

Aggregation by Movement

Ole Schügl

April 28, 2025

1 Introduction

Spatial pattern formation is a fascinating phenomenon that can be observed in a wide range of ecosystems, such as mussel beds, arid bushlands and ... [5, 10]. Non-spatial modeling approaches can not capture the potential effects of spatial patterns on the system behaviour and it is therefore necessary to model space explicitly, when spatial structure is relevant for system functioning [3]. Besides being able to model the spatial arrangement of organisms in ecosystems, spatial modeling approaches have been able to show that pattern formation can have a positive impact on the stability of ecosystems [3, 12, 9, 4]. In mussel beds, for example, spatial patterns enable a combination of high growth and high survival which can not be achieved in homogeneous beds [12].

Although incorporating space is sometimes necessary, it introduces higher mathematical complexity and should not always be done. To make this decision, it is important to understand when spatial structure impacts system functioning and to be explicit about the modeling assumptions and their justifications. These considerations are explored in ref. [3], where four different approaches, two spatial models and two non-spatial models, for modeling the same system were analysed under three parameter choices and it was shown that no two models agree for all three parameter combinations. Their results highlight how modeling results depend on the choice of the model and that it is important to select the right level of detail, depending on the modeling goals.

When trying to understand the spatial arrangement of ecological systems, it is clearly necessary to model space explicitly. There are two main mechanisms that are usually put forward to explain the formation of patterns. The first mechanism is based on the activation-inhibition principle described by Alan Turing [11], which has been applied in ecological settings to explain the formation of similar patterns in a diverse range of ecosystems [8]. The principle behind this mechanism can be summarized as local excitation (e.g. reproduction) and global inhibition (e.g. long range competition for resources).

The second mechanism is based on the lesser known phase-separation principle developed by John Cahn and John Hilliard [2]. In ecological settings, this corresponds to density-dependent movement, where organisms switch between dispersive and aggregative movement as the density of neighbouring organisms

of the same species increases [5]. There is a diverse cast of organisms, among them certain species of mussels, slime moulds and bird- and fish flocks, whose spatial distribution can be explained by density-dependent movement [5].

In this report, mathematical and numerical approaches for the modeling of spatial systems are explored by studying a simple model that leads to pattern formation based on a discrete version of the phase-separation principle. Counter-intuitively, organisms form clusters only through diffusion and by moving faster when more organisms are around. The model is simulated numerically, and the results are compared to density-independent diffusion and to the analytical results of a stability analysis of a PDE, describing an approximation of the discrete system with point densities.

2 The Model

First, a model for the density-dependent movement of organisms will be presented. The space in the model is a square with sidelength 1 and periodic boundary conditions, which effectively means that organisms exist on a torus.

Since the focus is on movement, the number of organisms, N , is constant. Initially, the organisms are distributed randomly in the environment. The position of all organisms is then updated at discrete time-steps by drawing a displacement from a normal distribution. Mathematically, the change of the position $\mathbf{r}_i = (x_i, y_i)$ of individual i can be described by

$$x_i(t + dt) = x_i(t) + \sqrt{2D(N_R(\mathbf{r}_i, t))dt}u_{i,x} \quad (1)$$

$$y_i(t + dt) = y_i(t) + \sqrt{2D(N_R(\mathbf{r}_i, t))dt}u_{i,y} \quad (2)$$

where $u_{i,x}$ and $u_{i,y}$ are independent, normally distributed random numbers $\sim \mathcal{N}(0, 1)$ and $D(N_R(\mathbf{r}_i, t))$ is a density-dependent diffusion rate. The argument $N_R(\mathbf{r}_i, t)$ is the number of other organisms within a radius R around the organism. For the diffusion rate function, we will use

$$D(N_R(\mathbf{r}_i, t)) = D_0 \left(\frac{N_R(\mathbf{r}_i, t)}{N\pi R^2} \right)^p. \quad (3)$$

For a fixed interaction radius $R = 0.1$, there are three parameters that influence the diffusion rate: p , N and D_0 .

In Fig. 1, the density-dependent diffusion rate as a function of the number of neighbours $\sqrt{2D(N_R)dt}$ is shown for a couple of different parameter values. This corresponds to the variance of the increments of the organism positions. It can be seen, that the diffusion rate increases with the number of organisms in the neighbourhood. Depending on the parameter p , this increase can be linear ($p = 2$) or grow polynomially ($\mathcal{O}(n^{p/2})$). For larger values of p , the diffusion rate grows very fast, which will be important later. The parameter D_0 shifts the curves up or down, and could be understood as a baseline diffusion rate.

