin et al, *Shared decision-making drives collective movement in wild baboons*, Science. 348(6241), 1358-1361 (2015).
- [74] I. Pinkoviezky, I.D. Couzin, and N.S. Gov, *Collective conflict resolution in groups on the move*, Physical Review E. 97(3), 032304 (2018).
- [75] P. Romanczuk and B.C. Daniels *Phase Transitions and Criticality in the Collective Behavior of Animals - Self-organization and biological function*, 2022.
- [76] C.H. Lee and A. Lucas, *Simple model for multiple-choice collective decision making*, Physical review E. 90(5), 05284 (2014).
- [77] B.C. Daniels and P. Romanczuk, *Quantifying the impact of network structure on speed and accuracy in collective decision-making*, Theory in Biosciences. 140, 379-390 (2021).
- [78] Sarfati R. et al , *Self-organization in natural swarms of *Photinus carolinus* synchronous fireflies*, Science Advances, 7(28), 1-6 (2021).
- [79] Y. Kuramoto , *Chemical Oscillations, Waves, and Turbulence*, Springer-Verlag, New York (1984).
- [80] H.A. Acebòn et al, *The Kuramoto model: a simple paradigm for synchtonization phenomena*, Reviews of Modern Physics. 77(1), 137-185 (2005)).
- [81] Rahmani P. et al *Flocking in complex environments—Attention trade-offs in collective information processing*, Computational Biology, 2020.
- [82] Rosenthal S. et al *Revealing the hidden networks of interaction in mobile animal groups allows prediction of complex behavioural contagion*, Proceedings of the National Academy of Sciences (PNAS), 2015.
- [83] Zhang Y. et al *Higher-order interactions shape collective dynamics differently in hypergraphs and simplicial complexes* , Nature communications, 2023.
- [84] A. Vespignani et al. *Absorbing-state phase transitions in fixed-energy sandpiles*, Physical review E 62, 2000.
- [85] A. Barrat, M. Barthélemy, A. Vespignani *Dynamical Processes on Complex Networks*, Cambridge University Press, 2008.

- [86] de Arruda, Guilherme Ferraz, G. Petri, and Y. Moreno, *Physical Review Research* 2, 023032, 2020.
- [87] D. P. Maki and M. Thompson, *Mathematical models and applications* (Prentice-Hall Inc., Englewood Cliffs, N.J., 1973.
- [88] S. Clar, B. Drossel, F.Schwabl. *Self-Organized Critical and Synchronized States in a Nonequilibrium Percolation Model*, Physical Review Letters, 1995.
- [89] W. Poel *Visual interactions and spatial group structure in collective information processing*, Dissertation, 2022.
- [90] Strandburg-Peshkin A. et al *Visual sensory networks and effective information transfer in animal groups*, Magazine, 2022.
- [91] Albert-Laszlo Barabasi *Network Science*, Cambridge University Press, 2016.
- [92] Iacopini et al *Simplicial models of social contagion*, Nature Communications, 2019.
- [93] Centola D. and Macy M. *Complex Contagions and the Weakness of Long Ties*, The university of chicago press journals, 2007.
- [94] Bonabeau E. et al *Self-organization in social insects.*, Reviews, 1997.
- [95] Lukas J. et al *Acoustic and visual stimuli combined promote stronger responses to aerial predation in fish*, Behavioral Ecology, 2021.
- [96] Bierbach. D, et al, *An interaction mechanism for the maintenance of fission–fusion dynamics under different individual densities*, Zoological science, 2020.
- [97] Klamser P. and Romanczuk P. *Collective predator evasion: putting the criticality hypothesis to the test*, 2021.
- [98] Jian Gao, Tao Zhou and Yanqing Hu *Bootstrap percolation on spatial networks*, scientific reports, 2015.
- [99] Lukas Riedel, Benjamin Herdeanu, Harald Mack, Yunus Sevinchan, and Julian Weninger. *Utopia: A Comprehensive and Collaborative Modeling Framework for Complex and Evolving Systems*, Journal of Open Source Software 5 (53): 2165. DOI: 10.21105/joss.02165, 2020.