In simulations, the impact of these parameters on pattern formation will be studied.

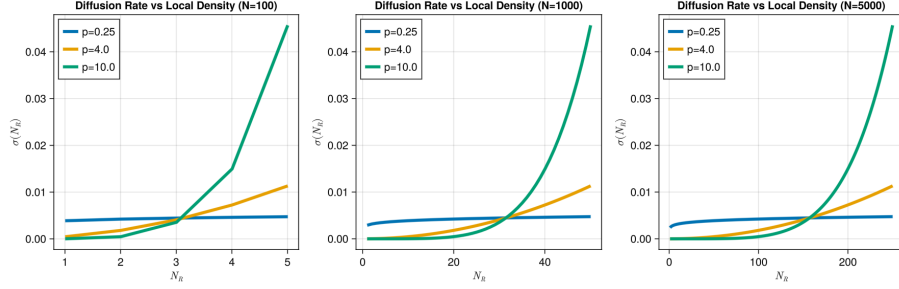


Figure 1: Diffusion rate of an organism based on the number of other organisms within distance R for $N = 100$, $N = 1000$, $N = 5000$ and $p = 0.25$, $p = 4$, $p = 10$ and $D_0 = 0.001$. Note that while the shape of the curves is the same, the number of organisms that can be within the radius to reach a certain diffusion rate increases with the total number of organisms.

2.1 A Single Organism

Before moving on to the simulations, the diffusion of independent organisms in space will be analysed. As will be shown below, regular diffusion leads to organisms spreading out in space. In general, diffusion is a process in which propability density evens out and it is therefore not obvious at all that clustering can be achieved through diffusive processes alone.

To see how orgnaisms disperse under regular diffusion, consider a simple environment with only a single organism and a constant diffusion term. This can be recovered from the density-dependent diffusion rate by setting $p = 0$. For simplicity, it will be assumed that the environment is infinite, as opposed to the torus used in the simulations.

The walker position will change similiarly to before, but with a constant diffusion term

$$x(t + dt) = x(t) + \sqrt{2Ddt}u_x \quad (4)$$

$$y(t + dt) = y(t) + \sqrt{2Ddt}u_y \quad (5)$$

where u_x, u_y are independent random numbers $\sim \mathcal{N}(0, 1)$.

Let's consider the increments $dx(t + dt) = x(t + dt) - x(t)$. Then

$$x(t) = \sum_{\tau=0}^{t/dt} dx(\tau) + x(0),$$

that is, the position of x at time t can be described by adding up the increments. From Eq. 4, we can see that the increments are normally distributed with variance $2Ddt$. The distribution of the displacements $x(t) - x(0)$ is therefore a sum of normally distributed random variables, thus being again normally distributed with the variance being the sum of the variances

$$x(t) - x(0) \sim \mathcal{N}(0, \sum 2Ddt) = \mathcal{N}(0, 2Dt).$$

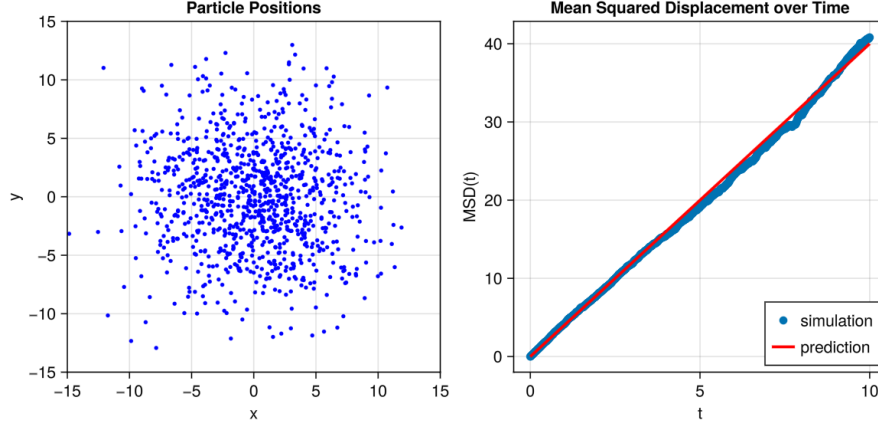


Figure 2: MSD of organisms diffusing in space for a constant diffusion rate of $D = 1$. The left panel shows the final distribution ($t = 10$) of organisms and the right panel shows the MSD calculated at each point in time.

To make the following equations clearer, we will use $\tilde{x}(t) = x(t) - x(0)$ and $\tilde{y}(t) = y(t) - y(0)$. Using this, we calculate the mean squared displacement

$$MSD(t) = \langle \tilde{x}^2(t) + \tilde{y}^2(t) \rangle = \langle \tilde{x}^2(t) \rangle + \langle \tilde{y}^2(t) \rangle.$$

Again we focus on the term involving x .

$$\langle \tilde{x}^2(t) \rangle = \int_{-\infty}^{\infty} \tilde{x}^2 p(\tilde{x}, t) d\tilde{x} = \int_{-\infty}^{\infty} \tilde{x}^2 \frac{1}{\sqrt{4\pi Dt}} e^{-\frac{\tilde{x}^2}{4Dt}} d\tilde{x}$$

Solving this integral requires a few steps, that will be only briefly mentioned. First, we do a variable transformation by setting $u = \frac{\tilde{x}}{\sqrt{4Dt}}$. The integral then becomes

$$\frac{4Dt}{\sqrt{\pi}} \int_{-\infty}^{\infty} u^2 e^{-u^2} du.$$

This integral can be solved by partial integration and is equal to $\sqrt{\pi}/2$. Thus, the final result is

$$\langle \tilde{x}^2(t) \rangle = 2Dt.$$

For $\langle \tilde{y}^2(t) \rangle$, we get the exact same result, and the mean squared displacement is therefore

$$MSD(t) = \langle \tilde{x}^2(t) \rangle + \langle \tilde{y}^2(t) \rangle = 4Dt.$$

These theoretical results are confirmed by simulations, where 1000 independent organisms are initialized at the origin and the mean squared displacement is calculated over time. The results can be seen in Figure 2 and the MSD is in very good agreement with the theoretical prediction.

2.2 Stability Analysis

Assuming that the population is large enough, we can approximate the dynamical system described by Eq. (1), (2) with a continuous density equation given by

$$\frac{\partial \rho(\mathbf{r}, t)}{\partial t} = \nabla^2 [D(\rho_R) \rho(\mathbf{r}, t)], \quad (6)$$

where $\rho(\mathbf{r}, t)$ now describes the point density at position \mathbf{r} and time t and

$$D(\rho_R) = D_0 \left(\frac{\int_R \rho(\mathbf{r}', t) d\mathbf{r}'}{\pi R^2 \tilde{\rho}} \right)^p. \quad (7)$$

The derivation of these equation from the discrete description are presented in [7]. The integral is taken over a disk of radius R centered at \mathbf{r} . We will write the argument $\rho_R = \frac{\int_R \rho(\mathbf{r}', t) d\mathbf{r}'}{\pi R^2}$, which will make the subsequent analysis easier, because the function $D(\rho) = D_0(\rho/\tilde{\rho})^p$ is simpler and does not contain non-local terms.

Following ref. [7], we do a linear stability analysis to understand under which conditions a uniform density is stable. This result will be used to find a suitable parameter range for simulations and is compared with the simulation results.

A small perturbation of the uniform density ρ_0 will be written as $\rho_0 + \epsilon \Psi$, giving

$$\frac{\partial [\rho_0 + \epsilon \Psi](\mathbf{r}, t)}{\partial t} = \nabla^2 \left[D \left(\frac{\int_R [\rho_0 + \epsilon \Psi](\mathbf{r}', t) d\mathbf{r}'}{\pi R^2} \right) [\rho_0 + \epsilon \Psi](\mathbf{r}, t) \right] \quad (8)$$

$$= \nabla^2 \left[D \left(\rho_0 + \frac{\epsilon \int_R \Psi(\mathbf{r}', t) d\mathbf{r}'}{\pi R^2} \right) [\rho_0 + \epsilon \Psi](\mathbf{r}, t) \right] \quad (9)$$

Expanding D in ρ_0 and setting $\Psi_R = \frac{\int_R \Psi(\mathbf{r}', t) d\mathbf{r}'}{\pi R^2}$, we get

$$D(\rho_0 + \epsilon \Psi_R) = D(\rho_0) + \epsilon \Psi_R D'(\rho_0) + \mathcal{O}(\epsilon^2). \quad (10)$$