- [100] Yunus Sevinchan, Benjamin Herdeanu, Harald Mack, Lukas Riedel, and Kurt Roth. *Boosting Group-Level Synergies by Using a Shared Modeling Framework.*, Computational Science – ICCS 2020, edited by Valeria V. Krzhizhanovskaya, Gábor Závodszy, Michael H. Lees, Jack J. Dongarra, Peter M. A. Sloot, Sérgio Brissos, and João Teixeira, 12143:442–456. Lecture Notes in Computer Science. Cham, Switzerland: Springer International Publishing. DOI: 10.1007/978-3-030-50436-632, 2020.
- [101] Yunus Sevinchan, Benjamin Herdeanu, and Jeremias Traub *dantro: a Python package for handling, transforming, and visualizing hierarchically structured data.*, Journal of Open Source Software 5 (52): 2316. DOI: 10.21105/joss.02316, 2020.
- [102] István Z. Kiss, Joel C. Miller, and Péter L. Simon, *Mathematics of Epidemics on Networks*, Cham: Springer 587, 31 (2017).
- [103] Lee, H., *Persistent brain network homology from the perspective of dendrogram.*, IEEE Trans. Med. Imaging 31, 2267–2277 (2012).
- [104] Sizemore, A.E., et al, *Cliques and cavities in the human connectome.*, J. Comp. Neurosci. 44, 115–145 (2018).
- [105] Estrada, E., et al, *Centralities in simplicial complexes. Applications to protein interaction networks.*, J. Theor. Biol. 438, 46–60 (2018).
- [106] Sizemore, A.E., et al, *Knowledge gaps in the early growth of semantic feature networks.*, Nat. Hum. Behav. 2, 682 (2018).
- [107] Patania, E., et al, *The shape of collaborations.*, EPJ Data Sci. 6, 18 (2017).
- [108] Giulio Burgio, Sergio Gomez, and Alex Arenas *A triadic approximation reveals the role of interaction overlap on the spread of complex contagions on higher-order networks*, Physical Review Letters 132(7), 077401 (2024).
- [109] G. Burgio, A. Arenas, S. Gómez, and J.T. Matamalas *Network clique cover approximation to analyse complex contagions through group interactions*, Communications Physics 4, 111 (2021).
- [110] Dai, Q., Gao, Y. (2023), *Mathematical Foundations of Hypergraph*. In: Hypergraph Computation. Artificial Intelligence: Foundations, Theory, and Algorithms. Springer, Singapore.

- [111] Cockrell Barbara J. (1984), *Effects of Temperature and Oxygenation on Predator-Prey Overlap and Prey Choice of Notonecta glauca*, Journal of Animal Ecology, Vol. 53. No. 2.1
- [112] Tobler M. (2011), *Evolution in extreme environments replicated phenotypic differentiation in livebearing fish inhabiting sulfidic springs*, Journal of Animal Ecology, Vol. 53. No. 2.
- [113] Altarelli F. et al, *Containing Epidemic Outbreaks by Message-Passing Techniques*, Physical Review X, 4, 021024 (2014).
- [114] Nicholas W. Landry and Juan G. Restrepo, *The effect of heterogeneity on hypergraph contagion models*, Chaos: An Interdisciplinary Journal of Nonlinear Science 30, 103117 (2020).
- [115] Sevinchan, Y. et al, *Spatio-Temporal Dynamics of Social Contagion in Bio-inspired Interaction Networks*, From Animals to Animats 17 (2024), pp 133–144.
- [116] Hernández, A., et al, *Multilayer adaptive networks in neuronal processing.*, The European Physical Journal Special Topics 227, 1039-1049 (2018).
- [117] Dietrich Stauffer and Amnon Aharony, *Introduction To Percolation Theory.*, CRC Press, 1994.