Substituting this in Eq. (9), gives

$$\nabla^2 [\rho_0 D(\rho_0) + \rho_0 \epsilon \Psi_R D'(\rho_0) + D(\rho_0) \epsilon \Psi + \epsilon^2 \rho_0 \Psi_R D'(\rho_0) \Psi]. \quad (11)$$

The first term is 0, since the density is uniform and we discard the last term because it is in $\mathcal{O}(\epsilon^2)$, leaving us with

$$\frac{\partial \Psi(\mathbf{r}, t)}{\partial t} = \nabla^2 [D(\rho_0) \Psi(\mathbf{r}, t) + \rho_0 D'(\rho_0) \Psi_R(\mathbf{r}, t)]. \quad (12)$$

Assuming a harmonic perturbation $\Psi(\mathbf{r}, t) = \Psi_0 e^{(\lambda t + i\mathbf{k} \cdot \mathbf{r})}$ and using Fourier analysis, we derive λ . Depending on the sign of λ , the perturbations will either grow or decay over time. We substitute Ψ into Eq. 12 to get

$$\lambda \Psi(\mathbf{r}, t) = D(\rho_0) \nabla^2 \Psi_0 e^{(\lambda t + i\mathbf{k} \cdot \mathbf{r})} + \rho_0 D'(\rho_0) \nabla^2 \frac{\int_R e^{(\lambda t + i\mathbf{k} \cdot \mathbf{r}') d\mathbf{r}'}{\pi R^2}. \quad (13)$$

The first term on the RHS is $-|\mathbf{k}|^2 D(\rho_0) \Psi(\mathbf{r}, t)$ while the second is more complicated and is only sketched below. We have

$$\nabla^2 \frac{\int_R e^{(\lambda t + i\mathbf{k} \cdot \mathbf{r}')} d\mathbf{r}'}{\pi R^2} = \nabla^2 e^{(\lambda t + i\mathbf{k} \cdot \mathbf{r})} \frac{\int_{\mathbb{R}^2} \chi_{|\mathbf{r} - \mathbf{r}'| < R} \cdot e^{i\mathbf{k} \cdot (\mathbf{r}' - \mathbf{r})} d\mathbf{r}'}{\pi R^2} \quad (14)$$

The integral is now independent

From [7], we find that

$$\lambda(\mathbf{k}) = -D_0 k^2 \left(1 + \frac{2pJ_1(kR)}{kR} \right), \quad (15)$$

where J_1 is the first-order Bessel function. The paper also states a critical value of $p \sim 7.6$, for which λ becomes positive, though there is no discussion about the value of k for which this is attained. This means that the onset of pattern formation for the PDE occurs around $p_c \sim 7.6$ and independent of D_0 .

2.3 Pairwise Correlation Function

The pairwise correlation function (PCF) and the closely related Ripley's K function are popular tools for quantifying the characteristics of point patterns [13]. They are based on the pairwise distances between all points and provide information about the spatial scales of the patterns as well. The PCF, denoted by $g(d)$, measures the average number of points at a distance d from a randomly selected point. Because there is a finite number of points, we look at rings of finite width and calculate the PCF based on the number of points within a small ring, instead of using exact distances. The PCF is normalized such that a completely random distribution of points corresponds to $g(d) \equiv 1$.

To calculate $g(d)$, the pairwise distances between all points is calculated and results are binned into a histogram, in our case with 200 evenly spaced bins, which corresponds to the rings mentioned above. The normalization is then

$$g(d) = \frac{W(d)}{2\pi d N^2 \delta d}, \quad (16)$$

where $W(d)$ is the average number of points at a distance between d and $d + \delta d$ from a randomly chosen point [1]. Because the simulation is on a torus, no edge effects occur for $d < 0.5$.

The PCF will be calculated for the final distribution of points in the simulations (see below), and used as an indicator to classify the observed pattern. While I have not found any source for this and the statistical significance of this approach is not clear, the MSD between the PCF and $f(d) \equiv 1$ seems to be a good measure of how different a pattern is from a completely random distribution and will be denoted by $MSD(g)$. Since no clear regular patterns were noted in the simulations, a high value of $MSD(g)$ will be interpreted as an indication of cluster formation, while a low value of $MSD(g)$ will be interpreted as an indication of random distribution.

2.4 Simulations

Numerical simulations of the system with density-dependent diffusion rates were implemented in the programming language julia. To keep things manageable, three population sizes were simulated, $N_1 = 200, N_2 = 1000, N_3 = 5000$. For each population size, the simulation was run for a number of values of the parameters p and D_0 . Based on the stability analysis, for population sizes N_1 and N_2 , 40 evenly spaced values between 5 and 10 were simulated for the parameter p and 10 values were used for the largest population size N_3 . The parameter D_0 was simulated across a broader range, from 10^{-1} to 10^{-4} and for population sizes N_1 and N_2 , 40 logarithmically evenly spaced values were used, while for population size N_3 , 10 values were used again. A smaller number of parameters was simulated for population size N_3 , because of the much higher computation time. The initial positions of the organisms was drawn from a uniform distribution on a square of sidelength 1.

To allow patterns to form, the simulation was run for $n = 2000$ steps. At the end of the simulation, the PCF is calculated and $MSD(g)$ is recorded as a measure to indicate pattern formation and plotted as a heatmap for all parameter values (see section 3, 5).

Because it is very insightful to watch the simulation play out live, the code provided alongside the report has the option to see the simulation play out in real-time.

3 Simulation Results

Depending on the simulation parameters, the system ends up in one of two possible configurations: clumped or randomly distributed. A random configuration is shown in Figure 3, alongside the PCF for the largest population size $N_3 = 5000$. No clear pattern is visible and the PCF is close to the line $y = 1$. For distances below the interaction range $R = 0.1$, the PCF lies a little bit below 1, which could indicate a slightly overdispersed pattern, though I am not sure if this is significant. In Figure 4, a clumped configuration of organisms and the PCF can be seen, as before. The clumps are very evenly spaced, forming a grid-like pattern. The PCF has clearly distinguished peaks and valleys, from which characteristic lengths in the system can be determined. The first time the PCF drops below 1 is for a distance slightly below 0.05 and the PCF rises above 1 for distances slightly above 0.1, which is also the interaction range, R , of the organisms. While for $N_2 = 1000$ a similar structure can be seen just with fewer individuals per clump, for $N_1 = 200$ cluster sizes are so small (at most 3 organisms), that no real patterns can be observed and the PCF is very noisy.

To find the range of parameters where clumped patterns occur, the indicator $MSD(g)$ is calculated for a large number of parameters and plotted as a heatmap in Figure 5. A slightly noisy transition occurs around $p = 7$ from small values of $MSD(g)$ to large values of $MSD(g)$, but in contrast to what the stability analysis of the density-based approximation of this system predicts,

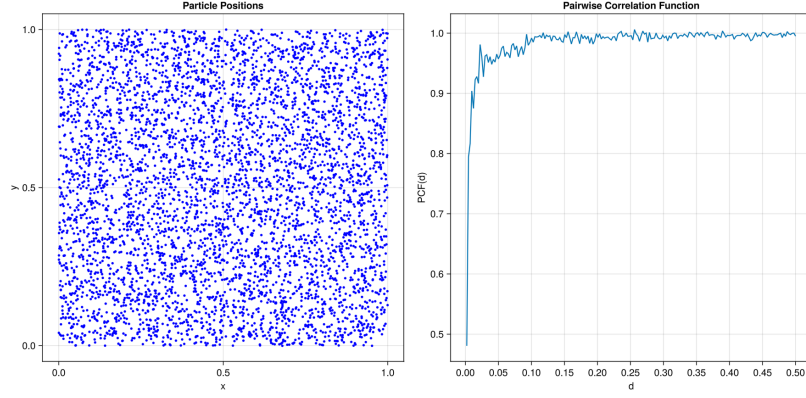


Figure 3: Results of the simulation with $N = 5000$, $p = 4$, $D_0 = 0.01$. The left panel shows the final distribution of the organisms and the right panel shows the PCF of this configuration.

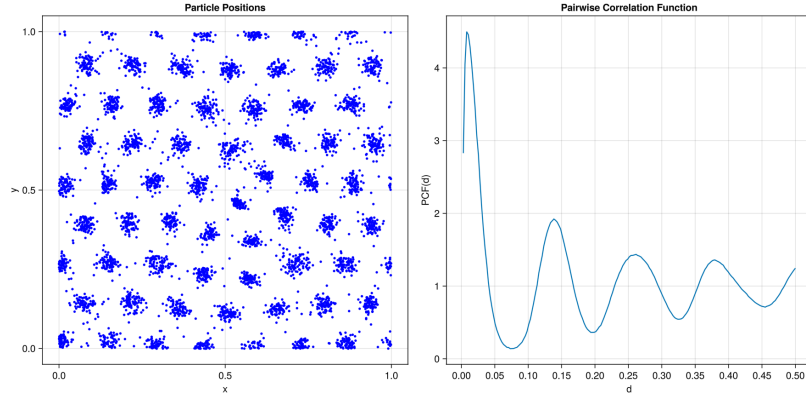


Figure 4: Results of the simulation with $N = 5000$, $p = 8$, $D_0 = 0.01$. The left panel shows the final distribution of the organisms and the right panel shows the PCF of this configuration.

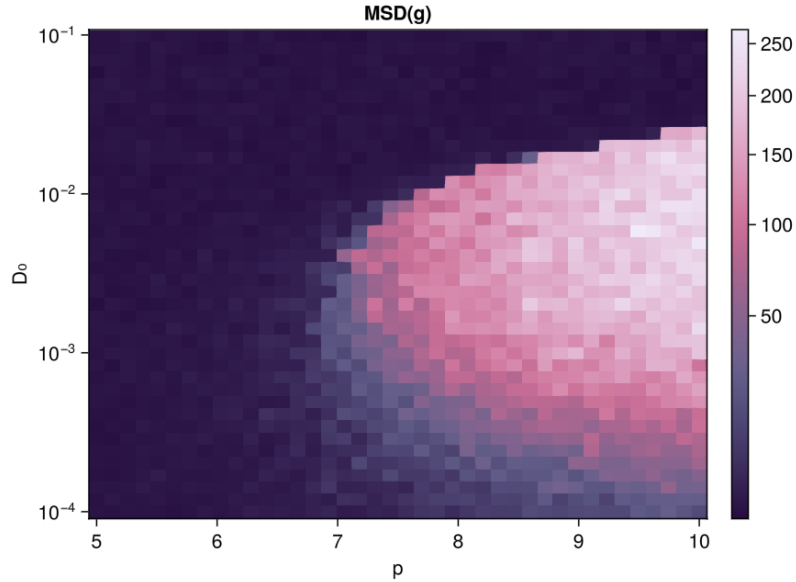


Figure 5: Value of $MSD(g)$ for a grid of parameter choices and $N = 1000$. Brighter values indicate clustering while dark purple indicates random distributions.

there is clearly a dependence on the parameter D_0 as well. For large value of D_0 , no clumped patterns occur anymore. Additionally, very small values of D_0 also seem to lead to random patterns, however, this is an artifact of not running the simulation for long enough. In this scenario, the organisms move very slowly and take much longer to settle into clumped patterns. For $N = 5000$, the 10×10 grid of parameter choices seems to behave similarly, although the resolution is not high enough to say this for sure. For $N = 200$, the heatmap is extremely noisy, making it difficult to point to a clear transition.

There is a transition region where neither a random nor a clumped distribution is stable, for example for $N = 1000, p = 6.5, D_0 = 0.001$.

4 Discussion

We have seen a counter-intuitive mechanism that leads to pattern formation by organisms that move faster when more other organisms are around. Additionally, we have seen some theoretical and numerical approaches for analysing the distribution of spatial patterns.

A speculative explanation of what is required for pattern formation is provided, building upon the observation of real-time simulations and the previous analysis. It boils down to a combination of a low baseline diffusion rate and a

strong increase in the diffusion rate with a growing number of organisms within the interaction range.

A low baseline diffusion rate ensures that the clumps don't just diffuse away because the organisms move only slowly within clusters. As we have seen in an earlier section, the mean displacement of an organism due to a constant diffusion rate grows like $\sqrt{4Dt}$. Thus an organism inside a cluster with an approximately constant number of organisms would leave the cluster after some amount of time, but this could take long time provided D is small enough.

The second aspect allowing pattern formation is a rapid increase in the diffusion rate with an increasing number of organisms within the interaction range. As we have noted, clumps are slightly further apart than the interaction range, meaning that organisms within a cluster only interact within that cluster. However, an organism that leaves the cluster through diffusion suddenly is within the interaction range of multiple clusters and has more organisms within their interaction range. Because the diffusion rate grows rapidly with an increase in neighbours, the organism moves very fast until it jumps into a cluster again where the diffusion rate becomes much lower again. Since organisms take much longer to diffuse out of a cluster, than it takes them to be captured again, the clusters are stable.

While this does not entirely explain the initial formation of patterns, it explains why patterns are stable once they emerge.

The continuous approximation (Eq. 6), predicts that the uniform solution loses stability for $p = 7.6$ and independent of D_0 . The discrete dynamics on the other hand, shows a dependence on D_0 and an earlier transition, with the random distribution becoming unstable as early as $p = 6.5$. While the continuous approximation can provide some insights into the behavior of the discrete model, there are qualitative differences in the resulting dynamics, again highlights the findings of ref. [3].

While this model is unlikely to be a realistic description of any real ecosystem, the continuous description used in the stability analysis, and which is derived from the model studied in this report, has been modified and applied in ecological settings to mussel beds and bacteria, among others [5, 6]. Additionally, the methods presented here can be used to study a variety of spatial systems.

References

- [1] *5.2 Pair Correlation Function*. <https://homepage.univie.ac.at/franz.vesely/simsp/dx/node22.html>. (Visited on 04/25/2025).
- [2] John W. Cahn and John E. Hilliard. "Free Energy of a Nonuniform System. I. Interfacial Free Energy". In: *The Journal of Chemical Physics* 28.2 (Feb. 1958), pp. 258–267. ISSN: 0021-9606. DOI: 10.1063/1.1744102. (Visited on 04/25/2025).

- [3] R. Durrett and S. Levin. “The Importance of Being Discrete (and Spatial)”. In: *Theoretical Population Biology* 46.3 (Dec. 1994), pp. 363–394. ISSN: 0040-5809. DOI: 10.1006/tpbi.1994.1032. (Visited on 03/31/2025).
- [4] Quan-Xing Liu et al. “Pattern Formation at Multiple Spatial Scales Drives the Resilience of Mussel Bed Ecosystems”. In: *Nature Communications* 5.1 (Oct. 2014), p. 5234. ISSN: 2041-1723. DOI: 10.1038/ncomms6234. (Visited on 04/06/2025).
- [5] Quan-Xing Liu et al. “Phase Separation Driven by Density-Dependent Movement: A Novel Mechanism for Ecological Patterns”. In: *Physics of Life Reviews* 19 (Dec. 2016), pp. 107–121. ISSN: 15710645. DOI: 10.1016/j.plrev.2016.07.009. (Visited on 04/04/2025).
- [6] Quan-Xing Liu et al. “Phase Separation Explains a New Class of Self-Organized Spatial Patterns in Ecological Systems”. In: *Proceedings of the National Academy of Sciences* 110.29 (July 2013), pp. 11905–11910. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1222339110. (Visited on 04/12/2025).
- [7] Cristóbal López. “Macroscopic Description of Particle Systems with Non-local Density-Dependent Diffusivity”. In: *Physical Review E* 74.1 (July 2006), p. 012102. ISSN: 1539-3755, 1550-2376. DOI: 10.1103/PhysRevE.74.012102. (Visited on 04/16/2025).
- [8] Max Rietkerk and Johan Van De Koppel. “Regular Pattern Formation in Real Ecosystems”. In: *Trends in Ecology & Evolution* 23.3 (Mar. 2008), pp. 169–175. ISSN: 01695347. DOI: 10.1016/j.tree.2007.10.013. (Visited on 04/25/2025).
- [9] Max Rietkerk et al. “Evasion of Tipping in Complex Systems through Spatial Pattern Formation”. In: *Science* 374.6564 (Oct. 2021), eabj0359. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.abj0359. (Visited on 04/25/2025).
- [10] Max Rietkerk et al. “Self-Organization of Vegetation in Arid Ecosystems.” In: ().
- [11] Alan Turing. “The Chemical Basis of Morphogenesis”. In: *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 237.641 (Aug. 1952), pp. 37–72. ISSN: 2054-0280. DOI: 10.1098/rstb.1952.0012. (Visited on 04/12/2025).
- [12] Johan Van De Koppel et al. “Experimental Evidence for Spatial Self-Organization and Its Emergent Effects in Mussel Bed Ecosystems”. In: *Science* 322.5902 (Oct. 2008), pp. 739–742. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1163952. (Visited on 04/12/2025).
- [13] Thorsten Wiegand and Kirk A. Moloney. “Rings, Circles, and Null-Models for Point Pattern Analysis in Ecology”. In: *Oikos* 104.2 (2004), pp. 209–229. ISSN: 1600-0706. DOI: 10.1111/j.0030-1299.2004.12497.x. (Visited on 04/25/2025).